Supplementary Figure 1: A) Multiple alignment of Cas1 representative sequence set; B) Multiple alignment of Cas2 representative sequence set; C) Multiple alignment of Cas4 representative sequence set; (Please see Supplementary Materials and Methods for details); Sequences are denoted by Genbank identifiers and complete organism name.

A

```
407462288_Candidatus_Nitrosopumilus_koreensis_AR1                 --------------------NQKNHYNVKFLKGY-GHSISVKDSKII---L----KSNHDPFSPP----------EQE-EWFV-------------KNMPYEKIVLSGK-GYVSTEALSLLSENNRNVILL--DNH-G--KP-------V
21229351_Methanosarcina_mazei_Go1                                 -------------------------MKLLLLNGH-GINMHVDGAKLH---I----K-DGRFSTTE----------EPQ-EYVF-----------SPKRIDIDGIIIYGKSGNLTLEAIRWLIKHNVQVSIL--DWN-G--KL-------L
147921327_Methanocella_arvoryzae_MRE50                            -------------------------MNPLLIQGF-GTRISVEKRRLK---I----S-T-----ET----------NDY-EFYP-------------HQIDHDTIVIDGFTGNISFEAMRWIMKHKINLTLL--DWN-G--NL-------L
124485266_Methanocorpusculum_labreanum_Z                          ------------------------MIPWVTVWGY-GAEIKSTRDSLI---V----R-Q-----KG----------STT-Q--Y-------------PLDDMRHLLIAGG-HSLHTSVLERLADRGIAVSFF--TAH-G--KP-------V
307352568_Methanoplanus_petrolearius_DSM_11571                    --------------------MFSPDIPWHVVWGF-GGHIKSTATTLI---I----L-N-----KS----------DVE-E--I-------------PIKSVDHLLVVGG-HNIHTSAIIHLLRQNSSVSFF--DAD-G--TP-------V
154150254_Methanoregula_boonei_6A8                                ---------------------MTRAIPWVAVYGF-GAHIKSTQKKLS---I----L-A-----KG----------KVE-E--Y-------------PLEEIQNLLIVGG-HHLNSTTFSHLLRNGSYISFF--EPD-G--SP-------L
219852754_Methanosphaerula_palustris_E1-9c                        --------------------RSCSKLPWLVVSGY-GAHIKSTPRLLI---V----Q-K-----NG----------TTT-E--Y-------------PIGDVHHLLVVGG-HTIHSAVLQHMQNAGNWVSFF--AAD-G--TP-------V
126178701_Methanoculleus_marisnigri_JR1                           --------------------QMAATEPWLPVFGY-GGHIKATTRELI---I----A-R-----GS----------DTR-R--Y-------------PIQAVKHLLIVGG-HTLHTSAVTNLLKAGAAITIF--DID-G--TP-------V
397780289_Methanoculleus_bourgensis_MS2                           ---------------------MTATQPWLPVFGY-GGHIKATTQELV---I----A-H-----GS----------TTR-R--Y-------------PLQAVKHLLVIGG-HTLHTSAVTNLLKAGAVITFF--DID-G--TP-------V
156937861_Ignicoccus_hospitalis_KIN4-I                            --------------------------MILVVRGH--AFVGRKGYMIT---V----RYKK----DG----------KEV-TEAF-------------PALDIEMAVFVGKGITVSTAALQLLEEQNVPTLFH--GVD-W--S--------F
70607609_Sulfolobus_acidocaldarius_DSM_639                        -------------------------MSILVIKNA---TIMRKGSDLQ---M----TWS-----VG----------KNL-T--V-------------SSLDLELVVIVGNNVRLTSEVILFLSSLNIPVLVH--GKR-N--DV--------
187736488_Akkermansia_muciniphila_ATCC_BAA-835                    -----------------------MSYHILSIDAY-TCHLSCDKGQLR---C----A-D-----GE--N-------SPR-T--I-------------PLEDVGAVVLSSFKATLTSNLLIELARKRIGFVLC--ESY----RP-------A
436839745_Desulfovibrio_hydrothermalis_AM13_-_DSM_14728          ----------------------AQLKRTVYVCNH-FASLGVDIDCLA---V----K-K-----NG--T-------AIA-R--I-------------PFARTNEIVVYGN-SSVSTKLIQKCSYEKIPVTFC--TPS-G--YY-------I
71894593_Mycoplasma_synoviae_53                                   ----------------------MMAKKVVEIKET--EYLFFHLNSLI---V----KKE-----NN----------KIV----L-------------PLSDIDTILISNPYCTISVPLINAIVSNNINLIIC--NKD-F--EP-------N
193216853_Mycoplasma_arthritidis_158L3-1                          -----------------------MEKRVIEISES--NYTYLFLNNLV---V----N-K-----DG--T-------KIT----I-------------PTSQIDTVLFENDKLTITLPLINDLVDHGINIIVC--GRN-H--MP-------K
187250661_Elusimicrobium_minutum_Pei191                           ------------------------MWRVLDIPGD-GYHLCVKNNNFS--AV----K-D-----RE----------EKL-H--C-------------LFDDINSIILYGNNITISNTCIQKCLEHKVPVIFC--DKT-Y--NP-------A
325972002_Sphaerochaeta_globus_str-_Buddy                         ------------------------MWRVIDLGGT-GYYLHTVHNKLI---V----E-K-----DG--E-------ESL----I-------------PFDDIHSIICHGIGFRYSDKFFKSCLHHKVPVTFC--DEK-H--IP-------L
47458867_Mycoplasma_mobile_163K                                   -----------------------MGWKIVEINTD--EYVHLYLNNLY---I----K-------RK--N-------EKI-L--L-------------NIRDIDTILFNNQYSTISIRLLTFLAKNNVNIIFM--NEK-N--EP-------N
269123825_Streptobacillus_moniliformis_DSM_12112                  -----------------------MSFRQVLITQK--SYIHFENDNLI---V----E-------KD--D-------KKL-S--V-------------PISDISIIVFESLESYISLRIISELSLRNITMIFC--DYR-H--MP-------V
57865886_Staphylococcus_epidermidis_RP62A                         ------------------------MKDVIYVENH--YFVTAKENSIKFRNV----I-D-----KS----------EKF----Y-------------LFEEIEAIIFDHYKSYFSHKLVIKCIENDIAIIFC--DKK-H--SP-------V
319957207_Nitratifractor_salsuginis_DSM_16511                     ----------------------MAAWKTIRITKP--CRLSIENGNLV---V----E-------DE--K-------VRF-K--L-------------SLSDTDSIIFEGDRFTLSAKVLAALSRHKVATLFC--DEY-Y--MP-------T
310286727_Bifidobacterium_bifidum_S17                             -----------------------MAWRNVVVTRH--CKISTKMRLLV---I----Q-T-----DD----------DVC-Q--L-------------PMDDIGMLMIETTRAVITTCAIADLLKRDVKIVFC--DDK-H--LP-------I
90961086_Lactobacillus_salivarius_UCC118                          -----------------------MGFRNVIITQH--SKLSYSNNAMI---V----QNK-----DG----------INQ----I-------------PLVDMDILLISTTQAVITSALVSKLAESGIKVIFT--DNK-N--EP-------V
406600270_Leuconostoc_gelidum_JB7                                 -----------------------MAWRTVVVTKH--SKLNYKMNHLV---V----Q-T-----AE----------KTY-Q--I-------------PMEDIQIILITTTQAVVTAYVIAETIKRNIKIIFS--DIN-G--QP-------I
386318631_Staphylococcus_pseudintermedius_ED99                    -----------------------MSFRTVIITKE--SKLSLRMNQLI---V----K-S-----E-----------NLS-R--I-------------PLNEILCLVIEHPNVSMTGHLMNALSDQKIVTILC--NTK-H--LP-------N
302336019_Olsenella_uli_DSM_7084                                  -----------------------MGFRNILIESP--CKCTYKGGYLV---V----R-K-----ED----------EIV-K--V-------------HLSEISSVTLQSTRAFISAYLLSELAKAKISFVVS--DEK-C--NP-------I
15675042_Streptococcus_pyogenes_M1_GAS                            ----------------------MAGWRTVVVNTH--SKLSYKNNHLI---F----K-D-----AY----------KTE-L--I-------------HLSEIDILLLETTDIVLSTMLVKRLVDENVLVIFC--DDK-R--LP-------T
352684362_Acidaminococcus_intestini_RyC-MR95                      -----------------------MSYRTIVITQR--CKLDFCMNYVE---V----Q-T-----AV----------SKK-R--I-------------FIDEIKTLIIENTGVAITAYLLSELMNRKVNLIFC--DKA-H--NP-------Q
42525844_Treponema_denticola_ATCC_35405                           -----------------------MSWRTVVISNR--AKLDLHLNHLV---V------------RG--E-------KTQ-K--V-------------FIEEISVLIIETTAVSITAALLNELIKQKVKVIFC--DEK-R--NP-------A
169823756_Finegoldia_magna_ATCC_29328                             -----------------------MGWRTVLISGR--AKLDLKLNNLV---I----R-------KE----------EIV-K--I-------------HIPEIAVLIIDSTMVSFTTALIEKLIAEKVKIIFC--GKD-H--NT-------I
117929157_Acidothermus_cellulolyticus_11B                         ---------------------MTGPWRVVDLSEL-SGEVHAAQGALL---V------------GD----------ERV------------------PLVDV-AMMLTGPYVSLHGSVIDRAAAFGVGVVHC--DWR-G--VP-------V
283456134_Bifidobacterium_dentium_Bd1                             ---------------------MQSQWRVIDCTNL-DGTISYERGRMK---V----R-K-----HG----------ESE-SVCV-------------PLAQAAVVLIGLR-VCCSSAVLHEMAKAGVSVMLC--DWR-G--IP-------D
38232679_Corynebacterium_diphtheriae_NCTC_13129                   ---------------------MNPGWRVVDLIDF-DGKVSYQRGQLA---I----T-S-----DS--G-------ELRAT--L-------------PLAQIAVVLIGNK-LIISGAVLVKLSEYDIAVLVC--DWR-R--VP-------V
238924076_Eubacterium_rectale_ATCC_33656                          -----------------------MAFRVILIENE--VTIKVKLNNLI---I----T-------KC--G-------EDI-W--I-------------PLDDISMIVLDNLASNLSTRLMCQLAEQGIGLMLY--NQQ-H--LP-------T
116627543_Streptococcus_thermophilus_LMD-9                        -----------------------MTWRVVHVSQS--EKMRLKLDNLL---V----Q-------KM--G-------QEF-T--V-------------PLSDISIIVAEGGDTVVTLRLLSALSKYNIALVVC--DNE-H--LP-------T
57238545_Campylobacter_jejuni_RM1221                              --------------------SYDEAFKTLLISSN--AKLNLELNHLV---I----K-Q-----DE----------NIA-K--L-------------FLKDINIIVLESLQVSISSALFNAFARHKIILLTC--DET-H--SI-------N
291276264_Helicobacter_mustelae_12198                             --------------------MFDEAFRTLFMSKP--AKLSHKNNNLI---I------------HG--K-------EDI-G--I-------------PLRDIHTIVIDTPQITLTSSLMKELAQYKILVFFC--DDS-H--LP-------S
294086112_Candidatus_Puniceispirillum_marinum_IMCC1322            ---------------------------MIEISGQ-PRHISVRRGAII---V----S-D-----NE--T-------ELG-Q--I-------------DLDGVLSVIVTSRGATVTTPMITEAAARNIPIIIC--NDR-F--QP-------V
189485226_uncultured_Termite_group_1_bacterium_phylotype_Rs-D17   -----------------------MSWRTVFFSKP--CSLSVKNSQLY---C----R-F-----QD--S-------TVH-D--I-------------PIEDISVIVLESNRINLTSALISECSQSNIVIFSC--DSS-H--IP-------C
222109284_Acidovorax_ebreus_TPSY                                  -----------------------MVWRSIVINRP--AKLKREHFALV---V----E-------QE----------QSA-R--V-------------PFEDIAVIVLNHREITLTHPVLSACGEYGIGLYST--GDN-H--QP-------N
15602991_Pasteurella_multocida_subsp-_multocida_str-_Pm70         --------------------GVIMSWRSILISQG--GKLSLQRKQML---I----Q-------QS--N-------NSF-T--V-------------PLEDIAVIVIESKETLITAPLLSALAQYGVTLLTC--DEQ-F--LP-------C
310780383_Ilyobacter_polytropus_DSM_2926                          -----------------------MGWRNIFITSS--TKISTRDSGLL---I----T-------KE--G-------GEV-N--I-------------PLEDISSIILDNNSVTITSRVLEEIASRDIVLLTS--DEK-H--LP-------N
220930481_Clostridium_cellulolyticum_H10                          -----------------------MGWRNIIVSNP--TKLKLKQNNLW---V----E-Q-----SD----------GFS----I-------------PIDDINTIVLDSADVTITSALLSKLAEEDIALYSC--DGK-H--TP-------N
34557789_Wolinella_succinogenes_DSM_1740                          ----------------------MAAWRTVVIATS--CHLRVRDRQLV---I----E-S-----RE--G-------ERT-Q--L-------------PLADIGVVIVENPQVTLSAALLSALAQEGTALMSC--DSS-H--LP-------D
90425960_Rhodopseudomonas_palustris_BisB18                        -----------------------MAWRGLHLTQA--ARLSLADGQVC---V----R-Q-----DA--G-------EVR----L-------------ALEDIAWIVIDTPQATLSSALMSACMDAGVVLIFT--DER-H--TP-------S
209542526_Gluconacetobacter_diazotrophicus_PAl_5                  -----------------------MAWRGVHISHP--SRLTHRNRQLV---V----A-Q-----DG--G-------EVS----L-------------AVEDIACLILDTRQVSITGSLLSALAENGVAMIVP--DAR-H--HP-------A
121608212_Verminephrobacter_eiseniae_EF01-2                       -----------------------MIGRTVEIADD-RRHLFVNRGFMV---IKDTEG-E-----RK----------ELG-Q--V-------------PLDDIAAVIANAHGLTYTNNLLVALAERGAPFVLC--GPN-H--NA-------V
60683388_Bacteroides_fragilis_NCTC_9343                           -----------------------MIKKTLYFGNP--VYLSLRNAQLV---I----K-L-----PD--V-------EKA-T--VLPETLKKQAEVTKPIEDIGIVVLDNKQITITSGVLEALLENNCSVITC--DSR-S--MP-------V
154250554_Parvibaculum_lavamentivorans_DS-1                       -----------------------MIRKTVEFATP-GTRLSVANKQLV---I----E-R-----PD--L-------PKA-T--L-------------PIEDLGVVIVDDLRATYTQAVFIELLEAGATVMVT--GRD-H--LP-------A
283778924_Pirellula_staleyi_DSM_6068                              --------------------------PGTLIAGPGIEEILVHGDVLQ---F----K-H-----QD--QR------GIE-S--I-------------PLHTLREVVVLGA-VSLSHRVLSAIQQNAISLLLL--DES----AN-------R
154174048_Campylobacter_curvus_525-92                             ---------------------MQKSDRTHFILSS--GRLRRQDNNIY---F------DKFDETGG----------VTA-SKIL-------------PINAIDEIYILTR-VELDTYTLAFLADNNILLHVF--SPF-Q--SF-------R
292492938_Nitrosococcus_halophilus_Nc4                            -------------------------METLFVSRE--ARLRRRENTLS---I----T-I-----EG----------KTR-A--F-------------PIETIGHLVLLAE-STLNSKLLTLCGKHGVRLSVF--DYY-G--YF-------K
111224947_Frankia_alni_ACN14a                                     ---------------------MPAAGRTYWLTEP--CRIRREDNSLR---I----ERA-----DS----------TPV-R--L-------------PITDIRDLVTFDH-VDINTAAVSLLGRHGVVVHIL--DHY-G--NY-------A
385809790_Ignavibacterium_album_JCM_16511                         --------------------RGGKMKNNIYIFSN--SILSRKDDSLV---L------DSIPQTDAIHETDFYETEGII-L--PAPKTDGSGNPRHIPAESVEAIYTFGE-VRFNTQFFKCVGFYEIPVHIF--NYY-G--NY-------V
150021526_Thermosipho_melanesiensis_BI429                         -------------------------METLYIFTD--GELRRKNNALY---L----K-P-----KD--N-------NMK-PLFV-------------PLKSVSSLMIFSE-VSMNKRFLDLLSKEGIPAFFY--GYY-E--NF-------I
28211139_Clostridium_tetani_E88                                   ------------------------MGSTRYITSV--GELKRKDNSLC---F----R-K-----NN----------KNV-Y--I-------------PVENTKEIYCMSE-VNINSKLLDFLSQNNIIMHFF--NYY-E--GY-------S
48477075_Picrophilus_torridus_DSM_9790                            -------------------------MQTYYIISS--GTLNREMDSLR---F----S-N-----SN----------ENK-I--I-------------PLENIESIIISGN-VSITKPAISILSKKNIPVFFM--SMY-D--NY-------I
41614813_Nanoarchaeum_equitans_Kin4-M                             -------------------------MKTIYILSI-GKLYRGKNGLYF---I----N-K-----DK----------KKS-P--I-------------PLESIKEIFILNK-VSVTYNALKLLADRNIPIHFFYENTKKGISYY-------L
55376278_Haloarcula_marismortui_ATCC_43049                        ------------------------MNDNYHVFTD-GRLERHNDTIRL---V----T-E-----DG----------EKK-Y--L-------------PIENAEALFLHGQ-IDYNTRLVSFLNDVGVAIHVF--GWN-D--YY-------A
389844445_Mesotoga_prima_MesG1-Ag-4-2                             ------------------------MKKNFYLFAS--GQLSRKDNTLL---L----E-R-----EG----------RRK-F--I-------------PIEKIESISTFGE-IDLNTHLLSFLSKKEVALHVF--SHN-G--WY-------S
390962116_Thermococcus_sp-_CL1                                    ------------------------MKYPLFITDH--GRLERQSGSLV---F------------VG----------RLE-KRTI-------------PLAQVSEVHCLAR-VSLTSGAVELLSERRIPVHFY--STR-G--DY-------R
13540937_Thermoplasma_volcanium_GSS1                              -------------------------MYNFYITQD-GTITREGNTLYF---I----G-P-----DF----------KRH----L-------------PVMNVQEIIISAK-VSLSSWALDYLSKLGIIVHII--STN-G--KY-------M
289596694_Aciduliprofundum_boonei_T469                            -------------------------MDSLYITKE--AIIKREANTIY---L----V-R-----KG----------EKR-S--I-------------PIHNLRDITCIAP-VSFSSGAIKHVLNSGVVVHFF--DMY-G--NY-------E
14590110_Pyrococcus_horikoshii_OT3                                ------------------------MKSPIYITQP-GILERKANTLFF---V----N-E-----EM----------KRA----L-------------PINTISEIHCFAP-VTLTSGAIKLLSDNDVPVHFY--NKY-G--YY-------R
84489236_Methanosphaera_stadtmanae_DSM_3091                       --------------------GCDTIKEPLYINSN-GILYRKENTLKF--------K-N-----KE--V-------DTS----I-------------PIHAINEINCYGK-VSLRSGAISLLQKEKIIINFF--NKY-G--YY-------E
268317185_Rhodothermus_marinus_DSM_4252                           ------------------------MKRPYYIFSS-GRLRRRQNTLFF---EKAAGE-RIPDDQDETGVPSGTPTGEST-P--F-------------PVEQVESLYFFGE-VDLNSKLLTFLARHDIPAHFY--DYY-G--NY-------T
18977490_Pyrococcus_furiosus_DSM_3638                             -----------------------MRKKSLTIFSD-GTLLRRENTLYF--------E-N-----VN--G-------RKP----L-------------AIEGIYDIYIYGH-VNITSQALHYIAQKGILIHFF--NHY-G--YY-------D
427708213_Nostoc_sp-_PCC_7107                                     -------------------------MSTIYITEQ-DVTFQIQHQYLK---V----F-Y-----QQ--T-------QRI-C--I-------------PIRNLSQFIIFGN-INLPKEVIQIVRSHNIPVLYL--TQT-G--EY-------L
119720226_Thermofilum_pendens_Hrk_5                               -------------------------MRLLVLRGVDEVTVSSRSTVVI---K------------SG--N-------RVF-E--R-------------ALRDVDAVLVVGSGIKISSSLPPVLALHGIPLSIL--AK-----GH-------V
118497969_Francisella_novicida_U112                               --------------------NDIESKNIVFVNIFDGVKLSLSLGNIV---I----K-D-----KE--T-------DEV-KTKL-------------SVHKVLALFIVGN-MTMTSQLLETCKKNAIQLVFM--KNS-F--RP-------Y
332661941_Haliscomenobacter_hydrossis_DSM_1100                    --------------------------MEIFIQSE-NVYLDYHEGTIL---V----V-N-----GE----------KRT-P--V-------------AFKQIERIWIHPS-TSIAAQLLVKLLEVQIELVIC--SGE-G--SP-------L
326793969_Marinomonas_mediterranea_MMB-1                          --------------------EREKLGTLLIIAGD-IAMLSSEKQRLI---V----E-Q-----YD----------ELH-T--Y-------------PWATLSSVLLVGP-HHITTPALKSAMFHNVPVHFA--SQY-G--RY-------Q
220934970_Thioalkalivibrio_sulfidophilus_HL-EbGr7                 --------------------QAAAEGTQLLLAGE-PARVGLDGGRLR---V----Q-------RG----------EAE-LLSA-------------PLETLAGVTLLGP-HQVSTQLLGALLDRGIPLALA--TGQ-G--RL-------R
288947693_Allochromatium_vinosum_DSM_180                          -----------------------MNAQLLLVIDHRDTQLTLDGGALR---L----V-T-----PG--T-------KPR-H--I-------------PLGVLGLVVVHGR-ALVGCDVWRALAERGIPAVMQ--PGR-G-----------R
431929780_Thioflavicoccus_mobilis_8321                            --------------------------MILVVDRR-DSVLRHDSGVLC---L----E-R-----EG--G-------ETR-R--A-------------PINQLELVVVYGN-PLAETAVWRKLAAAGVPTVLL--PVR-G--RD-------G
206890539_Thermodesulfovibrio_yellowstonii_DSM_11347              -------------------------MSTVFIDRK-DIEIRVDGNSIS---F----YAK-----GK----------KDG-S--L-------------PLSPLKRVVIVGN-VKIETSVLYKLVNHGITVLFL--TGK-L--KY-------S
52425690_Mannheimia_succiniciproducens_MBEL55E                    -------------------------MPTLYIDRR-TTELKVNGDVLI---C----Y-E-----KG--E-------RIA-T--I-------------PLASVDRLYMKGD-INLQISLLSKLGEKGIGVVFL--QGR-KNKPM--------
30248142_Nitrosomonas_europaea_ATCC_19718                         -------------------------MTSLFVDRR-GVELGLESGAIV---F----R-E-----NG--E-------RIG-T--V-------------PIAPLTRVFLRGD-VNLPAALLGKLGERGVGVVIL--SGRTS--RP--------
121997419_Halorhodospira_halophila_SL1                            -------------------------MGTLYIDRR-RTRLELAHKALT---I----R-E-----PE--A-------QPR-S--V-------------PLSLIDRLIVIGQ-VELSSGVLTTLAESGVSLVFM--PSR-G--QR-------R
430760780_Thioalkalivibrio_nitratireducens_DSM_14787              -------------------------MATLFLDRA-QLELRTDGNVLA---V----Y-E-----GG--R-------RRG-T--V-------------PLSLLERVVVQGRQTRFESGVLLKLAESGTATVFM--GAR-S-SRQ-------V
```

```
292492856_Nitrosococcus_halophilus_Nc4                            -----------------------MEMTTLYLDRK-GLELVRDGGRMM---L----Y-------PA----------GIA-GSGI-------------PLTGISRVVIRGD-VTLSSGLLGVLAEQGVGVLML--SGRHG--RR-------T
34541577_Porphyromonas_gingivalis_W83                             ------------REQQKRQLQERGEIFDLVVTEP-GAFIGISRNHVL---V----R-K-----YG--K-------TIC-K--Q-------------PAAQIEQISIISDGVSLSSNVTKYCRKKNIRVIFY--NAT-G--QA-------Y
88602029_Methanospirillum_hungatei_JF-1                           -------------------------MNSVLITGA-GYRIRKRGDVLT---I----E-TGKDSDTA--E-------PPR-T--L-------------SPLGLDLLAIAGD-HSISTAAVRLVTSHGGAIALM--DGL-G--NP-------F
325980970_Nitrosomonas_sp-_AL212                                  --------------------AEAKEKDVLTLAGN-GCSLQVEKGQLV---LRSGYS-CSTVTERE--K-------RIT-R----------------GVHAVRAIIVINTTGNLSTAAIQWCSDQRIALYVL--DRD-A--HL-------T
269798874_Veillonella_parvula_DSM_2008                            ------------------------MLETLYVMTP-GISIRQDGGLLV---L----E-K-----DH--S-------VIK-E--I-------------PIATVSNLVLGRT-IQISTQVMFSLVKQGSLIQFV--DHK-Y--QL-------V
332668016_Haliscomenobacter_hydrossis_DSM_1100                    --------------------------MQLVLDTH-GLQLTKKGGVFE---V----I-G-----EK----------SKR-E--I-------------SPAKLSSIAVTAT-VILHSDAVVLAIKNQIPILFF--DRI-G--KV-------R
379728404_Saprospira_grandis_str-_Lewin                           --------------------------------------------MFW---F----K-P-----KH----------SEG-R--A-------I-----ACQRINCIFLTKG-VRISSDAILLALSHNIPMVFT--DQM-G--RS-------Q
327314603_Prevotella_denticola_F0289                              --------------------------MELILNTY-GVSLNRDNEGFV---I----S-T-----SD----------GKQ-R--I-------------PAVGIDSIQISRG-AQITSDAVMLAVEHEIEVMFM--DRT-G--KP-------I
392395717_Flexibacter_litoralis_DSM_6794                          --------------------HLIIKKMQIVIHTK-GTYLKRSQNMFE---I----K-T-----EN----------DIQ-R--I-------------SAEDIDTIFLAKG-TLLTTDVILFALSHQIEIIFT--DYH-G--TA-------K
150398872_Methanococcus_vannielii_SB                              --------------------------MKCVVNKY-GSKISKSGNRIV---I----E-S-----KD----------GVF-E--S-------------SIENIEQVLICAP-SSITTEFLELSVKNNVDVVLL--NKY-G--KP-------V
150401497_Methanococcus_aeolicus_Nankai-3                         --------------------------MDLLLKTN-GFFISKKEDRFR---V--ELE-K-----DG----------KII-KTDI-------------PATKIERIILSNN-GTITTGAISLAIENNIPIIFL--KWE----EP-------T
284040851_Spirosoma_linguale_DSM_74                               --------------------------MQLHISTY-GTYLHVKDAMFD---V----R-R-----KG----------EDG-K--V-------ISATY-SAEKVTHILLATG-TSLSTDAVRLAMRHNVDIVFI--EQQ-G--DP-------I
257125862_Leptotrichia_buccalis_C-1013-b                          --------------------------MELYITDL-GTVVKKRDDLFE---I----T-T-----SE----------KKV-A--V-------------APQKIKSLVLSKG-IFLTTDVIKLAVDNNIDIVIV--DDF-G--NP-------Y
21226661_Methanosarcina_mazei_Go1                                 --------------------------MQLVINTR-GSFLKKQNNCFL---V----K-V-----DD----------KTF-E--V-------------SEKKVDSILITTS-ATITTDAIKFAVENNIDIVFL--DNF-G--DP-------Y
218961577_Candidatus_Cloacamonas_acidaminovorans_str-_Evry        -------------------------MMEVHLNTY-GSYLRKKDEMFE---L----S-I-----ED----------RKT-K--L-------------SPEKISSIVISNA-ATITTDAIQLAMDYNIDIVFL--DKY-G--SP-------Y
319790504_Thermovibrio_ammonificans_HB-1                          ------------------------MKRPIVVASY-GAEVSKRANRLL---V----S-V-----NG----------RNK-P--V-------------PLGVIELLILFPG-VSITSSAIRFLSHHNRPVIFV--NGF-G--KV-------V
337288662_Thermodesulfobacterium_geofontis_OPF15                  ------------------------MKQTIIIYSN-YVEVSRKGGRLY---I------------ET----------PKT-KCSI-------------PLSNIDGILIFGK-SYLTSEAISLCMKKQIPIILF--TIY-G--KI-------K
20094748_Methanopyrus_kandleri_AV19                               --------------------MGAPNDTVIFLRR---GKIERREDAFR---I------------------------GKS-K--Y-------------SAVRTTGIIIAGG-AQITTQAVRLALRNEVPIVYL--GGN----RI-------L
386750121_Helicobacter_cetorum_MIT_00-7128                        --------------------------MSLCIATF-GVNLSLSKDKLI---A----K-K-----QG--K-------IIQ-Q--T-------------PLNKLEQIIIENPSTNLSSLVIYACAKRKIPIIFI--EDL----KP-------L
126465226_Staphylothermus_marinus_F1                              --------------------VFLTDWGHLVVDDY-GVFLRKKHGRIV---V----V-S-----RG----------GRE-E--F-------------PVRRVREVIISGKA-GISTELLKLLADSGIDVLVT--SYT-G--RP-------V
170290046_Candidatus_Korarchaeum_cryptofilum_OPF8                 --------------------METTQYGHLLIDGF-GVSLRKRRGRIL---I----L-S-----KG--E-------KKE----I-------------PMKSVKEVVIIGK-AALSSELLKALAQSGTDLLIA--TPT-G--RP-------V
385805115_Fervidicoccus_fontis_Kam940                             -------------------------MRLLVITGY-GKRLSVEGKVFK---I----E-D-----RN------K---SSS-K--V-------------NPADIESIVIESSGVSITSAAIELASMAGIPIFII--RGR-G--SP--------
15898245_Sulfolobus_solfataricus_P2                               --------------------NLPMDKKIAFVKDY-GAYLKVEKGLIT---C----K-I-----KN----------QVKWS--I-------------APTELHSIVVLTN-SSISSEVVKVANEYGIEIVFF--NKH----EP-------Y
156937921_Ignicoccus_hospitalis_KIN4-I                            -------------------------MNVFVISEP-GKLFVKDGGLAF---A----N-S-----KG----------EVA----Y-------------LANLYDVIILATSKVSVTGAALRAMGRLGVDLVVL--EWN-G--RP-------S
124027534_Hyperthermus_butylicus_DSM_5456                         ------------------------MPEVLLVATP-GTRIYVRRGVVY---A----E-A-----PS----------GEK-A--V-------------VTADTELVVLATGSVSVSGRALRRLAELGVRLVVL--GQR-G--QV-------V
118431380_Aeropyrum_pernix_K1                                     --------------------------MMMVVAEP-GSRIRVARGALV---V----E-T-----KA----------GKK-V--V-------------VESSVERVIISSSRVSISSAAVRAAAKMGIDLVFL--DWD-G--SP-------V
307595807_Vulcanisaeta_distributa_DSM_14429                       ------------------------MANEVFITEP-GTRIVARKGSVI---I----M-G-----RN----------GNR-Q--V-------------LPPNTEEVIIATSRASITTKAIRILAMKGTNLVIL--NPL-G--DP-------V
305663330_Ignisphaera_aggregans_DSM_17230                         -------------------------MKTIFIAEP-GSRLYCRRGGVY---I----A-T-----RD--G-------RKI----A-------------ITPDVDQIIIASSRVGISAKAIRYLAREGIDVVIL--DVN-G--YP-------I
347523053_Pyrolobus_fumarii_1A                                    --------------------GGDGAPRVLVVSAR-GARLRVHRGAIV---V----E-S-----VD------GGGRSRS-V--V-------------PIAEVDRIVIVTSQVSITSSAVRRLARLGIDILFL--DAR-G--NP-------V
302347808_Acidilobus_saccharovorans_345-15                        -------------------------MRALLLTGY-GLLLGASGSSLV---V----R-S-----KD--G-------RRE-E--V-------------PLSQVDLVVVASGGVSVTSRALRLMAAAGVEMVVL--DRG-G--MP-------V
218884093_Desulfurococcus_kamchatkensis_1221n                     -------------------------MKTLIVSGY-GVRLRYRKGLML---I----E-S-----KN----------EKN-E--V-------------PLTDLEQVVIETGGVWFSSKLIRKMVEYGIDFIVL--DHR-G--YP-------A
389861365_Thermogladius_cellulolyticus_1633                       -------------------------MRDLVVSEP-GTVLRFRKGVFV---V----E-K-----KG----------SKT-E--I-------------PPLEVERVFVASSRVGLSSKVLLKAVEYGIDVVLL--DSR-G--VP-------V
15898286_Sulfolobus_solfataricus_P2                               ----------------------MISVRTLVISEY-GAYVYVKKNMLV---I----K-K-----GD----------KKV-E--I-------------SPSEVDEILITVS-CSISTSALSLALTHGISVMFL--NSR-E--TP-------W
15922975_Sulfolobus_tokodaii_str-_7                               --------------------RGVLEMKSLEIADY-GVKLSTKGKTFV---I----SKK-----DG----------KKV-N--V-------------SPAEVDQIVIMTSGVTVTSKAIRLALDHGIDIIFL--DSR-G--NP-------F
18311772_Pyrobaculum_aerophilum_str-_IM2                          --------------------------MQIVVASY-GARIRAKKGLLI---V----EGR-----EG----------RRE----Y-------------PLHQVDEVLLLTGGISISTRALRALLRAGAVVAVF--DQR-G--EP-------L
18311699_Pyrobaculum_aerophilum_str-_IM2                          --------------------------MEVVVKDW-GVSLGYSRGALV---I----K-K-----RG----------GVD-R--I-------------PLFQVDRIWILTGGVSISSRLVRALSHHFIDVVFF--DAK-G--NP-------V
54296139_Legionella_pneumophila_str-_Paris                        -------------------------MKHLTLSEQ-GSFCKLQGERIL---V----F-N-----KD------Q---MVF-E--G-------------TLNRLKTISIVKQGVTLTSNLLVACAMRGIQIFIL--DYS-G--KA-------V
34557933_Wolinella_succinogenes_DSM_1740                          ------------------------MTKHLYIDEY-GSFVGKTSERLT---V----S-Q-----GG------E---LKA-E--I-------------PLSRLKSVNILKGGISLSSDIIEACSLRGIKIFFL--DFR-N--VA-------H
118497353_Francisella_novicida_U112                               -------------------------MKHLIISEY-GIYLGLESGRLV---V----K-N-----KD--D-------KKY----F-------------PLNRLATLSIAKKGVSFSSDLVEQFSLRGIKLFFL--DFR-G--VA-------H
328954440_Desulfobacca_acetoxidans_DSM_11109                      PQAALAQADQAVARKKRKYIKKLAVETEVAVTSP-GTFVGKQGGRLF---I----RRE-----RK----------QVY-S--I-------------PLIKVRGLNVHSSGVTLSSDLIKALSDKKLAIHFC--TPG-G--EP-------Y
408403000_Candidatus_Nitrososphaera_gargensis_Ga9-2               -----------------------------MLSEF-GLFLGRKRNRYV---V----K-------NG------N---AKN-E--I-------------LSTDVEQLHILNPGVSLSTSALRLALSNQTLVVLG--SRN-G--WP-------H
330509017_Methanosaeta_concilii_GP6                               -------------------------MSRIILSDR-GTFLGKSSEQFR---I----T-R-----KD------L---PDV-L--V-------------PARGVEQILVLGSGISVSSDAIQMADELGVEVVFA--SYY-G--KP-------M
148656004_Roseiflexus_sp-_RS-1                                    --------------------------MHLIVDQY-GVFVSKHQGRIR---V----V-K-----DK------E---RLA-E--V-------------PILHLEQIMICGDGIGLSSDVVRVCAEEGIPIHFV--DSI-G--ND-------Y
386854797_Deinococcus_gobiensis_I-0                               --------------------------MNLTISER-GTFLALRSERLQ---L----R-H-----PG------Q---ETQ-E--V-------------ALRDLETVTVTTTACTLSAEAIRACAKFGVQIDLV--DGL-G--AP-------Y
167628642_Heliobacterium_modesticaldum_Ice1                       --------------------ASESEKQHIIISQF-GASLGKKSGRLV---V----R-E-----KG------Q---VVT-E--V-------------PFDRVEQVTVITSGASLSTDAIEECVRHGIEINLL--DFR-G--SP-------Y
148642083_Methanobrevibacter_smithii_ATCC_35061                   --------------------------MKLIIDGY-NKTLRKKDNQIV---V----M-E-----KD--N-------EIY-R--I-------------SANKVSDITIMAK-GHITFDALNLMAKNNIKLISI--NYF-GQINY--------
288560430_Methanobrevibacter_ruminantium_M1                       --------------------------MKLVVDGF-GKSVAKRDNQIV---I----K-E-----NG----------KEINY--Y-------------LAKDISQILLTGK-GSITFDALTLLAENDVDCVSI--NWK-GHVDY--------
11499462_Archaeoglobus_fulgidus_DSM_4304                          --------------------------MRLVVDGF-GKYLGIENGLIV---V----K-E-----KG--K-------ALR-K--V-------------RPEDLKQVLIIGK-AAISSDAIKLLLKNRVDVVFL--DFN-G--EI-------L
383319612_Methanocella_conradii_HZ254                             -------------------------MERLVIDGW-GRYIGVDHEQIV---V----R-E-----RK--DGSYK---EVH-R--C-------------LPSSLRQVVLSGR-GSISIDAIELLADNGVDIVLI--DWR-G--QV-------T
294676823_Rhodobacter_capsulatus_SB_1003                          --------------------GYDPGARVLYLVEP-ERRLSLRAESFA---V----L-D----GGG--S-------LLL-S--L-------------PPARVNRIELGPG-AVVDDPVWRHALATDTTLAFL--DGH-G--KT-------Q
302343124_Desulfarculus_baarsii_DSM_2075                          -----------------------PRLRVMYLSRK-GCRLDVRGRAFV---V----R-S-GPEPDA--P-------ELM-V--V-------------LPSQLDRVELWPG-CDISQKAQRFALECRTPVAYV--DGW-G--RT-------L
389875622_Tistrella_mobilis_KA081020-065                          -----------------------PILRPLYVVEP-GRRLSVRNQAFT---V----E-E---IETG--R-------ELI-A--L-------------PAARVGRIELGHR-VAADARSLRHAAAARIPVALL--DGL-G--GV-------E
374998939_Azospirillum_lipoferum_4B                               --------------------GLDPALRTLHVQTP-RRRLALRNEAFT---V----V-ERDDLGGE--R-------ELI-A--I-------------HHHHLDRIDLGPE-ADADAEALRWALATDTELAFV--DGH-G--AT-------H
288957883_Azospirillum_sp-_B510                                   -----------------------PGLRVLYVTRR-GRRLTRRNQSFA---V----T-E-----DG--A-------ELL-A--I-------------PPARLDRIELGPG-CEADDDVLRHALAHGVDVALV--DGW-G--DS-------V
116627764_Streptococcus_thermophilus_LMD-9                        -------------------------MSDLYSQRS-NYYLSLSEQRII---I----K-N-----DN--KE------IVK-E--V-------------SISLVDNVLLFGN-AQLTTQLIKALSKNKVNVYYF--SNV-G--QF-------I
15609954_Mycobacterium_tuberculosis_H37Rv                         -------------------------MVQLYVSDS-VSRISFADGRVI---V----W-S-----EE--L-------GES-Q--Y-------------PIETLDGITLFGR-PTMTTPFIVEMLKRERDIQLF--TTD-G--HY-------Q
348026605_Megasphaera_elsdenii_DSM_20460                          -------------------------MSFAYITEK-GATISKKDGRFV---V----GRN-----HE----------TLL-E--I-------------PEETLEGLLVTDT-VQLTSHAIVSLLHLGIPVTWL--SSH-G--KY-------F
296133514_Thermincola_potens_JR                                   -------------------------MSFLYVCEP-DTRVRIKDGRIT---A---E--Q-----KD--G-------MEV-S--I-------------PLELLEGVVLMGS-AQMTAACSVELLEKGIPVTFL--SRS-G--FF-------Y
114567264_Syntrophomonas_wolfei_subsp-_wolfei_str-_Goettingen     -------------------------MSFLYVYER-SAKIGVQENCVV---V---ES-K-----KE--N-------LKR-I--L-------------PIEGVENVIIFGD-ASLSSNCVKQFMERDINLTWL--SSR-G--KF-------Y
332296542_Thermodesulfobium_narugense_DSM_14796                   --------------------------MILYIVEH-GTRIHKDLDRFD---I----F-R-----GS--E-------LIR-S--Y-------------PASQIESIVIMAN-ASFTPHAINYILKNNIDAVFL--SSK-G--LY-------R
218442809_Cyanothece_sp-_PCC_7424                                 -------------------------MATLYLMEQ-GTWVQKEQERLI---I----Q-V-----SK--T-------QKM-E--V-------------LMREVERIMIFGN-VQLSTPAINACLKHNILVLFL--NQA-G--QY-------N
17228961_Nostoc_sp-_PCC_7120                                      --------------------YWQETMTTLYITDQ-GAYLSVKNQQFQ---V----F-H-----QG--E-------LRI-K--V-------------PVMRVSNVVLFGC-CNVSHGAVSMALRRRIPIMYL--SQK-G--RY-------F
119357853_Chlorobium_phaeobacteroides_DSM_266                     --------------------EASLFLRTLYIQEQ-GSMLRVENGCFR---VTCGHD-D-----DV--A-------ELL-E--V-------------QSIKVGQIVLFGA-CMITPAAIRHCLMNRIPVVLL--SQH-G--EY-------F
119357846_Chlorobium_phaeobacteroides_DSM_266                     --------------------LYSPFLNTLYLQEQ-GSLMRKDGERFS---I----E-K-----DG--S-------VIN-E--V-------------IVRRIEQVVIFGN-VALTTPVMQYCLQNEIPVTFL--SQH-G--KY-------F
336116789_Microlunatus_phosphovorus_NM-1                          --------------------VAVPDRRVLYAGLQ-GGRVRVKQGRVI---V----E-T-----AD--D------VVAL-D--V-------------PTDHVSRVVCFGS-VGFSAGARSWAMSKDVDVVFV--SRR-G--TY-------Q
37677204_Vibrio_vulnificus_YJ016                                  ----------------------IARLKLVYIHKQ-GAAVHKRGNRIN---V----FFD-----GK----------PLN-S--I-------------PANTIDMMLCFGA-VHLTQAVQRFCLENDISVVWL--SQS-G--QY-------L
154250301_Fervidobacterium_nodosum_Rt17-B1                        --------------------------MTIYVLEQ-GTVLAKKDGRMI---I----T-K-----AK--Q-------VLD-E--I-------------PLKKIERINLLGN-ITLTSQMINYCLDNKIEVIFM--TQH-G--RY-------R
38505760_Synechocystis_sp-_PCC_6803                               -------------------------MKTLYVSQS-DCYVSLRKEFLL---V----KRQ-----DR----------ILA-E--V-------------QLPFLEQVLIFGK-AQVTTQAIHACLSRNIPILYL--SRM-G--FC-------Y
83592490_Rhodospirillum_rubrum_ATCC_11170                         --------------------PRSSPMTTLYVTQP-GSVVRSEGGSLT---V--WVE-TEADDPGP--NDSPVRRKRLA-S--V-------------EPHRLESLVLLGF-TTITANAMRLCMANKIAVSLL--DGG-G--GL-------A
261415948_Fibrobacter_succinogenes_subsp-_succinogenes_S85        --------------------------MHLYVTEQ-GTRLGKNGGHLV---V----Q-R-----DG--C-------TID-D--I-------------LLSEVDSLSLFGA-VHPTTDAMLALLDKGADIAFL--SSG-G--RY-------R
76802276_Natronomonas_pharaonis_DSM_2160                          --------------------EGMFDESVVYVTKQ-GTQVRTDGGQII---V----W-D--VDGDD--G-------ELA-A--Y-------------PKEKLDTINVFGG-VNFSTPFVAEANEHGIVLNYF--TQN-G--KY-------R
312114617_Rhodomicrobium_vannielii_ATCC_17100                     --------------------RFSPAAAPVHVLSG-AAVVRVNNSTLL---V----E-R-----PG--E-------PVF-E--R-------------PIELVSTLHIHGW-ARVTGACIGRLTAQGATVVWR--GLH-G--YP-------V
219683006_Bifidobacterium_animalis_subsp-_lactis_AD011            --------------------VKVPDHAVTHLATP-GARAYLKSGRMH---V----S-K-----NG--D-------EIT-S--V-------------PLDSIQALQIHGN-VDVSSGLMRELMWRNIPILWC--SGT-G--RL-------M
116748794_Syntrophobacter_fumaroxidans_MPOB                       -------------------------MERTYILEQ-GAYLRKAGNHLV---V----T-K-----NR--E-------IIA-E--I-------------PLEGLSQLTLVGF-SSLSGAVLEVLIRHRIETVLL--SPR-G--QF-------R
317051217_Desulfurispirillum_indicum_S5                           -------------------------MATVVVDEQ-GSHIQYKTNQLI---I----T-S-----KD----------DNL-SWSI-------------PKMNIERLLIVGR-TTLAQSAMDFLLDASIPTHFA--TTS-G--CY-------K
328949009_Treponema_succinifaciens_DSM_2489                       -------------------------MSEIYILSD-YGKLSKKDETIV---F----S-Q-----PD--G-------TST-I--L-------------FPFKTEHLVLVGK-VSISADALRLLTKYKIGTTFL--SSN-G--RF-------N
333996012_Treponema_azotonutricium_ZAS-9                          -------------------------MPVVYVISD-HGRLHKANEIFE---F----T-D-----AN--G-------DTK-K--I-------------FPHTMESLIITGM-VTITGEAMRLLMQHQINTLFV--SPN-G--KY-------N
313673559_Calditerrivibrio_nitroreducens_DSM_19672                --------------------------MVVYINEQ-GAVINISGERII---L----K-K-----AG--Q-------IIA-V--F-------------LTKDLDQVVLMGN-IGLTTQSIRHFLQNKIDVVFT--SAS-G--KY-------E
218961167_Candidatus_Cloacamonas_acidaminovorans_str-_Evry        -------------------------MPVVYIADQ-GSHICKKGDRLY---V----Y-R-----GN--Q-------LLR-W--F-------------HTKDIIQLILVGN-IGLTSQVVTYLLKNRIDTVFL--SYY-G--KF-------K
294828260_Leptospira_interrogans_serovar_Lai_str-_56601           -----------------------REKTTLHVFGH-DSRIKKSDNVLL---V----E-KVTETGEK--S-------KSE-K--I-------------PIQEIESVNIHGN-CQISSQMIKFLVSEEIPVHWF--SGG-G--NY-------I
347756303_Candidatus_Chloracidobacterium_thermophilum_B           --------------------PERDERRILHITEP-GTRVERAAEELV---V----E-I-----PG--R-------VPV-R--F-------------PGRDVGAVVLHGA-VQITTQAIHYCAANDVAIHWL--TGG-G--NV-------V
162457462_Sorangium_cellulosum_So_ce56                            --------------------PPDTERRSLHVVSH-GARIGRASDAFE---V----TER-----EG----------ERT-R--I-------------GAREVSDIVVHGH-AQITTQALRLCAAEEIAVHFV--GAG-G--AH-------V
86742032_Frankia_sp-_CcI3                                         ---------------------MAELLNTLYATTP-GTSLHLDGDAVR---I----W-H-----PD--N-------DKG-RRLL-------------PLVRVDHIVVFGG-VTITDDLLQRCATDRRSVTWL--TGN-G--RF-------R
433631663_Mycobacterium_canettii_CIPT_140070010                   --------------------LIADSRRSLYLTTV-GTFVALDHDAFE---V----R-R-----PE--L-------PAV-R--A-------------PVRSIESIVCFGN-VTLSTSAMARCAQDGIDVSWL--THG-G--RF-------R
15612904_Bacillus_halodurans_C-125                                ---------------------MKKLLNTLYVTQP-DTYLSLDGDNVV---L----L-K-----EQ--E-------KLG-R--L-------------PLHNLEAIVGFGY-TGASPALMGYCAERNISITFL--TKN-G--RF-------L
108763792_Myxococcus_xanthus_DK_1622                              ---------------------MTTALNTLFITAE-GTRLNKEGECVV---V----T-V-----QD--Q-------KKA-E--V-------------PLRHLRSVVCLTR-AWLTPELMESCLEAGIHVSFF--GMT-G--RF-------L
344996859_Caldicellulosiruptor_lactoaceticus_6A                   -------------------------MQTLFIDTQ-EGQLGKKDERLI---F--KNY-S-----TG--E--------ER-D--I-------------KILDIHDIVVLGR-ATITTPAIHLLMEKKVPVHYV--GLD-G--KY-------K
85858452_Syntrophus_aciditrophicus_SB                             --------------------------MVVYVRTQ-GARIIKEGRHLL---V----R-K-----GD--A-------VYH-T--L-------------FTYKLDQLVIFGN-VEITHQALAQLMRYGIDVVFL--SFR-G--RY-------L
148657121_Roseiflexus_sp-_RS-1                                    -------------------------MPTLYIQEQ-GVMVRKRDNQVL---I----T-K-----DG--Q-------TLS-E--V-------------PLAKIDQVVLMGRGVQLSTALLIDLLERGIPVTFT--NQH-G-SRH-------Y
159898907_Herpetosiphon_aurantiacus_DSM_785                       -------------------------MPTLYLNEQ-GTRLGKKDERLI---I----L-R-----GQ--E-------LIN-D--I-------------PVIKVDRVIVMGQGVQVSHAAIVFLAQRGIPLIFT--TQS-G-GSQ-------K
17227877_Nostoc_sp-_PCC_7120                                      -------------------------MTTLYLTEP-GTILRYRNESLI---I------M-----KQ--E-------KSH-N--C-------------RLAEITLIVVLPG-VQLTDVVISQLLDRGIETIFL--RQD-G--QF-------R
268318378_Rhodothermus_marinus_DSM_4252                           --------------------PPVRQARTLYVDEI-GAVVRRKGRQLV---V----T-VSRDGRRQ----------ELL-R--V-------------PALLVDQVVLVGP-VQITSQALRMLLRRNVDIVYL--SGE-G--RF-------E
55978328_Thermus_thermophilus_HB8                                 --------------------------MTLHLTRQ-GATLRLRQGRLL---L----E-E-----EG--R-------EVA-G--F-------------PARQVRSVALWGN-VRLSTPALVFLLRQGVPVFFY--SLE-G--FL-------H
297622675_Truepera_radiovictrix_DSM_17093                         -------------------------MAVLHLLEQ-GATARLRAGRIC---V----E-L-----DG--H-------ALG-E--V-------------PARKVRQVALHGN-VRLTTPALQFFLGNGVPVLYV--SLK-G--AL-------H
328953000_Desulfobacca_acetoxidans_DSM_11109                      -------------------------MAFLYLSEQ-GACLQKTGERLV---V----AKE-----GE----------TLL-D--L-------------PVGKVEAVLIFGN-VQFTTQAALLLLQQGVEMALF--TRR-G--RL-------V
```

```
337286709_Thermodesulfatator_indicus_DSM_15286                    --------------------------MALYITEQ-GLKVRKEGQRLQ---F----Y-K-----EK--N-------VVR-E--F-------------RLDDLDEIYVFGR-LNFSAAALQALLKHEIKVHFL--TAS-G--KY-------L
328953422_Desulfobacca_acetoxidans_DSM_11109                      -------------------------MSILYLTDQ-GATLTKKGNRLV---I----E-K-----KG--N-------LVH-W--V-------------HAFKVEQVVVMGQ-VQLSPGAIAFLLQEGIDTVFL--SYY-G--KY-------R
38505684_Synechocystis_sp-_PCC_6803                               -------------------------MSTLYLTQP-DAVLSKKHEAFH---V-ALKQ-E-----DG----------SWK-KQPI-------------PAQTLEDIVLLGY-PSITGEALGYALELGLPVHYL--TQF-G--KY-------V
134299482_Desulfotomaculum_reducens_MI-1                          --------------------PGINLGRVLYVDEQ-GASLYKKGERVL---V----T-------KD----------QIK-FKDI-------------PLCNLDQVVLVGN-VNLSSQLIKLFLGRGTEVHFI--STK-G--KY-------Y
17229060_Nostoc_sp-_PCC_7120                                      -------------------------MGTVYVTQD-DSFISKVDERLN---V----K-F-----EK--K-------VIL-D--V-------------PLIKVDGLVVMGR-ASISPAAIFELIDKKIPLTFL--TNN-G--KY-------L
302391602_Acetohalobium_arabaticum_DSM_5501                       -------------------------MPTVYVTQE-DTALRKKGERLI---V----K-Q-----GK--E-------KIA-D--V-------------PLVKVDQVVVFGN-ASISGSLISLLLKAEVPIAYL--SYY-G--KY-------K
159898755_Herpetosiphon_aurantiacus_DSM_785                       -------------------------MQTLYLSEQ-YSIVKREGEALR---V----E-I-----PE--DQQLGRQRQVV-R--V-------------PLNVIERVVVQGE-ITLTASALACLLERRICTHFL--SYS-G--RS-------Q
320161859_Anaerolinea_thermophila_UNI-1                           --------------------LGGQNMPPLYVVQQ-NSKLRLNNRRVQ---V----E-QETDEGIQ----------VLA-Q--I-------------PIGQVSEIILFGN-VGLTTPLMDALLYEGIPVIFL--TRD-G--DY-------R
156741961_Roseiflexus_castenholzii_DSM_13941                      -------------------------MDTLYVTEQ-GSEIGCDGERLA---V----R-R-----DN--A-------IIA-S--I-------------PLIKIEDIVIIGN-VGLSTPAIKRMLDNGINVTFL--TVH-G--RY-------Q
383762167_Caldilinea_aerophila_DSM_14535_-_NBRC_104270            ------------------------MRAPLYLLEQ-GSKLSCQSRLLV---V----T-K-----DG--E-------TLA-R--L-------------PMIQVSQVLVFGN-VEITTPALKRLLDEGIEVVLL--SQR-G--RF-------Y
46255267_Thermus_thermophilus_HB27                                -------------------------MGVVYVLEN-EAYLSKEGGTLK---V--SRR-A-----GR--E-------VLL-Q--K-------------PLIAVEEIVILGN-AVVTPALLKHCAQEGVGIHYL--SPT-G--TY-------Y
134101617_Saccharopolyspora_erythraea_NRRL_2338                   -------------------ARAPARDPVYYVTEP-GTTVGIRSERLA---V----R-K-----DH--E-------ELL-S--C-------------RLRDVLHLVAAGP-VQVTSQAVHALAEQGSPVVWT--STT-G--RL-------K
378719041_Gordonia_polyisoprenivorans_VH2                         --------------------PRDPDSRPLYVTTH-GAAVGLKGGRIR---V----V-Q-----KG--S-------KPA-E--I-------------LLIDVSQLCVFGN-VQVSSQVLAELWARGIPVLWF--SYN-G--WL-------R
320102061_Isosphaera_pallida_ATCC_43644                           --------------------TPRDDLKSLYLNTQ-GTFVSKSGAVLK---I----R-K-----DD--Q-------ILQ-E--V-------------RIQETCQVSVFGN-VQLSSQAIQTLCEHEVPIAYF--SQG-G--WF-------Y
268318441_Rhodothermus_marinus_DSM_4252                           --------------------PSRNDRLPLYVLSQ-GSVVKRKGAQLL---V----Q-T-----QD--D-------KDQ-H--F-------------RLIDLSRVSIFGN-VQITTQAIRALVEHNIPVFFH--SYA-G--RM-------L
39995168_Geobacter_sulfurreducens_PCA                             --------------------PADGRGLPLYVQSP-KAYVRKDGDCLV---I----E-E-----ER--V-------RVA-E--A-------------RLGETSQVALFGN-ATLTTAALHECLRREIPVTWL--SYG-G--WF-------M
83591519_Rhodospirillum_rubrum_ATCC_11170                         --------------------PPASPALPLYGQTP-GARIGKKDYTLV---I----Q-V-----EG--E-------ADR-S--L-------------ALDEISEVVLAGP-VSLTTPAIHELLRREIPVAWM--SS--G-FWF-------L
50085564_Acinetobacter_sp-_ADP1                                   --------------------LRQERSRPVMLSKR-ACVFYLEKVRVV---L------------KD--D-------RIV---YLTDNSQPVEYFYNIPEKNTAFLLLGKG-SSLTDAAARRLAESNVMVGFC--GSG-G--SP-LFSSLDL
77166205_Nitrosococcus_oceani_ATCC_19707                          ---------------------LPSHRQGLYYLEH--CRVMAKDERVV---Y----ACQ-----EG----------AFTKFFAI-------------PPANTNVILLGSG-TSLTQAAARLLASEQVMVAFV--GGG-G--SP--------
15602176_Pasteurella_multocida_subsp-_multocida_str-_Pm70         ------------IPSSDLKTIMHSKRANIYYLEH--CRVLLNGGRVE--YV----T-D-----EG--R-------ESL-YWNI-------------PIANTSCLLLGSG-TSITQAAMRELSKAGVMVGFC--SGG-G-TPLFNGTEAEI
38234755_Corynebacterium_diphtheriae_NCTC_13129                   ----------------PASHQIRLEDRLSFLYLE-YCLIRQDRTGVI---AIQNGHNESSDEEEKLEP-------QIL-RIQL-------------PVASLAVLCLGPG-TSISNAAMTSCTRSGCTVIFT--GGG-G-----------V
21674788_Chlorobium_tepidum_TLS                                   ------------------LPQVKEKYPFLYLEKG---RIEIDDSSIKWIDC------------DC----------NVV-R--L-------------PVAMLNCILLGPG-TTVTHEAVKVMAAANCGICWV--GDD-SLMFY-------A
53804737_Methylococcus_capsulatus_str-_Bath                       -----------LGLAESR-IPHADRHGLLWLTFG---NLTVEDGTLH---F----R-AAPSEWMD--AG------DYA----I-------------PYQGLSMILLAPG-TTVSHDALRLLARHGTLLAAV--GDG-G-VKF-------Y
209966592_Rhodospirillum_centenum_SW                              ------------LGLDAARIPQKSRNGLIYVERC---RLSIDNGSLV---I--AFD-D-----RG--E-------ELE----L-------------PYQRLNAVLLGPG-SSITHDAVRHCSGHGTCLAFV--GSD-G-TRL--------
89885834_Rhodoferax_ferrireducens_T118                            --------------RVPAPVRMAHKNRIPYLFLE-KGILRVDGHCLL---L----C-Q-----AE----------SAI-E--I-------------PGSMVSCLMIEPG-VSVTHEAMKLCGENGTLLMWV--GEG-GTRFY--------
16130662_Escherichia_coli_str-_K-12_substr-_MG1655                ---------------LPLNPIPLKDRVSMIFLQY-GQIDVIDGAFVL---I----D-K-----TG----------IRT-H--I-------------PVGSVACIMLEPG-TRVSHAAVRLAAQVGTLLVWV--GEA-G-----------V
68535718_Corynebacterium_jeikeium_K411                            --------------------LPRVQDRISFLYAE-RCVVNRDNNAVT---F----A-D-----ER----------GIT-H--V-------------PAASIAALLLGPG-TKVTYAAMALLGDSGVSVVWV--GEK-G-VRY-------Y
86738746_Frankia_sp-_CcI3                                         ------------------LHRVEDRVSTLYVEKC---HVDRDDNAVV---L----V-N-----KE--R-------TVR----V-------------PAAFVATVLLGPG-TRITSAAVRLLADSGTALCWV--GDR-G-VRM-------Y
51891801_Symbiobacterium_thermophilum_IAM_14863                   ------------------LPKVRDSWSYLYVEHA---RIDQEDKGIC---I----HDA-----KG----------RVH----V-------------PCASLSLLMLGPG-TSVTHAAVMAMVEAGCLVAWC--GEQ-G-VRF-------Y
29834079_Streptomyces_avermitilis_MA-4680                         ------------------LPRIADSLSFLYLDIV---RIHQDDTGVC---A----EVTSERRGTE----------KVY----V-------------PTAALGCVLLGPG-TSITARALATFARHGTTVLIT--GAG-G-VRC-------Y
```

Continued

```
407462288_Candidatus_Nitrosopumilus_koreensis_AR1                 T--FCNGMMD-SLTATKYRMA-QYDTFRNPEKCEYLRKQIISAKKESQLKLL-----------------------------------------------RLIGS-----EI------------------------------SELPDSEHI
21229351_Methanosarcina_mazei_Go1                                 T--TMLP--P-ESTNLRTKFA-QYHAFEDKEARLEIAKKFIEAKFYKSKAVL---DFLS------QRYPE--------INFDILDGL------------TKLKD-----VK------------ST----------------REILGVEGT
147921327_Methanocella_arvoryzae_MRE50                            G--TWMP--K-ETSVGKLRVK-QYAKYLDPTTRYNIAYQIIKEKVKKSCNLL---TELS------DYYEELDKSEIEEAFINEYTGF------------VAYNK-------------------KEQHDI------------NKIMVYEAK
124485266_Methanocorpusculum_labreanum_Z                          G--GIYG-----KGAPSLASQ-QRDIPIHK-----FAMASIRSSLDERLRYI---NELA------EFDPEGL------YFKGEFDIL------------TAARE-----EL------------EY------------LITLPEIGRAFSL
307352568_Methanoplanus_petrolearius_DSM_11571                    G--VLRP----FGRKPDE----EMAELQSEVSSYNSAAEIVRSSIKARIMMI---EKAG-----GEIDRELYYKGEDEILFTSLEEI------------DYLIK------M------------------------------EELRRIQKF
154150254_Methanoregula_boonei_6A8                                G--ILKP----WGSTCDT----TIRALQEGAPRDRYATALAQAALKSRLIAI---EHIQ------DQQGSSL------FYEGELQIL------------HNALH-----NL------------EY------------MVKLDEIRRLSDL
219852754_Methanosphaerula_palustris_E1-9c                        G--LIRP----PEDRVDE----QVRAIQRHAPAHSYALGITRAALGRRLQVI---GETT------VVTGESP------LYQGELEVL------------QDARQ-----EL------------EY------------LVTLDEIRRLHRL
126178701_Methanoculleus_marisnigri_JR1                           G--YLYP----YGYRPDE----SVRLAQERAGPHRFAQPLARASLQSRLLLL---EELY------DHAGHDI------FYAGELDFL------------HQARE-----EL------------SASVT---------M---ENLRRLSRL
397780289_Methanoculleus_bourgensis_MS2                           G--YLYP----YGYRPES----GVRTAQEHAAPHRFAQPLATAGLQSKLLLL---EELC------DCTGQDI------FYTGELDFL------------YQARE-----EL------------AVSVT---------M---EGLRRLSRL
156937861_Ignicoccus_hospitalis_KIN4-I                            V--TINP--V-KVGWSRARKN-QYSMGETELG-VKVAKEFIFGKLEGMSNVA---KNLS-------YKGKKPTPNSDYWRSEGRGEL-------------ASCK-----NL------------------------------DCVKKLEAE
70607609_Sulfolobus_acidocaldarius_DSM_639                        ---VLVS--PFLNSLVNVRRK---FYTLSDDMKLYLARKFIKGKILGMINVA---KYFM-------------------YLTKVPVEI------------DLELK-----QI------------DNTKSI------------DELRSVEAE
187736488_Akkermansia_muciniphila_ATCC_BAA-835                    V--LLLP----ADRSTDTGLL-RHLADMPARLRNRLWQKTLDAKCGNQTALA---QAWN-------------------PHHPAIAEL----------------K-----RM------------AV----------------TEKTAREAE
436839745_Desulfovibrio_hydrothermalis_AM13_-_DSM_14728           N--TLYPDSRQYHITSGKHMA-RYESMSDTNKAK-AARKLVSAKINNYRQWL---SSRG-------------------AACDISDKL------------KSIEK-----SV---------Q--DASGN-------------STLMGIEGN
71894593_Mycoplasma_synoviae_53                                   V--QLLS----ISGYYSNKNF-LSQINWTQEFKDKTWEKIIKLKTTNYVNLI---YFFG------------------LLNKEDVEKF------------NFYYK-----KI------------IP----------------GDKNSMEGH
193216853_Mycoplasma_arthritidis_158L3-1                          A--QIIP----FQGYYNAKIL-QKQIKWDNDFKERVWRRIIKLKIRQTMVML---EHLA------------------ILRDDDKVKL------------EDYAN-----SV------------KA----------------FDITNCEGH
187250661_Elusimicrobium_minutum_Pei191                           G--MLLS--S-FTTNIYGRRL-QLQINASKPQIKQAWQQIITSKLNNQAEVL---KRFD-------------------TL-KAAETI------------FNMAR-----EV------------RS----------------GDATFKEGV
325972002_Sphaerochaeta_globus_str-_Buddy                         G--MLLP--M-LQHKDTSQRM-DIQLKASLPRRKQAWQQIITEKLKNQRDFL---QHLQ--------------------I-PGVKRL------------DVLSQ-----TV------------LS----------------GDSSNNEAQ
47458867_Mycoplasma_mobile_163K                                   S--YLIP--I-EGNHNSLKVV-AEQVKWTKEYKAILWKDIIKNKIHNQKNLLIKNKLFN-------------------NSEFGIEYF------------DDLIN-----NV------------NL----------------FDISNREGH
269123825_Streptobacillus_moniliformis_DSM_12112                  A--YSLP--I-NQHYRIPYVH-SLQGQQTSKSKEFVWEKLLKAKIRNSKKVL---ELEN-------------------ASSDLIELM------------QKYED-----EV------------IG----------------SDVQNREGT
57865886_Staphylococcus_epidermidis_RP62A                         T--QLIS--S-YGMVNRLKRI-QSQFQLSGRTKDRIWKKIVINKIFNQTRCL----ENN-------------------LHNENVKLM------------LGFAK-----EV------------SS----------------GDKSNKEAH
319957207_Nitratifractor_salsuginis_DSM_16511                     A--ILHP--Y-HQSSLATETL-KAQLSLPDDLRDRLWQEIIRVKILNQSAVL---SHFA----------------------KKNPKL------------ERYAE-----QI------------RP----------------GDPYRTEAK
310286727_Bifidobacterium_bifidum_S17                             G--ETSP--Y-ETESSRRAYI-IQQIRWDEARKDLLWQHIIREKIAHQVSVL---EKNC--------------------DSLDLEKM------------KSLVV-----DV------------KP----------------GDSDNREAV
90961086_Lactobacillus_salivarius_UCC118                          T--ETVN--Y-YPNNRSLDTY-LQQYEWNDHVKEILWTKIVRSKIINQIKVL---KNYQ----------------------IDCQDL------------KIELD-----KL------------EI----------------NDMTNREAV
406600270_Leuconostoc_gelidum_JB7                                 G--EINQ--Y-HTNGRHNRNI-SEQIIWTLQRKNLLWGEIVKRKIENQMQVL----FAY--------------------QSADREKF------------QALIA-----GI------------KV----------------GDETNREAV
386318631_Staphylococcus_pseudintermedius_ED99                    T--LLLP--L-YGHHRQARLI-QNQMNWKNERKGQLWQLLVKHKILNQKQVI---QECF-------------------KE-RDSELF------------DAFAN-----NV------------EL----------------DDSTNREGH
302336019_Olsenella_uli_DSM_7084                                  G--QYLP--L-YGAHNTSKRI-AEQLEWSEPSKKRVWQHVVKDKISHQARLL---ELRG-------------------RE-GFSDQL------------QGMVD-----EV------------RS----------------GDTTNREAA
15675042_Streptococcus_pyogenes_M1_GAS                            A--MLMP--F-YGRHDSSLQL-GKQMSWSETVKSQVWTTIIAQKILNQSCYL---GACS-------------------YF-EKSQSI------------MDLYH-----GL------------EN----------------FDPSNREGH
352684362_Acidaminococcus_intestini_RyC-MR95                      S--SLLP--L-HARFDSIRKI-KQQMIWSQEIKDEVWDCIVKAKIRQQALFL---DELE--------------------KKEQSKML------------MSYLD-----DV------------VS----------------ADAHNREGH
42525844_Treponema_denticola_ATCC_35405                           S--ELIG--Y-YGSHDTSEKI-RLQIKWDKNIKQLVWTEVVTEKIRQQKYLL---EKLN-------------------L--PQASLL------------AEYIT-----DI------------DI----------------NDKTNKEAH
169823756_Finegoldia_magna_ATCC_29328                             G--EIVN--Y-YNKHNSPLMI-RKQISWQNNSKQKTWREITKEKINRQALLL---KFIG---------------------NENANLL------------LEYKS-----QV------------QD----------------GDVSNREGH
117929157_Acidothermus_cellulolyticus_11B                         A--ATLPWST--HNRVAARHR--AQAELSLPRQKNAWMNIVKTKIRNQAAVL---RALR---------------------RDGVAQL------------ERLAA-----QV------------RS----------------GDASNAEGA
283456134_Bifidobacterium_dentium_Bd1                             A--ALHSWTN-VPSEVAVRQI--AQSEMTLPRKKNAWAKIVKAKIRGQASCL---DSLG-------------------I--EGGGAL------------RGIAS-----SV------------RS----------------GDTSNYEGY
38232679_Corynebacterium_diphtheriae_NCTC_13129                   A--GSFSWNE--HTRIAARQR--AQASLSLPRQKSAWAQIIKAKILGQARTA---AQLG----------------------FDATDL------------KNLAR-----AV------------RS----------------GDVDNREAM
238924076_Eubacterium_rectale_ATCC_33656                          G--FYSA--Y-DNHSRASKVI-GFQIDKEQDYYGKLWQQIVKIKIENQAKAY---QIMT-------------------RDSDGPEKI------------IEFSK-----NI------------LI----------------GDKSNREAH
116627543_Streptococcus_thermophilus_LMD-9                        G--IYHS--Q-NGHFRAYKRL-KEQLDWSQKQKDKAWQIVTYYKINNQEDVL---AMFE-------------------KSLDNIRLL------------SDYKE-----QI------------EP----------------GDRTNREGH
57238545_Campylobacter_jejuni_RM1221                              G--VFTP--F-LGHFQSAKIA-KEQMNVSAQKKAILWQKIIKNKILNQAFIL---KKYN--------------------KIEQSNEL------------INLAK-----KV------------SL----------------NDSKNIEAV
291276264_Helicobacter_mustelae_12198                             G--VFSP--F-LAHYRSSAIL-DLQIHFSKQNKSILWQKIIQAKIQNQTNLL---MQEN-------------------KT-STATKL------------QALQK-----AV------------KL----------------ADGTNNEAK
294086112_Candidatus_Puniceispirillum_marinum_IMCC1322            S--LSLP--L-VQHSDQTRRF-ETQALAKKGIKNRIWQKIIKGKVKNQATLL---VLYQ-------------------SA-S-VERL------------NRLHT-----LI------------KS----------------GDPQNIEAQ
189485226_uncultured_Termite_group_1_bacterium_phylotype_Rs-D17   G--IYVP--F-NQHSRFTQTA-NSQVKWDTAFKNRIWQKIVKQKICNQAEII---KKYS--------------------F-ANYVGL------------KGTCD-----RV------------QS----------------GDKTNCEAF
222109284_Acidovorax_ebreus_TPSY                                  G--VFMP--F-LQHSRATRMQ-RLQLDLDKPSAKRAWAHIVQVKIGNQARCM---ELLG---------------------TVGTDRL------------ASYAR-----RV------------RS----------------GDGGNLEAQ
15602991_Pasteurella_multocida_subsp-_multocida_str-_Pm70         G--QWLP--F-GQYYRQLKTL-KLQLETTLPQKKQLWQKIVQQKIRNQGFVL---EQVG-------------------YF-NETKRL------------NRMAN-----QV------------KS----------------GDKENVEAQ
310780383_Ilyobacter_polytropus_DSM_2926                          G--IFQG--Y-GRHSRQLAVA-KAQINMSVPRKKRLWQEIVTKKLENQGRVL---ELMR--------------------C-EGSKYL------------YDLSK-----RV------------KS----------------GDDGNLEAL
220930481_Clostridium_cellulolyticum_H10                          G--VLLP--F-SCHSRQYKIV-KTQINLSAPFKKRCWQRVVQQKIENQAFCL---NILE-------------------L--KGRDEL------------INLSK-----SV------------LS----------------GDSTNVEAH
34557789_Wolinella_succinogenes_DSM_1740                          G--VFAP--F-GTHSRHTRAA-RVQAVWSEPFKKRCWQKIVRAKVSSQAEVL---RRVG-------------------AE-DAARRL------------ENLVG-----KV------------TS----------------GDTTGVEAQ
90425960_Rhodopseudomonas_palustris_BisB18                        G--VALP--F-HRHHRQGAIA-KLQFDAKDGVKRRLWQAIIRRKILNQAASL---SVLN---------------------RQNSETL------------AEIAR-----HV------------EP----------------GDPENVEAR
209542526_Gluconacetobacter_diazotrophicus_PAl_5                  G--ILLP--F-HQHHAQAHIA-HAQISISQPLKKRLWQTLVVAKIRNQAALL---DQLG---------------------RPQGQTI------------AAMAG-----RV------------AS----------------GDPGNVEAQ
121608212_Verminephrobacter_eiseniae_EF01-2                       G--MLLP--L-DGHHVQAKRI-EAQIAAGLPMHKRLWAAVVKSKLEQQAAAL---EAVA----------------------APTAPL------------QALVA-----KV------------RS----------------GDPENIEGQ
60683388_Bacteroides_fragilis_NCTC_9343                           G--LMLP--L-YGNTTQNERF-RKQLDTSLPLKKQLWQQTIQAKINNQASVL---KDCM-------------------DE-E-VKCM------------RIWAT-----NV------------RS----------------GDPDNLEAR
154250554_Parvibaculum_lavamentivorans_DS-1                       G--MMLP--L-DAHHIQTERH-RAQVEASEPTKKRAWQALIRSKIAQQGIVL---AHFT----------------------GEHGGL------------LPMAR-----RV------------RS----------------GDPDNLEAQ
283778924_Pirellula_staleyi_DSM_6068                              T--AYIDCQN-AEPDSAGIEA-QVDLIRNPESSLAIARQLISAKLHNYATLA---DAYP------PKSHSGN------AH-RSLMRL------------AKDAQ-----QA---------------SS---------L---PELLGIEGS
154174048_Campylobacter_curvus_525-92                             G--NFYPSTSNSVNKSGFALLSQLRAFDDPVKRVYIAREITRAHMLNDAANC---KKHG-------------------VKFDPAPHI------------AALDA-----AA------------DV----------------GQIMAAEGA
292492938_Nitrosococcus_halophilus_Nc4                            G--SFEPFNR--NPSGQVKLA-QAEILLNETQRMALARELVRGAAHNMQANL---LYYR-----YRGKTGLE------AMLKEMKRL------------VNLIP-------------------KAQDS-------------STLMGIEGN
111224947_Frankia_alni_ACN14a                                     G--ALMPAED--MSSAQVVRA-QVDLTADPVRRAAIARSLIAATAANVRWAL---ETDL-------------------LD-GPLDAL------------GTRLA-------------------AAESS-------------EELMGAEGT
385809790_Ignavibacterium_album_JCM_16511                         G--SFLPASA--NGSGEIQIL-QYKAYIDNYKRITIAKSIIDAAIKNILFNL---KSYT------YSGIDLD---------ETINTI------------ASIQE-----QV------------RFTDSI------------QELMGFEGT
150021526_Thermosipho_melanesiensis_BI429                         G--TFYPKQV--NKTGEMLVL-QYKHYEDINKRLKIASEILNAAVDNMVNLL---SSFK------------------EITHTEIEYI------------TKLKA-----SF------------KKQRNI------------PALMAIEGN
28211139_Clostridium_tetani_E88                                   G--TFYPREH--YNSGKLLVK-QVETYEN--RRLEVAKSIVEAIGDNIYELL---YHYY------KHDKKEVKETLDWIKNHSKINL-------------KKAN-----DI------------------------------KQIMQVEGE
48477075_Picrophilus_torridus_DSM_9790                            S--SLIPEDY--LLSGKVIMN-QAIKAYNIDERIKIARIFVYAAARNMAIVL---KR------------------------GNMGKI------------KIPYK-----EI------------MESKSI------------NELMSYEGN
41614813_Nanoarchaeum_equitans_Kin4-M                             G--SFLPRQK--TKSGLVLVK-QVEAYKDIEKRTEIALEIVDAIRYNCIKVL---EKYH--------IDEVKELRKIDVWKMFEESL------------NDWKD-----AI------------------------------NIIRGIESN
55376278_Haloarcula_marismortui_ATCC_43049                        G--SIMPERG--QTSGQTLVE-QVRAYDDSTRRRAIARQFIEGSIHNMRANV---KYYD------DRGYDFD---------AIIDRL------------EAQAE-----SL-----------SDEMDT-------------NELMGVEAT
```

```
389844445_Mesotoga_prima_MesG1-Ag-4-2                               G--SFVPRKQ--KVSGKLLVE-QIKTFLDPIKRLKLSREFIDGAMHNMLRVL---QRSS------SKIPGYK---------DKIEGL-----------KNTRE-----RL-----------QNTTGV-----------NSVMAVEGA
390962116_Thermococcus_sp-_CL1                                      G--SLINDA---SPRGRLHLA-QAEHHLDPEKRLFIARAIVQGMQNAMAFTL---SRWG------------------VNPVKLNSV-----------KVDGG-----SV-----------------------------DELMGKEAE
13540937_Thermoplasma_volcanium_GSS1                                S--SLIPGNR--NERGNTTVQ-QVKAYLSDK-RFKIAAEMVNGIKNNILANL---RYYS--------------DNNEIT-KYIKAI-----------ESYNPDYNSNDI-----------------------------NKILGVEGN
289596694_Aciduliprofundum_boonei_T469                              G--TLYPRER--SISGEVVVN-QAKHYIFWEKRKYIAKEMIEGIKHNILRNL---KKSN------KELEEII---------EKIERV-----------EVEGD-----SI-----------------------------EELMNREAQ
14590110_Pyrococcus_horikoshii_OT3                                  G--SYLPAES--QISGSIVVA-QASHYIDNEKRLYIAREILEGTRASMISLL---KSQR------------------AEYKDLADI-----------DLKGE-----SI-----------------------------EELMGIESQ
84489236_Methanosphaera_stadtmanae_DSM_3091                         G--SLYPKIA--LNSGVIVVK-QALTYNNENKRCFIAKEMVNGMKHNMIKTL---KYYK------KKGKDVD------EHIQNLENE-----------SILDG-----NI-----------------------------NRILSSEGK
268317185_Rhodothermus_marinus_DSM_4252                             G--TYIPRDY--LHSGRLRIE-QVLHYVRPKRRRYLARAIVEAATYNLLRVL---RYYV------NRLEGERREAVA----EAIATI-----------EQERT-----QL-----------RSAEK---------I---PELMGIEGR
18977490_Pyrococcus_furiosus_DSM_3638                               G--TFYPRET--LLSGDLIIK-QAEHYLDKNKRLFLAKSFVVGGAKNMEKNL---KNWG------------------ISSDFSNHL-----------KELQG-----AK-----------K-----------V---TEIMNVEGR
427708213_Nostoc_sp-_PCC_7107                                       G--RLEN--P-SQRQAKYLTY-QHRRAQDKEFNRATAESIIWAKLHNQHTFL---QNWT------RHYTNYI------IQ-RALNYL-----------TLLMD-----NL-----------SVAPG---------I---EELHEYSEE
119720226_Thermofilum_pendens_Hrk_5                                 A--VLLN--P-VGTKYNNYRALQYTLPKNKA--LAIALEYLKSRVRGMASII---RNRG------GRLPALPEPPDPALY-EDPARL---------------------------------E-------------------SDIRSWEAA
118497969_Francisella_novicida_U112                                 L--CFGD--I-AEANFLARYK-QYSVVEQD---ISLARIFITSKIRNQHNLV---KSLR------DKTPE--------QQ-EIVKKN-----------KQLIA-----EL---------E--NT-TS---------L---AELMGIEGN
332661941_Haliscomenobacter_hydrossis_DSM_1100                      G--TLSP-WV-NPIGARARRR-QLLFANSPDAQRLAATWIVEKQVRQYQLLA---MHLD------QSGLLNFEKLKAIAL-NNQAQL-----------ELQLP-----EL-----------TF-----------V---PELQ--EAV
326793969_Marinomonas_mediterranea_MMB-1                            G--VSAGAAP-SVFGADFWLL-QAQYLQQETNALNISQVLIQARIEGIRAVI---SRRE------------------KDAPELNKI-----------QRLDE----KRL-----------RAETL------------DQLRGYEGQ
220934970_Thioalkalivibrio_sulfidophilus_HL-EbGr7                   G--VLWNGVP-GDAGPGLWMR-QAACFEDDARALEAARAVVDARLRQQREVL------------RNRMSPE------RCDDLLPRL-----------DRLIA-----KT-----------AAASDR-----------ASLNGLEGQ
288947693_Allochromatium_vinosum_DSM_180                            GVCAWMGPAL--GATGVLRAA-QHRAAEQVERRLMLARDLIAAKLLAQARVI---ERLP--------VASDPPETRTAVRRQQDLAL-----------ARLGQ-----AS-----------ST---------------TEVMGLEGS
431929780_Thioflavicoccus_mobilis_8321                              A--AVLGNGL--ATQLPLRRL-QHRRAAEPARALELARWILAQKLASYDLPL---QPLR------RRHGADEAACAGLLR-QRDRTL-----------AALAT-----AR-----------SR---------------DELMGLEGQ
206890539_Thermodesulfovibrio_yellowstonii_DSM_11347                G--ILNG--P-LHNNGLLRVK-QYQKSLSG-FSLKFAKELIKRKIVSQRDFLSEIREIK------KALAMQADRAIEILN-KAISNI-----------EVTPI-----SI-----------------------------DSLRGIEGA
52425690_Mannheimia_succiniciproducens_MBEL55E                      ---QFLP----QPHNDAYRRVTQTYLADNKLFCLTLAKNIVLNKCIKQCQFL---AKFI------EHNPKII---------TFIAEL-----------QKLFN-----LI-----------VKQEN---------I---DSLRGIEGR
30248142_Nitrosomonas_europaea_ATCC_19718                           ---SLLLARP--HNDAARRVA-QVRLSLDEPASLIIARELIERKLTRQIEWF---TELR----------ENDIQARYELS-RALRGL-----------EEHRA-----RL-----------GNINNA-----------ASLRGIEGS
121997419_Halorhodospira_halophila_SL1                              S--AFLR--S-EGHGDAVRRLGQYRLIHLEAERQAWARRLVRLRAGQQRLLASALYRR------PDQRQPLTAAHREIE-AAQATV-----------RREAP-----------------AG-----------------EQLRGQEGT
430760780_Thioalkalivibrio_nitratireducens_DSM_14787                A--LMLGPRH--NDAA-VRIA-QTLRIRDADYCARWARDLVAAKLRRQYRLL---IQCE------RQRPDARKA----LF-DARETL-----------ARIRA-----QL----------TDGSCDI------------GRLRGLEGA
292492856_Nitrosococcus_halophilus_Nc4                              A--CLLG----RPHADVARRLGQYRLFFDPRQRLAWSWRLIAAKLTAQHRGL---EQAL------AGRPDQRR-----VLTGALRSI-----------ESIRD-----KV-----------KDAENL-----------EVLNGYEGA
34541577_Porphyromonas_gingivalis_W83                               A--SLNG--M-NTILPSVMEA-QMRLSEEKKREFILT--LIKNKVRNQGKLL---RYYH------KYYRHDKELKEP-LS-NAIAEL-----------KQLEG-----IP-----------IA-EGSSLADF---R---QHAMLHEAR
88602029_Methanospirillum_hungatei_JF-1                             G--HFLP----LGRSALIEQY-EAQASAPEERRLEIARSICTGALENKRTLL---SNLE------RIRGFD------LS-REIRLV-----------EDAQD-----KA---------L--EC-QS---------L---DSLRGVEGS
325980970_Nitrosomonas_sp-_AL212                                    A--LTHA--P-TPHVVSLRRL-QYSV-----ETLTLAKEILYRKFDACIKCM---PILA------QS-----------FT-QSADSI-----------LYAN------TL-----------------------------EEMRLIEAR
269798874_Veillonella_parvula_DSM_2008                              G--TL----G-DEHTPLQRLLWQVACFQNQEFALDGAKYIVRRKIKGQIALL---NQYK------KSKCIPN------FV-VVHRTM-----------QALLK-----RV---------E--RT-KK---------V---ETLRGIEGL
332668016_Haliscomenobacter_hydrossis_DSM_1100                      A--RLWS-PY-FESLATLRRQ-QVRFAESVEG-GNWMVDVFMLKAEGQVQSL---RFLA----SKVAGQKPV------LD-SVIGLI-----------KQQSR-----QF---------EKYRE-QMPE-----EMR---NQMMGTEGA
379728404_Saprospira_grandis_str-_Lewin                             G--YLWS-GQ-YGSISTVRKK-QARFADHPLG-MDWVRDQLLQRAKNQLTTL---ELLQ--EQPYARKDKQG------QS-DFVKDQ-----------KSLSK-----LI--------QRWEKL-PISPRANLKELA---ASFRGWEGG
327314603_Prevotella_denticola_F0289                                G--RIWS-PK-YGSISTIRKG-QLNFVYSHRA-VDWIKEVLLRKIENQQALM---MLFD-----TSDHTDEP------TE-KNVARL-----------EDYRN-----KI--------RVLDG-DIVN-----DIA---PMLRGWEGQ
392395717_Flexibacter_litoralis_DSM_6794                            G--RIWN-NK-FGSTPLIRRN-QAIYATHSKP-IQWIKNTLIEKISNQISFL---HQLSYGVLPNGETTQRR------LA-DAIEQM-----------NAIQE-----NL--------QAISVK-RIKKEN--DWK---DSFRGHEGS
150398872_Methanococcus_vannielii_SB                                G--RFFG-NG-IEKTA-LKRK-QLKLTENWAG-ISIIKEFLSEKIENQIEHL---NNIS----KEN-ISEGS------LS-KSISKI-----------NKSKN-----ML--------LEIKGE-NIA------EIR---DQIQGLEGS
150401497_Methanococcus_aeolicus_Nankai-3                           G--MIWHCRI--GKTAKTRRS-QLKFSESING-IRFASNWIKQKMGNQLNYL---KDLK----KNHNHKGDFDTVIKNIN-NYIMEL-----------TQYTN-----NI--------DDKLPHREIK------------DTIIGIEGI
284040851_Spirosoma_linguale_DSM_74                                 G--RVWH-AK-LGSTTKIRKR-QLEASLGPDG-LRWVRAWLLAKLDNQMGFI---RSLK----KHRPQHAGY------LD-DKLVRI-----------EAMAL-----SISTLASVGEQTPATT-CVA------DVA---DTLRGLEGT
257125862_Leptotrichia_buccalis_C-1013-b                            G--RFWQ-SK-FGSTANIRRK-QLEIFGTQKG-IELAKQILIQKIKNCAEHL---EDLK----IKREAKKAF------LD-KQIKEM-----------KRYIY-----QI---------KLVEG-NVS------EKR---GTLMGYEGN
21226661_Methanosarcina_mazei_Go1                                   G--RVWH-SK-LGSTTYIRRR-QLQISEEPEG-FRLAKGWIEQKIDSQILFL---KDLK----KNRPDQKDE------LD-DYISGM-----------EDMKF-----QL---------EALEG-KLD------EMR---GKIMGLEGM
218961577_Candidatus_Cloacamonas_acidaminovorans_str-_Evry          G--RIWF-PK-IGSTVLIRRR-QLEMLSDNVG-LQFIDKWIAIKIMNQYRFV---QRLI----SKRDCDKSI------FQ-TRMQNM-----------QEAAI-----CI---------MQAEG-KLE------ELS---GSFMGWEGG
319790504_Thermovibrio_ammonificans_HB-1                            S--VSLPEHF-GSSTVNLRAA-QYRLFPEEAFANALIRELLLKKCNQLKEVI---PDIS------------------EKITEVEEK-----------IPVAD-----TL-----------------------------NTLLGLDGT
337288662_Thermodesulfobacterium_geofontis_OPF15                    G--QILP--PANSETTNKRIK-QFGLYF--FKRFEIAKYLIKRKYNEIERTF-NVELWE------------------LK-PKIDEV-----------TDY-------------------------------------------TSLLGIEGI
20094748_Methanopyrus_kandleri_AV19                                 G--VTVP--F-SERYATLRLK-QYEIASQPSARLAFARPLIASSILARAAVL---EFLA------NETGITG------LE-DAADEV-----------RSEAE-----RA---------L--NAGST-------------DALRGYEGR
386750121_Helicobacter_cetorum_MIT_00-7128                          A--TIFT--Y-QASNTQIIHK-QS-LLLHTSKHLYLAKQFIRSKSLNQLNFL---KYLN------KYHKN--------LD-EIIFKM-----------QSLTP-----KV---------S--KA-KE---------V---SELFGIEGI
126465226_Staphylothermus_marinus_F1                                A--LFVH-AR-SGGSVRNRIE-QYRSLEDGRACK-AAAMIISGKLSNQVSNL---KYYS------KPRKNISQESKI-LY-EKAEEI-----------KKLRE-----KL---------R--DI-ETSDLGKC---R---ENIMSIEAQ
170290046_Candidatus_Korarchaeum_cryptofilum_OPF8                   A--RLIP-AK-AGGTARNRYE-QYKSLEDRRG-IEIARAVIVGKIRNQASNL---SYYS------KARRMDEELSSE-LY-DAAQQL-----------KREME-----EL---------------KNEEFPDIDEAR---KRIMAREAK
385805115_Fervidicoccus_fontis_Kam940                               T--VISP-IY-SVKTAETRRK-QYSATTGKSGCT-LAKSLVRSKISNQAECL---RNLN------RYFND----ED--VR-KEVEKL-----------RENSK-----RA---------Q--EL-QC--SENW---S---SQLRQIESS
15898245_Sulfolobus_solfataricus_P2                                 A--KLIP--AKYAGSFKVWLK-QLTAW--KRRKVEFAKAFIYGKVHNQWVTL---RYYE------RKYGYN------LTSQELDRL-----------AREIT-----FV-----------NTA--------------EEVMQKEAE
156937921_Ignicoccus_hospitalis_KIN4-I                              G--RFSSPVP--NKSALARLK-QYEVVLKGEG-LKYAKPMIVRKIIEQGRTL---RYFA------KTKRMKW------LR-EASYEL-----------SKLSA-----DA-----------SSAGSP-----------EALRAVEAQ
124027534_Hyperthermus_butylicus_DSM_5456                           A--EHRPVDR-VNRTIEARME-QYRVKATGEA-LYYAAEMVYAKIVNQAKLL---RYLA------KSRRE----PW--LR-DAGYRV-----------EGHAD-----RL---------R--QI-IENEEPTT---P---EVIRSIEAQ
118431380_Aeropyrum_pernix_K1                                        A--RLYP-PI-INKTVATRIG-QFSA--NERLRRLIAAELVSAKIYNQGQTL---KYIA------RQRAD----ER--LR-EAGYEV-----------ELLSG-----EP---------L--RI-ADEDGPGF---R---DKLLSIEAR
307595807_Vulcanisaeta_distributa_DSM_14429                         A--IITP-PI-INKTVITRIR-QYELFLNNEARLELARRIITCKIENQARLI---RYFA------KSSRE----QS--LY-DIAYAV-----------SSIAT-----QL---------M--SI-ETNKLN-----P---EIIESMEAQ
305663330_Ignisphaera_aggregans_DSM_17230                            A--RLYM--PYINKTVATRVG-QYEKILSGLG-IEIAKELIYSKIMNQAQLL---KYLA------KSSREEW------LR-DLGYEV-----------EAIAT-----DL---------YSAKDLDR------------DRIMGFEAH
347523053_Pyrolobus_fumarii_1A                                      V--TLTP-PW-VSRTSETRIA-QYRAATDPSSRLAVAKSIVYGKMLNQAGFL---RSLA------RSSE-----AW--LR-DEAERI-----------EEYAR-----RA---------L--GE-DS---------V---DRLRVLEAH
302347808_Acidilobus_saccharovorans_345-15                          SVTYMSH----YSRTPETRRA-QYMAASGELGGS-MASSLVACKVLSQASTL---RRLA-----LRAGATVK---------STVSRM-----------RELAR-----EV---------MELRGTPNDLR---------PRLLELEAA
218884093_Desulfurococcus_kamchatkensis_1221n                       G--RLYP--PFISRTVETRA-QYAAYESWRG-VHVMRELVYSKLANQAGLL---KRYY------------------MY-TIMEEL-----------GDAYK-----KI---------IELAYRARMLEAEFEEAR---EKLRQLEAE
389861365_Thermogladius_cellulolyticus_1633                         A--RVHP--LFNNKLTETRKA-QYRALLDGRWAL-VAREVVYSKVMNQAGLL---KRYY------SYTRDQA------LR-EGYEKL-----------STLAS-----EA-----------RLKEGVALSELK------EWLRGVEAE
15898286_Sulfolobus_solfataricus_P2                                 G--ILLPSI--VTETVKTKKA-QYEAIV-VRKDNRYGEEIISSKIYNQSVHL---KYWA---------------------RVTGT-----------KNDYK------------------------------------------ELLDKDEPA
15922975_Sulfolobus_tokodaii_str-_7                                 G--RLFHSE--PIKTVETRKA-QYLAILKGEE--EIPREIIKSKIKNQANHI---KFWF---------------------KKLGIE-----------GNDYK-----LI-----------------------------EGKDDDEAT
18311772_Pyrobaculum_aerophilum_str-_IM2                            G--IFMK--PVGDATGAKRLC-QYAAATDGRG-LQLAKKWVWLKIRGQLENL---KRWR------RRLGKYG---------TYAESI-----------SKAIN-----AL-----------ASAATP-----------REVMEAEAA
18311699_Pyrobaculum_aerophilum_str-_IM2                            A--RLFP-PE-ANGTVTHRRA-QYEAYLTGRGFE-LAKLVTYGKLINQARAL---RRLG------QWKRE--HYGA--LA-EAASKI-----------ADLAG-----RI---------P--SC-AD---------V---QCVLGHEGA
54296139_Legionella_pneumophila_str-_Paris                          C--SLQG--G-LHRTAKVRES-QFKYIYTKEASN-MAARFILAKIKKQSAVL---KYFA------KAKNRPENEKII-LA-AASQKM-----------AGIVA-----NI---------T--ET-NWIKLQDW---R---FTLLGFEGS
34557933_Wolinella_succinogenes_DSM_1740                            T--AITS-TA-NHAVANVRKA-QFLTIESNHGRE-IARTLIVGKLKNQRATL---QYFY------KYLKNTDKQSKK-LE-SAIEKI-----------KETIE-----SL---------K--NI-ELSNNKHW---S---QIIMGKEGG
118497353_Francisella_novicida_U112                                 S--MLVGAN--QHAVVQARIN-QYRYI--DRNALTLSIKLIAAKIKNQRATL---NYFN------KHHKSIN------LL-NAIEEL-----------KRVAQ-----LI-----------KNAKT---------L---NDVLGYEGY
328954440_Desulfobacca_acetoxidans_DSM_11109                        A--LLHA--P-ISPQATLSRL-QLTASQSRTG-LELAKEFVDGKIRNQLNLL---KYYL------RSRQAETDEYLPSSKSSIL-----------QRCQQ-----EL---------GKLQLDEDYDTGR--------KRLFALEGQ
408403000_Candidatus_Nitrososphaera_gargensis_Ga9-2                 G--FVIP-AK-ISGTIRAKRE-QFLAYHDFRGAV-LAKKFASGKSFNQRNLL---KLLW------KNRVKTEPGLAEKMY-KASEEV-----------ERLAR-----EI---------D--SI-EGSSVDKI---R---TDVMNREAR
330509017_Methanosaeta_concilii_GP6                                 A--RLIP-AS-LGGTVKTRRE-QYNAYNDSRGTN-LARSFVRGKLKNQASLL---KSFA------KKWKGERQDLWGEFR-NNSNQI-----------EELIP-----RL---------N--LI-KGK-VDDV---R---GEIMGVEGI
148656004_Roseiflexus_sp-_RS-1                                      G--ALMH-GG-ITGMALTRRA-QLRAGDDERGLM-LAQAFASGKIQSQANLL---RYAA------KNRKESDPDLHHDLM-RTATEI-----------LDTLP-----SV---------R--AM-RGVLTEET---R---AALMGFEGM
386854797_Deinococcus_gobiensis_I-0                                 A--KFSS-PY-LVGTVSTRRA-QMAAYLTPAGVE-VARHAIRARLRNQSSLL---KYFG------KYRKEADPAAYEAIQ-RALPGL-----------TTLEA-----EL---------D--ML-QGECIDDL---R---DVLLGVEGR
167628642_Heliobacterium_modesticaldum_Ice1                         A--KLFA-PG-LTATVKTRRE-QLLAFNDQRSIF-LAKAFVRGKIQNQINTL---KYFA------KYRKSARQEVYAYLQ-DAALLM-----------EKNLQ-----EL---------T--GI-DGLNIDAV---R---GPLMSVEGR
148642083_Methanobrevibacter_smithii_ATCC_35061                      ---ILES--P-NQNNIVLKKL-QYQASENHKG-LLISKEFITSKMKNQKSTI---KTLN------KNKKIKEVKIIENKIKQNIKDF-----------KNFKI-----TL---------NDNIFTSK------------NKIMGFEGI
288560430_Methanobrevibacter_ruminantium_M1                          ---RLSA--P-DRKNAIVKKE-QYFALTDSRSGY-LAKAFVRAKIENQKAVL---GTLA------KSREEKD------YIIEQREKV-----------SEHIG-----KI-----------EKLSNINSDNIR------NNILGIEGQ
11499462_Archaeoglobus_fulgidus_DSM_4304                            G--RLSH--P-LIGTAKTRRE-QYLAYGDKRG-VHLAKEFIAKMANQMAIL---TNLA------KARKDSNPEVAESLL-KAKKEI-----------DACLN-----EL-----------------DGVEAEMIDKVR---ERLLGIEGK
383319612_Methanocella_conradii_HZ254                               A--YVSP--P-QMRTVNTRRE-QYRAYDSPAGAV-LAKEFIYAKMKNMSATL---GTLA------KTRKDTSPESAESLR-QARDEI-----------NMRIASLETLDI---------KKKKCDDI---------R---DTIMGLEGS
294676823_Rhodobacter_capsulatus_SB_1003                            G--YLAP--P-QPAEAQLQLA-QAALVLHPERRLALARVLVGARLRNQRARL---AVLN------RSAKATE------VD-RAIAAL-----------GRAIR-----GL---------A--APMPD---------L---DTLRGHEGR
302343124_Desulfarculus_baarsii_DSM_2075                            G--VLEP--M-VADKAALHLA-QAAVALDETKRLALARLICAGRVRGQRALL---MRLN------RRRKNSD------IE-SNLAAF-----------KQLPR-----RI---------A--TA-TT---------I---SELLGLEGE
389875622_Tistrella_mobilis_KA081020-065                            A--WLEP--D-GGARGQRHLA-QARHILNPDSRLTLARLLVDARLQTTQALL---KRLN------RRKRLAT------VD-AAADRL-----------KRIRR-----KL---------P--LA-AD---------I---PQLMGHEGE
374998939_Azospirillum_lipoferum_4B                                 G--RLTR--P-EARRAALHLD-QARHALDEGLRLDLARRIVDGRLRNQRALL---RRLN------RSRKLGV------VE-DAVLAI-----------NALLR-----RL---------D--TA-AD---------V---AALLGFEGQ
288957883_Azospirillum_sp-_B510                                     G--TLER--P-AGPRAGLHLA-QARLAIDPAARLDLARRIVDGRLRNQRALL---RRLN------RRKADAD------VT-DAALAL-----------NRLIR-----KL---------P--VA-ED---------V---PALMGLEGA
116627764_Streptococcus_thermophilus_LMD-9                          S--SIET--H-RQDEFQKQEL-QAKAYFEEDFRLEVARSIATTKVRHQIALL---REFD------TDGLLDTSDYSR-FE-DSVNDI---------------------------------Q--KA-YS---------I---TEIMGYEGR
15609954_Mycobacterium_tuberculosis_H37Rv                           G--RIST--P-DVSYAPRLRQ-QVHRTDDPAFCLSLSKRIVSRKILNQQALI---RAHT----------SGQD------VA-ESIRTM-----------KHSLA-----WV---------D--RS-GS---------L---AELNGFEGN
348026605_Megasphaera_elsdenii_DSM_20460                            G--RLES----TRHVSVFKQK-QQFLLQDQPFSLEMSRRVLLAKVHNQLTLL---RRYN------RDRKIPS------VM-IDIHNM-----------MTMAD-----HL-----------KIAEDC-----------ESLMGYEGM
296133514_Thermincola_potens_JR                                     G--RLES--T-RHVNILRQRK-QFRAGDDEEFCFKFTCMIVAAKIHNQAVIL---RRYN------RHVNSPA------VD-ECISRM-----------QLLEE-----NI---------A--RA-GS---------I---AQVMGFEGA
114567264_Syntrophomonas_wolfei_subsp-_wolfei_str-_Goettingen       G--RLES--T-RNVNIYRQRK-QFACGEDDEFCLALKRIILAKVKNQITIL---RRYR------RNRPEKS------VQ-KIIDAM-----------AKLLP-----IM---------E--RV-HN---------K---DELMGHEGM
332296542_Thermodesulfobium_narugense_DSM_14796                      G--RLVG--G-MPRNVELRIK-QFKLYMNEKFKVEFSKKIVFGKANNCKYLL---QRYN------WERRNSK------VK-DCIIKI-----------KSLIA-----KL---------E--FT-DD---------I---DQIRGIEGK
218442809_Cyanothece_sp-_PCC_7424                                   G--HLWS--L-GSIHLNNEMV-QIKRHQEHEFQVKISKAIVYGKLMNSKRLL---MRLN------RKQVPD------MD-KVIEGI-----------NSDIL-----SL---------E--SV-DN---------L---DQLRGYEGI
17228961_Nostoc_sp-_PCC_7120                                        G--RLQT--E-GDAKVEYLML-QVERCQNHEFTRKQAEAIVRAKLHNSRALL---LKLN------RRHPSKI------AA-TAISGI-----------AELME-----KL---------S--LA-ES---------M---DSLRGYEGK
119357853_Chlorobium_phaeobacteroides_DSM_266                        T--RLES--T-DDVNIDLERL-QFQRSAEESFPLECSRTIVRAKLHNSGVLL---RRHA------ESSGSEA------LR-HAATQL-----------RQLEE-----HV---------D--RA-DS---------V---DAVRGYEGS
119357846_Chlorobium_phaeobacteroides_DSM_266                        G--RLEA--T-TADNAEMQRF-HFLRSIDEPFALETARSIVAAKISNSKTMI---RRRK------TVVQDRDSTLQNKMA-YNLDIM-----------ADLAL-----KA-----------EASTD---------I---DALRGIEGK
336116789_Microlunatus_phosphovorus_NM-1                             G--QLLP--A-SSPTRAERLRGQLAFSDDPTRTVPLALAIVEAKIRKQIVLL---QRFG------RRAHAEL------VQ-EAVAEM-----------RNVLR-----LL---------P--DC-TS---------T---EEAMGVEGA
37677204_Vibrio_vulnificus_YJ016                                    G--CMSA--P-FQGDPLLTAA-QVNSRADPDVRNHLAKFFIRSKINNALGVM---RRQG------KKQQRNK---------EQEKRI-----------VHLRE-----QL-----------DAHQTL-----------ASIRGTEGA
154250301_Fervidobacterium_nodosum_Rt17-B1                          G--KLYT--D-EYRNVLLRLK-QYERATDKQFQLEISKSIVQGKLQNYYNFL---TQKS------KNLPKGL------LS-EERAGI-----------RTVIE-----KV---------N--KA-KT---------V---DEVRGYEGI
38505760_Synechocystis_sp-_PCC_6803                                 G--RIIA-----IERGYRHLS-RYQQELGWQQRLITARTIVAAKIKNARVFL---QRQH------RKHPTPS------LQ-LAIDGL-----------KHFVE----LTV-----------KADKL-------------EQLLGYEGT
83592490_Rhodospirillum_rubrum_ATCC_11170                           A--RVVP--P-EARSADLRLH-QYALHLDPPERLIRARAVVTAKLRNAAAVL---RGIR------SNQASSA----A-LA-SAITQT-----------EASAE-----AA---------A--AA-VS---------A---ESLLGIEGN
261415948_Fibrobacter_succinogenes_subsp-_succinogenes_S85          G--RLVS--A-VGKNVPLRLC-QYDVFRDDDRAFALAKSCVVRKLENGLRVL---EAYS------KNP-HNSFRFEN-RD-EYLRNL------------NAVR-----RL---------Q--GF-DR-------------DELRGFEGN
76802276_Natronomonas_pharaonis_DSM_2160                            G--SFVP--E-KNTIAEVRRA-QYAL--TPAKELAIAKEMIRGKIRNQRTLL---SRKG------VTG------------------------------TDVLK-----DL--------GV--RV-DN---------IKTKDDLRGAEGE
312114617_Rhodomicrobium_vannielii_ATCC_17100                       A--LAQP--M-HGAGLDIRRA-QY-FEAAGERGLAIARALISAKIQNMRGLV---RRRA--------------NIE-GR-DCLTAL-----------AALAK-----KA---------K--HA-SR---------------ESLLGIEGS
219683006_Bifidobacterium_animalis_subsp-_lactis_AD011              G--WSVS--S-YGPNGETRVA-QHVASHEG--RLDLAREFISAKIHNQIVLL---RRSD------KNN------------NVLFDM-----------KHIEK-----SV---------V--NA-NR---------I---QDILSLEGQ
116748794_Syntrophobacter_fumaroxidans_MPOB                         A--RLMV--D-EHKHVQRRQG-QYVKLSGADFALRTTQSIVRGKLRNTARFL---ALRG------SRY-GSE----A-LH-RAAAQI-----------KGLSA-----LV---------D--RQ-KD---------M---DLLRGIEGH
```

```
317051217_Desulfurispirillum_indicum_S5                           G--QLCG--M-EGKNVFARMK-QYERYANPAYRNGLARVIVRQKVNNLRYFL---AKHN------RYHENRV------IE-DAIVQI------------KATLQ-----GL------------DKATNT------------DEIRGIEGA
328949009_Treponema_succinifaciens_DSM_2489                       G--KLSF--G-DGKNIFLRQK-QFRILDNPQKSLEIAKSIVSGKIRNEISFM---QRIK------RKN-SNE--NEA-VE-NAISSV------------KNTLS-----NA---------E--KT-DD---------I---EKLRGYEGL
333996012_Treponema_azotonutricium_ZAS-9                          G--KLNF--A-NAKNIFLRKK-QYAISDNDALALPIAKSIVLAKIKNEISFV---QRIK------RKNGYDD------EFSQAIQAL------------KAALE-----GA------------ECAENH------------GELRGHEGI
313673559_Calditerrivibrio_nitroreducens_DSM_19672                G--RLVQ--E-LGKNVALRRL-QFEATQDEKFKFKIAKSIIYGKYKNSISAL---RKLN------Y---YHK--SKE-VS-DILNQL------------TARHN-----DL---------T--AA-ET---------Y---ESLLGYEGI
218961167_Candidatus_Cloacamonas_acidaminovorans_str-_Evry        G--RLVG--E-FGKNVMLRVN-QFQYLNEPENRVYLANLIIRGKIENSLLHL---AKRK------K---RNK--AES-LT-FAYIKN------------KALLA-----QL--------KT--QI-LP---------K---DILLGYEGI
294828260_Leptospira_interrogans_serovar_Lai_str-_56601           G--GINI--N-PSG-VQRRIR-QFKALTKETIRLNLAKKLVSAKCESQLRYL---LRAT------R---GKDETRNE-TE-SYLATI------------RSGLK-----NI---------E--SA-DS---------P---SQLLGIEGS
347756303_Candidatus_Chloracidobacterium_thermophilum_B           G--SFRG--G-TPG-VQRRHR-QYAALSDPDFCLALARRLAMAKVENQIRYI---LRAT------RPMLTEAGRDPA-LE-QHLTQM------------RHALA-----QI---------S--RA-VG---------P---DEIRGYEGS
162457462_Sorangium_cellulosum_So_ce56                            G--VFSG--G-SSG-VQRRIR-QFRGLTDDVFAIGLARRLVMAKIEMQLRHV---LRAS------RKDEALRAGLDEPIE-SLRGGL------------RRAAK-----AA------------DR----------------TSLMGQEGN
86742032_Frankia_sp-_CcI3                                         A--RVEG--P-TGGNPHLRIA-QHDHFRDDERRLTLAMSYIAGKLQNSRQLL---LRAA------RDATGTR--QTA-LR-DTAAHL------------ADALP-----TL---------R--DT-TN---------V---AEAMGVEGQ
433631663_Mycobacterium_canettii_CIPT_140070010                   F--GLRA--P-THGNVLLRME-QWRAAVDRDRCLDVAKPVVAAKLLNSRVVL---LDAA------KDR-SSR--SKQ-LR-SAADEL------------GGFAE-----QV---------K--AA-RD---------L---DTVRGLEGI
15612904_Bacillus_halodurans_C-125                                A--RVVG--E-SRGNVVLRKT-QYRISENDQESTKIARNFITGKVYNSKWML---ERMT------REH-PLRVNVEQ-FK-ATSQLL------------SVMMQ-----EI---------R--NC-DS---------L---ESLRGWEGQ
108763792_Myxococcus_xanthus_DK_1622                              A--RVEG--V-PGGNVLLRRQ-QYRAADDIARSVAISRALVVGKLGNARQFV---LHAR------RDAAPER--QEA-LS-ETARRL------------SEHLR-----AL---------T--RV-ED---------L---VQVRGLEGI
344996859_Caldicellulosiruptor_lactoaceticus_6A                   G--SIIP--A-LGKNINLRIL-QYKL--DSGKKLELSKLFIAGKVKNQITML---FRWQ------KNY-GIR--LEE-E----IEQI------------KKIKK-----KL---------P--SA-QD---------I---AELMGYEGA
85858452_Syntrophus_aciditrophicus_SB                             G--RISP--P-ESKNVFLHKR-QYSLLGNETFTLRQVRAIVAGKLANMATLL---MRIK------RSR-NVS--LAS-Q---KAHEI------------QSLIR-----LL---------F--QA-ES---------V---DSLRGYEGR
148657121_Roseiflexus_sp-_RS-1                                    A--TLTA--G-PSRFGDLRIR-QMQFVGAPDRALRLAKDIVSAKLTNQRRLL---AATG--------W-PAA--ATA-IA-Q-IDAA---------------LT-----AA---------A--NA-PH---------V---DMLRGHEGA
159898907_Herpetosiphon_aurantiacus_DSM_785                       A--MVSA--G-LGNNAALRLA-QCRIVDNPHLAVPLVQAIVVGKVANQIQLL---ERYG------SDW-GGM--GLR-AK-QTMQHV---------------IQ-----QT---------Q--HM-PD---------I---EQLRGLEGA
17227877_Nostoc_sp-_PCC_7120                                      G--RLQG--H-FATNMTIRLA-QYRT-VETTFGMALAQKLVIGKVRNQRVLL---QRRN------RAT-NGQ--ISE-LT-EAIDLI------------SVYAS-----QL---------N--NT-TT---------PLNRNELMGVEGI
268318378_Rhodothermus_marinus_DSM_4252                           G--RLAA--E-FHPHVALRLA-QYEAFRDPERTLTLARLFVRGKLQNMAGLL---RRYA------DEYGSASLRAAASEINRDLERL------------EQVTT------L------------------------------EALRGVEGT
55978328_Thermus_thermophilus_HB8                                 G--VAGA--Y-PDPHPAHLRA-QFAA-----EGLPLARAFVVGKLRSALALL---ERHR-------------------LP-EAGGVV------------EALARAEGASEL------------------------------ERLRGAEGE
297622675_Truepera_radiovictrix_DSM_17093                         G--VASS----CALAPPARLR-AQFAAATSAVGAGIAQLFVLAKLRSSLAVL---TAFR------RTRPELA---------AAAAQL------------HALGG-----KV------------AACTE---------L---ERLRGFEGL
328953000_Desulfobacca_acetoxidans_DSM_11109                      G--QLTS--P-FTKNVTLRQA-QYDRAADPEFALDLAKIIVGAKLTNSRGLL---QEFA------RNHPESG------LK-GEIERL------------TELIL-----QI------------GGSPNL------------AALLGLEGA
337286709_Thermodesulfatator_indicus_DSM_15286                    G--RLAP--P-RGKNVELRLA-QFRAFDNEKRRLEIARAVIAGKIRNQKNFL---RRQN------RKLKNEK------IG-QAILKL------------RHKIK-----EA------------EDAQS---------L---ESLRGIEGQ
328953422_Desulfobacca_acetoxidans_DSM_11109                      G--RLIS--E-FGKNIELRRR-QFQKLDDPGIRLALARSYVKGKLNNCRIFL---RRQN------QGLNQAD------LT-NTIHQL------------RRLAN-----QA------------DACDS---------I---ESLMGMEGA
38505684_Synechocystis_sp-_PCC_6803                               G--SALP--S-ESRNGQLRLA-QFRAHEDPIQRLDIVKAFVKGKVHNQYNLLYRRGQVD-------------------NPLKGRGKL------------VMRQQ-----TL------------------------------EQVRGIEGL
134299482_Desulfotomaculum_reducens_MI-1                          G--CLQA--A-LSKNSVLRIA-QHRAYQKQEERLLYASEFVRGKLSNMRTNL---LRYN------RSLNNHS------ID-EAVSRI------------KNIIK-----RL------------EKAKDL------------NELMGLEGA
17229060_Nostoc_sp-_PCC_7120                                      A--SLEP--E-MGKNIFVRAA-QWKAAGESEQAIHVTQGFIRGKLKNYRHTL---LEAQ------RRYEVDLNSNISQLS-NAIASI-------------DKAN-----SI------------------------------NTIRGLEGA
302391602_Acetohalobium_arabaticum_DSM_5501                       G--RLVP--E-YSKNSLLRLK-QFEAHKDMKVKLELTRKIVKGKLTNMRTLL---MRDT------RETRSEE------VK-QLCQKM------------KNIIN-----KI------------ERTES---------I---DKLRGFEGM
159898755_Herpetosiphon_aurantiacus_DSM_785                       G--ALTP--D-PTRNASLRLA-QYAAHTSIQHRFSLARTFVDGKLRNLRTQI---LRFN------RSQ-REP--TLT-Q---AIERLRDAHRDLHGLSIPEYVD-----PL---------D--RM-HG---------M---GQILGCEGQ
320161859_Anaerolinea_thermophila_UNI-1                           G--ILSG--G-LTPHVPLRRA-QYRALEKPSFSLEMAKGFVRAKLRHQRTLL---QRQN------RPPKQDA------SLSEVIERM------------EHAID-----EV------------QRKTS---------L---SSLRGLEGS
156741961_Roseiflexus_castenholzii_DSM_13941                      G--RLVG--S-VSAHAALRAA-QYRRADDRAWSLRLAQRFVEGKLRNCRALL---RRFA------RNR-ADA--PAE-AG-QAADDL------------DRFID-----RV---------P--RT-TT---------L---NALMGVEGS
383762167_Caldilinea_aerophila_DSM_14535_-_NBRC_104270            G--RLVG--A-TGGNGALRVA-QVLRSQDSLFALSLARHFVRAKLHNCRVLL---QRYA------RRQEEPA------IQ-AAVEGL------------ERAME-----QV---------ERCRT------------L---NSLLGVEGQ
46255267_Thermus_thermophilus_HB27                                A--GLTR--T-PSKNAPARVA-QFKAYLEPTWKLALAQRFVLGKIRNGLVFL---RRN--------------------GA-EGWERL------------KEALL-----EA---------E--RA-QD---------E---EALRGAEGR
134101617_Saccharopolyspora_erythraea_NRRL_2338                   S--VDIP--T-VGKHVELRRR-QF--TATPNTALDFARRIVGGKIRNARTLL---RRN-------------P--HVE-EP-DLLNRL------------DADAV-----RA---------E--RA-PT---------R---STLLGIEGA
378719041_Gordonia_polyisoprenivorans_VH2                         G--WAQG-----EMSKYVQLR-TAQACLDDATALGISRSIISGKIKNARTLL---RRNS------RTDAE-----------QSVQAL------------AGMAD-----SA------------TGAASP------------AELLGIEGA
320102061_Isosphaera_pallida_ATCC_43644                           G--LTRG--L-DLRNVMIRRE-QFRRADDPAFCLALAKRLVAGKARNQRTML---QRN-------------H--IEP-PA-GAIQRI------------RAIIQ-----EV---------E--QA-DS---------L---ESLLGLEGT
268318441_Rhodothermus_marinus_DSM_4252                           A--RLVS--M-YDVNAPVRVA-QFEVAADETKSLAIARAIVTGKIKNQRTLL---RRNQ------RTRSE-----------RVLREL------------SRLAR-----EA------------RRASS---------L---DELLGIEGA
39995168_Geobacter_sulfurreducens_PCA                             G--HTVS--T-GHRNVETRTY-QYQRSFDPETCLNLARRWIVAKIANCRTLL---RRNW------RGE-GDE--AKA-PP-GLLMSL------------QDDMR-----HA---------M--RA-PS---------L---EVLLGIEGA
83591519_Rhodospirillum_rubrum_ATCC_11170                         G--STGG--Q-GPRSAAVRTA-QYALAGDERRRQAFARDLVSAKIRNGRTLL---RRNW------RGAEAERQ--------IALDRL------------ARLAE-----RA------------TTAET---------T---ACLLGIEGE
50085564_Acinetobacter_sp-_ADP1                                   T--FLAPQSE--YRPTEYMQI-WMKAWLDDDARLLMAKILLLERIVLVSQSW---KKNL-----------DLIQHGIYIDENALNTF------------KQDID-------------------RSENQ-------------QELLAAEGR
77166205_Nitrosococcus_oceani_ATCC_19707                          ---LFLASQN-EYRPTEYCQA-WMRLWQDNDQRLKVAKTFQRNRAEFLMQQW---PKLA------EPKPHK----------ASLEKL-----------AERYLA-----DI------------ELAGDN------------GTILAQEAK
15602176_Pasteurella_multocida_subsp-_multocida_str-_Pm70         GCEFFSPQSE--YRPVMYLQQ-WCKFWFDDEKRLAVAKALQQQRLALLEPIW---SKLE-------------------WV-YSCEKL------------TALVT-----QY-------QHEFAQAAHS-------------QALLAAEGR
38234755_Corynebacterium_diphtheriae_NCTC_13129                   N--AYSHATP-LTSTAKWAIA-QACLVSNTEYQKKAALAFYKRQFGG----------------------------------NAITGG----------------------SI------------------------------SVMRGLEGR
21674788_Chlorobium_tepidum_TLS                                   S--GQTP-----TSNTRNMTH-QMKLAANPAKALEVARRLFAYRFPD----------------------------------ANLENK----------------------TL------------------------------PQMMGMEGL
53804737_Methylococcus_capsulatus_str-_Bath                       T--APPM--G--QGRSDVARA-HARLWADEKARLDVARHMYAFRFGRI---------------------------------LPHKDI------------------------------------------------------AVLRGIEGA
209966592_Rhodospirillum_centenum_SW                              ---YTAPPLF--DRDSTLARQ-QATWWAGESTRIMVAKRMYAKRFGETPRAT---------------------------------SL------------------------------------------------------DSLRGMEAA
89885834_Rhodoferax_ferrireducens_T118                            ---AAAH-----AHQDASRVLRQAAIHTNQRERIAAASRLYGLMFDDHM--------------------------------PPSFTI------------------------------------------------------EKLRGLEGS
16130662_Escherichia_coli_str-_K-12_substr-_MG1655                R--VYASGQP--GGARSDKLLYQAKLALDEDLRLKVVRKMFELRFGEPA--------------------------------PARRSV------------------------------------------------------EQLRGIEGS
68535718_Corynebacterium_jeikeium_K411                            A--SGRS--L--SNSSAAAEA-QASIVSNQRRRLFCARQMYAMRFPG----------------------------------EDISRA----------------------TM------------------------------AQLRGREGS
86738746_Frankia_sp-_CcI3                                         A--AGLG--P--SRGAGLLMR-QAYLVTRTSERLDVARRMYAKRFPDDD--------------------------------VTTATM------------------------------------------------------QQLRGREGA
51891801_Symbiobacterium_thermophilum_IAM_14863                   A--QGMG--E--TRSAVNLMQ-QARFWADPQRRMQVVRRMYEMRFGEPL--------------------------------DPSLSL------------------------------------------------------RQIRGKEGA
29834079_Streptomyces_avermitilis_MA-4680                         A--AALP--D--SLTTTWLER-QTRAWADDTRRLAVATAMYEKRFGEATV-------------------------------PPGTTL------------------------------------------------------AQLRGFEGQ
```

Continued

```
407462288_Candidatus_Nitrosopumilus_koreensis_AR1                 SAKVYWTEFAK-FI----PE---K-----------------YQ---FHS--R----NQSHITSSKNNATDIINALLNYGY-SVLAGEISKFVCGFGLDPYLGFMH----RSHT-GFQPLVYDIIEPFRWL----VDYTVYSMAN-HSSTR
21229351_Methanosarcina_mazei_Go1                                 LAGKYWIEFSK-AV----PK---E-----------------YD---FCN--R------IDQFRRAMGSGDMINTMLNYGY-SLLEAECLKAINSVGLDTHVGFLH----EMAP-SKNSLAYDLQEPFRFI----VDLAVISLIE--S---
147921327_Methanocella_arvoryzae_MRE50                            TAKAYWDRLTK-VF----NKLYPD-----------------FR---FTG--R----RNKSNSWNMNA-SDEINALLNYGY-AILEAQIRRPINAMGLDPAMGYLH----EMKG-SGAPLVYDLQELYRWL----IDLSVIQLLE--E---
124485266_Methanocorpusculum_labreanum_Z                          TKTMYYEIIGR-KL----PK---V-----------------LG---YRR--R----CQ------PPF-MDPVNVMMSHGY-AVLYANFALACTGAGLDLSRGALYGEIVSAPG-GRGGCVLDLMEPATVSM---VDRVIIQMAA--E---
307352568_Methanoplanus_petrolearius_DSM_11571                    TSDMYYEIMGR-CL----TP---S-----------------HG---FKR--R----TS------RPH-VDPVNSLLSLGY-SILFGNCCVPLVGAYLDVDKGILR-------E-GTNSLILDLIDPLKPFM---VDSVVFSIAR------
154150254_Methanoregula_boonei_6A8                                TANMYYEILSR-DI----PR---E-----------------FD---FRR--R----TV------RPQ-CDPINAMLSFGY-AMLFGNCCVPVIGARLDPDLGILY-------E-GSGALVQDLMASFKAQM---IDGVIWSIAR--D---
219852754_Methanosphaerula_palustris_E1-9c                        ATDMYYEIMAR-TI----PK---G-----------------TG---FRR--R----TA------RPY-MDPVNTMLSFSY-GILSGVCAVHLAGAHLDANIGLLH-------Q-GERALVRDLTELFKPQM---VDQPIFALVR--Q---
126178701_Methanoculleus_marisnigri_JR1                           TTDMYYEILSR-TL----PP---E-----------------LG---FRR--R----TS------RPY-LDPVNAMFALGY-AMIYGNCCVSVVGAHLDPDLGMLH-------E-GAGSFVHDLIEPQKASM---VDRAVIRFAR--E---
397780289_Methanoculleus_bourgensis_MS2                           TGDMYYEILSR-TL----PP---E-----------------LG---FRR--R----TN------RPY-RDPVNAMFSLGY-AMLYGNCCVSLVGAHLDPDLGMLH-------E-GAGSLVYDLIEPQKAVM---VDHTVIRFAS------
156937861_Ignicoccus_hospitalis_KIN4-I                            WSSKLWKDIVQ-FV-----P---G----------------------MRS--R------------VPRGNDPPNRTLDYLY-ALLYSVCNHALVGAGLDPYAGLIH----RERA-GKLSFVYDFSEMFKPMA---IYVMATAIRT------
70607609_Sulfolobus_acidocaldarius_DSM_639                        MSKACWEQLRK-FL----PP---S----------------------FTG--R------------KPRNEDEINRAVDYAY-SVIYALCTHSLIASGLDPYGGLMH----VEYP-GRTALTYDFSEMFKPVA---IHSVIAVSRK-VK---
187736488_Akkermansia_muciniphila_ATCC_BAA-835                    CARLFWSVFAD------------T-----------------WANSDFRR-------GR------H---EEGFNNLFNYAY-AILLSCILQYLFALGLDPCFGIFH----QSRE-HAAPLAYDLMEPFRPA----FDANVARWIHLCLRE-
436839745_Desulfovibrio_hydrothermalis_AM13_-_DSM_14728           AARLCFSVFNSFIMDSTMRS---H-----------------VR---LPR-------SK----------PDQLNALLDFCY-TLLFNRINSLLRVRGLNPYLGILH----SSKN-NYESLVADLQEPFRAR----MDRFALRLLN--K---
71894593_Mycoplasma_synoviae_53                                   IAKLTFKNLYG-------ST--------------------------FNR--T----DK----------ENEINKFLNYGY-TILMTYVSRNLVKKGYDNRIGVFH----KSFN-NHFALATDLMEPFRFL----IDKLVYELL-------
193216853_Mycoplasma_arthritidis_158L3-1                          CAKLNFKILFG-------EK--------------------------FVR--K------------TNTMEDTYNAYLNYGY-SVLLSYVARAICSKGYDNRLGIFH----RSYS-NFYPLACDLMEPFRCI----VESLVYKHVR------
187250661_Elusimicrobium_minutum_Pei191                           GAKVYFENLFN-------------------------------D---FHR---------------NTDDKDIINSALNYGY-AIVRSSIARAVVSAGLNPAIGIFH----SKNH-NPFCLIDDLIEPLRPL----IDFMVKNKLD------
325972002_Sphaerochaeta_globus_str-_Buddy                         GARVYFESIYG-------TS---------------------F----FRS-------DE----------DNAINGKLDYGY-SILRSAVARAVVDTGLLPSCGVFH----SSRI-NPFCLVDDLMEPFRPM----VDSLVLAMCS--K---
47458867_Mycoplasma_mobile_163K                                   AAKVYWNMTYS-------KE--------------------------FIR--N------NSATKLKFDKFEIVNAILNYGY-SILRSSAIQSIIKKGLDPRFSIFH----KSFS-NFFALASDLIEIFRPL----VDEIAY----------
269123825_Streptobacillus_moniliformis_DSM_12112                  AAKVFFNYLYG-------TN--------------------------FCR--Q----NE----------RDSINMALDYGY-GVFRSAITRLLCTYGFATYIGVHH----SSMM-NAFNLTYDFIEPYRPI----IDYYVFNHLY------
57865886_Staphylococcus_epidermidis_RP62A                         ATRIYFKDLFG-------KQ--------------------------FKR--G----RY----------NDVINSGLNYGY-SILRSFINKELAIHGFEMSLGIKH----QSKE-NPFNLADDIIEVFRPF----VDNIVYEIVF-KK---
319957207_Nitratifractor_salsuginis_DSM_16511                     SARLYWNKLFD-------------------------------------GLKR------------EADSLDVRNQALNYAY-AILRSLITRDLSAAGFLPALGLWH----DNRY-NAFNLSDDLMEPFRPI----VDAAVYLLLP------
310286727_Bifidobacterium_bifidum_S17                             AAHMYFPRLFT------------------------------YE---FVR--S----DE----------ENGINGILDYGY-AIILSEVARKITEIGYLTEFGIHH----DNEK-NPFNLACDLMEPFRPF----MDERVFQLGA------
90961086_Lactobacillus_salivarius_UCC118                          VARKYFEKLFG-------NK--------------------------YSR--K----DF-----------TPMNAALNYGY-SILLSAVNKEIVSAGYVTYLGIHH----QSQE-NMFNFGSDLMESFRPV----VDYWLADKNF------
406600270_Leuconostoc_gelidum_JB7                                 AARMYFNRLFG-------NE--------------------------FVR-------HE----------AGIVNAMLDYGY-QILMAAVAREIVLQGYLSELGIHH----NGAQ-NDYNLASDLMEPFRPL----VDAVAKKHEH-ET---
386318631_Staphylococcus_pseudintermedius_ED99                    AAKVYFNIVLD-------PG--------------------------YAR-------GD----------DTAQNAAMDYGY-QVLLAIVARTIVSKGYLTELGIKH----HNEF-NIYNLASDFMEIFRPL----IDYLVLKNVE-QQ---
302336019_Olsenella_uli_DSM_7084                                  AARLYFPVLFG-------NG--------------------------FSR-------DI----------DCPINAALNYGY-AVILSAVSREIVSRGYLTPLGICH----RSEY-NQFNLACDLMEPFRPI----VDRLVFDNVE-VD---
15675042_Streptococcus_pyogenes_M1_GAS                            AARIYFNTLFG-------ND--------------------------FSR-------DL----------EHPINAGLDYGY-TLLLSMFAREVVVSGCMTQFGLKH----ANQF-NQFNFASDIMEPFRPL----VDKIVYENRN------
352684362_Acidaminococcus_intestini_RyC-MR95                      AAKVYFNALFG-------NS--------------------------FTR-------DL----------DSPINAGLNYGY-SLLLSLFNREIVSAGYLTQLGIFH----ENTY-NPYNFSCDLMEPFRIL----VDRYVYNMNP------
42525844_Treponema_denticola_ATCC_35405                           AAKAYFAALFG-------AG--------------------------FSR-------SL----------DIPINAALNYGY-SILLSAFNREIIANGYITQLGIFH----DNMF-NPFNLGSDLMEPFRPL----VDAEVFKLNP------
169823756_Finegoldia_magna_ATCC_29328                             AAKVYFNSLFG-------KD--------------------------FSR-------SQ----------DNPTNAMLNFGY-SILLSIFTREITSSGYLTQLGIFH----ENQF-NYYNLSSDLMEPLRPM----IDNIVYFYEP-EK---
117929157_Acidothermus_cellulolyticus_11B                         AARVYWARLFQDKH--------------------------------FRR---------------VPRARDVVNGLLDYGY-AILRGCCLRAVVGAGLAPSLGLWH----RRHD-NPFTLVDDLIEPFRPA----VDKTVIEIVT------
283456134_Bifidobacterium_dentium_Bd1                             AAREYWKRIFIGDG----KR--------------------------FKR---------------IPGDGTGRNAQLDYAY-TILRGFAVKAILSAGLIPTLGVNH----HGRS-NYFCLADDLLEVYRPA----IDYWVAQLEP------
38232679_Corynebacterium_diphtheriae_NCTC_13129                   AAKRYWEIISA-------ED---D----------------------FRR--L------------PGLAATGWNGALDYAY-TVLRGHGMRAICSAGLVGTLGVFH----HGRG-NQFALVDDLIEPFRPA----IDYAVFSIVS------
238924076_Eubacterium_rectale_ATCC_33656                          AAKVYFNLLMG-------TS--------------------------FSR--G----NE----------DILLNSGLDYGY-AVIRGYIARACVGYGLNTQIGIHH----KSEY-NRFNLVDDLMEPLRPI----VDIVAYNSMK-ND---
116627543_Streptococcus_thermophilus_LMD-9                        AAKVYFNELFG-------KQ--------------------------FVRVTQ----KE----------ADVINAGLNYGY-AIMRAQMARIVAGYGLNGLLGIFH----KNEY-NQFNLVDDLMEPFRQI----VDVWVYDNLR-DQ---
57238545_Campylobacter_jejuni_RM1221                              AAALYFKTLFG-------TS--------------------------FSR-------DE----------LCFENSALNYGY-AIIRACIIRAVCISGLLPWLGIKH----DNIY-NSFALCDDLIEVFRAS----VDDCVLKLKG-ES---
291276264_Helicobacter_mustelae_12198                             ASALYFPSLFG-------RG--------------------------FSR-------KD----------FSPINAALNYGY-AIIRGCIARNLVGSGLIPSLGIFH----SNQF-NPFNLADDLIEPYRPF----VDSRVLAL--------
294086112_Candidatus_Puniceispirillum_marinum_IMCC1322            AAQVYWPALFG-------RD--------------------------FRR-------DR------H---KGDVNALLNYGY-AIIRSAMVSAVLSAGLHPTFGMFH----KNKN-NPLCLVDDLMEPYRPL----VDQLVKRLVE------
```

```
189485226_uncultured_Termite_group_1_bacterium_phylotype_Rs-D17   AAKIYWESIFE------------N----------------------FQR--N----KN----------SDIRNSALNYGY-AIVRGVVARSLASSGFITCFGVHH----SNDL-NAFNLADDIIEPFRPF----VDDIAIDIFK-NS---
222109284_Acidovorax_ebreus_TPSY                                  ASAYYFPQVFG-------RS--------------------------FHR-------SQ----------TGWSNAALDYGY-AVMRGACARALVAHGMLPSLGLFH----RSEQ-NAFNLADDLIEPYRPV----VDLHVAQHRP--A---
15602991_Pasteurella_multocida_subsp-_multocida_str-_Pm70         AALIYFQTLFG-------KT--------------------------FKR-------SE----------ENNINAHLNYAY-TVLRSAIARALVLYGWLPQLGIFH----RSEV-NPFNLADDFVEPFRPL----VDLMVWQLWH-TN---
310780383_Ilyobacter_polytropus_DSM_2926                          GAKYYFQKLFG-------DD--------------------------FTR-------RD----------HDIKNFALNYGY-AVIRGRIARSLTSYGYICSLGIQH----KNEY-NAFNLADDIIEPYRSV----VDLWVANNID------
220930481_Clostridium_cellulolyticum_H10                          AAKYYFSVLFT------------N----------------------FKR-------GM----------QDNTNYALNYGY-SILRGAVARTIASYGFIPSIGIHH----RSEL-NNFNLADDFIEPFRPI----VDMWVKQNIN-ED---
34557789_Wolinella_succinogenes_DSM_1740                          AAQVYWRSLFV-NF----KR---------------------WD-------------GG----------LDARNGGLNYGY-AVLRGALGRAIAGAGLIPAFGLHH----AGEL-NSFNLADDLIEPFRPF----VDLLVWGLFL------
90425960_Rhodopseudomonas_palustris_BisB18                        AARFYWGRLFG-------------------------------D---FVR--------DD------EG--DLRNKMLNYGY-AVMRAGVARALVACGFLPAFGLKH----ESAA-NAFNLADDIVEPFRPF----VDGLAWTTLG--D---
209542526_Gluconacetobacter_diazotrophicus_PAl_5                  AARAYWASLFS------------D----------------------FTR--A----NE----------NDRRNALLNYGY-AIMRAAIARACVALGLLPAFGVHH----ASKT-NAFNLVDDLIEPFRPF----VDRMAHDRAL--E---
121608212_Verminephrobacter_eiseniae_EF01-2                       GARRYWGLLFG-------AE--------------------------FRR---------------DQS-GDGLNALLNYGY-TIVRSACARAVVAAGLHPSIGLHH----SNDA-NAMRLVDDLMEPFRPI----VDLKVWQLHK------
60683388_Bacteroides_fragilis_NCTC_9343                           AAAYYWKSLFS-DV----DS--------------------------FTR-------ER------EGI---PPNNLLNYGY-AILRAVVARGLVISGLLPTLGIHH----HNRY-NAYCLADDIMEPYRPY----VDELVFSLIQ------
154250554_Parvibaculum_lavamentivorans_DS-1                       AAQRYWPRLFG-------KD--------------------------FRR-------DR---------DLEGVNALLNYGY-AVVRAATARATVAAGLIPSLGVFH----RNRA-NPFCLADDLLEPYRPY----VDWRVRLLAN------
283778924_Pirellula_staleyi_DSM_6068                              GAALWYGEIGM-RL----SP---G-----------------FH---WER--R----VA------PNA-HDPVNILLNLAQ-TVLYRMTQHAIAQAKLVDTLGFLH----QPRA-GHAALASDMQEPFRHL----MDRAVLETVR------
154174048_Campylobacter_curvus_525-92                             FQKLYYEKWNE-II--ADQR---S-----------------FK---FTV--R----SK------RPP-ADKINSFISYVN-TRIYNVCLSEIYKTELDPRIGFLH----EPNY-RALSLHLDLAEIFKPIL---GDTLIFAMLN-KK---
292492938_Nitrosococcus_halophilus_Nc4                            LHQWYYKAWKL-ID----PV---------------------LD---FTP--R----VR------RPP-NNPINCLLSFLN-QLTYTIVRHELFKTHLEETFSLLH----SPGN-GRSSLSLDLAEPFKPIL---ADSLIFRMLG-RK---
111224947_Frankia_alni_ACN14a                                     FRRSAWGVFDT-IL----PA---------------------WLR--LDG--R----TR------RPP-TNAGNAFISYAN-SVVYGRVLTALRATPLHPAVGFLH---ADTDR-RRNTLALDLAEPFKPLF---AERLLRRAAA-QK---
385809790_Ignavibacterium_album_JCM_16511                         ARNTYYSAWQN-IF----KQ---E-----------------TD---FEK--R----IK------NPP-VGIINSLISFGN-SLLYAVCLSEIYRTRLNPFIGFIH----ETGD-NKHPLVYDISEIFKPVI---VDKVIFRVIN-LN---
150021526_Thermosipho_melanesiensis_BI429                         IRKVYYSAIGK-IV---GKK---D-----------------FE---FVE--R----TK------QPP-KDKINALISFGN-VVLYNIVLAEIYKTSLEPRISFLH----EPNK-RNFSLQLDISEIFKPIV---VDKVVLKLIN--N---
28211139_Clostridium_tetani_E88                                   TWQRFYGEFKN-IL----PE---E-----------------FV---MNK--R----VK------RPP-DNPINALISFGN-TLLYGKTITAIYNTHLDQRISFLH----EPSE-GRFSLSLDISEAFKPVI---VFKTIFDLVN--N---
48477075_Picrophilus_torridus_DSM_9790                            FRNNFLNIVDQ-YL----PN---N-----------------YK---IIK--R----SR------RPP-RNKMNALISYLN-MLLYSITESQIFLTHLNPAISFLH----EPFE-RRNSLSLDVSEIFKPLI---CDRLAIKMVK-LK---
41614813_Nanoarchaeum_equitans_Kin4-M                             IWNLFYQGLDK-IL----KL---------------------YK---LER--R----TR------RPP-KNEANTIVSFAN-TLLYGVTLTEIYKTHLDPTISFLH----ELRD-TRYSLALDLSENFKPII---TFRILIWLVN-QG---
55376278_Haloarcula_marismortui_ATCC_43049                        ARKAYYSIFDD-VL----PD---G-----------------FD---FGG--R----QY------NPP-DNEVNSLISFGN-SLVYANVVSAIRATALDPAVSFLH----EPGE-RRYSLALDIADLFKPLL---ADRIIFRVVN-RN---
389844445_Mesotoga_prima_MesG1-Ag-4-2                             FRQYYYPILAG-LT---------G-----------------QE---FVT--R----IR------QPP-LGILNCMISFGN-SLLYTRTLSKIYQTQLDPTVSFLH----EPCE-MRFSLCLDISEVFKPVI---VDRLIMTLLN--K---
390962116_Thermococcus_sp-_CL1                                    LWNHFYRYFGDALG----VE--------------------------FRR-------TR------RPP-EDELNAMISYAN-AVIYGLALSSAMKVGLDPAIGYLH----AVND-RRFSLPLDLADIFKPLF---VFSTVKSLWG------
13540937_Thermoplasma_volcanium_GSS1                              IWSTYYSAFPF-IY-----K---K-----------------YP---DFK--R----EF------YPP-KDELNAMISYGN-ALLYATVLTKIFITGLNPSISFLH----EPSE-RSFSLALDIADIFKPVI---VERTIANLVNKNI---
289596694_Aciduliprofundum_boonei_T469                            IWGYYYKSLDY-TL----KK---------------------FQ---LER--R----DY------RPP-INELNALISFGN-TLLYSAVLTEIYHTHLNPSISYLH----EPSE-RRFSLSLDIADVFKPTM---VYRHIHDIVN--H---
14590110_Pyrococcus_horikoshii_OT3                                LWKTFYAHFSH-LL----KF---------------------FN---FDE--R----NR------RPP-RDEINAMISYGN-SVLYTVTLSEIRKTYLHPGISYLH----EPRE-RRYSLALDLAEIFKPIV---VFRVILRLVN------
84489236_Methanosphaera_stadtmanae_DSM_3091                        LWQSYYPSFDN-IT----KK---------------------FP---IEK--R----EF------KPP-KNEMNSLISYGN-SLLYTTTLSEIYHTYLHPSISFLH----EPRE-RRFSLACDLADIFKPLI---ISRTIFKLVN-TN---
268317185_Rhodothermus_marinus_DSM_4252                           SREAYYSAWPS-ILADGPGE---A-----------------FV---FEK--R----ER------RPP-SNEINALISFGN-SLCYTTTIRQIHRTALDPTISYLH----EPGA-RRFSLALDLSEIFKPIL---VDRAIFRLVK--T---
18977490_Pyrococcus_furiosus_DSM_3638                             IRQEYYARWDE-SL----PG---E-----------------FR---IGK--R----TR------RPP-KNEMNALISFLN-SRLYATMISEIYNTQLAPTISYLH----EPSE-RRFSLALDLSEIFKPII---ADRIANRLVK-KG---
427708213_Nostoc_sp-_PCC_7107                                     ADNIYYCALTA-LL----SF--------------------------YNG-------SP------YIN-SKRISGLIDLGN-QLLHQYIYTLLNTVGLDPNYAILH----RNNH-NELPLAWDFTAEFRTLM---VDDLVLNFIR------
119720226_Thermofilum_pendens_Hrk_5                               ASNTLWDEVFK-LL----DP---SAARELRER---------YG---FAG--R------------KPGHPDPLNKAISAMY-AVLYTLSTKALVAAGLDPTYGFLH----RTQY--SVPLAFDYAEAFKPLA---VE-AALDLVN-EE---
118497969_Francisella_novicida_U112                               VAKNFFKGFYG-HL----DS--------------------------WQG--R----KP------RIK-QDPYNVVLDLGY-SMLFNFVECFLRLFGFDLYKGFCH----QTWY-KRKSLVCDFVEPFRCI----VDNQVRKSWN--L---
332661941_Haliscomenobacter_hydrossis_DSM_1100                    FGRLFFGLLSD-LL----PV---P-----------------FQ---FEK--R----SR------QPA-LDYFNAGLNYGY-GIFYGMLQGQVQRVGLDPYIGFFH----KERD-GNPVLVYDLIESFRPW----VEQACLHLAK-SA---
326793969_Marinomonas_mediterranea_MMB-1                          ASKQLWAFFQR-IL----EE---D-----------------WG---FTG--R----NR------RPP-KDPINALLSLGY-TYLYSLVDSVNRTVGLYPWQGALH----QRHG-YHHTLASDLMEPWRYL----VEHVVLTLIN-RH---
220934970_Thioalkalivibrio_sulfidophilus_HL-EbGr7                 AARLYFGALAE-LL----PP---E-----------------LG---FTG--R----NR------RPP-RDPFNVLLSLGY-TVLHAHVDTVVRLNGLYPWRGFYH----QPHG-LHPALASDLMEPFRHL----VERVALNVVA--R---
288947693_Allochromatium_vinosum_DSM_180                          AAAAWFRWLTL-WL----SP---T-----------------WG---FQG--R----NR------RPP-RDPVNALLSLGY-TLLGGEMLSVIQQQGLDPARGLLH----ELVP-GRESLVLDLIEPLRPS----VDLVLLGMLD------
431929780_Thioflavicoccus_mobilis_8321                            LAHAWFALLAK-RL----AP---A-----------------WG---FTG--R----NR------RPP-QDPINALFSLGY-TLLGAEVHQGVVAAGLDPSLGFLH----QPVP-GREAMVLDLTEPFRSA----VDHFVLDWIA-VD---
206890539_Thermodesulfovibrio_yellowstonii_DSM_11347              ASSIYFITYSK-IF----PN---S-----------------LK---FVR--R----IK------RPP-KDPVNAMLSLCY-TLLHYEIVREIQLIGLDPTIGFYH----QFEY-GRESLACDLVELFRVN----VDRFVYELFK-AK---
52425690_Mannheimia_succiniciproducens_MBEL55E                    MGAIYFAAFAD-IL----PR---S-----------------LG---FNG--R----NR------RPP-KDPVNAVLSLTY-TLLYSEATLAVYGAGLDPYIGFFH----TLHF-GRKSLSCDLMEPIRPS----VDEWIAECFT-AE---
30248142_Nitrosomonas_europaea_ATCC_19718                         AAARYFTGLQA-VI----PG---S-----------------LH---FHG--R----NR------RPP-RDPFNALLSLTY-TLLHSEITIALHGAGFDPYIGFYH----RLDF-GRESLASDLLEPLRPL----ADRFAFALVH-RR---
121997419_Halorhodospira_halophila_SL1                            AAAAFFRGYGA-LF----AE---A-----------------LG---FSG--R----NR------RPP-RDPVNAVLSLGY-TLAHGDALRAVTAAGLDPAIGVLH----EPAW-GRDSLACDLTEIARAR----VERLTWELFA-SE---
430760780_Thioalkalivibrio_nitratireducens_DSM_14787              AARAYFAGLGA-VF----PP---A-----------------VG---FRG--R----NR------RPP-RDPVNACLSLGY-TLLHFDAVRAAHAAGLDPLIGFYH----RPSF-GRESLASDLIEPLRPV----VDAWVWELFR-AE---
292492856_Nitrosococcus_halophilus_Nc4                            AAAAYFRGYGA-LF----PL---E-----------------LT---FSK--R----TR------RPP-TDPVNACLSLAY-TLVHFEGVQACQGAGLDPFLGFFH----QPAY-GRESLACDLIEPLRPR----IDNWVWRLFR-QR---
34541577_Porphyromonas_gingivalis_W83                             CAQVYWRAFAL-LV----HR---SG----------------HE---FEG--R----EH------KGA-EGLVNQMLNYGY-AILRSYVMKTIVLWQLNPNIGILH----STQD-NKPALCFDLMEQYRAFV---VDRSILALLA--K---
88602029_Methanospirillum_hungatei_JF-1                           GAHAYFQGFSL-AF----DE---E-----------------WG---FLG--R----SQ------NPA-TDPVNSLLSYGY-GMLYIQARQALVLSGYSPYYGAYH----ETYK-KQEALVYDLVEEFRQPV---VDRTVVTFLA-KH---
325980970_Nitrosomonas_sp-_AL212                                  AAIIYWSTWNNCVI----KW---KEKDVPIE----------WR--GFSQ--R----ASGISGKGYKA-THPVNAILNYAY-AILAGQVERALQIVGLDVAVGSLH----ADQD-GRASLVYDLMEPLRPV----IDKIIFAWVT--S---
269798874_Veillonella_parvula_DSM_2008                            ASRTYFSVLGH-VL----SE---P-----------------WE---FSG--R----KR------HPS-PDPVNAILSYGY-SFLEREVRACLLTAGLDVRIGVLH----STNN-RKDSLVYDVMDIFRQDI---IDRFVLKLLN--R---
332668016_Haliscomenobacter_hydrossis_DSM_1100                    IARVYWQGLGA-AL----PR---Q-----------------YA---FNK--R----TR------QPA-EDVFNAALNYLY-GMLYSVVEGGIFAAGLDPYLGILH----VDEY-RKPTLSYDLIEPFRPWI---DQLLVLECLE------
379728404_Saprospira_grandis_str-_Lewin                           ASRTYFQAFAR-VL----PE---K-----------------WG---FAG--R----MR------RPA-YDPFNALLNYLY-GMLYPLVELSLIKAGLDPYMGVLH----TDRY-NRPTLVFDCIEIYRHW----AEWTALQLVF-EK---
327314603_Prevotella_denticola_F0289                              ASRIYFSTINA-FL----PE---Q-----------------FR---FNV--R----SQ------HPA-LDVANAFLNYGY-GILYGRIEGALIKAGVDPYIGVLH----RDDY-NRPVLVYDVIEMYRIW----VDYVVYTLLA-QN---
392395717_Flexibacter_litoralis_DSM_6794                          VSRVYFKALAY-SL----PH---S-----------------YY---FES--R----SM------RPA-HDKFNCLLNYGY-GMMYGYVQGALIRAGLDPSLGILH----ADQYNSAPTLAFDLIEKYRVW----VDTVAFRLCQ------
150398872_Methanococcus_vannielii_SB                              VSKIYFRVLSK-SL----PK---K-----------------YQ---FNG--R----SR------KPA-KDYFNCMLNYGY-GMLYSEIEKICIISGIDPTIGILH----TDGQ-NRKSFVFDYIEKYRNI----IDKIVYESFL-NE---
150401497_Methanococcus_aeolicus_Nankai-3                         ASKYYFEGINH-AL----PK---N-----------------YK---FKE--R----SR------RPA-RDKFNALLNYGY-GMLYPMVEKCLIVAGLDPYVGFIH----ADNY-NKTTLVYDVIEMYRAH----IDRGVVNFIN--S---
284040851_Spirosoma_linguale_DSM_74                               AGRLYFETLSY-VL----PK---E-----------------YQ---FSG--R----SS------RPA-QDAFNAFLNYGY-GMLYGKVEKTLMMAGLDPYVGFLH----RDDY-NQLSMVYDFIEPYRGW----TDEVVFRLFT--A---
257125862_Leptotrichia_buccalis_C-1013-b                          AAKIYYQTLSE-LI----PE---G-----------------FK---FEK--R----SM------HPA-EDEFNAMLNYAF-GILYSKVEKACIIAGLDPYVGIIH----TDNY-GKKSLVFDLIESYRHL----ASRTVFSLFT-QK---
21226661_Methanosarcina_mazei_Go1                                 ASRHYFAALSF-LL----PD---K-----------------WK---FNG--R----SR------NPA-KDEFNCLLNYGY-GVLYSKVEKACIIAGLDPYVGFNH----TDDY-NKKSFVFDLIEMYRIH----IDRTVFNLFS-KK---
218961577_Candidatus_Cloacamonas_acidaminovorans_str-_Evry        ASKNYFSILAE-LI----PD---A-----------------YK---FEG--R----SS------RPA-KDAFNAMLNYGY-GILYSKVERALIIAGLDPYLGLLH----SDNY-NKKSFVFDFIEPYRIL----VDEPVFYIFS------
319790504_Thermovibrio_ammonificans_HB-1                          LGKLMYGYYSQKDL----GE---------------------FS---FKS--R----EY------YPP-PDPVNAVLSLAF-SYFYNLITANVIGSGFDPYMGFFH----RRRG-THASLASDVIEIVRPK----IANFVFELFE-TK---
337288662_Thermodesulfobacterium_geofontis_OPF15                  ATKVMFQKFES-MI----KD---TE----------------FK---FWD--R----NY------RPP-KDEVNALLSFIY-TLGYNLATGIIVLKGFDPFISFLH----FKRG-KHAAFASDLMEIIRPK----LTLFSAELIT-KN---
20094748_Methanopyrus_kandleri_AV19                               AACRYFRALAE-VL----PD---------------------WA---FSG--RR---TR------RPP-RDPFNAAISFGYAGVLLPVLLSRTVAAGLEPFLGFLH----GPRG-RRPGLILDLMEEWRALA---VDVPVLRRFL--D---
386750121_Helicobacter_cetorum_MIT_00-7128                        ISANYWQGISQ-VI----KL---P-----------------FK--------R----VK------KGA-NDAYNSAFNYAY-AILYSKVQVYLLQAGLSLHISFLH----AINS-LKPTLVFDVIEEFRTFM---VDRVIVSMIN--K---
126465226_Staphylothermus_marinus_F1                              AANTYWDGMRI-VL----GK---------------------YG---FKERVK----RG------RGKEVDPVNLCLNIIY-NKLAGTVWKYVLRFSLDPFQGFLH----ARRP-GKLSLVYDLMEPFRPI----ADRFIARFLYRLDQ--
170290046_Candidatus_Korarchaeum_cryptofilum_OPF8                 CANIYWEKIAS-IM----EE---------------------WK---FRG--RE---KRTDL---EGNVIDPVNLCLNVCY-NLLSAQIWKNVLRFGLDPFLGYLH----VERP-GRISLVYDLMEPFRPM----VDRFVFSYLR------
385805115_Fervidicoccus_fontis_Kam940                             AARAYWGSLAV-FL----PE---S-----------------IG---FDG--R----D-------VDS-GDAFNCSLNYIY-ALLYSECFKKLSIAGLDPYGGIYH----VDKA-GRMSLVYDFSDPVKPF----LDYSLTKERQLLLN--
15898245_Sulfolobus_solfataricus_P2                               AAKVYWRGVKS-LL----PK---S-----------------LG---FKG--RM---KR------VSDNLDPFNRALNIGY-GMLRKVVWGAVISVGLNPYIGFLH----KFRS-GRISLVFDLMEEFRSPF---VDRKLIGLAR------
156937921_Ignicoccus_hospitalis_KIN4-I                            AARLYWGLLSE-AF----PE--------------------------FPG--R----EH------EG--CDPYNSAINYSY-GILYSYAFKALSVTGLDPYAGLFH----AIKS-GREALVYDFSEQFKPL----VDRRALPLAH------
124027534_Hyperthermus_butylicus_DSM_5456                         AARDYWDAIAQ-IA----PT---P----------------------FPG--R------------QPRGEDHLNMALSYGY-AILYSIAHDALTVAGLDPYAGFLH----ADRS-GRPSLTYDYADTYKPIA---VDKPLLTAPR------
118431380_Aeropyrum_pernix_K1                                     ASRRYWQCIAE-IL----PG---R-----------------LG---FSG-------RD------RGA-LDPFNAALNYGY-GMLYSIVEKSLLLVGLDPYLGVFH----SEKS-GKPSLTLDAIEPFRAPI---VDRILALKAG--R---
307595807_Vulcanisaeta_distributa_DSM_14429                       AANKYWDAVAT-LL----PS---Q-----------------LG---FHG--R------------NPEGTDPMNLALNYGY-GILYSTMERYLIMAGLDPYLGILH----TIKS-GKPSLVFDIVEVFRPVA---VDKPLITEIN-DI---
305663330_Ignisphaera_aggregans_DSM_17230                         ASRKYWQAIAS-LI----PE---N-----------------LG---FRG--R------------DPDSLDPFNMALNYGY-GILYSVCEKALVLAGLDPYLGVLH----VAKS-GRPSLTLDFVEMFRPIA---IDKPLVIESK--R---
347523053_Pyrolobus_fumarii_1A                                    AARRYWGALAA-LL----PG---D-----------------VG---FQG--R----NH------DS--PDPMNTILNYLY-ALLYAWVHRYLVMYGLDPYVGFIH----ADRS-GRPTLVYDFAELFRVSA---VDRLAYKLLA------
302347808_Acidilobus_saccharovorans_345-15                        AAREYWGFVAS-LL----PR---D-----------------LG---FDG--R----DQ------EG--DDPVNACLNYMY-AVLYSLGLRALLLAGLDPYAGFLH----VDRS-GRPVLVYDYVEQFRAPV---VDEPLVALVT------
218884093_Desulfurococcus_kamchatkensis_1221n                     AARIYWPSLSI-LI----PK---E-----------------LG---FNS--R----DQ------DS--EDLVNTSLNYAY-GILYGESWKVLVLAGLDPYAGFMH----TDRS-GKPVLAFDLIEMFRFT----ADSTLLAMYR--H---
389861365_Thermogladius_cellulolyticus_1633                       AARVYWGLYAT-LL----PR---E-----------------LG---FQG--R----DQ------EG--GDPVNASLNYAY-GILYSESYSALVIAGLDPYAGFLH----EDRP-GSTSLVFDFVEQFRFI----ADAALLDLLR------
15898286_Sulfolobus_solfataricus_P2                               AARVYWQNISQ-LL----PK---D-----------------IG---FDG--R----DV------DG--TDQFNMALNYSY-AILYNTIFKYLVIAGLDPYLGFIH----KDRP-GNESLVYDFSEMFKPY----IDFLLVRALR-SG---
15922975_Sulfolobus_tokodaii_str-_7                               AARYYWHALGR-II----P---------------------------MKG--R------------DPESTDPFNVSFNYAY-AILYSNIQRVLQLVGLDPYAGFIH----KDRS-GKPSLVYDFSEMFKPVL---VDYPLVSLFI--N---
18311772_Pyrobaculum_aerophilum_str-_IM2                          AAEAYWAAYRE-IT----------------------------G---FPG--R----DQ------EG--RDPVNAGLNYGY-GILKALCFKSILLAGLDPYVGFLH----ADKS-GRPSLVLDFMEQWRPR----VDAVVAQLAE------
18311699_Pyrobaculum_aerophilum_str-_IM2                          AASLYWDAVSK-TT----------------------------G---LPG--R----NP------EA--ADPLNLALNYGY-GVLKYAVWRQAVIHGLDPYAGYIH----ADKS-GRPSLVLDLMEEFRPH----VDLLVIRLRP--S---
54296139_Legionella_pneumophila_str-_Paris                        AARIYWETVKE-VN----LV---C-----------------SD---FEG--R----RP------RVN-KDITNSMLNYAY-SILSSWVWQAITNAGLELYAGILH----TSKP-GKPSLVLDLMEEFRPWC---ADRIIFKLHA--R---
34557933_Wolinella_succinogenes_DSM_1740                          MAAIYFDAIGA-EL----LE---E-----------------CD---FTT--R----IG------RGA-EDIMNSALNYGY-AILTSFVWQSIINAGLEAYAGVLH----TNRP-GKPSLVLDVMEEYRAWV---VDRNIIKLRS------
118497353_Francisella_novicida_U112                               AANIYFSSLAK-DK----FL---S-----------------AS---FAN--R----EG------RGS-QEIANSMLNFGY-AILSSYILNAITNAGLEPYLGFLH----QKRP-GKMSLVLDLMEEYRAWV---VDRVVIKLRE------
328954440_Desulfobacca_acetoxidans_DSM_11109                      AGAAYWAMIKT-LL----AP---E-----------------ID---YPG--R----ER------RGA-VDLVNSLLNYGY-GMLYPRVWRALLLAGLNPHISFLH----SFQE-QKPTLAFDLIEEFRSQV---VDRAILSMLT--R---
408403000_Candidatus_Nitrososphaera_gargensis_Ga9-2               ASKEYWDAVSN-LF----PP---E-----------------LK---FPG--R----KT------RGA-TDPVNAMLNFGYKAILFVECWKAVYYSGLDPYAGYLH----ADRP-GKPSLALDLMEEFRQHV---VDRVLFTIFS--K---
330509017_Methanosaeta_concilii_GP6                               GAEIYWRTWAK-LI----PD---D-----------------WS---FPG--R----DY------PAA-RDQINSLLNFGY-YLLEQEVWAAILYAGLDPYAGFLH----ADRP-GKEKLVYDMMEEFRPLV---VDRVVVSLAR------
148656004_Roseiflexus_sp-_RS-1                                    SSARYWAAVAR-II----PD---D-----------------LA---WPG--R----ET------RGA-RDRFNQALNYGY-GILQTQVRTALILAGLDPHAGFLH----ADRP-GKPSLTLDLIEEFRQAV---VDRTLIGLVN-RQ---
386854797_Deinococcus_gobiensis_I-0                               GGVVYWGAVAT-ML----PA---D-----------------LA---FPG--R----LG------RGA-TDPVNMTLNYGY-GILYARVAGVLATAGLEPFAGFLH----TDRP-GKPSLVLDFVEGFRAAC---VDRPVLALLG------
167628642_Heliobacterium_modesticaldum_Ice1                       AATRYWDAVAF-LL----KG---Y-----------------VV---FPG--R----EN------RGA-TDPVNSLLNYGY-AVLEARVLGAVLQAGLDPYAGFLH----VDRP-GKTSLVYDFIEEFRQPV---IDRPVLAMVI--A---
148642083_Methanobrevibacter_smithii_ATCC_35061                   ASVNYWEAVSL-LL----PD---E-----------------IN---FKK--R----NQ------KPE-NDVVNSMLNYGY-AILASEIAKNIVTLGLDPYCGLLH----ADLK-RRQSLIFDLIEEFRQQI---VDKTVFKLIN-TN---
288560430_Methanobrevibacter_ruminantium_M1                       ASHEYWSAFAS-VL----DE---K-----------------WE---FFG--R----SG------RGA-KDPVNSLLNYGY-AVIESEIWKSIYLAGLDPYCGFLH----SERY-GRASLVYDLIEEFRQQI---VDKTVLSIVN--R---
11499462_Archaeoglobus_fulgidus_DSM_4304                          ASKHYWDAISL-VI----PE---E-----------------YR---FNG--R----RGIEIGSPRYA-KDIVNAMLNYGY-SILLAECVKAVELAGLDPYAGFLH----VDVS-GRSSLAIDLMENFRQQV---VDRVVLRLIS-YR---
383319612_Methanocella_conradii_HZ254                             ASAAYWSSIVT-II----PE---E-----------------FE---FKE--R----SG------RYA-TDPLNAMLNYGY-GILEGECRRAIHYAGLDPYGGFLH----VDRP-GKASLVYDLMEEFRQQV---VDKSVLKIFS--L---
294676823_Rhodobacter_capsulatus_SB_1003                          AARIYWPALARLCT----AH---K-----------------GP---FLC--------E------RPA-KTPLNAAINYLT-ALLARDIRAALIGAGLHPGFGVLH----QAQD-RGDAAVYDLMEVFRATL---SEGLAVSLFN--R---
302343124_Desulfarculus_baarsii_DSM_2075                          AAKRYWASLAL-LL----DK---S-----------------WG---FSS--R----QR------RPP-RDGVNMVISYVA-SMLYRDLRCLAARHGLHPGFASLH----GSLD-GKPGCISDLVEEFRAPL---CEGLAVYLAN-NH---
389875622_Tistrella_mobilis_KA081020-065                          AAAIYWPALGA-LI----DH---G-----------------WV---FRR--R----LR------RNA-ADPVNLILDWLA-SLLAREIKALATRHALDPGLGVLH----EVRD-GHDALVSDLIEIFRAPL---AEACTVYLFN-NR---
374998939_Azospirillum_lipoferum_4B                               GAARYWPALAA-QI----EG---E-----------------WE---FEG--R----RR------RPP-PDPVNAVLSYLS-GLLERDVAALIARHGLHPGFGVLH----SPQD-RHDAGIYDLMEEFRAPL---MEGLAVTLFN-RR---
```

```
288957883_Azospirillum_sp-_B510                                   AAAAYWPALGR-CL----EQ---G-----------------WR---LER--R----RR------RPP-PDPVNLVISYLS-AVLHRDLAAFVVRHGLHPGFGALH----SARD-GAQACVSDLIEEFRAPL---VEGLAVYLFN-NR---
116627764_Streptococcus_thermophilus_LMD-9                        LAKSYFYYLNL-LV----PD---D-----------------FH---FNG--R----SR------RTA-EDCFNSALNFGY-SILYSCLMGLIKKNGLSLGFGVIH----KHHQ-HHATLASDLMEEWRPII---VDNTLMELIR--N---
15609954_Mycobacterium_tuberculosis_H37Rv                         AAKAYFTALGH-LV----PQ---E-----------------FA---FQG--R----ST------RPP-LDAFNSMVSLGY-SLLYKNIIGAIERHSLNAYIGFLH----QDSR-GHATLASDLMEVWRAPI---IDDTVLRLIA--D---
348026605_Megasphaera_elsdenii_DSM_20460                          AAKIYFSALGK-LV----DP-----T---------------FA---FEK--R----SK------RPP-LDPFNSLLSFAY-TLIMYELFTAITNEGLHPYVGFLH----TLKE-HHPALASDLLEEWRAVL---ADSFVMSLVQ--H---
296133514_Thermincola_potens_JR                                   ASKHYFKALSL-MV----DR-----R---------------FA---FSG--R----NR------MPP-LDPFNSLLSLGY-TLLLYETYTAVVNKGLHPYAGLMH----RDRQ-GHPALASDLMEEWRPVI---VDSLVMSIVQ--G---
114567264_Syntrophomonas_wolfei_subsp-_wolfei_str-_Goettingen     AARYYYQGLAE-LV----EP-----D---------------FA---FSG--R----NR------QPP-RDPFNSLLSFAY-TLLMYDLYTAAVNRGLNPYASFLH----SIRR-GHPALCSDLMEEWRAIL---ADSLALYVTS--K---
332296542_Thermodesulfobium_narugense_DSM_14796                   IASIYFEALAN-LF----IH---K-E---------------FK---FFG--R----NR------RPP-LDPINALMSFLY-VLLYQRVESVLYECSLDPYLGFFH----TTDY-SKPSLALDLMEEFRAL----VDGLVLRLIN--K---
218442809_Cyanothece_sp-_PCC_7424                                 AAARYFPAFGQ-LI--TNAA---------------------FS---FSL--R----NR------QPP-TDPVNSLLSFGY-TLLFNNVLSLIISEGLSPYFGNFH----YGER-DKPYLAFDLMEEFRAII---VDGMVLRVIN--N---
17228961_Nostoc_sp-_PCC_7120                                      AATLYFQGLGS-LF----TG-----A---------------FV---FEK--R----TK------RPP-TDPVNSLLSLGY-TLLSQNVFSFVQVIGLHTHFGNLH----VPRD-NHPALVSDLMEEFRAQL---VDSLVAYLIN--S---
119357853_Chlorobium_phaeobacteroides_DSM_266                     GAATYFGVFED-FF----DT---G-G---------------FI---FRE--R----VK------RPP-TDPVNAMLSFGY-SLLFNNIFSMARLHRLHPYVGFLH----ADKP-AHPALVSDLIEEFRTL----VDGLVIALIN--K---
119357846_Chlorobium_phaeobacteroides_DSM_266                     ASALYFECYGM-LF----SK---NLP---------------FH---TRSFLR----VR------RPP-TDPVNSLLSFGY-TMLHTNIFSMVQASGLNPYIGFLH----AERK-GNPALVNDLVEEFRTI----VDSLVLYTLN--R---
336116789_Microlunatus_phosphovorus_NM-1                          AARVYFPCLGQ-LF----PA---E-----------------LQ---FTL--R----SR------QPP-MDIANSALSYLY-TLLTAECVTALVAAGLDPAIGFLH----ADDK-GRPSLALDLVEEFRPHV---ADQVVITAAR--N---
37677204_Vibrio_vulnificus_YJ016                                  AAREYFSWFRQ-QF----SA---D-----------------WG---FNS--R----NR------QPP-KDPINAMLSYGY-TLLFHNIRALVEARGLLPHLGYLH----GSQP-RQPSLVLDLMEGYRPWL---VDELVLQLAG--K---
154250301_Fervidobacterium_nodosum_Rt17-B1                        GSKIYFSGFKK-CI----RT---EE----------------LT---FNG--R----TA------HPP-KDEINAMLSLGY-YFLYVEMLLAINAVGLDPYFGNLH----TIDV-SKQSLLFDLVEEFRCVI---IDNFVLNLIN--L---
38505760_Synechocystis_sp-_PCC_6803                               AANLYFQAWAD-CI--ENSN---------------------FI---FFG--R----SR------RPP-GNPVNAMLSFGY-QVLWNHLLTLIELQGLDPYQACLH----QGSE-RHAALASDLIEEFRVPI---VDSLVLYMIN--R---
83592490_Rhodospirillum_rubrum_ATCC_11170                         GAHQYFAGLRT-AF----VG---G-----------------IP---FLG--R----AQ------RPP-PDPANSLLSFGY-VLLGNRLTGLLEARGVDPCLGFFH----DLRP-GRPSLALDLLEELRHPV---VDRLALRICN--L---
261415948_Fibrobacter_succinogenes_subsp-_succinogenes_S85        GARIYFENFGR-CL----AC---G-----------------LD---FPG--R----KY------HPS-TDPVNALLSFGY-TLTARSLESLLESYGMDPMLGYLH----EPSY-GRNSLAQDMLEEFRHPL---VDRLVLFLFN--R---
76802276_Natronomonas_pharaonis_DSM_2160                          AAERYFDRLDE-TL----VD---G-----------------WT---FEK--R----TK------RPP-EDHINSLLSLTY-VFMKNEVMSALRQYNLDPFLGVLH----ADRH-GRPSLALDLQEEFRPIF---CDALVTRLVN-RR---
312114617_Rhodomicrobium_vannielii_ATCC_17100                     ATAFYFSAWPH-MF----AA---RAGD--------------VE---FEV--R----SR------RPP-QNAVNATLSYAY-AVLSAECVCALAAVGLDPRLGVFH----QPRS-GRASLALDLMEPFRPLI---ADQAVLTGFN--T---
219683006_Bifidobacterium_animalis_subsp-_lactis_AD011            AAALYFSQFHH-LI----SV---NKRNE-------------WP---WLE--R----MR------HPA-PDPLNALLDYTY-SLLLSDCIRAIVSCGLDAHAGFLH----SSKR-NKPALALDLMEEFRAPI---ADSVVQTVVN--N---
116748794_Syntrophobacter_fumaroxidans_MPOB                       AANLYFEVFPL-LV--RVPG---------------------FE---FNG--R----NR------RPP-LDPLNALLSFVY-TLLTQEVLTAIKVVGLDPYLGCLH----AVDY-GRPSLACDLVEEWRTFL---GDRLVLALVN--R---
317051217_Desulfurispirillum_indicum_S5                           CAATYFQAFGS-IF--EDRE---------------------FQ---FRK--R----SR------RPP-LDPINAMLSLGY-TVLLAEVITALQSTGLDAYVGYLH----SAEY-GRPSLACDLQEEFR-FV---VDELVVKLVN--L---
328949009_Treponema_succinifaciens_DSM_2489                       AARKYFEVFKF-NL----QD---E-----------------WA--EFKT--R----SK------NPP-RTNVNAVLSFLY-TLLMYRVESAIESQGLDSCCGNLH----AVNY-GKTALVFDLMEEFRSPV---CDSVCCSLFN--L---
333996012_Treponema_azotonutricium_ZAS-9                          AARNYFSVFRH-NL----IP---E-----------------WA--QFPR--R----SK------NPP-LTNINSVLSFLY-TLLMYRVETAIEITGLDPMAGFLH----AAEY-GKNALVFDLMEEFRTPI---ADTLCCALCN-LG---
313673559_Calditerrivibrio_nitroreducens_DSM_19672                ISSIYFSAFDCLLN----KE---------------------IK---FEK--R----TR------RPP-QNEFNALLSFGY-TILSNLIRTEVYITGLDPYFGVLH----AEDY-GRPSLVLDLMEEFRAIA---VDMPAINAVN--K---
218961167_Candidatus_Cloacamonas_acidaminovorans_str-_Evry        AAKNYFAAFPD-LI--ANPD---------------------FP---FSG--R----NK------RPP-KDEVNAMLSLSY-TFLMNQVMCAAYICGLDPYYGALH----DLDY-GRQSLVLDIMEEFRPL----IDNMVISLIN--R---
294828260_Leptospira_interrogans_serovar_Lai_str-_56601           SARAYFSGLPA-LLKNSDPF---------------------LV---PNG--R----SK------RPP-KDPFNATLSFLY-SLLYKSVRQAIIAVGLDPSFGFYH----TPRS-SAEPLVLDLMELFRVSL---CDMTLIGSIN--R---
347756303_Candidatus_Chloracidobacterium_thermophilum_B           AGQAYFAAFAS-LI----RT---SPSEPGSPAG--------FE---FVG--R----NR------RPP-RDRVNALLSFFY-ALVYKDCLAALIAVGLEPSFGFLH----TPRS-AAHPLALDLMELFRVPL---ADMPVVASIN--R---
162457462_Sorangium_cellulosum_So_ce56                            AARGYFGALAALVH----PDAGDA-----------------LR---PRG--R----SR------RPP-EDRFNALLSFGY-TLLYRDVLSAILRVGLEPGFGVLH----QPRS-AAFPLALDLTELFRVPV---VDMAVLGAVN--R---
86742032_Frankia_sp-_CcI3                                         AARRYIATWPH-LL----TP---HATV--------------TA---PAG--R----TS------RPA-TDPVNAALSFGY-GILRIAVHGALDHVGLDPHIGYLH----GIRP-GKPALALDLMEEFRALL---VDRLVFTAFN--Q---
433631663_Mycobacterium_canettii_CIPT_140070010                   GAKRYFAEWSS-LL--SPQG---------------------FF---FTA--R----IR------RPP-MDPVNALLSFGY-AMLRTRCQAAAEHIGLDPQVGFLH----PVRP-GRPSLALDLMEELRAPF---VDRLVLTLIN--R---
15612904_Bacillus_halodurans_C-125                                AAINYNKVFDQ-MILQQKEE---------------------FA---FHG--R----SR------RPP-KDNVNAMLSFAY-TLLANDVAAALETVGLDAYVGFMH----QDRP-GRASLALDLMEELRGLY---ADRFVLSLIN--R---
108763792_Myxococcus_xanthus_DK_1622                              AARDYFESFPA-LLKKSAQG---------------------FE---FDG--R----NR------RPP-RNPLNAMLSFGY-ALLAQDCAGALTGVGLDPAVGFLH----EDRP-GRLSLALDLMEEFRAPV---VDRLVFSLVN--R---
344996859_Caldicellulosiruptor_lactoaceticus_6A                   ISKIYFEGFKK-VL----PP---E-----------------FC---FEE--R----NR------RPP-RDVVNALLSFTY-TIFLTQCITGAFLSGLDPYFGYLH----SNVY-GRPALALDLLEEIRP-V---ADAFVLSLLR--N---
85858452_Syntrophus_aciditrophicus_SB                             GSALYFEAFGR-GF----IE---N-----------------QG---FFR--R----VR------RPP-TDPVNSVLSLLY-TFLMNRVYAAVRVAGLDPYPGFLH----ALDY-GRYSLVLDLMEEFRTII---ADTLTLSLFN--L---
148657121_Roseiflexus_sp-_RS-1                                    AAAAYFGAWRA-SL----PP---V-----------------WG---FGG--R----AF------YPP-PDPINAMLSFGY-TLALHDVITAVQITGLDTYLGVFH----VIEP-GRPSLALDLLEEFRPLI---VDRLVIDLVR--T---
159898907_Herpetosiphon_aurantiacus_DSM_785                       GAAAYWGTWSA-VF----KT---A-----------------WG---FAG--R----AY------RPT-PDPLNALLSFGY-TLLLNDLMTAVQALSFDPYLGVFH----TVQF-GRPSLALDLEEEFRPCI---VDRMVLDVLD-AG---
17227877_Nostoc_sp-_PCC_7120                                      CARTYYQALKH-WF----PT---Q-----------------WN---FNG--R----NR------RPP-LDPINALLSWGY-GVLLARVFSACVQAGLDPYLGFFH----AIEP-YRPNLVLDLMEEFRPVV---VDQAVISLIQ--S---
268318378_Rhodothermus_marinus_DSM_4252                           ASRRYFSVFGE-ML----RA---EAYAP-------------TGWPAFPG--R----HR------RPP-TDPVNATLGYLY-ALLLGNVVAACALAGLDPYVGYLH----APAY-GRPSLALDLMEEFRAPA---ADRLALRLFN--R---
55978328_Thermus_thermophilus_HB8                                 GSRVYFQGLAR-LL----GP---------------------YG---FGG--R----TR------RPP-RDPVNAALSYGY-ALLLGRVLVAVRLAGLHPEVGFLH----AEGR-RSPALALDLMEEFRVPV---VDQVVLSAFR--R---
297622675_Truepera_radiovictrix_DSM_17093                         CARLYYAALQR-PL----AA---------------------YG---FTG--R----NR------RPP-RDPVNAALSYGY-ALLLAQVQLATLSAGLHPEVGLLH----AESR-RNPAMSLDLMEEFRVPV---VDLTVFRAFL--S---
328953000_Desulfobacca_acetoxidans_DSM_11109                      AAHTYFQGLAR-MV----RH---G-----------------FG---FSG--R----QH------HPA-PDPVNALLSLGY-TLVYNEISSLLDGMGFDPYMGFYH----QPRY-GHATLASDLLEEFRALL---VDRLTLSLIN--N---
337286709_Thermodesulfatator_indicus_DSM_15286                    AAQVYFDVFGK-LF----QV---EG----------------LK---FPG--R----IR------RPP-PDPINALLSLGY-TLLFAQIWSVAESTGFDPYLGFLH----VPEY-GRPSLVLDLAEEWRPLI---VDSLVVRLFN-WK---
328953422_Desulfobacca_acetoxidans_DSM_11109                      GAASYFGCFQH-LF----RV---DD----------------IT---FSG--R----NR------RPP-QDPVNVLLSLGY-TLLANVVQTQVHVAGLDPYLGCLH----SVEY-GRPSLVLDLMEEFRPVL---VDAVVVQVVN-KR---
38505684_Synechocystis_sp-_PCC_6803                               AAREYFASWQE-ML----GH---E-----------------WT---FTG--R----FR------RPP-TDPVNALLSFGY-GLLRTQVTAAVHIAGLDPYIGFLH----ETTR-GQPAMILDLMEEFRALV---ADSVVLTVLK-QR---
134299482_Desulfotomaculum_reducens_MI-1                          GSRDYFSVFGL-LIKDRVP----------------------FD---FNK--R----SR------RPP-EDPANALLSFSY-SLLLKDVITAVQVVGFDPFIGFLH----RSDF-GRPALALDIIEEFRPVV---ADSVVLTALN-KG---
17229060_Nostoc_sp-_PCC_7120                                      GSAAYFGCFNQ-LI--RVDN---------------------FS---FHT--R----NR------RPP-IDPVNSLLSLGY-SLLRHDVQGALNIVGFDPYLGYLH----TERY-GRPSLALDLMEEFRPLI---VDAVVLTAIN-RR---
302391602_Acetohalobium_arabaticum_DSM_5501                       GSRYYFSQFNEVLN----DS---------------------FN---FNG--R----NR------RPP-KDPVNCLLSFGY-SLLLNDVLAALYLVGFDPYIGFFH----SSQY-GKPALALDLMEEFRPVI---VDSVVKSVIN-KK---
159898755_Herpetosiphon_aurantiacus_DSM_785                       GSAAYWDCWGM-LL----NQ---P-----------------WE---WHG--R----RR------RPP-PDPVNALLSYGY-VILTSQVLSQLAIVGFDPYIGFLH----QSSF-GKPALALDLMEEFRPVI---VDSVVLTVLN-TK---
320161859_Anaerolinea_thermophila_UNI-1                           ATAAYFSGLRQ-LF----NP---E-----------------WK---FDA--R----LR------RPP-PDPVNVLLSLGY-TLLAQDCVAAVQAVGLDPYAGFLH----EVAY-NRPALGLDLLEEFRPL----VDGVVLWICH-SG---
156741961_Roseiflexus_castenholzii_DSM_13941                      ATARYFAGVRA-LI----GA---E-----------------WR---FEA--R----IR------RPP-PDRVNALLSFGY-TLLVHKMLGAVEAAGFDPYLGYLH----HIDY-GRPSLALDLIEEFRPIL---VDSLVIRCCN-DG---
383762167_Caldilinea_aerophila_DSM_14535_-_NBRC_104270            GTAIYFGAWKS-LL----KP---P-----------------WQ---FHT--R----QR------RPP-PDPVNVLLSFGY-TLLSQNLLGAVLAAGLDPYVGFLH----QLSY-NRPNLALDLMEEFRPVV---VDSVVLRCLN-NS---
46255267_Thermus_thermophilus_HB27                                AADLYFRAFAE-LL----PE---E-----------------FA---FGE--R----SR------RPP-RDPANSLLSLAY-TLLAKECESALLVAGLDPYVGYLH----EVRY-GRPSLALDLMEEFRSVL---ADSVVLSLLN--N---
134101617_Saccharopolyspora_erythraea_NRRL_2338                   AARTYFAGLVE-TF----RT---DHRLPGP----------AFD---TMG--R----TR------RPP-RDAVSCLLSFLY-CLLIKDITTACYALGLDPYFGFYH----QPRH-GRPALTLDLAEEFRPLI---ADSTALTLIN--N---
378719041_Gordonia_polyisoprenivorans_VH2                         AARVYFENFSG-LL----RG---ALDQIAS-----------EFD---TNG--R----KR------RPA-PDPVNALLGYVY-SLLVKDLVAVCIGVGLDPYLGVYH----RPRY-GRPALALDLMEEFRPLL---ADSAVVGMLN--N---
320102061_Isosphaera_pallida_ATCC_43644                           IARLYFQNFNG-LI----KP---DDNHTPHQPPSDDPRALTFH---FDQ--R----NR------RPP-RDPVNALLSLGY-ALLTKDMTIACHAVGLDPYLGFYH----QPRH-GRCGLALDLMEPFRPLI---ADSAVLFAIN--T---
268318441_Rhodothermus_marinus_DSM_4252                           AARLYFRQFSR-ML----RH------------------RIAFD---FKN--R----NR------RPP-KDPVNAMLSFLY-ALLLKDAMCALMATGLDPYRGIFH----QMRF-GRPSLALDLMEEFRPLI---ADSVVLRLVN-TG---
39995168_Geobacter_sulfurreducens_PCA                             SAGRYFQHFSR-ML----RG---GDGEGM-----------GFD---FTT--R----NR------RPP-KDPVNALLSFAY-AMLTREWTVALAAVGLDPYRGFYH----QPRF-GRPALALDMMEPFRPLI---ADSTVLMAIN-NG---
83591519_Rhodospirillum_rubrum_ATCC_11170                         AAAVYFRAFPQ-LF----TQ---AVTTLP-----------AFA---FER--R----NR------RPP-ADPVNACLSLCY-AVLTRTLSSALSIAGLDPWKGFYH----TERP-GRPALALDLIESFRPVL---ADSTVLMVLN-NG---
50085564_Acinetobacter_sp-_ADP1                                   WAKALYKELAKGFN---------------------------ID---FKR--DEGKNSH------DSL-TDIANNYLDHGN-YIAYGYAAVALNGMGISFALPILH-----GKT-RRGGLVFDLADLVKDAF---VMPTAFICAS--K---
77166205_Nitrosococcus_oceani_ATCC_19707                          FAKKLYKFWAN-CT----ET---------------------EN---FTR--D----PG------KRDFNDPFNSYLDHGN-YLVYGIAAAVLWVLGIPHSLPVIH-----GTT-RRGALVFDVADIIKDTC---VMPIAFQHAA------
15602176_Pasteurella_multocida_subsp-_multocida_str-_Pm70         LTTKLYALANQATM--NEPD--------------------------FKR-------SE------RGKSIDLVNQFLDHGN-YLAYGLGATACWVLGLPHGLAVLH-----GKT-RRGGLVFDAADIIKDSM---ILPQAFISAQ------
38234755_Corynebacterium_diphtheriae_NCTC_13129                   IMRNTYRENAKKAG-------IRG----------------------FKR---------------DTKAADPVNVGLNISN-SILYGAAATVCTAIGVNPALGIIH------RG-DTRSLLFDLADLYKASI---VIPLVFSHAK------
21674788_Chlorobium_tepidum_TLS                                   RVRKLYEEMAVKYK----VG---------------------WKGRRFEP-------GK------FEM-SDTTNKILTASN-AALYSIILSAVHSMGYSPHIGFIH----SGSP---LPFIYDLADLYKQQV---SIDLAFSLTA-DM---
53804737_Methylococcus_capsulatus_str-_Bath                       RVKETYQLLANQHG----VE---------------------WKGRRYDR--N------------NPNAADIPNQAINHAA-TFVEAAADVAVAAVGGLPPLGFIH----EESS---NAFTLDIADLFRAEV---TLPLAFSVAK------
209966592_Rhodospirillum_centenum_SW                              RIRHSYELIAAQAG----IV---------------------WRGRRFDR--S------------DPDGDDLPNQAINHVV-TAVEACVAIAVQATGTLPPLGFLH----EDSA---KSWTLDICDLYRTSV---TVPLAFRCVK------
89885834_Rhodoferax_ferrireducens_T118                            RVKEIYVNLADKLG---------------------------MV---WQG--R----EE----------KSALNTSIGFAT-SCLYALCEVAILAAGYHPGIGVVH----SGNP---RSLVFDLADTVKFKT---VVPLAFEIAA------
16130662_Escherichia_coli_str-_K-12_substr-_MG1655                RVRATYALLAKQYG----VT---------------------WNGRRYDP-------KD------WEK-GDTINQCISAAT-SCLYGVTEAAILAAGYAPAIGFVH----TGKP---LSFVYDIADIIKFDT---VVPKAFEIAR-RNP--
68535718_Corynebacterium_jeikeium_K411                            RMKKVYLAEAERTG----VP---------------------WKKRMYDP-------QN------FSS-GDPINRTLTAAN-AALYGVAHAVIASLGFIPSLGVIH----QGTE---RAFVYDIADLYKAEI---SIPAAFDAVA-EN---
86738746_Frankia_sp-_CcI3                                         RIKKIYRDHATRTG----VT---------------------WNKRVYTH-------GD------PFADSDDINRLLSAGH-SCLYGICHAAIVGIGASPALGFVH----TGAA---TSFVLDIADLYKADY---TIPLAFDLAA------
51891801_Symbiobacterium_thermophilum_IAM_14863                   RVRTAYARASETYG----VP---------------------WSGRNYDRNDW----DR----------SDPVNRALSAAN-ACLYGVCHAAIVAAGYSPALGFIH----TGKM---LSFVYDVADLYKTEI---TIPAAFMEVA-KGV--
29834079_Streptomyces_avermitilis_MA-4680                         HVKAHYRVLAQQYK-------------------------IGRFRRAYDPASW----DT----------QDPVNLALSAAN-TCLYGIVHAAILALGCSAALGFVH----NGNQ---QAFVYDIADLYKADL---TIPLAFSLHA------
```

Continued

```
407462288_Candidatus_Nitrosopumilus_koreensis_AR1                 QR----IKLKEYSFTKDG----------------------------------------T-VV----------------LEYSLIKRFLEMLERQFSQERKYSFRHGKKTKDGLKSVQEITVVKIIIQ---NLVEYS--TGK---QKSLES
21229351_Methanosarcina_mazei_Go1                                 -G----AMESK-DFIRTE--------------------------------NY------N-LR----------------LKPTGARKIVNEFSNT-LNKKVSYQGK---------ESTWSYVIFLKVR---ELAHYL--TSK-KEKLDFTK
147921327_Methanocella_arvoryzae_MRE50                            -K----KLKKN-DFIVTE--------------------------------NY------H-LR----------------LREATAKKLIERIKLN-FNLKAPYKNQ---------NYTYENILIDQVQ---QFANFI--QDK-NKTVEFTT
124485266_Methanocorpusculum_labreanum_Z                          -G-----RLDG-AYEVTT----------------------------------------R-CL----------------LSNELKEEFMKRLHGS-IN---------------------ITLIEENVN---RYAESV--KDG-RDIVFH--
307352568_Methanoplanus_petrolearius_DSM_11571                    -E----YLTDDMYETNTK----------------------------------------R-CH----------------LDQGTLRLLSNALKSS-ID---------------------QRKIEENVM---SYRDSL--INR---------
154150254_Methanoregula_boonei_6A8                                -S-----LNVG-DFEITS--------------------------------N-------R-CI----------------LSDNLIQNLMSSFRKS-ID---------------------NQKLDCQVH---NFLQAL--SNK-EDLTIL--
219852754_Methanosphaerula_palustris_E1-9c                        -G-----ITAS-DYEIGE--------------------------------S-------R-CT----------------LSDALIRRMLLHLQTS-IE---------------------VTAIGRQVE---MLADAL--IRN-REFQILDP
126178701_Methanoculleus_marisnigri_JR1                           -E-----ISSG-DYECGE--------------------------------K-------R-CY----------------LGGDLSSRLAAALRDS-ID---------------------QARIDAQVR---IVRDAL--LAN---------
397780289_Methanoculleus_bourgensis_MS2                           -E----EISED-DYECSE--------------------------------K-------R-CY----------------LDANLSTRLVQEFRDS-ID---------------------QSRVDTQVQ---ILRDAL--LTN---------
156937861_Ignicoccus_hospitalis_KIN4-I                            -Y----KIELEGDF----------------------------------------------------------------LNKESLQKVTQLFYSI-FESK---------------KYSVRKWVYAKAW---QLRDAI-ESGK---------
70607609_Sulfolobus_acidocaldarius_DSM_639                        -L----KLDNS-------------------------------------------------GY----------------LSKESLGHLTKHLYET-LHKG---------------KRSVRGQIYSMGM---KTKNFI--AEG---------
187736488_Akkermansia_muciniphila_ATCC_BAA-835                    -G----KTEER-----------------------------------------------A-GE----------------ITREFRQHITATLQAS-VMYRDK-------------QLPLKAAVEAVCR---SFRKAV--LAG-QSE-----
436839745_Desulfovibrio_hydrothermalis_AM13_-_DSM_14728           -N----IITAK-DFEPVT--------------------------------AK------K-FK----------------LNGRSMGKILESFEKE-MCIRIATE-----------AGTFAQLLDAQVQ---NLRNWV--MDR--DELKFFR
71894593_Mycoplasma_synoviae_53                                   ------IIEKNYDF------------------------------------------------------------------INFKKKVFLIFEEK-ILLNKS-------------PISVNEYI---CK---LIENFI--NKD---------
193216853_Mycoplasma_arthritidis_158L3-1                          ------ISRGEPNF------------------------------------------------------------------QEFKEELFTIFYKH-FNCQGE-------------NLMLIDCIDKLVV---LVLSDH-DIKG---------
```

```
187250661_Elusimicrobium_minutum_Pei191                           ------VLTQE-------------------------------------------------ES----------------LSPSAKKYMASVIESN-LYFEDG-------------AFNLTAGIQKYIQ---SYIAFL--EER------ENR
325972002_Sphaerochaeta_globus_str-_Buddy                         -Y----GVIED-------------------------------------------------------------------LESREKKELMRITARE-VVFEEQ-------------DVEFSYALRLYVL---SYLSFL--SRN------RDA
47458867_Mycoplasma_mobile_163K                                   ------LHKDE-------------------------------------------------TF----------------FDVKIKDELINVLNKK-VIYKGK-------------YFYVNDTFDQIAD---NLVNKR--------------
269123825_Streptobacillus_moniliformis_DSM_12112                  ------RFEK------------------------------------------------D-DE----------------LKTDTRKELISLLNAN-IKVNEK-------------EYTVLYSMELLIK---SYLNFL--EEG---NEMLAI
57865886_Staphylococcus_epidermidis_RP62A                         ------NIDT--------------------------------------------------------------------FDINEKKLLLNVLYER-CIIDKK-------------VVRLLDSVKIVIQ---SLIRCY--EEN--------T
319957207_Nitratifractor_salsuginis_DSM_16511                     ------LCDEE--------------------------------------------------Y----------------LTPEFKKAIISIFDSEYLIYDAG-------------RSTLRTVATRYIQ---GFKAVV--RSRKIENLHFPT
310286727_Bifidobacterium_bifidum_S17                             ------------------------------------------------------------EH----------------LDSKVKVELMRFMRSD-VPEL---------------GTTITQLISVFVR---DALRYL--SGE------ECL
90961086_Lactobacillus_salivarius_UCC118                          ------------------------------------------------------------LE----------------FTPDIKYGLVDLLNLE-IKYNKK-------------QMLLKNAITKYVR---DVIEYLNQETVEFDMEVEFT
406600270_Leuconostoc_gelidum_JB7                                 -A----------------------------------------------------------------------------LNLDIKLALVDIINQV-IEINGQ-------------QMLVVNAITILVR---DAMNYL--NSG---------
386318631_Staphylococcus_pseudintermedius_ED99                    ------------------------------------------------------------------------------FETEQKRAILDILNRK-LKIDNK-------------KYYFLNAVEIYVD---SLFKYL-SNGN------REV
302336019_Olsenella_uli_DSM_7084                                  ------------------------------------------------------------------------------FTKETRRLLADVLNVS-ICYRDG-------------KYKVSSVISLYVQ---DCLNAL--CRR------LAV
15675042_Streptococcus_pyogenes_M1_GAS                            -Q----------------------------------------------------------------------------PFPKIKRELFTLFSDT-FSYNGK-------------EMYLTNIISDYTK---KVVKAL-NNEG---------
352684362_Acidaminococcus_intestini_RyC-MR95                      ------------------------------------------------------------TT----------------FAKDEKREIINLFQEI-LYIDDS-------------RQFLANAINIYAN---SIFQALEKCDT---------
42525844_Treponema_denticola_ATCC_35405                           ------------------------------------------------------------QK----------------FEHEEKLKIVSVINKK-VLINNK-------------EHYLNKAIEIFVH---SIFDAL--NEK-------DI
169823756_Finegoldia_magna_ATCC_29328                             ------------------------------------------------------------------------------FDKEEKIATLKIMDEE-VFIGKT-------------RYTLLNAVNEYCK---SVFLAL-EKGE------EQL
117929157_Acidothermus_cellulolyticus_11B                         -------AGAS-----------------------------------------------G-------------------LDRPTKRLLVAVLDHQ-FDAS---------------GATVGTAVERFAQ---QVGRYV--EGE--------I
283456134_Bifidobacterium_dentium_Bd1                             ------EDGPS-------------------------------------------------------------------DKNVKRYLADSVNQQ-FDSSGLTIPSS--------ISDFAQQFGLYCE---KKIDAL------------EV
38232679_Corynebacterium_diphtheriae_NCTC_13129                   -------NSQE-------------------------------------------------------------------LDKETKRQLVAAVEEP-FNSA---------------GQSIPTVFTAFAQ---QYGRYV--EGD---VEKLVP
238924076_Eubacterium_rectale_ATCC_33656                          ------------------------------------------------------------KY----------------FTPEHRRKLVSILNMK-IMYREK-------------KMYMCNMIENYIE---QYASLI--MKRTDNIVFPDV
116627543_Streptococcus_thermophilus_LMD-9                        ------------------------------------------------------------EF----------------LKYEYRLGLTDLLNAK-IKYGKE-------------TCSVTVAMDKYVK---GFIKYI--SEK-------DS
57238545_Campylobacter_jejuni_RM1221                              ------------------------------------------------------------EF----------------LSKDDKRALIGNLQSK-INFDGQ-------------NYPLNRAINHYVA---NFKNAL-LYED---------
291276264_Helicobacter_mustelae_12198                             ------RIENE-------------------------------------------------------------------LTLSHRVELVNLLNTK-VLLQGK-------------FYPFHRSITRMVW---SISNFICDRGKMIELPILDK
294086112_Candidatus_Puniceispirillum_marinum_IMCC1322            ------RGHSN-------------------------------------------------------------------VTTEVKRCLASIVTSD-QAVVGA-------------TSPLFQHMSQLSY---SLWENL--DGQ------KTT
189485226_uncultured_Termite_group_1_bacterium_phylotype_Rs-D17   ------ETSKE-------------------------------------------------------------------LLKEHKQQLLSVLSTD-CIFNDK-------------NSNILTAASQTTESFMIASREL--NPD--------K
222109284_Acidovorax_ebreus_TPSY                                  ------DEDAE-------------------------------------------------------------------LQPSDKVALVGLLNVD-VAMPRG-------------QMSVLASIEQAAE---SLARLY--DGG--SEQVLEL
15602991_Pasteurella_multocida_subsp-_multocida_str-_Pm70         ------QLTSG-------------------------------------------------------------------LSPNTKQKLISLLHHQ-MSFQEE-------------TFSVLAAIDRTVG---SFQAAL-TTPNANLLKLPEI
310780383_Ilyobacter_polytropus_DSM_2926                          ------KLSDE-------------------------------------------------------------------LTTDNKVSLVEILNAD-VKLTGK-------------KYSVNISIDEVIK---RYTKCI--RGKGGNLILPSI
220930481_Clostridium_cellulolyticum_H10                          ------------------------------------------------------------TL----------------LTPKHKLNLISLLGYE-CVFEGK-------------IISIRSAIEKVIS---SFSSSC--AKN--DYSLLKL
34557789_Wolinella_succinogenes_DSM_1740                          ------ERTGD-------------------------------------------------DP----------------LTREERMRIAQILGES-CIMTERYETLLSATQIIA-RSLVSAGEQKESR---RLILPW--MGE-------DD
90425960_Rhodopseudomonas_palustris_BisB18                        ------RVAKNGD-----------------------------------------------------------------LTLDDRRAMAGVLLMN-GRVGDA-------------KVSLLVAAEMASA---SLCRAL--EFE---------
209542526_Gluconacetobacter_diazotrophicus_PAl_5                  ------HVGDT-------------------------------------------------------------------LSIEDRRQMSTILNDN-AAIGRE-------------RMTVLAATEAVAM---SVVRAI-EHGS------AAL
121608212_Verminephrobacter_eiseniae_EF01-2                       ------AGESH-------------------------------------------------------------------VTPDTKRALVRVLYDD-MQTAAG-------------VTPVMVCMQRLAT---SLAQVY--LGE-RAKLDLPL
60683388_Bacteroides_fragilis_NCTC_9343                           ------EYGKN-------------------------------------------------AE----------------LTKEMKTRLLTVPTLE-VIIGGK-------------RSPLMVAVGQTTA---SLYKCF--NGE------LRR
154250554_Parvibaculum_lavamentivorans_DS-1                       -Q----MGEEA-----------------------------------------------P-SL----------------DDRDTRAALLSIFNET-VLVGGR-------------RMPLLLALHASAA---SLCRAL--TGG---EAALAL
283778924_Pirellula_staleyi_DSM_6068                              ------RIRPE-EFEPDE------------------------------------RGPFP-VR----------------MKPRALRESLAWIHRI-LALPCN-------GAGQSSPTAYRLQIFKQVR---SFRRFL--MDP------GG-
154174048_Campylobacter_curvus_525-92                             ------EITAK-DFQTDAGR----------------------------------------IK----------------FSNDAIQKIEMKMISR-LSETIALNGQ---------NLTWRQVIRREAN---QLKKCI--CED---------
292492938_Nitrosococcus_halophilus_Nc4                            ------MLDAS-WFEQYEGV----------------------------------------CL----------------LSETGRRNIAEQFSIR-LEERYQ-------------DRSFREWIYREAL---AIERHF--LGV---------
111224947_Frankia_alni_ACN14a                                     ------TLRPS-DFEADVGA----------------------------------------AS----------------LSREGRKKVSMMMREE-LAATVFHRTLKR-------KVSYEELIHLEAL---KLVRLC--LED---------
385809790_Ignavibacterium_album_JCM_16511                         ------IIKPD-DFKQTD---------------------------------------KG-LR----------------LKDDARKKFVEEFENK-LSSIIHHKRINR-------RISLRSLIRMECY---NLINYL--TGK--------L
150021526_Thermosipho_melanesiensis_BI429                         -G----MIKKK-DFENVK---------------------------------------DG-VY----------------LSKEGKKKFVQELENR-LEKKIFLYKDQ--------RVTLKTVILHECY---KLIRHL--KGE---------
28211139_Clostridium_tetani_E88                                   -K----RIQVSKHFDKKL------------------------------------N---Y-CL----------------LNDEGRNIFITAFEER-MESIFL-NEKL--KR----KISYKTAIKLDCY---KLIKFI--LEN---K-----
48477075_Picrophilus_torridus_DSM_9790                            ------IIKDS-DFLEEN----------------------------------------G-VF----------------LNENGRKKVVKMFDDK-MQETVYSSSLKR-------YVSYKYLIRLELY---KIEKYI--LEE---------
41614813_Nanoarchaeum_equitans_Kin4-M                             ------IIKDT-HFVKGL---------------------------------------NG-VL----------------LNEQGKKLVIKEFNKR-LDETIKLKSGL--------KRSMRWYIKAQAY---NLERFL-LDGR---------
55376278_Haloarcula_marismortui_ATCC_43049                        ------QLQLD-DFESDL------------------------------------N---G-CL----------------LNENGRKTFSKAFEES-LEQTVD-HPRLNR------KVSYQYLLRVEVY---KLKKHL-LTGE---------
389844445_Mesotoga_prima_MesG1-Ag-4-2                             ------AIITEDDFDKEMNS----------------------------------------TL----------------LTNEGMKKYVEHFEKF-MEETVF-HPTLKR------KVSYNTLLLTECY---KLIRSI--LSE---------
390962116_Thermococcus_sp-_CL1                                    ------ERGKG-DFTRKG------------------------------------KA----VY----------------LSRAGRRKLLKALTGT-LRRTVYYKPWKR-------SVSYRSIMEMEAR---RLRRHL--LGE---------
13540937_Thermoplasma_volcanium_GSS1                              ------MAHDSFRFEDNG------------------------------------------CY----------------LSLNGRRKFIEYYNEK-LNTVIKIGNN---------YLSYNNIIEREAY---NLLNAI--RND---------
289596694_Aciduliprofundum_boonei_T469                            ------GIITEDDFRKEFNG----------------------------------------VF----------------LNEQGKRKYIRKWEDR-MSSTVYHRRLKR-------SVSQRGLLRLEAY---KLVKHV--MND---------
14590110_Pyrococcus_horikoshii_OT3                                -K----RIIREEHFVRDV----------------------------------------G-VL----------------LNSEGVKIFLGGINDE-LSRKVL-HPTLRR------KVSVRYLIRLEAY---SLLKHL--IGD---------
84489236_Methanosphaera_stadtmanae_DSM_3091                       ------IINEK-HFKKDV----------------------------------------G-VY----------------LNEKGRQKFIQEYNNK-LKTTIK-HPQIKK------KVSYRYLIRLEGY---KLIKHI--LED---------
268317185_Rhodothermus_marinus_DSM_4252                           -G----EITPR-HFEERL------------------------------------G---G-VY----------------LKEEGRRIFVRHWDER-LRQTVYHRRLER-------HVSYERLIRLECY---KLIRHL--LDP-------KN
18977490_Pyrococcus_furiosus_DSM_3638                             ------IIKKD-HFREDL---------------------------------------NG-VL----------------LTDEGMKIVTKAYNQE-LEKTVK-HPKLGK------GVTRRRLIRLEAY---KIIKHL--VGV---------
427708213_Nostoc_sp-_PCC_7107                                     -N----LANTNGNGNGKS------------------------------------------------------------HSHKLLQKFLQHWEGK-LRTFVL-HPYAG-------EISYRQCLDLQVR---EYLASL--LGD--------V
119720226_Thermofilum_pendens_Hrk_5                               -G----LPTLSED-----------------------------------------G---D-------------------LDKDSLNKAMKRLYRY-LSAKHRET-----------GKTPYQQIHLKAF---CLAKHL---EG---KCSREK
118497969_Francisella_novicida_U112                               -G----QFSVE-DFGCKN------------------------------------E---Q-FY----------------IKKDKTKDYSKILFAE--------------------IISYKLEIFEYVR---EFYRAF--MRG---K---EI
332661941_Haliscomenobacter_hydrossis_DSM_1100                    -Q----WKAPE-DFALQP-----------------------------------EG---G-YW----------------LGPTGKEKLIKAIFEI-LEEPGAQP-----------KTSKQTEMFQRVK---ELQQLI--------------
326793969_Marinomonas_mediterranea_MMB-1                          ------QIHKD-DFVIKE---------------------------------------NG-CE----------------MSSGARKTLLKELLVQ-LTKVPKG------------GNSLLTEMSNQSY---RLALSCKMQQR---------
220934970_Thioalkalivibrio_sulfidophilus_HL-EbGr7                 -G----RIRVS-DFAQQG----------------------------------DA------CR----------------IEAGARRRYLADLSER-LLTPVRARGDSE-------ARSLHDHLHRQAR---SLIAWI--REE--------A
288947693_Allochromatium_vinosum_DSM_180                          -R----LLTPE-DFTTSP--------------------------------------EDG-CR----------------LSKEARGRFYQAWAQA-RQDWPDLQATLEDERAGT-VVSLTQLCRRQTEVLRQVLKPV-MTGQCSQG-----
431929780_Thioflavicoccus_mobilis_8321                            -G----PRQADYYYREAD----------------------------------------G-CR----------------LSKTARPGFFQAWASY---RHHWPYAVRLDPSADWPASTLREQIRGQIE---RLRSAM----K---------
206890539_Thermodesulfovibrio_yellowstonii_DSM_11347              ------HLGNR-DFMKDE-------------------------------------ESGG-VY----------------LKKTGRKKFYPLYEQW--------------------VQQQRTIWRGEVQ---GFARRI--LEE---------
52425690_Mannheimia_succiniciproducens_MBEL55E                    ------VLKID-QFSQTN---------------------------------------EG-CI----------------LGKEGRVIFYTAFEKV--------------------VSEWRKIFEKQAY---ELVHLI--CGY-QTE--YHQ
30248142_Nitrosomonas_europaea_ATCC_19718                         ------VLDKD-HFTTTE--------------------------------------S-G-CL----------------LGKAGRVRYYAAYEEH--------------------SEILRKGINQEIE---QLAEQV---GSALTPESGNT
121997419_Halorhodospira_halophila_SL1                            ------TLQRT-DFTNST---------------------------------------EG-VR----------------LGKAARQTFFGCWERH--------------------AGLHRRWQRRAAQ---ALAAEC--AHH-GAQ--TIP
430760780_Thioalkalivibrio_nitratireducens_DSM_14787              ------VLRAH-DFSRDH---------------------------------------EA-CL----------------LNKAGRARFYGRWEPF--------------------ARTHRRWLRLRCG---QLARDF--RKQ-GET--FLD
292492856_Nitrosococcus_halophilus_Nc4                            ------TLRAD-NFRQED----------------------------------------SACL----------------LGKAGRQRFYRDYEAT--------------------APAWRRWLRQQSR---IMARRVLALSPQEILPAVKE
34541577_Porphyromonas_gingivalis_W83                             -G----E-----DVGQNS----------------------------------------K-GL----------------LDMPTRSRIISKINER-WFATEYYRSG---------EKLFSDIMKLQTK---DVSAFC--CGK---V--KRI
88602029_Methanospirillum_hungatei_JF-1                           ------MATPD-DFTYPD--------------------------------------EGG-CM----------------IGTMAKKKYAAAVLTR-IHGKVKYE-----------EQTFQDIFKRQAE---RIGKAL-TEGD---------
325980970_Nitrosomonas_sp-_AL212                                  -Q----QWRRA-DFV---------------------------------------------------------------IDRQGVIRVHPQLARI-VVTKAILP-----------DKVIREEINAYVG---LLKRLV--------------
269798874_Veillonella_parvula_DSM_2008                            -H----MILPE-DFDLSE------------------------------------R---G-CF----------------LSKEANKKWVELYEDY-MKAELSRID----------NLAPRKWIQQEIQ---AFISYL--KSL---DETVDI
332668016_Haliscomenobacter_hydrossis_DSM_1100                    -K----RLQTT-FFTRNQ---------------------------------------HG-LF----------------LNKEGKAFIIPLFNNF-LRSTRTYFDR---------ESTVKNHIYFLAH---LLAQRI--RGTV--------
379728404_Saprospira_grandis_str-_Lewin                           ------GLDPQQHFEERS-------------------------------------IREG-IR----------------LAPSGKKKVIAAMLGH-LEERTIYQGQ---------KRKRKTQIDLGLQ---SLASYL--KTL---------
327314603_Prevotella_denticola_F0289                              ------VITEEFYSVREDGS----------------------------------------YW----------------LEALGRRVLIQSLNDY-MDEVVTVKGT---------TRSRLTQLTLYVQ---SLAQKF--KEE---------
392395717_Flexibacter_litoralis_DSM_6794                          -G----YSFEEKDFIPKG--------------------------------------EQG-VW----------------LHGNGKRKFVKALHEY-LHEVIPQKNV---------RRSRLHHIQLDAQ---SLAQLF--LGL---------
150398872_Methanococcus_vannielii_SB                              ------TIKEN-FFKEVE---------------------------------------RG-YL----------------LDEEGRKHIISKFNGY-LNMKVKKYGK---------TYTMADVMQKDAY---KLSERI--LRS---------
150401497_Methanococcus_aeolicus_Nankai-3                         -K----KVKGS-HFNILE------------------------------------N---G-IS----------------LSDDGKGEFAPYIKDI-VFGKEFYFKNK--------KYLLPDFIQKECY---EIAKSI--RGK---IL----
284040851_Spirosoma_linguale_DSM_74                               -K----KVNKA-HIGEVS------------------------------------GSRTG-VS----------------LNADGKALLVNAFNEC-MDNDPIRYRGR--------NLVRSHCMQLDAH---QFANEL--IGK--TGGLPDL
257125862_Leptotrichia_buccalis_C-1013-b                          ------RVQKY-FFKREG---------------------------------------NS-VM----------------LVGDGKKVLVESFYNR-LEKKVRYNNR---------NISSLDKLQFECH---ELANFL--ISK---------
21226661_Methanosarcina_mazei_Go1                                 ------QVSDS-FFDEIP---------------------------------------KG-MT----------------LNKEGKAVLIQAVNET-FDKEISYRGR---------NIKIGNTIQFDCH---RIANSL--IGK---------
218961577_Candidatus_Cloacamonas_acidaminovorans_str-_Evry        ------RHKFSPYFIEPV--------------------------------------HQG-VL----------------LSTTGKKFLAPLLLEH-FDEIIRYRNK---------NRKRLDMIQTDAH---AFANFLIDKRE-NYLETSIN
319790504_Thermovibrio_ammonificans_HB-1                          ------IFKKE-DFEQTI---------------------------------------RG-IT----------------LREDKLKFLLLLLTEF----------------------QRKKKVLNTAN---SFIKKL--KEK---------
337288662_Thermodesulfobacterium_geofontis_OPF15                  ------KIIKE-DFVKENNY----------------------------------------FY----------------LKKQAIYKILEEFNQL--------------------KEELITSMKKFLT---DLENF---------------
20094748_Methanopyrus_kandleri_AV19                               -G----SLSRE-------------------------------------------------MF-----------------RWKGDAVLLRDLDAV-SAPVLTVLSRV--------RGGLLEAVDRRIR---EVRDGV---SR---QSPPEP
386750121_Helicobacter_cetorum_MIT_00-7128                        -K----ESLRL------------------------------------------DD---N-KR----------------LSLESRKRISQNMRQK-FQSLSTYK-----KQ----PLTCEKIIELECN---ALAKFI--QED---S--K--
126465226_Staphylothermus_marinus_F1                              -G----FLRRA----------------------------------------------SG-AE----------------LAKLLTKKYYEEFTRQRIEYKAR-------------KMTMEAAVFWFVQ---EIVRFL--NNR---TSSIQ-
170290046_Candidatus_Korarchaeum_cryptofilum_OPF8                 ------GMSPS-LFSSNI--------------------------------------VSG-------------------TIASLRSRFFSDFMNWRLDYKGR-------------KLGMETIMFLYVR---DIVSFL--RGG--------K
385805115_Fervidicoccus_fontis_Kam940                             ------IKVSD-------------------------------------------------GV----------------LDEDARKNIILLYDGA-RKSKVRIYDKT--------SVTFQEALTSFAF---NLAESF-RKSE---------
15898245_Sulfolobus_solfataricus_P2                               ------------------------------------------------------------------------------ESADKVTDLKTVYSLF--------------------SDVKEDEIYTQAR---RLVNAI--LND---------
156937921_Ignicoccus_hospitalis_KIN4-I                            ------ELEVD-----------------------------------------------G-CS----------------LTYSSRKALGEEIKKL--------------------LESCDKTLLAEAW---NLASSM--REG---------
124027534_Hyperthermus_butylicus_DSM_5456                         ------KTDCLDTYMGA-------------------------------------------------------------LTYNARRCIATLVLEN-IYKTPYPDSRGR-------KKTLRDHIYTYAW---NLAAAI--RQH---------
118431380_Aeropyrum_pernix_K1                                     -M----YLKLE-----------------------------------------------A-GR----------------LDYKSRKEVAKAVASS-LSMKAAVRGLGR-------RIRLEDAIMVQAR---WLAEAF--RGS---------
307595807_Vulcanisaeta_distributa_DSM_14429                       -K----LSTVN-------------------------------------------------GF----------------LDYDSRKAIARVILNN-MNRDYI-NPRTGR------KASLTEIMKHEAW---NLASTI-REGN--------V
305663330_Ignisphaera_aggregans_DSM_17230                         -L----KIDVV-----------------------------------------------G-NR----------------LDYESRKRIAEIVLSN-LSAKYRCSKIDR-------SMSLEEHIRREAW---DLAKFF--REG---------
347523053_Pyrolobus_fumarii_1A                                    -G----GWRPVRDPQTGL------------------------------------------------------------MDAETRSRLVAEYARW-VRERVVRDGATGK------SDTLERHMARHAR---LLAQAV-RSGS---------
302347808_Acidilobus_saccharovorans_345-15                        -S----GWRPRLE---------------------------------------------G-GL----------------LDVGSRRALLEALFNR-LKDQALGAYR---------PMTYSEALREYAY---RLASSL-RSGQ---------
```

```
218884093_Desulfurococcus_kamchatkensis_1221n                    ------GWKPR----------------------------------------------VLNGL----------------LDYESRARIVESIMKT-LENTKTRYIDET-------PITLRQAMKKVTM---RLASFL--RGE---------
389861365_Thermogladius_cellulolyticus_1633                      -H----GWRPS----------------------------------------------LSGGL----------------LDYESRRRVVAGVLER-LEKAKTRRFGEA-------PVSLRGAMRRSAL---SLASYLRGEGGLPGLRLRVV
15898286_Sulfolobus_solfataricus_P2                              -F----RLKVK-----------------------------------------------G-GL----------------IEENSRGDLAKLIRKG-MEENVKEESDHN-------PKTLIQAIRAHAV---KLASSI-REGK---------
15922975_Sulfolobus_tokodaii_str-_7                              -G----FIPNVKD-----------------------------------------------GI----------------LDAESRKKIAEAVINA-MNGKVKDEGGD--------VRTGIQAMRAYAL---KLASAL--RGE------E--
18311772_Pyrobaculum_aerophilum_str-_IM2                         ------ELEAE-----------------------------------------------N-GL----------------LTHKSRLRLAAAVLEE-FNAT---------------GRPLSAEIHREAR---SIAKAL--CTS---------
18311699_Pyrobaculum_aerophilum_str-_IM2                         -A----DWADG-------------------------------------------------GV----------------LKREIRAMLVEEWTGE----------------------RLEPVIARQVG---LAVAHL--HGQ------R--
54296139_Legionella_pneumophila_str-_Paris                       -A----QKQNE-------------------------------------------------------------------LTPALKKTLIGEMANL-YAKPLPYHKR---------RISLESIIQRQVY---HLSAVF--SGE---------
34557933_Wolinella_succinogenes_DSM_1740                         -S----LAGKK-----------------------------------------------E-------------------LDAELKKKIITEINAT-MNKKHLYNGK---------NLSLQGVMQRQVY---RLVGAF-MQQN---------
118497353_Francisella_novicida_U112                              ------QYKNK-------------------------------------------------QY----------------IDTKLKSILISEIQAT-IAKKYIYNGK---------KLKLEHIIQRQVY---RLSGEF--AGE---------
328954440_Desulfobacca_acetoxidans_DSM_11109                     -GEKLTQIKSG-------------------------------------------------AL----------------LTQETRRKVITQVVER-LGSLHPYRGQ---------KVSLEEIIRRQAR---LLADHL--EGK---------
408403000_Candidatus_Nitrososphaera_gargensis_Ga9-2              -K----MLKID-EIMQFD---------------------------------EVKN---Q-QR----------------LSKQTIQVLIGEITGQ-LESEVAYGRNSGE------RCLMKNVIQNQSR---ALVSYL--MNP-----SK--
330509017_Methanosaeta_concilii_GP6                              ------EMRQG-HFQDD-------------------------------------------CR----------------MTKDGIKIASKAFYGR-LDERITYREK---------SQILRNIIRSQAS---SVATFL--RGE--------R
148656004_Roseiflexus_sp-_RS-1                                   -V----EIGQGDD-----------------------------------------------GL----------------LDAATRKRIAEKILER-LDSTEPYEGK---------RQPLRHILQCQAR---HIATFV--RGE--------R
386854797_Deinococcus_gobiensis_I-0                              -R----GWRPETDDN---------------------------------------------GK----------------LTPETRKRIATAVNDR-LDARDVVGGK---------RFRLQNIMVMQAR---RLATFL--RGE---------
167628642_Heliobacterium_modesticaldum_Ice1                      -G----VEIGM-E----G------------------------------------------ER----------------LADGTKRDLLERIQQR-LAATERFDGK---------KYPLQSILLRQAR---RMASYL--RGD-------Q-
148642083_Methanobrevibacter_smithii_ATCC_35061                  ------QINET-NIDKRN------------------------------------------NS----------------LKLESRKLLAGEIMAK-IHSDLTFENE---------KVTYAKIIENQTK---KLVKSL--LNN---------
288560430_Methanobrevibacter_ruminantium_M1                      -N----QITPD-DFEEDGNY----------------------------------------IK----------------IHERARRLLIAKILDK-LNSKIMFHSK---------NISYSDIILYQGK---LMADYL-TNGV---------
11499462_Archaeoglobus_fulgidus_DSM_4304                         ------QIKPE-DCEKRNMV----------------------------------------CQ----------------LSDNARRLLLASLLER-LDSKTQYRGR---------NLAYSSIILLHAR---DVVAFL--RGE---------
383319612_Methanocella_conradii_HZ254                            -G----QVKPA-DFTLEN----------------------------------------GVCK----------------MADSARKLLLNELLGK-LEDQIRYNDK---------SMKWTDLILGQAR---GIAQYL--RGE--------S
294676823_Rhodobacter_capsulatus_SB_1003                         -R----RLREE-MFEPQG----------------------------------------DSLR----------------LLTPGRRALITGYEEA-LERVIR-SSQTGQ------RHALRRIMLEEAR---ALARHC--RSP-------ET
302343124_Desulfarculus_baarsii_DSM_2075                         ------ILKKE-MFYKTD--------------------------KWP-------------CH----------------VTPEGRETIIRAYEAW-LDRPVK-SPRSGE------KVKWRGLLEEQVL---AYRDHV--MGR--------S
389875622_Tistrella_mobilis_KA081020-065                         ------ILGPDDIIRYPP-----------PPDAPDQAGVPGQPDGG--------------WR----------------VLPEAGRRVIRTWEGW-LDRPVT-GP-DGR------RRAWRGLIDTQLA---AWCRHI--EDD---------
374998939_Azospirillum_lipoferum_4B                              ------TLRPD-HFSRRE--------------------------TG-EGEIKG-------CR----------------IDPDAVGAIIRAYEQW-VRRPVA-SPRDGK------RTTWRGLIGFQAQ---ALAAHV--QGR--------E
288957883_Azospirillum_sp-_B510                                  ------ILSRA-DFLTTD--------------------------PP-DGDAPDGQPGAAVCR----------------IQSAGRDRLIRGYEGW-LDRAVK-GP-SGK------RMLWRRVMQAQVV---AYARHA--LGE--------A
116627764_Streptococcus_thermophilus_LMD-9                       -G----KLLLS-HFENKD----------------------------------------QDFI----------------LTDEGREIFAWALRSR-ILEVHR-YIELDKK-----RYSFLYTADRQIK---SLIRAF-RELD------PS-
15609954_Mycobacterium_tuberculosis_H37Rv                        -G----VVDTR-AFSKNS--------------------------D--------T----GAVF----------------ATREATRSIARAFGNR-IARTAT-YIKGDPH-----RYTFQYALDLQLQ---SLVRVI-EAGH------PS-
348026605_Megasphaera_elsdenii_DSM_20460                         -H----EIKEE-HFCCDE-------------------------------------ANHG-IY----------------LTPEGRKIFFRAYEKK-MRSINQ-YIDG--------KHSFRRSLNYQVA---QYGQAL--MAR-------EP
296133514_Thermincola_potens_JR                                  -G----ILAPG-DFYKDE--------------------------A--------S----GAVL----------------LKNDALRKFIKHFEQK-IRSEAN-YLSYLDY-----RVSYRRAVQHQAG---VLANCI-EDND------IT-
114567264_Syntrophomonas_wolfei_subsp-_wolfei_str-_Goettingen    -G----IIKRE-NFEKPN--------------------------E--------E----GGVY----------------LDGIGSKAFIAEYEKK-VRGRSN-YLAYVDY-----SVSFRRAMEMQCQ---RLAKAI-EEGD------PS-
332296542_Thermodesulfobium_narugense_DSM_14796                  -N----IFSSE-DFWYKE--------------------------EYLEDDNAKEVNSVEAVY----------------LNIEGVKKALVYFQEL-LEKNKY-ISTLGN------NINLFQIIREQAR---IFVRCL--SEN------E--
218442809_Cyanothece_sp-_PCC_7424                                -G----LLTLK-DFEPVA--------------------------S--------N----GGVY----------------LTDKGRRIFLKEFESR-INKLIS-HPDIQS------PVSYRQTIQLQIR---RYKQSL--LSD------V--
17228961_Nostoc_sp-_PCC_7120                                     -K----IFTFE-DFTPPD--------------------------E--------K----GGVY----------------LQPHALKKFLKHWEEK-LQSEVT-HPHTGY------QVSLRRCLELQVR---EYITSL--TGE------VE-
119357853_Chlorobium_phaeobacteroides_DSM_266                    -R----LISPE-EFTVAR---------------------------------HDDGKPKG-CY----------------LSDGARKTFLREFENL-MHRTTT-HPATGY------EVTCRRCLDLQVG---EFALYL--KGE---------
119357846_Chlorobium_phaeobacteroides_DSM_266                    -G----LLQEK-DFYYRK--------------------------D--------E----PGCF----------------LSNDARKRFLNIFETR-MWQESR-DGCTGK------TLNFRRHIEKQVR---IMREVI--AGT------RT-
336116789_Microlunatus_phosphovorus_NM-1                         -G----ELTAASGHTEPG---------------------------------------KSGVL----------------LTKAARTSLLAGYELR-MLRTTK-GALPDF------SGTLRRHLYRQAQ---RLLATV--TRG-------E-
37677204_Vibrio_vulnificus_YJ016                                 -K----LITPE-QFTTHK----------------------------------------NGVL----------------LNREGKRLFIQSLEQR-FLSQHQ-HPTLKI------KADTRRLIDIQVL---ELRHQL--LSG------QW-
154250301_Fervidobacterium_nodosum_Rt17-B1                       -K----TIKKE-DFEKRE---------------------------------------NDIYY----------------FTKDGMKKYITEYEQM-MKQKLK-YHLDSE------ENYIRTIFEKQAR---HYARVV--LGD------EE-
38505760_Synechocystis_sp-_PCC_6803                              -R----LIDVGEDFVFQN----------------------------------------GGCF----------------LNSSGRRKWLGAFVGR-MEEFLS-FDQGDF--------PRWHCLIDQVK---KYKAFI--YSP------SN-
83592490_Rhodospirillum_rubrum_ATCC_11170                        -R----KIQPQ-HFEPDA-------------------------------------ERPGGVK----------------LTVDGRKIFLEEWEGH-LARPLR-EPGVAAEH----RLDVHRLLQRQVD---RLVSDL--RGG-------E-
261415948_Fibrobacter_succinogenes_subsp-_succinogenes_S85       -R----VLVAD-DFEQRN-----------------------------------DENSSGQLF----------------LKPEKMRVFLHHYEDF-----VA-RPNGIYQGLA--NLPWRSVMRLRVE---AFRRLL--LEG-------V-
76802276_Natronomonas_pharaonis_DSM_2160                         ------TLTHD-DFNKDN-------------------------------------------R----------------LSDDAFKTYLSKFDEF-MQEEFT-HPHFEY------SVTRRKAVRQQAI---LLRKAI--TGE--------M
312114617_Rhodomicrobium_vannielii_ATCC_17100                    -G----QIRTG-DAAEAD----------------------------------------DGWR----------------LGETGKRTVIDLMEKR-LTTAIS-VSGTEQ------QVSYREAIGRQAR---GIATAL--QTG-------A-
219683006_Bifidobacterium_animalis_subsp-_lactis_AD011           -G----EIKRN-GFANVM----------------------------------------GSVR----------------LRDETRKTLIGAYERR-MATELK-HPVYAY------RASWRRIVEIQAR---MVLGRL--EGS------LE-
116748794_Syntrophobacter_fumaroxidans_MPOB                      -R----VIGLD-DFVYRP-------------------------TPCADAVDEEELKHRRPVE----------------MKPKIARAFIEAYEKW-MASRIL-DPGSRE------RTDYRGLIQRQVW---KFCHYL--VGD------RD-
317051217_Desulfurispirillum_indicum_S5                          -K----QVKAE-SFEYQE---------------------------------------DGAVI----------------LTEPAREVFYRAFERK-IRGGTI-YD--GK------RHSYRRIIQQQAQ---KLSKSL--FES------EP-
328949009_Treponema_succinifaciens_DSM_2489                      -G----TLKSE-DFEKKE-------FSAESEDYPLELEQGKGDSDSNLGDEENSPAKYG-IL----------------LTKAGIKKVISAFEAK-MNSLIL-YSPTAQ------KISYTKIIYQQVQ---HYKRVI--CGE--------E
333996012_Treponema_azotonutricium_ZAS-9                         ------ALLPG-DFEPADISGDTAA-----------------------------------IY----------------LNKEGLSKAISAFEKK-METKLH-YQPLNT------ELSLDEIIIEQVK---HFKRVL--MGE--------E
313673559_Calditerrivibrio_nitroreducens_DSM_19672               -G----LITKS-DFVL----------------------------------NDDGDEDHLPIQ----------------LTIFGRKKWLSILEKR-FNSFLW-FQPKNQ------KTTLRNIIKYQCY---LISKSF--IEK-------A-
218961167_Candidatus_Cloacamonas_acidaminovorans_str-_Evry       -K----EIRLE-HFLYNT---------------------------LPDDENSDLDNSTLPVS----------------LTKNGMKIIITAFSKL-INSKFN-YNQLSE------QWILKDIITLQIR---KLAAHF--DNK-------K-
294828260_Leptospira_interrogans_serovar_Lai_str-_56601          -K----SW-IDEDFEITK----------------------------------------NKVW----------------LSESGRKKATQLYETR-LDDTWK-HPVVNY------SLSYYRMIELEVR---LLEKEW--SGE------AN-
347756303_Candidatus_Chloracidobacterium_thermophilum_B          -K----QWDATEDFVVTP----------------------------------------HKVW----------------LSLEGKRKAIALYEER-KQESWQ-HPVLRY------SLSYQRAIELEAR---LLEKEW--TGA------PG-
162457462_Sorangium_cellulosum_So_ce56                           -R----TFDAERDFVITG----------------------------------------RQVW----------------LSDAGHRAFIEVYERR-KHEEYR-HDVLGF------SLSYARMMELEAR---LLEKEW--SGE------PG-
86742032_Frankia_sp-_CcI3                                        -R----QLTDA-DFEHHP---------------------------------------GGSCQ----------------LTESGRKNYLTLWSQA-RARTWP-HTLLTH------DTPAATLPLLQAR---ILARHL--RGD------IP-
433631663_Mycobacterium_canettii_CIPT_140070010                  -R----QLTTK-SFSSEP---------------------------------------TGAWR----------------LTDDGRATFLAAFDSH-LRTAVS-HRVIEQ------DIERRKVPGLQAL---VLARYL--RGD------LT-
15612904_Bacillus_halodurans_C-125                               -K----EMTAD-GFYKKE---------------------------------------NGAVL----------------MTDEARKTFLKAWQTK-KQEKIT-HPYLGE------KMSWGLVPYVQAL---LLARFL--RGD------LD-
108763792_Myxococcus_xanthus_DK_1622                             -G----QLKPG-DFRTES---------------------------------------AGAVL----------------LKDDARKTFLVAYQEA-KQVGVR-HAFLGQ------ETTWGMAPHLQAL---LLARHL--RGE------LD-
344996859_Caldicellulosiruptor_lactoaceticus_6A                  -E----QLNNE-NFENKF----------------------------------------GGVY----------------LTDEGRRTFLRKIRDR-LERQSS-HILMKG------NITFQKTVYLQFK---VLVKYL--IGD------IE-
85858452_Syntrophus_aciditrophicus_SB                            -K----ILKRE-DFYFEEPVREEDFVQDPGERIADVSRDPIGLIANGHGDSELLDLPEQRMAGEMPVEDLSTGKYPVKLRDAAFKRVIEAFEQK-LTTSFH-YPPAER------QLTYGEALIFQAG---HCRRVI--EGE------AD-
148657121_Roseiflexus_sp-_RS-1                                   -N----AIGRE-HFHHPQ------------------------------------ERPDA-VY----------------LDDVGRTLLVQRYESM-LQTKVR-LPGGE-------QTPLRRVILLQAQ---AIARIV--RGE--------Q
159898907_Herpetosiphon_aurantiacus_DSM_785                      ------LLQMS-NFSRTE---------------------------------------KG-FL----------------LNDRARKSFIQAYEQR-MQTPIR-YQGTGN------NEPMRRVLLLQTQ---HLARVL--QGE--------E
17227877_Nostoc_sp-_PCC_7120                                     -D----LLTQE-DFQPSP--------------------------------------DGVGIW----------------LGTMGKKLLLGELERR-FRTSVL-YPPQNR------QLSLNQILLEQAR---SLGRCL--LQS------KL-
268318378_Rhodothermus_marinus_DSM_4252                          -G----RLRPQ-HFEERN----------------------------------------GGVY----------------LNEAGRAVVLEAWQAH-RQQTSA-HPVLGM------ELSLARHFEAQAR---LLARAL--QEQ------GI-
55978328_Thermus_thermophilus_HB8                                -G----LLTPS-HAEVRE----------------------------------------GGVY----------------LNEEGRRRLIQLFEER-LLEGVS-HP-LGF------RKPLGETIEVQAQ---RLKAAL--LGR------G--
297622675_Truepera_radiovictrix_DSM_17093                        -G----ALHPTQHFRDEA----------------------------------------GGIY----------------LSDAGRKVLVPLLEAR-LDAEAK-PP-EGG------ACPYRTLIERQAR---RLAAAL--EGR------R--
328953000_Desulfobacca_acetoxidans_DSM_11109                     -R----VFGEQ-DFFRHE--------------------------------------PSGGMY----------------LGDEPRKRYFQEYERY-LTNVHH-CPENGG------ETDFRRLFRRQAE---RLMRAV--LLK-------G-
337286709_Thermodesulfatator_indicus_DSM_15286                   ------AVKPE-DFTEEP----------WDDEEDFTS-----------------------FK----------------LTPDGLRKFLAKFRER-LDEEALYAPLNK-------RLSYRYIMQQQVW---HLARVL--DGR--------E
328953422_Desulfobacca_acetoxidans_DSM_11109                     ------IIRSTDFYRPEDREPAAFDFAETESRRENYP-----------------------IL----------------LVHTGMKKFITHFEAR-LQQKVLYLPSGR-------RLTYRQICLEQVR---ALVRFL-NDGS---------
38505684_Synechocystis_sp-_PCC_6803                              ------EIQRQ-DFTESL----------------------------------------GAFR----------------LTDSATKTFLGAFDRK-LSSEFK-HPIFNY------KCTYRRAIELQAR---LLARHL--QEG---------
134299482_Desulfotomaculum_reducens_MI-1                         ------VIAEG-DF--EY----------------RMG---G-------------------CF----------------LSETGRKKMYRLYEER-RKEMIT-HPVFGY------RISYLRTIELQAR---FLAKVL--TKE------ID-
17229060_Nostoc_sp-_PCC_7120                                     ------MLSPK-DFITEP----------------ISN---A-------------------VS----------------LTKEGLHIFLRLYQEK-KQTQFK-HPVMQK------KYTYQETFEIQAR---LLAKYL--LGE------LD-
302391602_Acetohalobium_arabaticum_DSM_5501                      ------MIVSD-NFNKTC----------------------GT------------------VK----------------MKDKARNKFLEQYEKR-LREEFT-HPVFEY------KVTWRRCIELQAR---LISKTI-NEE---------I
159898755_Herpetosiphon_aurantiacus_DSM_785                      ------ILNQQ-HFQREP------------------G---S-------------------VQ----------------LSKEGRKLFLTKLEER-FSSEIQ-HPIFGY------RVSYRRCIELQAR---LLAKAL--MGE------IQ-
320161859_Anaerolinea_thermophila_UNI-1                          ------KICPQ-QFTPGPPERP--------------------------------------VV----------------LDDQGKRDFIKAFEER-MDSRFT-HPLAQQ------QLTMRQCLIEQAR---QIAHRI--LED--------R
156741961_Roseiflexus_castenholzii_DSM_13941                     ------RIAFD-DFTETP--------------DGDYP-----------------------VL----------------LSDDGKRRFVAAFEER-MRTEAT-HPDGADGRPG--KVSYLRCLALQAR---RLARAV--QGG---------
383762167_Caldilinea_aerophila_DSM_14535_-_NBRC_104270           ------ILQPE-HFQPGE-------------------------------------SPERPIV----------------LAEEGVKRFVREMEHR-LEQDIQ-HPQTGE------RVTYRRLFLQQVY---SLARTL--QGG------ERE
46255267_Thermus_thermophilus_HB27                               -R----RVTLE-DFDDSE----------------------------------------GFPR----------------LRKEAWPKFLRAWEER-LNERVQ-HPLLRK------RLAYREILLAQAR---ILVKHL--LGE------LP-
134101617_Saccharopolyspora_erythraea_NRRL_2338                  -L----QADPA-MFHVHS----------------------------------------TAVA----------------LTSSGRRDVIDAYERR-LTTEVI-HPVFRY------KTTYRRAIEIQAR---LFAAHL--LDE------IP-
378719041_Gordonia_polyisoprenivorans_VH2                        -G----EVTSG-DFERRS----------------------------------------GGCW----------------LTLQGRRKVIRAYERR-LDVQVT-HPVFKY------KVSYRRVLDVQAR---ILAGVI--LGE------LP-
320102061_Isosphaera_pallida_ATCC_43644                          -R----MVTPS-DFLRAG----------------------------------------DSVA----------------LTPAGRKAFIRAYETR-MDHLAT-HPLFSY------RLSYRRMLELQVR---LLARVL--TGE------IK-
268318441_Rhodothermus_marinus_DSM_4252                          ------AVTEA-DFIVRG----------------------------------------PACA----------------MKKSAMEKVIEAYEQR-MNTILR-HPVFGY------AISYRRTMEVQAR---LLSRVL--TGE--------L
39995168_Geobacter_sulfurreducens_PCA                            ------EIRTG-DFVRSA----------------------------------------GGCN----------------LTDSARKRFIAGFERR-MEQEVT-HPIFKY------TISYRRLLEVQAR---LLTRYL--SGE--------I
83591519_Rhodospirillum_rubrum_ATCC_11170                        ------EIGTN-DFLYAG----------------------------------------GGCA----------------LKPNARRGLIAAYERR-LDQETT-HPVFGY------QLSMRRLIQVQAR---LLARFV--SGD--------I
50085564_Acinetobacter_sp-_ADP1                                  ------------------------------------------------------------GL----------------NQKEFRMQLIEICQDQ----------------------DLLDYMFSFIT---DICSKI----K---------
77166205_Nitrosococcus_oceani_ATCC_19707                         ----------------------------------------------------------A-GR----------------SDQEMRQACIAWLDES-------------------------HAMTFLFQ---SIKRVA--------------
15602176_Pasteurella_multocida_subsp-_multocida_str-_Pm70        -------------------------------------------------------------------------------NGDSEKEFRQACIQNLLRFGAL-------------DIIIERLQQLALQ---GASDEY--FAG---------
38234755_Corynebacterium_diphtheriae_NCTC_13129                  ------DEDPVTNIRRH-------------------------------------------LR----------------REIHSRKIMAGMLEAL-MEVLTPYL-----------PNRNDDRLIGDSD---EVKGHI-QYGK---------
21674788_Chlorobium_tepidum_TLS                                  ------------------------------------------------------------------------------AGYYDRHKIASEFRKRVI------------------EIDLLGKIGPDIE---TILGKK-QCSS---------
53804737_Methylococcus_capsulatus_str-_Bath                      -R----VMDDP-----------------------------------------------S-LP----------------LERALRKEAARQFHKQ-------------------------KVIPKMID---RIKELL-HVDD---------
209966592_Rhodospirillum_centenum_SW                             -R----IDQGA-----------------------------------------------T-DS----------------LDRICRRAVSAHVRDT-------------------------GFIDTIID---DIKEVL--A-----------
89885834_Rhodoferax_ferrireducens_T118                           ------TSPSN-------------------------------------------------------------------LNMAVRHGCRDLFSRE----------------------SMFETLLGHLE---NIFGTDHDCHD---------
16130662_Escherichia_coli_str-_K-12_substr-_MG1655               -G----EPDREVRLACRD------------------------------------------IF----------------RSSKTLAKLIPLIEDV-LAAGEIQPPAPP-------EDAQPVAIPLPVS-----------LGD---------
68535718_Corynebacterium_jeikeium_K411                           ------ELTADVNIRRVI------------------------------------------------------------RDYVVKKNLIPTMIRD-LMFLMEMSSEAQNI-----DAELVLWSDLEVV---ASKRNW--SDD---------
86738746_Frankia_sp-_CcI3                                        -A----GLTDERDIRTAF---------------------------------------RD-KV----------------ADGHLMARIIHDIKDLLIEEGTRDNDED--------ALHLWDELDGHVPGGVNWAADL--ADQ-TDDTTILG
51891801_Symbiobacterium_thermophilum_IAM_14863                  -Q----RLESRVRHRCRD------------------------------------------LF----------------AERRLLERIVTDLARL-FDLDPNPEP----------EVDLEVALPGDLW---DPEGPV--AGG---------
29834079_Streptomyces_avermitilis_MA-4680                        ------STNPEAEARRSF------------------------------------------------------------RDGLRLFKLLPRIVGD-IQELLGLGS----------AYDIPDPEEQLVDLWDPISGPV--PGG---------
```

Continued

```
407462288_Candidatus_Nitrosopumilus_koreensis_AR1                 NNF-----AFFDF-----------------------------------------------------------------------------------------
21229351_Methanosarcina_mazei_Go1                                 PEYEIERIDSYDIRQKILSISYVDWKKLGFSKGTLHYMKQNAKSDKPFTLNAHVLERVNKWEALVSSQR---------------------------------
147921327_Methanocella_arvoryzae_MRE50                            PDIMVNRDDPSDVRDALLKMTPAERKKLGISKTTLWYMQKNLREGKRIKIYEKSKGKLN------STKT---------------------------------
124485266_Methanocorpusculum_labreanum_Z                          --Y---------------------------------------------------------------------------------------------------
307352568_Methanoplanus_petrolearius_DSM_11571                    --------RPLRFSY---------------------------------------------------------------------------------------
154150254_Methanoregula_boonei_6A8                                --Y---------------------------------------------------------------------------------------------------
219852754_Methanosphaerula_palustris_E1-9c                        AEF-----LPRIPLKNH-------------------------------------------------------------------------------------
126178701_Methanoculleus_marisnigri_JR1                           --------AEFHVLYW--------------------------------------------------------------------------------------
397780289_Methanoculleus_bourgensis_MS2                           ---------------AEFHVLYW-------------------------------------------------------------------------------
156937861_Ignicoccus_hospitalis_KIN4-I                            -EF-----KAFVFRP---------------------------------------------------------------------------------------
70607609_Sulfolobus_acidocaldarius_DSM_639                        IDF-----EPFVYKPKQ-------------------------------------------------------------------------------------
187736488_Akkermansia_muciniphila_ATCC_BAA-835                    -PY-----EPWLMTTIKWAG----------------------------------------------------------------------------------
436839745_Desulfovibrio_hydrothermalis_AM13_-_DSM_14728           SGE-----RNLSLGEVEII-----------------------------------------------------------------------------------
71894593_Mycoplasma_synoviae_53                                   FNF-----ESLEIDWNKK------------------------------------------------------------------------------------
193216853_Mycoplasma_arthritidis_158L3-1                          -PY-----FNWSAE----------------------------------------------------------------------------------------
187250661_Elusimicrobium_minutum_Pei191                           IIF------PAILK----------------------------------------------------------------------------------------
325972002_Sphaerochaeta_globus_str-_Buddy                         IQF-----PKYRYAWAV-------------------------------------------------------------------------------------
47458867_Mycoplasma_mobile_163K                                   -SW-----EWVELWD---------------------------------------------------------------------------------------
269123825_Streptobacillus_moniliformis_DSM_12112                  PEI-----IKIDFYDLFKEL----------------------------------------------------------------------------------
57865886_Staphylococcus_epidermidis_RP62A                         PTYLL---LPKMIEVGN-------------------------------------------------------------------------------------
319957207_Nitratifractor_salsuginis_DSM_16511                     IDF-----GALRECF---------------------------------------------------------------------------------------
310286727_Bifidobacterium_bifidum_S17                             PELGFVRCGTELCE----------------------------------------------------------------------------------------
90961086_Lactobacillus_salivarius_UCC118                          NEV-----PNNAINDNI-------------------------------------------------------------------------------------
406600270_Leuconostoc_gelidum_JB7                                 LKL-----PEWRFEI---------------------------------------------------------------------------------------
386318631_Staphylococcus_pseudintermedius_ED99                    IKI-----PTWTY-----------------------------------------------------------------------------------------
302336019_Olsenella_uli_DSM_7084                                  GEI-----EPFDIV----------------------------------------------------------------------------------------
15675042_Streptococcus_pyogenes_M1_GAS                            KGV-----PEFRI-----------------------------------------------------------------------------------------
352684362_Acidaminococcus_intestini_RyC-MR95                      --------SKIKIYEI--------------------------------------------------------------------------------------
42525844_Treponema_denticola_ATCC_35405                           SQI-----NFYRNEL---------------------------------------------------------------------------------------
169823756_Finegoldia_magna_ATCC_29328                             IKF-----IDYEL-----------------------------------------------------------------------------------------
117929157_Acidothermus_cellulolyticus_11B                         RSL-----RPPAMELSHA------------------------------------------------------------------------------------
283456134_Bifidobacterium_dentium_Bd1                             PEYR----GPYEKG----------------------------------------------------------------------------------------
38232679_Corynebacterium_diphtheriae_NCTC_13129                   PIWE----GPFDAEER--------------------------------------------------------------------------------------
238924076_Eubacterium_rectale_ATCC_33656                          YNL-----IGEEYG----------------------------------------------------------------------------------------
116627543_Streptococcus_thermophilus_LMD-9                        SKFHCPVVSSLEWRK---------------------------------------------------------------------------------------
57238545_Campylobacter_jejuni_RM1221                              -EL-----KIVKFDD---------------------------------------------------------------------------------------
291276264_Helicobacter_mustelae_12198                             ESNGREVYESFSDV----------------------------------------------------------------------------------------
294086112_Candidatus_Puniceispirillum_marinum_IMCC1322            MPM-----PQLIPELEIEAMVSEC------------------------------------------------------------------------------
189485226_uncultured_Termite_group_1_bacterium_phylotype_Rs-D17   LLL-----PEFKTA----------------------------------------------------------------------------------------
222109284_Acidovorax_ebreus_TPSY                                  PRLIGLRQHVFEI-----------------------------------------------------------------------------------------
15602991_Pasteurella_multocida_subsp-_multocida_str-_Pm70         LPL-----KEHQYE----------------------------------------------------------------------------------------
310780383_Ilyobacter_polytropus_DSM_2926                          IPI-----QEHRYV----------------------------------------------------------------------------------------
220930481_Clostridium_cellulolyticum_H10                          PEI-----IPLEVHANE-------------------------------------------------------------------------------------
34557789_Wolinella_succinogenes_DSM_1740                          VQG-----EVYEADGDV-------------------------------------------------------------------------------------
90425960_Rhodopseudomonas_palustris_BisB18                        --------KPALLELPELERIS--------------------------------------------------------------------------------
209542526_Gluconacetobacter_diazotrophicus_PAl_5                  LST-----PTLKARD---------------------------------------------------------------------------------------
121608212_Verminephrobacter_eiseniae_EF01-2                       PGLPLALAGSFEPD----------------------------------------------------------------------------------------
60683388_Bacteroides_fragilis_NCTC_9343                           ISY-----PER-------------------------------------------------------------------------------------------
154250554_Parvibaculum_lavamentivorans_DS-1                       PEG-----MPLAPDLLDNDGEG--------------------------------------------------------------------------------
283778924_Pirellula_staleyi_DSM_6068                              -RF-----EPFRHP----------------------------------------------------------------------------------------
154174048_Campylobacter_curvus_525-92                             MPY-----VGFVWG----------------------------------------------------------------------------------------
292492938_Nitrosococcus_halophilus_Nc4                            AEY-----ESFKRKV---------------------------------------------------------------------------------------
111224947_Frankia_alni_ACN14a                                     TPY-----TPFRPWW---------------------------------------------------------------------------------------
385809790_Ignavibacterium_album_JCM_16511                         KSY-----DPYRAN----------------------------------------------------------------------------------------
150021526_Thermosipho_melanesiensis_BI429                         EEY-----KGFVLKE---------------------------------------------------------------------------------------
28211139_Clostridium_tetani_E88                                   -EF-----KPFSLKERM-------------------------------------------------------------------------------------
48477075_Picrophilus_torridus_DSM_9790                            KRY-----KPYISKR---------------------------------------------------------------------------------------
41614813_Nanoarchaeum_equitans_Kin4-M                             -KF-----KAFRLIY---------------------------------------------------------------------------------------
55376278_Haloarcula_marismortui_ATCC_43049                        -EY-----VPFKRWW---------------------------------------------------------------------------------------
389844445_Mesotoga_prima_MesG1-Ag-4-2                             EDY-----KALHMWW---------------------------------------------------------------------------------------
390962116_Thermococcus_sp-_CL1                                    ETY-----GPFRPWW---------------------------------------------------------------------------------------
13540937_Thermoplasma_volcanium_GSS1                              EDY-----KSFRSRS---------------------------------------------------------------------------------------
289596694_Aciduliprofundum_boonei_T469                            KRY-----RAFRVWW---------------------------------------------------------------------------------------
14590110_Pyrococcus_horikoshii_OT3                                KEY-----KSLRAWW---------------------------------------------------------------------------------------
84489236_Methanosphaera_stadtmanae_DSM_3091                       KTY-----ESFKTWW---------------------------------------------------------------------------------------
268317185_Rhodothermus_marinus_DSM_4252                           EPY-----RGLHMWW---------------------------------------------------------------------------------------
18977490_Pyrococcus_furiosus_DSM_3638                             EEY-----SPLVAWF---------------------------------------------------------------------------------------
427708213_Nostoc_sp-_PCC_7107                                     EFY-----RPLALKFHPANSDFTNTIKPQTVPLKLVKR----------------------------------------------------------------
119720226_Thermofilum_pendens_Hrk_5                               LAF-----TWDKKRYIIHE-----------------------------------------------------------------------------------
118497969_Francisella_novicida_U112                               AEY-----PIFCYETRRVYVDSQL------------------------------------------------------------------------------
332661941_Haliscomenobacter_hydrossis_DSM_1100                    -HH-----LPID------------------------------------------------------------------------------------------
326793969_Marinomonas_mediterranea_MMB-1                          --F-----IAWSPKR---------------------------------------------------------------------------------------
220934970_Thioalkalivibrio_sulfidophilus_HL-EbGr7                 PAF-----EPFRTR----------------------------------------------------------------------------------------
288947693_Allochromatium_vinosum_DSM_180                          --------DEFGLEF---------------------------------------------------------------------------------------
431929780_Thioflavicoccus_mobilis_8321                            --------TPEAPS----------------------------------------------------------------------------------------
206890539_Thermodesulfovibrio_yellowstonii_DSM_11347              --------KDIISG----------------------------------------------------------------------------------------
52425690_Mannheimia_succiniciproducens_MBEL55E                    DQF-----DDYTINMAHILGNEKCDI----------------------------------------------------------------------------
30248142_Nitrosomonas_europaea_ATCC_19718                         PDH-----DSGDWE----------------------------------------------------------------------------------------
121997419_Halorhodospira_halophila_SL1                            EA----------------------------------------------------------------------------------------------------
430760780_Thioalkalivibrio_nitratireducens_DSM_14787              PDG-----DDEAGET---------------------------------------------------------------------------------------
292492856_Nitrosococcus_halophilus_Nc4                            PRS---------------------------------------------------------------------------------------------------
34541577_Porphyromonas_gingivalis_W83                             KFY-----TPKW------------------------------------------------------------------------------------------
88602029_Methanospirillum_hungatei_JF-1                           -EY-----VPYRYRT---------------------------------------------------------------------------------------
325980970_Nitrosomonas_sp-_AL212                                  -GS-----ELKVLR----------------------------------------------------------------------------------------
269798874_Veillonella_parvula_DSM_2008                            EKS-----ENIS------------------------------------------------------------------------------------------
332668016_Haliscomenobacter_hydrossis_DSM_1100                    ------------------------------------------------------------------------------------------------------
379728404_Saprospira_grandis_str-_Lewin                           -DF---------------------------------------------------------------------------------------------------
327314603_Prevotella_denticola_F0289                              --------TKTVKR----------------------------------------------------------------------------------------
```

```
392395717_Flexibacter_litoralis_DSM_6794                          --------SE--------------------------------------------------------------------------------------------
150398872_Methanococcus_vannielii_SB                              ------------------------------------------------------------------------------------------------------
150401497_Methanococcus_aeolicus_Nankai-3                         ------------------------------------------------------------------------------------------------------
284040851_Spirosoma_linguale_DSM_74                               VKL---------------------------------------------------------------------------------------------------
257125862_Leptotrichia_buccalis_C-1013-b                          ------------------------------------------------------------------------------------------------------
21226661_Methanosarcina_mazei_Go1                                 ------------------------------------------------------------------------------------------------------
218961577_Candidatus_Cloacamonas_acidaminovorans_str-_Evry        RKL-----QNFLNSNFADNGEEEC------------------------------------------------------------------------------
319790504_Thermovibrio_ammonificans_HB-1                          -LV---------------------------------------------------------------------------------------------------
337288662_Thermodesulfobacterium_geofontis_OPF15                  -DF---------------------------------------------------------------------------------------------------
20094748_Methanopyrus_kandleri_AV19                               LEF-----DPEDVGVVWDALEV--------------------------------------------------------------------------------
386750121_Helicobacter_cetorum_MIT_00-7128                        -SY-----KPFIGKY---------------------------------------------------------------------------------------
126465226_Staphylothermus_marinus_F1                              --------PPYMPW----------------------------------------------------------------------------------------
170290046_Candidatus_Korarchaeum_cryptofilum_OPF8                 EPT-----MPYIPW----------------------------------------------------------------------------------------
385805115_Fervidicoccus_fontis_Kam940                             EAR-----MDMSMKW---------------------------------------------------------------------------------------
15898245_Sulfolobus_solfataricus_P2                               EEY-----RPYLAK----------------------------------------------------------------------------------------
156937921_Ignicoccus_hospitalis_KIN4-I                            --------REYLPGWKGCT-----------------------------------------------------------------------------------
124027534_Hyperthermus_butylicus_DSM_5456                         KPY-----KPFIVGRL--------------------------------------------------------------------------------------
118431380_Aeropyrum_pernix_K1                                     GGF-----SAVRLGL---------------------------------------------------------------------------------------
307595807_Vulcanisaeta_distributa_DSM_14429                       NEY-----KCFTWVY---------------------------------------------------------------------------------------
305663330_Ignisphaera_aggregans_DSM_17230                         SVY-----KGFRVYL---------------------------------------------------------------------------------------
347523053_Pyrolobus_fumarii_1A                                    -MY-----APFREAVR--------------------------------------------------------------------------------------
302347808_Acidilobus_saccharovorans_345-15                        -AY-----SCYAGGEPL-------------------------------------------------------------------------------------
218884093_Desulfurococcus_kamchatkensis_1221n                     SGF-----EGFVHGW---------------------------------------------------------------------------------------
389861365_Thermogladius_cellulolyticus_1633                       SHD-----SPGRVRHRGRRDQGEAQRLPEVQGADEDPEERLRWEAAAVHGEGRGEGATQVREGGEGRDTLHPAPRVQPQVPQVLRPPPRPGVRGQGAPGGVI
15898286_Sulfolobus_solfataricus_P2                               -EY-----RGFKLVM---------------------------------------------------------------------------------------
15922975_Sulfolobus_tokodaii_str-_7                               -KF-----KGFVKVW---------------------------------------------------------------------------------------
18311772_Pyrobaculum_aerophilum_str-_IM2                          ------------------------------------------------------------------------------------------------------
18311699_Pyrobaculum_aerophilum_str-_IM2                          -AY-----TPHQL-----------------------------------------------------------------------------------------
54296139_Legionella_pneumophila_str-_Paris                        GKY-----KGYLFKW---------------------------------------------------------------------------------------
34557933_Wolinella_succinogenes_DSM_1740                          -RY-----KSYIFKW---------------------------------------------------------------------------------------
118497353_Francisella_novicida_U112                               HNY-----KPYIFKW---------------------------------------------------------------------------------------
328954440_Desulfobacca_acetoxidans_DSM_11109                      KRY-----RSFRGRY---------------------------------------------------------------------------------------
408403000_Candidatus_Nitrososphaera_gargensis_Ga9-2               -EY-----DPFELAW---------------------------------------------------------------------------------------
330509017_Methanosaeta_concilii_GP6                               PCY-----EPFTPRW---------------------------------------------------------------------------------------
148656004_Roseiflexus_sp-_RS-1                                    PTY-----EPFVMGW---------------------------------------------------------------------------------------
386854797_Deinococcus_gobiensis_I-0                               GDY-----PVYIASW---------------------------------------------------------------------------------------
167628642_Heliobacterium_modesticaldum_Ice1                       -KY-----KPFVCSW---------------------------------------------------------------------------------------
148642083_Methanobrevibacter_smithii_ATCC_35061                   TEY-----NGFYLNW---------------------------------------------------------------------------------------
288560430_Methanobrevibacter_ruminantium_M1                       -PY-----EGFSLRW---------------------------------------------------------------------------------------
11499462_Archaeoglobus_fulgidus_DSM_4304                          RRY-----EGFVQKW---------------------------------------------------------------------------------------
383319612_Methanocella_conradii_HZ254                             KSY-----TGFWLRW---------------------------------------------------------------------------------------
294676823_Rhodobacter_capsulatus_SB_1003                          VPY-----LPQLQDY---------------------------------------------------------------------------------------
302343124_Desulfarculus_baarsii_DSM_2075                          -VY-----APYDMKY---------------------------------------------------------------------------------------
389875622_Tistrella_mobilis_KA081020-065                          APF-----VPYHLDY---------------------------------------------------------------------------------------
374998939_Azospirillum_lipoferum_4B                               -PY-----RAYVMDY---------------------------------------------------------------------------------------
288957883_Azospirillum_sp-_B510                                   -PY-----RPFILDY---------------------------------------------------------------------------------------
116627764_Streptococcus_thermophilus_LMD-9                        -LY-----ETSYTGGH--------------------------------------------------------------------------------------
15609954_Mycobacterium_tuberculosis_H37Rv                         -RL-----VDIDITSEPSGA----------------------------------------------------------------------------------
348026605_Megasphaera_elsdenii_DSM_20460                          KLY-----EPIWIR----------------------------------------------------------------------------------------
296133514_Thermincola_potens_JR                                   -IY-----RPVRLR----------------------------------------------------------------------------------------
114567264_Syntrophomonas_wolfei_subsp-_wolfei_str-_Goettingen     -IY-----QPVVIR----------------------------------------------------------------------------------------
332296542_Thermodesulfobium_narugense_DSM_14796                   -TY-----KPFIQKQ---------------------------------------------------------------------------------------
218442809_Cyanothece_sp-_PCC_7424                                 -SY-----QSFVRDI---------------------------------------------------------------------------------------
17228961_Nostoc_sp-_PCC_7120                                      -LY-----RPMIWKL---------------------------------------------------------------------------------------
119357853_Chlorobium_phaeobacteroides_DSM_266                     KPY-----TPYLRR----------------------------------------------------------------------------------------
119357846_Chlorobium_phaeobacteroides_DSM_266                     -QY-----DPYKLPV---------------------------------------------------------------------------------------
336116789_Microlunatus_phosphovorus_NM-1                          -PW-----TGLSWR----------------------------------------------------------------------------------------
37677204_Vibrio_vulnificus_YJ016                                  -SL-----TQQRIR----------------------------------------------------------------------------------------
154250301_Fervidobacterium_nodosum_Rt17-B1                        -KY-----QPYYLEKI--------------------------------------------------------------------------------------
38505760_Synechocystis_sp-_PCC_6803                               -SY-----SPFLIR----------------------------------------------------------------------------------------
83592490_Rhodospirillum_rubrum_ATCC_11170                         -PY-----RPFRFGTSRPG-----------------------------------------------------------------------------------
261415948_Fibrobacter_succinogenes_subsp-_succinogenes_S85        -EM-----ADLQWNPEYSESA---------------------------------------------------------------------------------
76802276_Natronomonas_pharaonis_DSM_2160                          DQY-----HPLTFQR---------------------------------------------------------------------------------------
312114617_Rhodomicrobium_vannielii_ATCC_17100                     -GF-----EALERP----------------------------------------------------------------------------------------
219683006_Bifidobacterium_animalis_subsp-_lactis_AD011            -RY-----RGIRVR----------------------------------------------------------------------------------------
116748794_Syntrophobacter_fumaroxidans_MPOB                       -SY-----EPFIWSEVS-------------------------------------------------------------------------------------
317051217_Desulfurispirillum_indicum_S5                           -SY-----QPFVRPQN--------------------------------------------------------------------------------------
328949009_Treponema_succinifaciens_DSM_2489                       TEY-----KAYYFK----------------------------------------------------------------------------------------
333996012_Treponema_azotonutricium_ZAS-9                          SEY-----RGYLYK----------------------------------------------------------------------------------------
313673559_Calditerrivibrio_nitroreducens_DSM_19672                -IY-----QPFILI----------------------------------------------------------------------------------------
218961167_Candidatus_Cloacamonas_acidaminovorans_str-_Evry        -EY-----SPFLWL----------------------------------------------------------------------------------------
294828260_Leptospira_interrogans_serovar_Lai_str-_56601           -IF-----AQARLR----------------------------------------------------------------------------------------
347756303_Candidatus_Chloracidobacterium_thermophilum_B           -LF-----ARMRLR----------------------------------------------------------------------------------------
162457462_Sorangium_cellulosum_So_ce56                            -LF-----ARFRIR----------------------------------------------------------------------------------------
86742032_Frankia_sp-_CcI3                                         -RY-----IPWSPT----------------------------------------------------------------------------------------
433631663_Mycobacterium_canettii_CIPT_140070010                   -HY-----FPYRTTGR--------------------------------------------------------------------------------------
15612904_Bacillus_halodurans_C-125                                -EY-----PPFLWK----------------------------------------------------------------------------------------
108763792_Myxococcus_xanthus_DK_1622                              -GY-----PPFAMR----------------------------------------------------------------------------------------
344996859_Caldicellulosiruptor_lactoaceticus_6A                   -RY-----IPLTYK----------------------------------------------------------------------------------------
85858452_Syntrophus_aciditrophicus_SB                             -VY-----QPVLLK----------------------------------------------------------------------------------------
148657121_Roseiflexus_sp-_RS-1                                    EQY-----TGFSLNN---------------------------------------------------------------------------------------
159898907_Herpetosiphon_aurantiacus_DSM_785                       PRY-----QPYVWRD---------------------------------------------------------------------------------------
17227877_Nostoc_sp-_PCC_7120                                      -DY-----EAFVIK----------------------------------------------------------------------------------------
268318378_Rhodothermus_marinus_DSM_4252                           -AY-----TPFVA-----------------------------------------------------------------------------------------
55978328_Thermus_thermophilus_HB8                                 -RY-----TPFYLWR---------------------------------------------------------------------------------------
297622675_Truepera_radiovictrix_DSM_17093                         -PY-----SPFYLSARPT------------------------------------------------------------------------------------
328953000_Desulfobacca_acetoxidans_DSM_11109                      -TY-----QPYTFFW---------------------------------------------------------------------------------------
337286709_Thermodesulfatator_indicus_DSM_15286                    EKY-----QAFSLE----------------------------------------------------------------------------------------
328953422_Desulfobacca_acetoxidans_DSM_11109                      -AY-----LPFSYR----------------------------------------------------------------------------------------
38505684_Synechocystis_sp-_PCC_6803                               VVY-----EPLVIR----------------------------------------------------------------------------------------
134299482_Desulfotomaculum_reducens_MI-1                          -GY-----KPFLVR----------------------------------------------------------------------------------------
17229060_Nostoc_sp-_PCC_7120                                      -KY-----PPLVMR----------------------------------------------------------------------------------------
302391602_Acetohalobium_arabaticum_DSM_5501                       EEY-----PPLVVK----------------------------------------------------------------------------------------
159898755_Herpetosiphon_aurantiacus_DSM_785                       -HY-----IPFLVR----------------------------------------------------------------------------------------
320161859_Anaerolinea_thermophila_UNI-1                           PGY-----DDMGFR----------------------------------------------------------------------------------------
```

```
156741961_Roseiflexus_castenholzii_DSM_13941                      ADY-----EPFAVR----------------------------------------------------------------------------------------
383762167_Caldilinea_aerophila_DSM_14535_-_NBRC_104270            AGY-----VPYLIR----------------------------------------------------------------------------------------
46255267_Thermus_thermophilus_HB27                                -RY-----EPFAVR----------------------------------------------------------------------------------------
134101617_Saccharopolyspora_erythraea_NRRL_2338                   -HY-----TAFTSR----------------------------------------------------------------------------------------
378719041_Gordonia_polyisoprenivorans_VH2                         -TY-----TPMVTR----------------------------------------------------------------------------------------
320102061_Isosphaera_pallida_ATCC_43644                           -DY-----PVFTTR----------------------------------------------------------------------------------------
268318441_Rhodothermus_marinus_DSM_4252                           PRY-----IPVTTR----------------------------------------------------------------------------------------
39995168_Geobacter_sulfurreducens_PCA                             PAY-----PNFVTR----------------------------------------------------------------------------------------
83591519_Rhodospirillum_rubrum_ATCC_11170                         PRY-----PHYCPR----------------------------------------------------------------------------------------
50085564_Acinetobacter_sp-_ADP1                                   ------------------------------------------------------------------------------------------------------
77166205_Nitrosococcus_oceani_ATCC_19707                          -QL---------------------------------------------------------------------------------------------------
15602176_Pasteurella_multocida_subsp-_multocida_str-_Pm70         ------------------------------------------------------------------------------------------------------
38234755_Corynebacterium_diphtheriae_NCTC_13129                   -----------EIN----------------------------------------------------------------------------------------
21674788_Chlorobium_tepidum_TLS                                   ------------------------------------------------------------------------------------------------------
53804737_Methylococcus_capsulatus_str-_Bath                       --------GNGDA-----------------------------------------------------------------------------------------
209966592_Rhodospirillum_centenum_SW                              ------------------------------------------------------------------------------------------------------
89885834_Rhodoferax_ferrireducens_T118                            ------------------------------------------------------------------------------------------------------
16130662_Escherichia_coli_str-_K-12_substr-_MG1655                --------AGHRSS----------------------------------------------------------------------------------------
68535718_Corynebacterium_jeikeium_K411                            --------GPIR------------------------------------------------------------------------------------------
86738746_Frankia_sp-_CcI3                                         VTGPDTDQPPPPW-----------------------------------------------------------------------------------------
51891801_Symbiobacterium_thermophilum_IAM_14863                   --------RNFGGRS---------------------------------------------------------------------------------------
29834079_Streptomyces_avermitilis_MA-4680                         --------MNHGTDT---------------------------------------------------------------------------------------
```

B

```
50085558|Acinetobacter_sp-_ADP1                                       ------------------------MHVIIISAC---------------EKRALKKTRAILDSYAIRTGHSSWQA-PMTIEGLKE---IRSAL------KK--VA-T----RQ----TAVAAYINF------GMR--RM-KL-AWVV----
77166204|Nitrosococcus_oceani_ATCC_19707                              ------------------------MIILFVSEC---------------EKAAWKRSRRILSRYAVQIGRRTWLA-RISAEGLQN---IRREL------TQ--AA-S----RH----TSVACHRVK------GRY--QT-EL-AWVV----
15602177|Pasteurella_multocida_subsp-_multocida_str-_Pm70             ------------------------MNILLVSQC---------------QKNALTETRRILDQFAERCGDRVWQT-AITQAGLET---LYKML------RQ--TA-R----KN----TAVACYWTH------RKN--SA-EL-LWIV----
209966591|Rhodospirillum_centenum_SW                                  ------------------------MIVICLSDT--------------ADRFHGFLRSVMLN-----VHPGVYVSMDLDKGSRER---IWDIL------TR--WW-E-AE-PR----GMVLMIHRD------TRKS-MD-LD-LRSL----
21674789|Chlorobium_tepidum_TLS                                       ------------------------MLVVVANDL--------------PPAVRGRMKLWFIE-----PRANVFVS-GVRDSLARK---VVDYL------HQ--HC------PP----KSGLMIFNS------SNTC-PG-YE-IFGL----
86738747|Frankia_sp-_CcI3                                             ------------------------MTVIVLIAT--------------ADGLRGHLTRWMIE-----ISAGVFVG-NPSARVRDR---LWTLL------SQ--RV------GD----GQAIMIEPA------TNE--QG-WA-ARTA----
38234756|Corynebacterium_diphtheriae_NCTC_13129                       -----------------------MFAVLYLQAA--------------PDHLLGYVTRFLTE-----ADTSIYVG-NVSKNVASN---LWIRV------TE--AI------KD----AHATMIVSD------NSR-EQG-FS-IMTT----
90111482|Escherichia_coli_str-_K-12_substr-_MG1655                    ----------------------MSMLVVVTENV--------------PPRLRGRLAIWLLE-----VRAGVYVG-DVSAKIREM---IWEQI------AG--LA------EE----GNVVMAWAT------NTE--TG-FE-FQTF----
77128445|Methylococcus_capsulatus_str-_Bath                           ----------------------MSMTVMVTRNV--------------SLRMRGFLASSMLE-----LAPGVYSAPRLSTAVRER---IWAVV------ED--WF-T--AEQD----ASVVMVWVD------PAM--PG-GQNVRTL----
57865885|Staphylococcus_epidermidis_RP62A                             ------------------------MYLLVSFDL--------PRDTKLERRIASKYRLRLIELGFSMKQFSLYER-YVSNVQKKDK--ILEIL------KQ--EI-P----DT----GSITLYILP------DEV--NS-NQ-ITIL----
193216854|Mycoplasma_arthritidis_158L3-1                              -----------------------------MYDV--------SFKDENDTKEYTKFYTSLQKLGYYRMQYSIYAK-TISTHTKYPY--EKAKL------IS--IL-P----KN----ANVRILCIS------EQQ--YQ-GI-EILR----
71894594|Mycoplasma_synoviae_53                                       ----------------------------MMYDI--------YSVDE-KNKEYTNFKKHILSLGYFMIQYSVYCK-NISVHTQFDS--EIKKL------SS--KI-P----NR----ANIRILLLT------ENQ--YQ-NM-IILS----
189440766|Bifidobacterium_longum_DJO10A                               ---------------MKRDEDSGGMWCLVMFDL--------PVKTKRQRHAATVFRNMLLDMGYGMVQYSVYVR-YTPTQSGNRA--TVKMI------KD--NL-P----AN----GLVRILHIS------DHQ--WS-TA-ARFS----
38232680|Corynebacterium_diphtheriae_NCTC_13129                       ----------------MPKKGSDPVWCVVMFDL--------PVKTKTQRKQATAFRQNLLDLGFCMAQLSVYVQ-YLPLAAKLSN--LVKLI------KE--KL-P----PG----GDVRILSVS------DIQ--WS-KM--IRF----
187250662|Elusimicrobium_minutum_Pei191                               -----------------MISPYKTMWILIAFDL--------PVKTKTQMRIANKFRQTLKNSGFIMLQKSIYSY-YAFSKERCDI--IEKQI------KQ--IV-P----DN----GHICFLFFT------DRM--FKMTK--HFY----
117929156|Acidothermus_cellulolyticus_11B                             ------------------MLKDEPVWAIAAFDL--------PTQTREQRRDANRYRHHLLELGFSRVQLSVYAK-YMINGAGFRW--VGSAA------GA--VV-P----PG----GVLRVIPVS------DHE--WA-RA-LCYV----
121608213|Verminephrobacter_eiseniae_EF01-2                           ------------------------MWMLVMFDL--------PVVEKAERQAATQFRNALLDMGFEMSQFSVYMR-FCASQTQVDT--LCRTV------ER--AL-P----SG----GKVHIFQFT------DKQ--YE-RA-ITYR----
325972001|Sphaerochaeta_globosa_str-_Buddy                            -----------------MLGPYNAMWVFVGFDV--------PTLTKLDVKRANRFRNDLLDMGFSRFQLSFYTY-YVPSKQRAET--LAKQV------EH--KV-P----PN----GKVSIFFIT------DRQFGMT-RT--FYA----
319957208|Nitratifractor_salsuginis_DSM_16511                         ------------------------MWLVVMFDL--------PTQSKKDRKRYRWLSRYLDKQGFIRLQYSVYAK-VFNSLQSANH--GKKRM------KE--FL-H-INVKK----GNVRMLLFT------DAQ--FG-KM-EIVV----
294086113|Candidatus_Puniceispirillum_marinum_IMCC1322                -----------------MLSGYRLLWIMTLFDL--------PVVTKAERKRATVFRNKLLDLGFEMVQYSVYMK-HVAGKEQAEA--IQVKI------KD--AV-P----EY----GNVKLLLIT------DRQ--FG-NI--LHL----
60683387|Bacteroides_fragilis_NCTC_9343                               ---------------MDRFNEYRIMWVLVLFDL--------PTETKKEKKAYADFRKNLQKDGFTMFQFSIYVR-HCGSSENAAV--HIKRV------KS--FL-P----EF----GHVGIMCIT------DKQ--FG-DI-ELFY----
154250553|Parvibaculum_lavamentivorans_DS-1                           ------------------------MWVIAMFDL--------PTDTPKARKAYARFRKDLLEDGFTMMQYSVYSR-HCASIENAEV--HVKRM------GA--VV-P----AQ----GEVRFLTIT------DNQ--FG-RI-KVYV----
187736487|Akkermansia_muciniphila_ATCC_BAA-835                        -----------------------MGWLVVFFDL--------PTTTPEDKKNYQYFRKALLEDGYTMIQYSVYAR-ACVTQERMAT--HSRRI------QK--MI-P----PD----GSVRCLFVT------NIQ--WD-KM-FVFH----
116627544|Streptococcus_thermophilus_LMD-9                            -------------------MRYEALRLLCFFDL--------PMESKDEKRIYRNFRKELISNGFEMLQFSVYYR-TCPNRSFANK--FYKKL------KM--SN-L----PA----GNVRLLAVT------EKQ--FS-EM-TLII----
57238544|Campylobacter_jejuni_RM1221                                  ------------------MIEDKFMRVLLMFDV--------PTKSKKEQKLASKFRNNLIKLGYFMLQFSVYMR-ICKGLSSAKS--SIENV------KK--IL-P----PY----GNVRALIIT------EKQ--FD-KM-ELLL----
291276263|Helicobacter_mustelae_12198                                 -----------------------------MFDM--------PTKSKEDRKASAKFRSTLLKEGYFMLQFSVYVR-ICKGVKSANT--YLERL------HL--YL-P----KK----GHIRALILT------EKQ--FD-HM-KILI----
220930480|Clostridium_cellulolyticum_H10                              --------------MSDILGRNRFMRILVFFDL--------PVKKKMDRDNYTKFRRNLIKDGFCMIQYSVYCR-VCSCPEMAEK--HVNKV------KT--FL-P----AK----GSVRVLTVT------DRQ--YS-SM-IILL----
209542527|Gluconacetobacter_diazotrophicus_PAl_5                      ----------------MKAEDIRFMWLLVFFDL--------PVRTKEQRRRASRFRHFLKDDGFLMLQFSVYAR-ICRGQDAVEK--HVRRV------RS--NL-P----KE----GSVRTLQIT------DRQ--YG-RM-ELML----
90425959|Rhodopseudomonas_palustris_BisB18                            ----------------MKSEEVRFVWMFVFFDL--------PVGTKTDRRHATRFRKFLKDDGFMMLQWSVYSR-VCRGDDAADK--HGHRV------VK--NL-P----PR----GSVRTLQVT------DRQ--YA-RM-RLLL----
189485227|uncultured_Termite_group_1_bacterium_phylotype_Rs-D17       ----------------MKKEGTRFMRILVFFDL--------PTETAENRHNAAKFRNNLIKDGYYMLQYSVYCR-ICRGMDSSEK--HIRRL------MM--RT-P----ME----GNIRVLEIT------EKQ--YE-RM-KIIL----
310780382|Ilyobacter_polytropus_DSM_2926                              --------------------MYKFMRIIVFFDL--------PTAKKRDRKNYAIFRRFLLNDGYHMLQFSVYSR-ICNGYDSVKK--HIKKL------NN--NL-P----PN----GSIRVLTLT------EKQ--YV-GM-ELLV----
15602990|Pasteurella_multocida_subsp-_multocida_str-_Pm70             ------------------MSEGKFMRLIVFFDL--------PVTTKAKRKAANQFRQFLLKDGYQMLQLSVYAR-IIRGRDSMQK--HNKRL------SE--HL-P----EE----GSIRCLEVT------EKQ--FA-GM-KILL----
222109283|Acidovorax_ebreus_TPSY                                      ------------------------MRMLVFFDL--------PVVSKADRRAYTVFRRFLLNDGYDMIQFSVYGR-ILNGTDAAQK--HMQRL------LA--NL-P----SE----GSVRVLTVT------EKQ--FA-SM-KLLV----
238924077|Eubacterium_rectale_ATCC_33656                              ------------------MDKYNIMRMFCMFDL--------PVETEEQRRNYRIFRKELIQEGFVMIQYSIYVR-VCPSREFANR--IENRI------KK--NL-P----LE----GNVRLLCVT------EKQ--YD-DM-KLLV----
406600269|Leuconostoc_gelidum_JB7                                     ---------------------------MIFFDL--------PTLTDADRRNYRHFRKSLINEGFLMIQESVYVR-VTTNRQAAAL--LEKRI------QT--IT-P----VD----GLVQSLIVT------EKQ--YT-MM-KFLS----
90961087|Lactobacillus_salivarius_UCC118                              ------------------------MRLMIIFDL--------PVETSEERKEYRKFRKNLINEGFIMIQYSVYVR-VCVNKKSAEF--TEKRI------ES--FL-P----SK----GVVQSLILT------EKQ--YN-DM-HFLL----
269123824|Streptobacillus_moniliformis_DSM_12112                      ------------------------MRLLVYFDL--------PVSTKLERKQAQDFRKNLLSEGFIMMQFSCYSR-FCRNDAESSK--YFNRV------KK--LS-NKL--SG----GEVRILKIT------NKQ--YE-NM--LIL----
34557788|Wolinella_succinogenes_DSM_1740                              ---------------------MKFMRLMVMFDL--------PTGTKKERKAYTLFRKFLIKEGFEMMQFSVYVR-ICRGADMALA--YEDRI------RR--NI-P----PK----GHVRSLVLT------NSQ--YE-RM-HILL----
331702226|Lactobacillus_buchneri_NRRL_B-30929                         ------------------------MRLMVMFDL--------PVETSENRRNYRKFRKALINEGFLMIQFSVYVR-VCVNKKSADF--MEKRI------AA--IA-P----PE----GLIQSLIMT------EKQ--YN-DM-HFIA----
15675043|Streptococcus_pyogenes_M1_GAS                                -------------------MSYRYMRMILMFDM--------PTDTAEERKAYRKFRKFLLSEGFIMHQFSIYSK-LLLNNTANNA--MIGRL------RE--HN-P----NK----GNITLLTVT------EKQ--FA-RM-IYLH----
374307736|Filifactor_alocis_ATCC_35896                                ------------------------MRIIVFFDL--------PTITLEDKRDYRDFRKFLIKKGFLMMQESVYCK-LVLNQTVASS--VLSSI------RK--HK-P----KS----GLVQMIMVT------EKQ--FA-KM-EYLC----
302336018|Olsenella_uli_DSM_7084                                      ---------------MTMEARIRYMRLIVFFDL--------PIDTAAQRREYRLFRKYLLHDGYLMLQESVYAK-LVTNDGTAGA--AVMRL------RK--NR-P----SS----GLVQILKVT------ESQ--FA-TM-EYIT----
169823757|Finegoldia_magna_ATCC_29328                                 ------------------------MRTIVFFDL--------PTISLDDKKRYRAFRKFLLSEGFMMLQESVYCKLSLNQTNANT---LKDRI------EK--NK-P----SM----GIVQILTIT------EKQ--FQ-SM-EYIV----
42525845|Treponema_denticola_ATCC_35405                               ------------------------MRVIVFFDL--------PVITPENRHNYSVFRKYLIKSGFIMQQKSVYSKLVLNLTNRDS---IVKSI------EK--NK-P----PE----GLVEVLTVT------EKQ--YA-KM-EIII----
148642084|Methanobrevibacter_smithii_ATCC_35061                       ------------------------MKILVLFNL-------------KFKKDILKLKRILNYNAFKQISEKTYIG-NLDNNEILV---LKEKI------SK--IP-D----EN----GFIIIIPIC------GAC--YN-KI--TYF----
88602028|Methanospirillum_hungatei_JF-1                               ---------------------MTVKHLIVCYDV-------------EKTKDRNKVIKVLEYYGLIRVQYSVFMG-SLTETRLHQ---MNARI------KR--EF-T-K--PS----IKILVIEVC------NAC--ME-RA--LLV----
119720227|Thermofilum_pendens_Hrk_5                                   ------------------------MIVIVAYDI-------------SDEDRRGRLRRYLRRLGLARVNRSVYAG-PGTATTAEL---VAERA------KE--IV-E----EG----DSVFVIVVR------EDE--YQ-RA-HVFD----
70607607|Sulfolobus_acidocaldarius_DSM_639                            ------------------------MYVVVAYDI-------------SDEKVRGKVRRLLRRYGLSFISRSVYAG-RLSWNRALL---ISERV------AK--YL-E----ER----DSVVIVPVQ------NVD--FS-RT--LIV----
156937922|Ignicoccus_hospitalis_KIN4-I                                ---------------------MEGVYVIVSYDI-------------KDDRKRLKVMKRLLALGFSRLQKSVYVR-RGSRGDAKD---AFNAT------VK--YL-D----DG----DKLFIIVVS------DKE--FR-----EAW----
18311771|Pyrobaculum_aerophilum_str-_IM2                              ------------------------MYVVVAYDI-------------TEDEVRNKVADALKAYGLERIQRSVFVG-RINPALLKD---LVERL------KR--IT-K-G--AN----ADITIFKVD------RRA--ID-TA--IRI----
385805116|Fervidicoccus_fontis_Kam940                                 -----------------------MGFYIIVYDI-------------GDQGKRQRVVEILEKWGLRRIQRSAFVG-YVQEGRAKD---IAREI------EM--TI-D-L--EK----DVVHIVQIE------RKE--WN-RA--IVV----
15922976|Sulfolobus_tokodaii_str-_7                                   ---------------------MVKMWVIVVYDI-------------SDDEKRNKIARELQRLGLSRIQRSAFVG-DIDSQRFKD---LVRIM------GK--LV-T-G--KD----DILHIIPLG------LRD--WERRV--VIN----
347523052|Pyrolobus_fumarii_1A                                        ------------------------MIVIVAYDI-------------SDDYKRDRAAKLLLSMGFQRIQRSLYAA-RGGREKARQ---VAAAL------KR--II-D-P--VT----DRVDIILVH------DHT--WE-TR--IVI----
118431379|Aeropyrum_pernix_K1                                         -----------------------MVYVLIAYDI-------------SNDSKRLKAAQKLLQMGFARVQKSVYIA-KGGRSLAKE---AYRAL------QR--LA-D-S--GK----DKIMVMVIP------GDS--VR-----DAY----
305663329|Ignisphaera_aggregans_DSM_17230                              -----------------------MPYIVVAYDI-------------SDDRKRLEVCEKLKSMGFTRIQRSLYIA-RGGSSKAKD---VARAL------QR--YI-D-R--ST----DSVLVMVIP------REV--LE-KT--ITI----
307595806|Vulcanisaeta_distributa_DSM_14429                           ------------------------MYVLVSYDI-------------SDDRRRLEIMNKLKSMGYVRVQRSLYIA-RGGNALAKD---TARAL------TR--LM-N----RE----DSVVILIID------GQT--LN-NA--IKL----
302347807|Acidilobus_saccharovorans_345-15                            ------------------------MRVLVVYDV-------------SDDSRRLRLSRELERMGLSRVQRSAFVG-PGGLAVAKE---AARAA------AQ--LV-D-D--ST----DVVHVVLLS------EVE--WA-NA--LVV----
124027532|Hyperthermus_butylicus_DSM_5456                             -----------------------MPYVVVFYDV-------------SDNKRRDLLAKTLQSLGLVRVQRSVFMG-RGGYTKAKE---AIRAA------SR--IV-D-A--RT----DSVVALVVP------EDY--AR-RM--LVY----
18311698|Pyrobaculum_aerophilum_str-_IM2                              ------------------------MIWLAVYDI-------------EDDGERAKASAILQAWGFVRVQRSFYVG-RMPRGKAAD---LLKIL------QR--HV------KS----GHIALIPIT------DEL--LA-KA--LEL----
150398873|Methanococcus_vannielii_SB                                  ------------------------MIYFLIYDI-------------SGNKARNRVINYSKNIGLYRIQKSVFSG-NTSIENINV---LKEKI------RK--TI-D-L--ET----DSVFIFGVS------KDQ--MD-KF--EVI----
257125861|Leptotrichia_buccalis_C-1013-b                              ------------------------MMTWLIYDI-------------TDDRNRREIIKIAQKHGLYRVQKSVFLG-NMEKDEVDE---IIVET------KK--AI-D-L--RE----DSVYIFPVC------EAD--YK-KA--TLL----
150401498|Methanococcus_aeolicus_Nankai-3                             -------------------------MIYVIYDI-------------SDNKIRKKISDKCLNYGLYRIQKSVFAG-SLNNNRVDE---LRVFC------EN--IF-K-EGDEN----DKIYIIPVC------KKC--FG-SL--ICV----
21226660|Methanosarcina_mazei_Go1                                     ------------------------MLVWVIYDI-------------TENRIRQKVSDCCKSYGLYRVQKSVFLG-DLNSNDRDS---LALEC------EE--WI-D-T--EK----DSVYVFPMD------EQS--FK-QV--KLI----
218961576|Candidatus_Cloacamonas_acidaminovorans_str-_Evry            ------------------------MLTWVMYDI-------------VKDKIRNKVSKCCEQAGIYRVQYSVFLG-EISRTKCKE---LSCQI------EE--LI-N-P--EI----DRVYIFPMC------NED--FK-SC--QLL----
284040852|Spirosoma_linguale_DSM_74                                   ------------------------MIIWVLYDI-------------VKDKSRTKAAKRCQQAGLYRVQFSCFLG-TLTQNQKDT---LQLQL------EE--LI-D-E--ET----DKVYIFPMS------RGE--LQ-QT--ALL----
15898244|Sulfolobus_solfataricus_P2                                   ----------------------MAMLYLIFYDI-------------TDDNLRNRVAEFLKKKGLDRIQYSVFMG-DLNSSRLKD---VEAGLKIIGNRKK--LQ-E-D--ER----FFILIVPIT------ENQ--FR-ER--IVI----
```

```
332668018|Haliscomenobacter_hydrossis_DSM_1100                        ------------------------MPHLICYDI------------TQDSLRTRLAKKISESGLDRINKSVYLG-SITDTALNA---LEKDL------AK--FV-QEKGSPG----DSLIILPVY------LEQ--IN-QM--RIY----
386854796|Deinococcus_gobiensis_I-0                                   ------------MPRLLQDPRKAGHPILVLYDV------------GDDKCRTKVMTACKDYGLGRWQYSAYHG-RLTGAARRE---LEIRL------EK--AM-G-A--SP----GLIAILQLE------QWQ--LD-DA-TLIV----
288560432|Methanobrevibacter_ruminantium_M1                           ------------------------MLVFVMYDI------------VDTPTRTSLVKKLKHYGLHRIQKSIFCG-FLSIDERLN---LASEF------DF--FM-S-S--ER----DSIILIPTC------ESC--VD-SI--LIE----
330509016|Methanosaeta_concilii_GP6                                   ------------------------MDTIIIYDI------------SDNSLRARVAKVLLDFGCIRIQKSAFWG-IMNHNTREK---LRLRL------ER--MM-H-E--KE----GNIQFYPMC------SKCFSLKDSI---------
20094746|Methanopyrus_kandleri_AV19                                   ---------------MGVLGGPSPRLRLYVYDF---------KEPGGEAERRKLRELLESHGAFRLQYSTYALLAEPEVHARV---LRRVV------AR--VD-F-E--EG----DSLIVVPMC------RRC--LR-VA--RWV----
11499460|Archaeoglobus_fulgidus_DSM_4304                              ------------------------MLHLVVYDI------------SDDGSRARLAKLLEKFGLQRVQYSAFRG-ELNPNDREV---LARQV------GK--FV-R-D--DR----DCIFIIPLC------QRC--SS-TA--IVI----
383319614|Methanocella_conradii_HZ254                                 ------------------------MMCLVIYDI------------TDDKRRNKLAAMLKVFGLARIQYSAFRG-ELNSNDRMV---LAKKA------SQ--FV-E-D--EK----DSIFIIPLC------DRCLGTA-TV--LSK----
148656005|Roseiflexus_sp-_RS-1                                         ------------------------MRCLLIYDI------------PDDRVRQKIADACLDYGLQRIQYSAFAG-NLSRTHQRE---LMTEV------RR--RL-G-K--HA----GNIQLFTFP------DDI--WNARR-TIEQ----
408403001|Candidatus_Nitrososphaera_gargensis_Ga9-2                   ------------------------MRYIIVYDI------------TDDNLRTLTSETLKDYGLKRIQKSAFMG-SLKKHALNS---LLTDL------RR--MV------NV----DSVIVFPLC------DSD--FR-NM--VSI----
167628643|Heliobacterium_modesticaldum_Ice1                            ------------------------MKTLVIYDI------------ELDKVRTKVSELCKDYGLQRIQYSAFFG-ELNTNRREE---LRQRI------KK--TL-G-R--TV----GRIHFFPIC------DKD--LR-LV--MQVMNLP
170290048|Candidatus_Korarchaeum_cryptofilum_OPF8                     -------------------MRGGNLKVLVVYDI------------TDDSLRLKVAEILKDLGLFRIQKSAFIG-EMTSQEREN---MEEIL------RR--QNLG----PS----DRIDVFPIC------DRDLKMHSQI---------
126465228|Staphylothermus_marinus_F1                                  ------------------------MKTLIVYDI------------SSTKLRSKVAEILKDWGLTRIQLSAFIG-ELTPQERND---LAEYL------KR--LP-L-S--ER----DKIDIFPIC------IKD--LKLHM--RIV----
327314605|Prevotella_denticola_F0289                                  ------------------------MLFFVMYDI------------ESDKVRYHIAKYLERKGCTRIQRSIFLA-DLDKSVYDE---IKSDL------AE--VQ-S-LYDNH----DSIIVCPVS------TDQ--IR-AM--QII----
392395715|Flexibacter_litoralis_DSM_6794                              ------------------------MTYLILYDI------------SEDRIRTHISKHLEERGCRRVQKSVFMA-KMSIEKYHE---IRDYL------QQ--IQ-D-EYEND----DTILFIPIS------QQV--LR-ET--YMI----
34541576|Porphyromonas_gingivalis_W83                                 ---------------------MNDKKLLVAYDI------------SSNRRRRKVARILEQC-GIRINKSVFIC-SLRELTMDK---LVEAV------TS--QT-A----KR----DKVFFLPLC------QHC--YT-AA--WMS----
288947692|Allochromatium_vinosum_DSM_180                              --------------------MSQSGWYLLAYDI------------ADPRRLQRLHRAVRRE-GIALQRSVFLV-QGTESDIEA---LLDRL------ET--LM-D-V--AK----DDLRAYPVAAPSALWLQGQCVV-QG---HL----
379728402|Saprospira_grandis_str-_Lewin                               ----------------------------MSYDI------------SSNRLRLAVCKLLQRQGAERLQKSVFFAPNMLLEEIRE---LRADL------QQLMLHHP-EALPQ----DSLICFPIR------KQS--IE-EM--VWA----
431929778|Thioflavicoccus_mobilis_8321                                ---------------------MARKPAVIAYDI------------RCDKRRRRAARCLKGW-RLDGQYSLIEC-RLTEAEAQE---LFLQL------VD--II-D-R--DE----DALLLAWMD-----------GT-RE--SCP----
288947691|Allochromatium_vinosum_DSM_180                              ---------------------MSSRLAVFAYDI------------RDDRVRRHALKTLREW-RLDGQLSVHEC-QVDAIQARR---LFEQL------GD--EL-D-P--AT----DAWLFTWVE------GHR-----AV--LAR----
269798875|Veillonella_parvula_DSM_2008                                ------------------------MRYLISYDI------------SDTKIRNKVVKYLESF-CYRIQYSVFLC-VHAHVRVET---IAKRL------EV--LT---QGDET----RRLLIVSIS------DTA-----KT-PIWC----
30248143|Nitrosomonas_europaea_ATCC_19718                             ---------------------MNRNLYLVAYDI------------CNPRRLRQVCRYLTGY-KVSGQKSVFEI-WVTPTELHT---IRTEL------DK--LM-D-T--QA----DRLHILSLD------PRM-----KP--RCY----
206889468|Thermodesulfovibrio_yellowstonii_DSM_11347                  ----------------------MRVLYIIAYDI------------TDARRLGQIRYFLKGY-STGGQKSVYEC-FLEREELKF---IISKI------KR--LINP----NE----DRVHIFRID------GRS-----KV--ITL----
121997422|Halorhodospira_halophila_SL1                                ---------------------MRRKLYLAAYDV------------RCPRRLSRAVRVIKGY-ASGGQKSAYEC-WLTPAEHKA---LHTEM------AQ--VIDP----TE----DQFALLPLE------PRR-----PL--VTL----
292492855|Nitrosococcus_halophilus_Nc_4                               --------------------MASRQLYLAAYDI------------ASATRLRQALYCLKKY-ATGGQKSVFEC-YLTKTERRK---LLREM------EG--II-D-A--EE----DRFLLLRLD------PRS-----RV--RTL----
430760779|Thioalkalivibrio_nitratireducens_DSM_14787                 --------------------MSERDFFLAAYDV------------SRPSRLAAALKLVRAY-ATGGQKSVHEV-FLTPAERGD---LLHDM------AM--IL-D-E--EE----DRFLLLRLD------PRA-----RV--YTL----
317051220|Desulfurispirillum_indicum_S5                               -----------------------MDYIITCFVP--------------EKKRERLIRILQSF-HIPTDHNIVTC-NLTTLQAKQ---LRSAL------EL--VIEP----DL----DEIRIYTPC------PYC--EQ-QV--ITI----
118497968|Francisella_novicida_U112                                   -------------------------MLIVSYDF------------SNNKVRAKFAKFLESY-GVRLQYSVFEL-KYSKRMLDL---ILAEI------ENN-YV-P-LFTNA----DSVLIFNAP------DKD--VI-KY---------
76802277|Natronomonas_pharaonis_DSM_2160                              ----------------------MTMRLAVAYDV------------SDDTNRRHVYRTLQRY-GAWRQYSVFEL-EISKTDRVE---LEAEL------ES--HI-D-P--SA---GDRIRIYRLC------SAC--QD-AT---TD----
162457463|Sorangium_cellulosum_So_ce56                                -------------------MPEPRHWHLVVYDV------------SEPKALRKVHKVLSAW-GKPVQYSVFRV-RGTARELAR---LRFEL------TG--LI-A----PT----DRLMVIRLC------PDC--AT-RV--TVQ----
347756304|Candidatus_Chloracidobacterium_thermophilum_B               --------------------MGERRWYLVSYDV------------RDPKRWRKVYERVKGN-AERVQFSVFRM-YCTKTDLEQ---LRFDL------AK--LM-T----SE----DDLLVIHLC------PGC--AR-RV--VDT----
24215881|Leptospira_interrogans_serovar_Lai_str-_56601                ----------------------MKHWRLVSYDI------------REPKRLRRVAKIMEGF-GERIQYSVFRI-YSTDKELEK---LRWKL------AK--VT-E----EE----DNIFYLTLC------TKC--AS-GA--HTQ----
312114616|Rhodomicrobium_vannielii_ATCC_17100                         -------------------MSKAAHLYLVAYDI------------SCPRRWKRVQKVIKGL-CRRSQLSVHFC-RATPARIRR---LENRL------HR--IMHD----EE----DRLMVIDLG------PAH--TQAQL--SLL----
218884092|Desulfurococcus_kamchatkensis_1221n                         ------------------------MYALVIYDV------------SSNEKRYNLANYLKTKGLTRIQRSAFIG-RIPPSTLLD---VERVI------PK--LI-D-G--SI----DVVHVFQLY------DTA--IQ-RL---KV----
431929779|Thioflavicoccus_mobilis_8321                                ----------------------MKHWSLIAYDI------------RDPRRLRRVHLCLRRQ-ALAVQKSVFLI-EADERRLAE---ILAEL------RR--RA-D-N--RE----DDIRLYAIPDPAAIWAAGRQSE-RV--ANL----
39995169|Geobacter_sulfurreducens_PCA                                 ----------------------MEHLYIVSYDI------------RNQRRWRRLFKTMHGF-GCWLQLSVFQC-RLDRIRIIK---MEAAI------NE--IV-N-H--AE----DHVLILDLG------PAENVKP-KV--SSI----
83591518|Rhodospirillum_rubrum_ATCC_11170                             ---------------MGTRRSNAEHAYVVAYDI------------ADPKRWRQVFKTMKGY-GQWVQLSVFQC-RLDGGRRIA---MASIL------ES--LIDR----ET----DHVLMLDLG------PAE--DV-DLAVESL----
332661940|Haliscomenobacter_hydrossis_DSM_1100                        ------------------------MPYLVCYDI------------SSTKKRSKIAKELVNSGFVRIQKSVFIG-MISSTGVSL---LAKRL------QA--LM-G-H--KT-AGHDKIHFLKIH------EQA--LL-NM---LM----
261415949|Fibrobacter_succinogenes_subsp-_succinogenes_S85           ------------------------MKYLVAFDI------------SNSRRRGRMVKLCLAT-GFRVQKSVFES-FLDSVQVAS---FENEA------EK--II-D-P--DT----DSVRLYPID------ARA--DE-GV--RIV----
121997420|Halorhodospira_halophila_SL1                                -------------------MGERAIDHLVCYDI------------RKPARLRRVHRLMKEW-GVPLQYSVFYC-RLVPSARHQ---LAEAL------RR--EI-D----ERV---DDVRIYALQ-----PQGEGTYQGPA-PLPP----
302343125|Desulfarculus_baarsii_DSM_2075                              -------------------MSGAEMLVVFAYDV------------EDDSRRRRLARVLGNH-AVRVQKSVFEA-WLDEGAAKI---IASRA------AA--EL-G----PR----DSLRVYALD------ASG--VG-KT--LVF----
389875623|Tistrella_mobilis_KA081020-065                              -------------------MSETRFVMVYAYDI------------VADRDRDRIAGLLEDV-AVRVQRSVFEI-RQTREEADR---LIARL------TP--YL-A----PG----DKLRVYALG------ATG--LD-HS--RVF----
374998938|Azospirillum_lipoferum_4B                                   -------------------MSETPMVMVFCYDV------------ANDRRRRRVSAVLEEW-GVRVQKSVFEA-RLTEPQARA---LLARA------GK--EL-G----PR----DSLRMYALS------AHG--LA-HS--AAL----
294676822|Rhodobacter_capsulatus_SB_1003                              -------------------MPRAEMLMVFTYDV------------SQDRKRRRVARILEGA-ATRVQYSVFEA-RLSEQRAAA---LAQRV------AA--EL-A----PG----DSLRVYAIG------ASG--ET-RT--RTY----
288957882|Azospirillum_sp-_B510                                       -------------------MPAAEMLMVFCYDV------------ADDKRRRRVAKALEDA-AVRVQKSVFEA-RMTDRAAAR---LAERV------AA--HL-A----PG----DSLRVYAIG------AHG--LT-RS--RSY----
319790503|Thermovibrio_ammonificans_HB-1                              ------------------------MMYVVVYDI------------RDNKTRLRVANILEKY-GERVQYSCFEIGNYSQEIIKK---LALIL------SV--RI-N-Y--DT----DRVFFYPIS------NYCRSYILRF-PRKF----
108759816|Myxococcus_xanthus_DK_1622                                  ------------------------MTVLICYDV-------RVSDPQGPRRLRKIARACRDH-GVRVQYSVFEC-VLEPKDWVV---LRARL------LG--AF-D-S--EC----DSLRFYFLS------EDD--AR-KT--EHH----
15612905|Bacillus_halodurans_C-125                                    ------------------------MLVLITYDV-------QTSSMGGTKRLRKVAKACQNY-GQRVQNSVFEC-IVDSTQLTS---LKLEL------TS--LI-D-E--EK----DSLRIYRLG------NNY--KT-KV--EHI----
189485224|uncultured_Termite_group_1_bacterium_phylotype_Rs-D17       ------------------------MLVLISYDI-------STIDAGGQKRLRRVSKECLDY-GQRVQYSVFEC-EVDPTQWIF---LKDKL------LN--TV-N-L--KK----DSIRFYMLG------SDW--KR-KI--EHY----
86742033|Frankia_sp-_CcI3                                             ------------------------MDLLVTYDV-------DTTTPDGNRRLRKVAKICEGH-GIRVQKSVFEI-VCTEPQRVL---LEHKI------SQ--VI-D-E--TL----DSIRIYQMP------QHT--LD-NV--HHL----
433631662|Mycobacterium_canettii_CIPT_140070010                       ------------------------MEILVVYDI-------STVSSDGERRLREVAKVCEGY-GQRVQKSVFEC-RLEQREIRL---LIHDI------EK--VI-D-R--AA----DRVAIYRLR------EPY--QR-HV--VAL----
333996013|Treponema_azotonutricium_ZAS-9                              ------------------------MTYLFAYDI------------SDCHRRNLVAKKLEQF-GFRIQRSIFQC-DVSPSVAED---IKLAL------RR--YI-N-E--KE----DSLLIYPLC------ADC--LE-KA--CII----
328949008|Treponema_succinifaciens_DSM_2489                           ------------------------MIYFISYDV------------SNPKRLNKVAKTLENF-GIRIQFSFFEC-EMEREQLEE---LKSAL------LK--II-N-K--KE----DSLRIYPLC------QDC--AA-KT--TSL----
337288661|Thermodesulfobacterium_geofontis_OPF15                       ------------------------MDYLIVYDI------------SDHKKRLKVSRLLQKF-GYRIQYSAFYL-PETKGNIIDF---LFKEI------SK--LI-D-K--NT----DRVFFYPVH------PPE---------VFK----
319790501|Thermovibrio_ammonificans_HB-1                              ------------------------MMYLITYDI------------SNDKTRTKVADILEKF-GKRVQFSCFEV-ECTEKELSE---LFKRI------DK--LI-N-K--QT----DHVIAYPLT------KNS--LKGET--AIC----
283778925|Pirellula_staleyi_DSM_6068                                  -----------------------MSLYLASYDI------------SDDKRRAKVARFLLQF-GARLQKSVFEI-DLDPDEIAE---FRRDV------GS--IL-S----QK----DLFVLVPVD------ERG-----SRSTLSW----
386750120|Helicobacter_cetorum_MIT_00-7128                            ------------------------MKFLITYDI------------IHNKKRKLLAEELEKH-GTRVNLSVFEC-ELTETKFKQ---LVFKI------KK--MI-R-S--KN----GNVRFYRLC------SNC--ED-KS---FE----
220934971|Thioalkalivibrio_sulfidophilus_HL-EbGr7                      -----------------------MQTFIIGYDI------------SGNRARRRAMKALEAY-GTRVQRSVFEIHVRGEAQLRD---LLARL------RE--EL-A-E--EE-G--ADLRAWRLT------EEG--RR-QS--ATL----
15609953|Mycobacterium_tuberculosis_H37Rv                             ------------------------MFVLVIYDI------------SDNRRRASLAKILAGF-GYRVQESAFEA-MLTKGQLAK---LVARI------DR--FA-I----DC----DNIRIYKIR------GVA-----AV--TFY----
85858453|Syntrophus_aciditrophicus_SB                                 ------------------------MIVLITYDI------------TEPKRLTALHHFLKEF-GLNTQKSVFEC-DIDEAGIRK---IRAYC------RN--NL-N-I--AK----DSVRIYKIC------SRC--IN-KV---VL----
268318440|Rhodothermus_marinus_DSM_4252                               ----------------------MRRYYLVCYDI------------RQSKRLQRMHRLMAGY-GDPLQYSVFLC-LLSRAEVVT---MLDAI------KE--VI-N-Q--RE----DSVLVVDLGKVKGARPDC--IQ-TL---------
320102062|Isosphaera_pallida_ATCC_43644                               ---------------------MSRRTYLVCYDI------------CDPKRLRKVFQTLRDF-GDHLQYSIFEC-RLTPIDLAN---CRRLL------GD--II-H-H--GE----DQVLFVDLG------PTEQRSD-RV-ITAL----
378719040|Gordonia_polyisoprenivorans_VH2                             ---------------------MARRRHLIAYDI------------CEPARLRRVCKVMEAY-GTRLQYSVFLC-DLSELERTR---WRTDI------HG--VM-D-Y--GE----DSVVVIDLG------AVG--SAATV--TQL----
134101618|Saccharopolyspora_erythraea_NRRL_2338                       ---------------------MRRHRYLVAYDI------------RDPRRLRLVAKKMEDF-GDRTQYSVFLC-DLTRADVAD---MTRAL------LT--VI-D-S--TI----DRILIVDLG------SGD--SDSRF--EFL----
336116790|Microlunatus_phosphovorus_NM-1                              ----------------------MALTIVAAYDV------------SDDSRRARLAALLQTC-GDRIQLSVFVL-TLGDDELDE---LQARA------LS--II-D-P--EE----DSLYLFRQC------GGC--WE-AA--RCI----
268318377|Rhodothermus_marinus_DSM_4252                               -----------------------MAFWLVGYDI------------AEPRRLQRVARLLEDY-GQRLHYSVFLC-RLEADEMRR---LRQGL------LH--IV-D-L--EE----DRLFFLRLC------RGC--ID-AL--EQM----
83592489|Rhodospirillum_rubrum_ATCC_11170                             -----------------------MNRYVICYDI------------VDDKRRLKVAKCLDSY-GSRVQFSVFEV-LVSKPLMTR---MVREL------GA--LI-N-A--KT----DRISIYPQC------ATC--DA-RR---TD----
206890293|Thermodesulfovibrio_yellowstonii_DSM_11347                  ----------------------MRLPYLVCYDI------------SDEGRLNRVYRFMKGK-GFHIQYSVFYC-ILTDVELKE---MKAEI------LK--LI-H-S--RY----DDVRIYPLP------NNS-----LV--AVL----
52425689|Mannheimia_succiniciproducens_MBEL55E                        ------------------------MRYLIGYDI------------TDSKRLQRIYRRMIKF-ATPLQYSVFLF-NGTKEQLDK---YMQTV------LR--LYNK----KE----DDLRIYPLP-----------VQAKY--WQI----
30248141|Nitrosomonas_europaea_ATCC_19718                             ------------MPADHFLKRQAMSDFIICYDI------------TDPRRLGRLYRYLIKR-AVPLQYSVFLF-RGDDRQLER---CIQDA------IE--LI-D-E--KQ----DDLRVYPLP------GRG--LK-AR---------
292492858|Nitrosococcus_halophilus_Nc_4                               ------------------MPDNARRLHLICYDI------------ACPRRLGRIHRYLKGR-AAPVQYSVFVA-KLNRREISL---LAARL------AH--RIHP----RK----DDIRIYPLP------EKL-----AL--QRL----
430760782|Thioalkalivibrio_nitratireducens_DSM_14787                 ------------------MALNAPRAWLICYDI------------ADPRRLIRLHRFLKRF-AQPVQYSVFYY-EGSSAQLAR---CLHDI------AG--RI-D-H--RE----DDVRAYPIP------NPA-----QV--DTL----
348026606|Megasphaera_elsdenii_DSM_20460                              -----------MKNSQTESVDDKKYIVLIIYDI------------TDNKTRNKMVACLEKY-GVRVQKSAFEA-YITKRKYHK---LMQEA------PF--LI-D-T--DT----DSLRIYLLD------SYM-----AV--HSW----
317051218|Desulfurispirillum_indicum_S5                               ------------------------MIVIIAYDI------------PSTKRRTRLHKALKGY-GVNAQRSVFEC-DVQPEQMKG---LLTML------EF--MI-D-P--SE----DDLRVYPLC------KGC--QE-KV--MGM----
337286710|Thermodesulfatator_indicus_DSM_15286                        --------------------MAKRQFYIVSYDI------------SDEKRLREIHKFMKDF-GEWKQKSVFEC-WLTPEAYEN---MKAGL------KK--LF-K-P--RE----DRVRIYRLC------KTC--RE-RA--FSY----
85858454|Syntrophus_aciditrophicus_SB                                 ------------------------MNMIVAYDI------------ADPKRIALVARILKDY-GVRVQKSIFEV-DVNGGVFAV---MKERV------EQ--VIIP----AE----DGVKYFPLC------RKC--IG-MV--EII----
436839746|Desulfovibrio_hydrothermalis_AM13_-_DSM_14728              ------------------MFMSKITRHLIAYDI------------SCDKERYKVSKILEGF-GFRVQESVFEC-DLTTQLKGR---LIAKL------EH--LQ-I----ET----GFVSIYPLG------SHK--PK-DI---------
134299483|Desulfotomaculum_reducens_MI-1                              -----------------------MKQYIISYDI------------TNDKRRNKIRKLLCDY-AVPVQYSVFEA-RLSEEDYIM---LRHKL------SR--LL-R-K--DE----DSIIFYLQC------SRC--RD-GI--IRL----
116627765|Streptococcus_thermophilus_LMD-9                            -----MGLYFNLSEEEREFAKQKTMFCLIIYDI------------RSNKRRLKLSKLLEGY-GVRVQKSCFEV-NLSRNDYQS---LLKDI------EGFYKA-D----EE----DSIIVYVTT------KEE--VT-SF--SPY----
37677206|Vibrio_vulnificus_YJ016                                      ------------------------MLWLISFDI------------SDDKRRRRAVKLLLGY-GNRTQYSVFEC-LLKRSQLRQ---IQQKM------QA--VI-E----PS----DRVLYYPLC------GKD-----ALCRYAD----
37677205|Vibrio_vulnificus_YJ016                                      -----------MVSHSAEDPLMSREHYVICYDI------------SERQTRYRVEKALLGY-GERVQYSVFEC-RLTTGQFFA---LRHEL------RR--WL-E----DG----DRVNYFRLC------LHC--QR-RR--LTQ----
427708212|Nostoc_sp-_PCC_7107                                         ------------------------MFYLVCYDI------------VSDTRRNKVAKLLESY-GLRVQKSVFEC-VLDEKQYES---LSKYL------FR--LV-N-K--RE----DQVRFYPMS------AHN--RC-KV--AVL----
46255266|Thermus_thermophilus_HB27                                    ---------------------MRELYLVIAYDT------------PDDRRRARLAKLLKGF-GERRQYSVFEA-RLTREQWAH---LKGKL------EA--LV-N-K--EE----DVLAVYFLP------PEA--VG-RT--WRI----
159898908|Herpetosiphon_aurantiacus_DSM_785                           ------------------------MFILISYDI------------PHDKRRSKIAKTLENF-GKRVQYSVFEC-QLTDSQLAD---VRGRL------TA--LV-V-P--NE----DSIRFYSLP------KDA--VT-AM--LIL----
114567263|Syntrophomonas_wolfei_subsp-_wolfei_str-_Goettingen         ---MIHSTNRIFLFDEFIAEERRKYLVVVIYDI------------VDNRRRAAFAKYLKGF-GVRVQKSAFEC-ILPDAKYQK---LLKGI------PK--LI-D----KE----DQVRVYKLT------SNA-----DI--RAW----
296133513|Thermincola_potens_JR                                       -MVETRLLDDYFRFDDTEPEEMRRYLVVVIYDV------------IDNKRRNRLAKYLKRF-GFRVQKSAFEC-VLDSKNYKK---LTGGI------AK--YI-T----AD----DLLRVYKLA------GNA-----DV--QVW----
326793968|Marinomonas_mediterranea_MMB-1                              -----------------------MRIYLACFDI------------EDDKKRRKLSNLLLEY-GDRVQYSVFEISLKDENELHK---LRKKC------SK--YT-E----EA----DSLRFYWLN------KES--RK-HS-QDVW----
302391603|Acetohalobium_arabaticum_DSM_5501                           ------------------------MFVVISYDI------------SEDKRRNRIFKILKDF-GTWIQYSVFEC-ELEKKDYLK---LRDRL------QK--VL-D-E--DE----DNIRFYFLC------SKC--ED-EI--ERI----
75910375|Anabaena_variabilis_ATCC_29413                               ------------------------MYLVISYDV------------PEDKRRTKIHKVLKSY-GQWMQYSVFEC-DLTPTQYAK---LRSRL------AK--LI-K-P--DQ----DSIRFYHLC------ACC--QG-KV-ERIG----
119357847|Chlorobium_phaeobacteroides_DSM_266                         -----------------------MQFILVTYDI------------ENDRRRTKIHKVLEGY-GISVQYSVFEC-FLSDEDYAG---LRIRL------QK--LM-D-PKHPL----DSIRYYHLC------RGC--VE-KV--DVD----
38505685|Synechocystis_sp-_PCC_6803                                   ------------------------MLYLIIYDV-------P-ATKAGNKRRTRLFDLLSGY-GKWRQFSVFEC-FLSVKQFAK---LQTAM------EK--LI-K-L--DE----DAVCIYVLD------ENT--VQ-RT--ITY----
337286707|Thermodesulfatator_indicus_DSM_15286                        ---------------------MKATHYLVSYDI------------ADDKRLRKVARIMEDF-GERVLYSVFEC-HLTPVQLEA---LRRSV------EP--IL-D-P--LE----DSVRYYPIC------ERC--EK-HI--EHL----
119357852|Chlorobium_phaeobacteroides_DSM_266                         --------------------------MLVSYDI------------CDIKRLPRVAKLMEGY-GVRVQYSVFEC-SLTQRQLME---LQRRL------KR--LI-K-P--EV----DSVRFYPLC------ESC--RD-EI--VIL----
344996858|Caldicellulosiruptor_lactoaceticus_6A                       -------------------MKEKKLMYVVCYDI------------VDDRRRNKIANVLKSF-GNRVQYSVFEC-FLTPTQFQR---LKSRL------LP--LL-E----RN----DSIMCYKIT------PEL--QK-DI--ERI----
218442810|Cyanothece_sp-_PCC_7424                                     ------------------------MLVLVVYDI------------PDNKRRTKLSKFLEGY-GERVQWSVFEC-FLSLEEMRV---LYQKV------KK--RV-E-P--LE----DNVRFYWIS------NEA--VS-RV--LTI----
17227878|Nostoc_sp-_PCC_7120                                          ----------------------MTKLYLVCYDI------------PDDKRRTKLAKLIEQR-CQRVQYSVFEC-PLEESVLNYQ--LEKRW------LP--ILKL----SE----DSLRVYPLD------ATA-----KQQTRVY----
```

```
159898756|Herpetosiphon_aurantiacus_DSM_785                           -------------------MSDSTLYIIAYDI-------------PNDKRRTKVHKLLCGY-GYWTQYSLFEC-WLTKRHLVE---LRAKI------NR--LV-D-A--SL----DTIRLYRVC------GAC--VE-QA--ITI----
218961168|Candidatus_Cloacamonas_acidaminovorans_str-_Evry            ------------------MSLNKHLYVVAFDI-------------VSDKIRYRCVKILLSY-GQRVQKSVFEC-ILNDKQFLD---MKSKL------DN--TI-N-P--EI----DSIRYYLLC------ANC--EK-NI--SVS----
148657119|Roseiflexus_sp-_RS-1                                        -----------------------MFTIISYDI-------------VDDRRRTNVMKYLKGW-GTRVQYSVFEC-ILDARAFAT---VVREL------QR--MI-D-Q--NT----DSVRCYRLD------EAA--VT-RI--QII----
320161860|Anaerolinea_thermophila_UNI-1                               ------------------MNDGRSFYVLAYDI-------------SDDRRRLKLARLMESL-GVRVQGSVFEA-YLTATELER---LLRRC------SK--IL-K-K--EE----DSLRIYRLC------TAC--RE-NI--RTE----
297622674|Truepera_radiovictrix_DSM_17093                             ---------------------MAGLYTVSYDI-------------PNDNRRVKLANVLKSF-GERVQYSVFEC-WLEPRELRE---LRQKL------EA--RL-N-E--RE----DSVRIYPVG------SPV---------EVL----
383762166|Caldilinea_aerophila_DSM_14535_-_NBRC_104270                -----------------------MFVVISYDI-------------ADDRRRLKVAQTLLDFGGERVQRSVFEC-YITERNYER---LQRGL------RR--HL-D-I--EE----DSVRFYVLC------EHC--RA-QV--NLW----
156741962|Roseiflexus_castenholzii_DSM_13941                          -----------------------MLYVIAYDI-------------PDDARRLKLANVLEGF-GQRVQRSVFEC-DLTEREYRA---LIKKV------ERVVNL-N----EG----DSVRIYRLC------GAC--VA-NV--DVR----
38505761|Synechocystis_sp-_PCC_6803                                   ----------------------MFLYVIAYDI-------------PDDRRRKKMADLLEGY-GQRVQYSVFEC-TLSKSKFNE---LQKRL------RK--IYQS----EE----DSLRFYPLS------GHT--LT-QV--DIW----
116748795|Syntrophobacter_fumaroxidans_MPOB                           -----------------------MFYVVSYDI-------------SDDRVRYRVAKVLKAY-GVRVQKSVFECPDLTEKRFLK---MKDRL------EA--LI-D-F--TT----DSVRYYRQC------RGC--LA-QC--EVS----
332296541|Thermodesulfobium_narugense_DSM_14796                       -----------------MSIKAYSFYVISYDV-------------VEDRKRYKIAKLLEGK-GTRVQKSVFEC-YLTEKELIK---LNKEI------MK--KI-N-P--ET----DSVRIYLLC------KRC--KA-AI--SVL----
313673557|Calditerrivibrio_nitroreducens_DSM_19672                    ---------------------MEKFYVISYDI-------------ADDRKRYRVDKLLKNN-GMRVQKSVFEC-RLTDAQFLT---LKMEL------GK--IINY----NT----DSVRFYYLC------KHC--VD-MI--EYI----
328953001|Desulfobacca_acetoxidans_DSM_11109                          ----------------------MMFYAISYDI-------------RDNRRRLRVAKILKDY-GERVQLSVFEA-DLDEKSLAR---LKKRL------EK--CL-D-L--TA----DGLRLYPLC------GAC--RP-RI--EIM----
328953423|Desulfobacca_acetoxidans_DSM_11109                          -----------------------MFITISYDI-------------TDNRRRQRLAKMLSNY-GHRVQKSVFEC-RLDDRQYLK---LKKGI------EE--II-D---WDD----DSVRYYFLC------QHC--TD-NI--EIS----
118497354|Francisella_novicida_U112                                   -------------------MQLRKEYLIAYDI-------------EDNKTRTIIYKQLLAYGLKAVQKSVFWG-YVSIAELNA---IKRLF------DS--SL-T----IS----DKVFITRVN------MHEQKLDYSF---------
54296140|Legionella_pneumophila_str-_Paris                            -----------------MQIRLYKEYIVSYDI-------------RDNSIRRTIYERLKDIGLTPVQKSVFWG-YLNEPEKRY---LTNFL------KK--EL-A----KQ----DTAFITPCK-----LAQE--YEVNL--INL----
34557934|Wolinella_succinogenes_DSM_1740                              -------------------MKLFFEVIVSYDI-------------EDNRNRKKLFEELKDMGLSPIQKSVFWG-HLKLAEIKI---LPNLF------CK--YC-S----DG----DKAFFVKAR-----------LSSEI--VKN----
292492939|Nitrosococcus_halophilus_Nc_4                               -----------------------MYMLITYDV--------------DAKRTEKFKKLLRRY-LNHDQYSVFSG-DITEAEAIR---LHREL------SR--LM-V----PG----DRVTEVTAA------NRR--NI-NV--THL----
337288663|Thermodesulfobacterium_geofontis_OPF15                      ------------------------MRVIVYDY-------------ETEKIREKVRKNLKKL-GVHAQWSVFES-LESYENLIKV--LLEEE-------------G----EK----YRVAIFKIN------PKG-----EI--RKI----
52425691|Mannheimia_succiniciproducens_MBEL55E                        ---------------------MRDRYLIVYDI-------------SSSKRRYYTHKYLSAY-AVGGQKSFYEC-WLTNRELVE---FKQKL------IN--CI-D-K--QE----DKLFIFQLN------KDT-----QP--QLF----
154174506|Campylobacter_curvus_525-92                                 -----------------------MYAILFYDI--------AGAEQKEKNNANRIRKAVEKF-LPRVQFSVFEG-EIRASDFKK---LTNLL------QK--ECVQ----EL----DSIVIYTFN------SLK--YS-ER--IVI----
48477074|Picrophilus_torridus_DSM_9790                                ----------------------MVWVILIYDI--------------EKSRVNKIRKICEGY-LVRVQESAFEG-YISAALLNI---LIKRL------KK--EI-D----ESY---DKIKIYSFR------EDR--KG-KI--IEL----
328949007|Treponema_succinifaciens_DSM_2489                           ---------------------MEKHWLIVYDI-------------RDTKRLAKISKCVESY-GWRVQKSVFES-TASDETIGQ---LKSRL------EK--LM-D-I--EE----DFVLFFEVC------ERD--FQ-KK--QIF----
111224948|Frankia_alni_ACN14a                                         -----------------------MFVVLVYDT--------------AAERNPNALRTCRKY-LHWVQRSVFEG-ELSAAQYRA---LMTTL------RD--QL-D-L--TY----DSIRVYRTR------SPA--LV-ET--EWL----
319790502|Thermovibrio_ammonificans_HB-1                              ---------------MREKGRKNSTYVVIYDIFNFEEEYQGAWVSSDYRRRCRIAKELLKY-GVRIQRSVFEI-EISKTQIPK---LKATI------QK--YC-S----KR----DKIFIYLLE------DKV--KS-RV--KEK----
55376279|Haloarcula_marismortui_ATCC_43049                            ----------------------MVYAIIVYDV--------------QADRTPKFLKYLRRY-LTHVQNSVFEG-ELTEGTLVE---VKETL------QS--ML-E----EE----ESVMVYRMD------SGS--YV-ER--SVY----
268317184|Rhodothermus_marinus_DSM_4252                               ----------------------MPYYIAVYDV--------------NVRRVAKMLKLFRRY-LTWVQRSVFEG-ELTEAQFQR---LQHEA------GR--LM-N-P--DE----DVVIFYELR------DVR--YS-RR--LVL----
150021525|Thermosipho_melanesiensis_BI429                             -----------------------MYVIMVYDV--------------NKKRVQKVLKTARKY-LKWVQKSVFEG-KITEKNLKK---LRDQL------LE--LI-N-Q--EE----DSIYWYIME------PEF--KP-YR--KIL----
385809791|Ignavibacterium_album_JCM_16511                             --------------MIHTEPISQVFYIAVYDI-------------TNQKRLPKVLKTFRKY-MNWVQNSAFEG-ELTAKQFQS---LKDEL------LE--II-N-K--EK----DSIIFYYAE------EQK--HI-GK--KIL----
41614812|Nanoarchaeum_equitans_Kin4-M                                 -----------------MQYKINMYAIVVYDV--------------NVSRQNQIREFLRKY-LYHVQRSVFEG-EISPSSLYY---MKKIL------QS--YI-G----ET----DSLIIYVLR------DKSCLMD-KI---VL----
389844442|Mesotoga_prima_MesG1-Ag-4-2                                 -----------------------MWVIIVYDI--------------NEKRVAKVLKYLRRY-LNWIQNSVFEG-ELTYAELKK---VKSGL------KR--LM-N-T--QE----DSVIIFSSS------LSN--DY--K--KEI----
390962115|Thermococcus_sp-_CL1                                        ---------------------MARYYIVVYDV--------------NEKRVVRVHKILKAY-LQWRQRSVFEG-WLDGNEVAE---LKRKL------SR--AI-N-E--EE----DSILFYSLP------SDS--NL--V--TFH----
13540936|Thermoplasma_volcanium_GSS1                                  -----------------------MYLIVVYDV--------------DESRVNKVNKYLKRY-LLWRQNSVFEG-EISENLIDS---MIKGI------TK--II-K----DE----DSIIIYKFK------VKS--YI-DN--III----
289596693|Aciduliprofundum_boonei_T469                                -----------------------MHVIIVYDV--------------AEERVNKVNKFLRSY-LHWIQNSVFRG-DITQSQLQA---IKDGI------KG--II-E-E--DY----DSVTIYVLP------SEK--YL-KV--INI----
33359262|Pyrococcus_horikoshii_OT3                                    --------------------MVVMYVIVVYDV--------------NVERVNRVHKLLKTY-LFWRQNSVFEG-ELSKAQLYE---LEMRL------KR--IV-K----ED----DSVLIYIFP------GKN--FDLHV----V----
150021527|Thermosipho_melanesiensis_BI429                             -------------------MRSIKYILVVYDV--------------NEKRINKVHKVLKKY-IIWQQNSTFEG-NVTQANIKR---MIYEL------NK--KI-N-S--KE----DSILIYFFN------SKT--VI-RK--EVD----
84489237|Methanosphaera_stadtmanae_DSM_3091                           --------------------MVGIYILVIYDI--------------KIERINKVYKFLKTY-LIWIQNSVFEG-EITKAQYKT---MIIKL------KE--LT-D-E--DE----DSIIIYKI-------PKK--YL-NK--EVL----
33359542|Pyrococcus_furiosus_DSM_3638                                 --------------------MFHMYIVVVYDV--------------GVERVNKVKKFLRMH-LNWVQNSVFEG-EVTLAEFER---IKEGL------KK--IIDE----NS----DSVIIYKLR------SMP-----PR--ETL----
```

Continued

```
50085558|Acinetobacter_sp-_ADP1                                       ------GAK--------HK----FAHDGAYPVASIKKQQKLLMLDE---WVKASSL-----------------------------------------------------------------------------
77166204|Nitrosococcus_oceani_ATCC_19707                              ------GSR--------RH----FSRNGEFAFS--HSATPLPSVNEEAA-PAVRLLNH---------------------------------------------------------------------------
15602177|Pasteurella_multocida_subsp-_multocida_str-_Pm70             ------GDR--------RQ----FNDQGRVPTN--RTQRNILRAEEEMAWQDAYSLQI---------------------------------------------------------------------------
209966591|Rhodospirillum_centenum_SW                                  ------GAP--------KR---TIVDYDDHY----------------ALRRKATE------------------------------------------------------------------------------
21674789|Chlorobium_tepidum_TLS                                       ------GDT--------RK---EITEISGLPLVIEKSAASPPENQNRLTPEAPKVQ-----------------------------------------------------------------------------
86738747|Frankia_sp-_CcI3                                             ------GRN--------RT---QPVDYDGLI----------------LFARSRR-------------------------------------------------------------------------------
38234756|Corynebacterium_diphtheriae_NCTC_13129                       ------GDS--------TL---QVLDADGLS----------------VLASRPGRAAVKLHGLQ---------------------------------------------------------------------
90111482|Escherichia_coli_str-_K-12_substr-_MG1655                    ------GLN--------RR---TPVDLDGLR----------------LVSFLPV-------------------------------------------------------------------------------
77128445|Methylococcus_capsulatus_str-_Bath                           ------GLP--------PV---KFVELDGIV----------------LTRRP---------------------------------------------------------------------------------
57865885|Staphylococcus_epidermidis_RP62A                             ------GKE------VKLA----VNKEPKLI----------------FI------------------------------------------------------------------------------------
193216854|Mycoplasma_arthritidis_158L3-1                              ------GSK------SLNE---KYNDNERYI----------------KL------------------------------------------------------------------------------------
71894594|Mycoplasma_synoviae_53                                       ------GEK------SLNE---IYNERDKYI----------------EL------------------------------------------------------------------------------------
189440766|Bifidobacterium_longum_DJO10A                               ------SGK------REIE----EETPDFLT----------------LF------------------------------------------------------------------------------------
38232680|Corynebacterium_diphtheriae_NCTC_13129                       ------SSS-------AEVSGEEKPDQLAIF------------------------------------------------------------------------------------------------------
187250662|Elusimicrobium_minutum_Pei191                               ------GRL------------KVQAERPDFF--------------DTLL------------------------------------------------------------------------------------
117929156|Acidothermus_cellulolyticus_11B                             ------GER------LVPP----EDPPQQLA----------------IFP-----------------------------------------------------------------------------------
121608213|Verminephrobacter_eiseniae_EF01-2                           ------GRS------RQPT----QKAPDQFD----------------LF------------------------------------------------------------------------------------
325972001|Sphaerochaeta_globosa_str-_Buddy                            ------GQQ-----------VENECPEQGLL----------------FE------------------------------------------------------------------------------------
319957208|Nitratifractor_salsuginis_DSM_16511                         ------GEQ------SSQE----EITQPTLF----------------DF------------------------------------------------------------------------------------
294086113|Candidatus_Puniceispirillum_marinum_IMCC1322                ------GRH------SKRP---INADQLALF------------------------------------------------------------------------------------------------------
60683387|Bacteroides_fragilis_NCTC_9343                               ------GKK-------VQD---VNTPGQQLE----------------LF------------------------------------------------------------------------------------
154250553|Parvibaculum_lavamentivorans_DS-1                           ------GKK-------RQP---TTQSPSQLQ----------------LF------------------------------------------------------------------------------------
187736487|Akkermansia_muciniphila_ATCC_BAA-835                        ------GRS------CPRQ---EVSLPQQML----------------LW------------------------------------------------------------------------------------
116627544|Streptococcus_thermophilus_LMD-9                            ------GGK------TKQE---EIVSDNKLV----------------AI------------------------------------------------------------------------------------
57238544|Campylobacter_jejuni_RM1221                                  ------GGI------VFNE---KVNNETNLT----------------LFDIDSHGEFKYKNSNNEEIQLNKKQEKYHQQNLFEF-------------------------------------------------
291276263|Helicobacter_mustelae_12198                                 ------GEI------SQAQ---KVQKPTPLT----------------LF------------------------------------------------------------------------------------
220930480|Clostridium_cellulolyticum_H10                              ------GKK------TSNEWIDKKMIEGQGM----------------LI------------------------------------------------------------------------------------
209542527|Gluconacetobacter_diazotrophicus_PAl_5                      ------GLA----------PKTERIGPSQMV----------------LL------------------------------------------------------------------------------------
90425959|Rhodopseudomonas_palustris_BisB18                            ------GEA----------TFNEKKAPQQLV----------------LL------------------------------------------------------------------------------------
189485227|uncultured_Termite_group_1_bacterium_phylotype_Rs-D17       ------GKK------QLAE---Q-IDSRQLV----------------FF------------------------------------------------------------------------------------
310780382|Ilyobacter_polytropus_DSM_2926                              ------SKE------RSPE---EKIGTSKTI----------------II------------------------------------------------------------------------------------
15602990|Pasteurella_multocida_subsp-_multocida_str-_Pm70             ------GRL------RLQE---KKVNANQML----------------LF------------------------------------------------------------------------------------
222109283|Acidovorax_ebreus_TPSY                                      ------GLP------LFQE---KKVNAAQIA----------------LF------------------------------------------------------------------------------------
238924077|Eubacterium_rectale_ATCC_33656                              ------GSR------QTSE---TAIGTERLI----------------II------------------------------------------------------------------------------------
406600269|Leuconostoc_gelidum_JB7                                     ------GTP-------SDD---VRNSDEKMV----------------VI------------------------------------------------------------------------------------
90961087|Lactobacillus_salivarius_UCC118                              ------GKK-------IKE---VRNSAERTI----------------IL------------------------------------------------------------------------------------
269123824|Streptobacillus_moniliformis_DSM_12112                      ------VKE--------PKLSEIKLSKNPLV----------------IF------------------------------------------------------------------------------------
34557788|Wolinella_succinogenes_DSM_1740                              ------GGP-------KSQ---EKIGAKQLV----------------LF------------------------------------------------------------------------------------
331702226|Lactobacillus_buchneri_NRRL_B-30929                         ------GKP-------KED---VRNSSERTI----------------IL------------------------------------------------------------------------------------
15675043|Streptococcus_pyogenes_M1_GAS                                ------GER------NNCI----ANSDERLV----------------FLGEAFDES-----------------------------------------------------------------------------
374307736|Filifactor_alocis_ATCC_35896                                ------GEK------RSNV----VDSDDRLV----------------IL------------------------------------------------------------------------------------
302336018|Olsenella_uli_DSM_7084                                      ------GNR------SVYD---EVDTMEELL----------------II------------------------------------------------------------------------------------
169823757|Finegoldia_magna_ATCC_29328                                 ------GEN-------KSS---MLDTDEGLI----------------IL------------------------------------------------------------------------------------
42525845|Treponema_denticola_ATCC_35405                               ------GES-------KTE---YLNTDERLV----------------VL------------------------------------------------------------------------------------
148642084|Methanobrevibacter_smithii_ATCC_35061                       ------GKE-------------ITLEEEKYK----------------IL------------------------------------------------------------------------------------
88602028|Methanospirillum_hungatei_JF-1                               -------HE--------EL---PKVNRQFEV----------------I-------------------------------------------------------------------------------------
119720227|Thermofilum_pendens_Hrk_5                                   ------GRD-------YYI---VSERKYEVY------------------------------------------------------------------------------------------------------
70607607|Sulfolobus_acidocaldarius_DSM_639                            ------TQN------KISKNELQIYVAGEDI------------------------------------------------------------------------------------------------------
156937922|Ignicoccus_hospitalis_KIN4-I                                ------AVE------------------GKVG------------------------------------------------------------------------------------------------------
18311771|Pyrobaculum_aerophilum_str-_IM2                              ------GPP---------P---PARKNVDLY------------------------------------------------------------------------------------------------------
```

```
385805116|Fervidicoccus_fontis_Kam940                                 ------GTS--------KW---SAGSPEGSV----------------VLR------------------------------------------------------------------------------------
15922976|Sulfolobus_tokodaii_str-_7                                   ------GDS-------------VRREAVSLL-------------------------------------------------------------------------------------------------------
347523052|Pyrolobus_fumarii_1A                                        ------GEG------ANIP---GAKPKPGVT----------------LA-------------------------------------------------------------------------------------
118431379|Aeropyrum_pernix_K1                                         ------GLG------------GSLEDGKRVV----------------VV-------------------------------------------------------------------------------------
305663329|Ignisphaera_aggregans_DSM_17230                             ------GVN-------RVE---INEQSYGVI-------------------------------------------------------------------------------------------------------
307595806|Vulcanisaeta_distributa_DSM_14429                           ------GTA--------NL---AIPNEPTVI-------------------------------------------------------------------------------------------------------
302347807|Acidilobus_saccharovorans_345-15                            ------GRP------LAAP---RVAEVTLLG-------------------------------------------------------------------------------------------------------
124027532|Hyperthermus_butylicus_DSM_5456                             ------GGI------MSDP---KQKQAVRVV-------------------------------------------------------------------------------------------------------
18311698|Pyrobaculum_aerophilum_str-_IM2                              ------GRP--------PYAPLKPPKYAQIY----------------VV-------------------------------------------------------------------------------------
150398873|Methanococcus_vannielii_SB                                  ------GQD--------FD---RELINGEIT----------------TKFL-----------------------------------------------------------------------------------
257125861|Leptotrichia_buccalis_C-1013-b                              ------GLS------FEQD---IVKDEKPVL----------------FF-------------------------------------------------------------------------------------
150401498|Methanococcus_aeolicus_Nankai-3                             ------GEE------FDKD---LITDNKNTL----------------VL-------------------------------------------------------------------------------------
21226660|Methanosarcina_mazei_Go1                                     ------GQA--------FD---KELVSDEII--------------TKFF-------------------------------------------------------------------------------------
218961576|Candidatus_Cloacamonas_acidaminovorans_str-_Evry            ------GQA------FDPA---LITDEIKAF----------------FL-------------------------------------------------------------------------------------
284040852|Spirosoma_linguale_DSM_74                                   ------GQA------FDKK---LVTDEIRAL----------------FF-------------------------------------------------------------------------------------
15898244|Sulfolobus_solfataricus_P2                                   -----GYSG--------SE---REEKSNVVW-------------------------------------------------------------------------------------------------------
332668018|Haliscomenobacter_hydrossis_DSM_1100                        ------GSI--------EPDRDELSGSKSTL----------------IV-------------------------------------------------------------------------------------
386854796|Deinococcus_gobiensis_I-0                                   ------GEA-------------KNG-----------------------------------------------------------------------------------------------------------------
288560432|Methanobrevibacter_ruminantium_M1                           ------GNL-------------DLPKKSLYE----------------FL-------------------------------------------------------------------------------------
330509016|Methanosaeta_concilii_GP6                                   ------GEI-------------YEVEEEDVS----------------VF-------------------------------------------------------------------------------------
20094746|Methanopyrus_kandleri_AV19                                   ------DAE--------GV---RGLRF-----------------------------------------------------------------------------------------------------------
11499460|Archaeoglobus_fulgidus_DSM_4304                              ------SNT--------GV---ELVKEKGVE----------------FV-------------------------------------------------------------------------------------
383319614|Methanocella_conradii_HZ254                                 ------GGA-------------CLVKESKVQ----------------IV-------------------------------------------------------------------------------------
148656005|Roseiflexus_sp-_RS-1                                        ------GEE-------------S---------------------------------------------------------------------------------------------------------------
408403001|Candidatus_Nitrososphaera_gargensis_Ga9-2                   ------GKE-------FVE---EEKQDVQFF-------------------------------------------------------------------------------------------------------
167628643|Heliobacterium_modesticaldum_Ice1                           ELLVDGDDG-------------EASC----------------------------------------------------------------------------------------------------------------
170290048|Candidatus_Korarchaeum_cryptofilum_OPF8                     ------GRG-------------KFGRGPP---------------------------------------------------------------------------------------------------------
126465228|Staphylothermus_marinus_F1                                  --K---GSI-----------------KHGEP-------------------------------------------------------------------------------------------------------
327314605|Prevotella_denticola_F0289                                  ------GED------ISID---VIIRSRNTL----------------FF-------------------------------------------------------------------------------------
392395715|Flexibacter_litoralis_DSM_6794                              ------GNN------IPFL---DAVDDKKVL----------------IF-------------------------------------------------------------------------------------
34541576|Porphyromonas_gingivalis_W83                                 ------GHP--------TLPKSRRKRKSIVV-------------------------------------------------------------------------------------------------------
288947692|Allochromatium_vinosum_DSM_180                              --L---GDG------EAAT---ESRPSAPGW-------RALTTPTPPVLPFTRSQS------------------------------------------------------------------------------
379728402|Saprospira_grandis_str-_Lewin                               ------GES---------EDLRRLLDELLFI----------------LL-------------------------------------------------------------------------------------
431929778|Thioflavicoccus_mobilis_8321                                --V---TRC--------AR---IGFRKPALY----------------LD-------------------------------------------------------------------------------------
288947691|Allochromatium_vinosum_DSM_180                              ------GKG--------RT---TALQDGLLL----------------AA-------------------------------------------------------------------------------------
269798875|Veillonella_parvula_DSM_2008                                ------NKI-------------IPLTEDKML----------------II-------------------------------------------------------------------------------------
30248143|Nitrosomonas_europaea_ATCC_19718                             ------GNA-------------STFTVQHFC----------------IV-------------------------------------------------------------------------------------
206889468|Thermodesulfovibrio_yellowstonii_DSM_11347                  ------GIA------------VPPIDPEYFY----------------IG-------------------------------------------------------------------------------------
121997422|Halorhodospira_halophila_SL1                                ------GAA-------------EVPADPDFF----------------YFG------------------------------------------------------------------------------------
292492855|Nitrosococcus_halophilus_Nc_4                               ------GIA------VKPV------DPAFFY----------------VG-------------------------------------------------------------------------------------
430760779|Thioalkalivibrio_nitratireducens_DSM_14787                  ------GKA------QAPS------DPDLFY----------------LG-------------------------------------------------------------------------------------
317051220|Desulfurispirillum_indicum_S5                               ------GTG------------LSRMAKTQAP-------------------------------------------------------------------------------------------------------
118497968|Francisella_novicida_U112                                   ------GYA-------------IHREQEVVF----------------ID-------------------------------------------------------------------------------------
76802277|Natronomonas_pharaonis_DSM_2160                              --I---GND--------PP---D--GQSNVI-------------------------------------------------------------------------------------------------------
162457463|Sorangium_cellulosum_So_ce56                                ------GND--------LTPLDLEPPPFRIV-------------------------------------------------------------------------------------------------------
347756304|Candidatus_Chloracidobacterium_thermophilum_B               ------STK------TSWD---EERKRIEIL-------------------------------------------------------------------------------------------------------
24215881|Leptospira_interrogans_serovar_Lai_str-_56601                ------EKK------SAWP---EAPKTLKIL-------------------------------------------------------------------------------------------------------
312114616|Rhodomicrobium_vannielii_ATCC_17100                         ------NPI-------------NDIADLKAA----------------IL-------------------------------------------------------------------------------------
218884092|Desulfurococcus_kamchatkensis_1221n                         --F---GKP--------LA---EITVEHGVM----------------VIT------------------------------------------------------------------------------------
431929779|Thioflavicoccus_mobilis_8321                                --Y---GGA------PAEA---ERPRRGRWL--------------KGLFGREAA--------------------------------------------------------------------------------
39995169|Geobacter_sulfurreducens_PCA                                 ------GKT--------FD---PILRQAVIV-------------------------------------------------------------------------------------------------------
83591518|Rhodospirillum_rubrum_ATCC_11170                             ------GKA--------FE---TLERQAMII-------------------------------------------------------------------------------------------------------
332661940|Haliscomenobacter_hydrossis_DSM_1100                        --L---GQE------MDTR---LLLGKDRYL----------------IL-------------------------------------------------------------------------------------
261415949|Fibrobacter_succinogenes_subsp-_succinogenes_S85            ------GCG-------------KRIENPAFK----------------VL-------------------------------------------------------------------------------------
121997420|Halorhodospira_halophila_SL1                                ------GLV--------LPGLSLRPQFPEAD-------------------------------------------------------------------------------------------------------
302343125|Desulfarculus_baarsii_DSM_2075                              ------GVT-------------APPQSHDYY----------------FV-------------------------------------------------------------------------------------
389875623|Tistrella_mobilis_KA081020-065                              ------GGP-------------PLPEPADYW----------------LL-------------------------------------------------------------------------------------
374998938|Azospirillum_lipoferum_4B                                   ------GGA-------------PVPEQQDFW----------------LL-------------------------------------------------------------------------------------
294676822|Rhodobacter_capsulatus_SB_1003                              --D---DAA--------PI---E--PPGNYW----------------IV-------------------------------------------------------------------------------------
288957882|Azospirillum_sp-_B510                                       ------GGA-------------PLAEPQDFY----------------LL-------------------------------------------------------------------------------------
319790503|Thermovibrio_ammonificans_HB-1                              ------GEIYEDLMDWKEELGRLTRRKFFYFLRDTLEMEIKSGKFSKYLLVYDISDNLIRIKFSDILEEFSERVQLSAFEVPASPNVIARIAVLAMQFAEIDDNLKVYLYPIDKGSEKKIIRIGSSYLSFRDIL
108759816|Myxococcus_xanthus_DK_1622                                  ------GAR--------AP---LDVEGPLVL-------------------------------------------------------------------------------------------------------
15612905|Bacillus_halodurans_C-125                                    ------GAK--------PS---IDLEDPLIF-------------------------------------------------------------------------------------------------------
189485224|uncultured_Termite_group_1_bacterium_phylotype_Rs-D17       ------GCK--------KP---VDLHETLII-------------------------------------------------------------------------------------------------------
86742033|Frankia_sp-_CcI3                                             ------GAG--------VQ---PVHRADHII-------------------------------------------------------------------------------------------------------
433631662|Mycobacterium_canettii_CIPT_140070010                       ------GRG--------PE---VDWRSPIVL-------------------------------------------------------------------------------------------------------
333996013|Treponema_azotonutricium_ZAS-9                              ------GNR--------TL---PPVSSYEIL-------------------------------------------------------------------------------------------------------
328949008|Treponema_succinifaciens_DSM_2489                           ------GNG-------------NIFIPKTFE----------------IL-------------------------------------------------------------------------------------
337288661|Thermodesulfobacterium_geofontis_OPF15                      ------GYP-------------LEPWKIFII-------------------------------------------------------------------------------------------------------
319790501|Thermovibrio_ammonificans_HB-1                              ------GKR-------------EEKTPPM---------------------------------------------------------------------------------------------------------
283778925|Pirellula_staleyi_DSM_6068                                  ------QRE--------------LRRMDAVI----------------VL-------------------------------------------------------------------------------------
386750120|Helicobacter_cetorum_MIT_00-7128                            --I---GEK------RGIF---K--PYDDFI-------------------------------------------------------------------------------------------------------
220934971|Thioalkalivibrio_sulfidophilus_HL-EbGr7                     --S---GEP------LSPL---P-----AVI----------------IL-------------------------------------------------------------------------------------
15609953|Mycobacterium_tuberculosis_H37Rv                             ------GRG-------------RLVSAEEFV----------------FF-------------------------------------------------------------------------------------
85858453|Syntrophus_aciditrophicus_SB                                 --S---GLG--------LK---V--TQLEYT----------------VI-------------------------------------------------------------------------------------
268318440|Rhodothermus_marinus_DSM_4252                               ------GCQ-------------KLPEMKNVF----------------LT-------------------------------------------------------------------------------------
320102062|Isosphaera_pallida_ATCC_43644                               ------GLP--------YA---PLDAPCVVV-------------------------------------------------------------------------------------------------------
378719040|Gordonia_polyisoprenivorans_VH2                             ------GRP-------------RTLPASGPM----------------VV-------------------------------------------------------------------------------------
134101618|Saccharopolyspora_erythraea_NRRL_2338                       ------GRR-------------HGLPTTGHR----------------VL-------------------------------------------------------------------------------------
336116790|Microlunatus_phosphovorus_NM-1                              ------GQA-------------YPPTQEYFW---------------AVL-------------------------------------------------------------------------------------
268318377|Rhodothermus_marinus_DSM_4252                               ------GQG----------RGIDLRQDHWIV-------------------------------------------------------------------------------------------------------
83592489|Rhodospirillum_rubrum_ATCC_11170                             --L---GAT------VEKP---V--HE-PWI----------------IV-------------------------------------------------------------------------------------
206890293|Thermodesulfovibrio_yellowstonii_DSM_11347                  ------GVG-------------DRIPDGVEV----------------FY-------------------------------------------------------------------------------------
52425689|Mannheimia_succiniciproducens_MBEL55E                        ------GKN---------PMPEGIVLSTFVF-------------------------------------------------------------------------------------------------------
30248141|Nitrosomonas_europaea_ATCC_19718                             --I---GRP------T-LP---E-----GIQ----------------WS-----GLPAKW--------------------------------------------------------------------------
292492858|Nitrosococcus_halophilus_Nc_4                               ------GRG-------------AFPEGIYLFAGGVEDFLEALEQNEQLTR------------------------------------------------------------------------------------
430760782|Thioalkalivibrio_nitratireducens_DSM_14787                  ------GRG------ALPDGALLQSHENPGL----------------VTLLQALAE------------------------------------------------------------------------------
348026606|Megasphaera_elsdenii_DSM_20460                              ------GRG--------------DTHTEDVI----------------IL-------------------------------------------------------------------------------------
317051218|Desulfurispirillum_indicum_S5                               ------GIG-------------DFNRSEEII----------------VI-------------------------------------------------------------------------------------
337286710|Thermodesulfatator_indicus_DSM_15286                        ------GWG-------ELP---EEFEEDVIL-------------------------------------------------------------------------------------------------------
85858454|Syntrophus_aciditrophicus_SB                                 ------GQG------------FYTNPDEEYY----------------IY-------------------------------------------------------------------------------------
436839746|Desulfovibrio_hydrothermalis_AM13_-_DSM_14728               ------GNA--------PP---RINENYSFV----------------I--------------------------------------------------------------------------------------
134299483|Desulfotomaculum_reducens_MI-1                              ------GNC-------------SDPFGDGIY----------------IV-------------------------------------------------------------------------------------
116627765|Streptococcus_thermophilus_LMD-9                            ------HSA-------------EKLDDILFF-------------------------------------------------------------------------------------------------------
37677206|Vibrio_vulnificus_YJ016                                      ------GAG-------------QVYWPDALY----------------LIDE-----------------------------------------------------------------------------------
```

```
37677205|Vibrio_vulnificus_YJ016                                      ------GAM-------------VAWQSQPFD----------------YID------------------------------------------------------------------------------------
427708212|Nostoc_sp-_PCC_7107                                         ------GTQ--------PE---FVVDDAAFI----------------V--------------------------------------------------------------------------------------
46255266|Thermus_thermophilus_HB27                                    ------GHE--------GL---KRLEDPDFV-------------------------------------------------------------------------------------------------------
159898908|Herpetosiphon_aurantiacus_DSM_785                           ------GHG-------------VVTHDPSFY----------------WTD------------------------------------------------------------------------------------
114567263|Syntrophomonas_wolfei_subsp-_wolfei_str-_Goettingen         ------GTV-------------ELLEEEEVV----------------IL-------------------------------------------------------------------------------------
296133513|Thermincola_potens_JR                                       ------GSV-------------EKTEVDEVI----------------IV-------------------------------------------------------------------------------------
326793968|Marinomonas_mediterranea_MMB-1                              ------GNP------------IAVFPAAVLL-------------------------------------------------------------------------------------------------------
302391603|Acetohalobium_arabaticum_DSM_5501                           ------GSE-------------SSPAQQSII----------------I--------------------------------------------------------------------------------------
75910375|Anabaena_variabilis_ATCC_29413                               ------GEM-------------PIDTTVFFA-------------------------------------------------------------------------------------------------------
119357847|Chlorobium_phaeobacteroides_DSM_266                         ------GNR-------------SFEIASPFI----------------IT-------------------------------------------------------------------------------------
38505685|Synechocystis_sp-_PCC_6803                                   ------GTP-------------QPEKPGSII----------------I--------------------------------------------------------------------------------------
337286707|Thermodesulfatator_indicus_DSM_15286                        ------GRD------------KKFLRRKEFE----------------IL-------------------------------------------------------------------------------------
119357852|Chlorobium_phaeobacteroides_DSM_266                         ------GQG--------TV---S--RDEPYY----------------IA-------------------------------------------------------------------------------------
344996858|Caldicellulosiruptor_lactoaceticus_6A                       ------GQM------------RYYYEEDDII-------------------------------------------------------------------------------------------------------
218442810|Cyanothece_sp-_PCC_7424                                     ------GGE-------------SPSPPPNYY----------------LI-------------------------------------------------------------------------------------
17227878|Nostoc_sp-_PCC_7120                                          ------GGE-------------PPYEPPDYL----------------IL-------------------------------------------------------------------------------------
159898756|Herpetosiphon_aurantiacus_DSM_785                           ------GSP-------------EPREAVTII----------------I--------------------------------------------------------------------------------------
218961168|Candidatus_Cloacamonas_acidaminovorans_str-_Evry            ------GIG-------------LYSYFEDLI----------------IV-------------------------------------------------------------------------------------
148657119|Roseiflexus_sp-_RS-1                                        ------GIG-------------AVSKDPTHY----------------LV-------------------------------------------------------------------------------------
320161860|Anaerolinea_thermophila_UNI-1                               ------GKG-------------KITPPPQVT----------------VI-------------------------------------------------------------------------------------
297622674|Truepera_radiovictrix_DSM_17093                             ------GSG-------------SPTEDARLF----------------MF-------------------------------------------------------------------------------------
383762166|Caldilinea_aerophila_DSM_14535_-_NBRC_104270                ------GVA--------RP---IERPGLRII-------------------------------------------------------------------------------------------------------
156741962|Roseiflexus_castenholzii_DSM_13941                          ------GEG--------PP----VEKSVDIY----------------IV-------------------------------------------------------------------------------------
38505761|Synechocystis_sp-_PCC_6803                                   ------GEP-------------PLTKPPGSV----------------IV-------------------------------------------------------------------------------------
116748795|Syntrophobacter_fumaroxidans_MPOB                           ------GAG-------------ELPATTKFS----------------VV-------------------------------------------------------------------------------------
332296541|Thermodesulfobium_narugense_DSM_14796                       ------GNG--------VP---QELSEYDIL-------------------------------------------------------------------------------------------------------
313673557|Calditerrivibrio_nitroreducens_DSM_19672                    ------GNS-------------VFLEKNDFL----------------II-------------------------------------------------------------------------------------
328953001|Desulfobacca_acetoxidans_DSM_11109                          ------GQG-------------VVSQDPDYI----------------IL-------------------------------------------------------------------------------------
328953423|Desulfobacca_acetoxidans_DSM_11109                          ------GWG-------------VVRDDEDVI----------------VV-------------------------------------------------------------------------------------
118497354|Francisella_novicida_U112                                   ------GYD------DKTF---KDWDEYGHI-------------------------------------------------------------------------------------------------------
54296140|Legionella_pneumophila_str-_Paris                            ------ELD-------------ELRKPDGHI----------------IV-------------------------------------------------------------------------------------
34557934|Wolinella_succinogenes_DSM_1740                              ------GFG--------YTSKIFEIKEYDII-------------------------------------------------------------------------------------------------------
292492939|Nitrosococcus_halophilus_Nc_4                               ------SKSDSGKGEMKRVPVEDHRRDFSVL-------------------------------------------------------------------------------------------------------
337288663|Thermodesulfobacterium_geofontis_OPF15                      ------GQN--------------WEKIKFIF-------------------------------------------------------------------------------------------------------
52425691|Mannheimia_succiniciproducens_MBEL55E                        ------GCA-------SLP---KFNQPYLII-------------------------------------------------------------------------------------------------------
154174506|Campylobacter_curvus_525-92                                 ------GQD---------------KNGALFS-------------------------------------------------------------------------------------------------------
48477074|Picrophilus_torridus_DSM_9790                                ------GDV---------------SHKDNII-------------------------------------------------------------------------------------------------------
328949007|Treponema_succinifaciens_DSM_2489                           ------GKN------IEKD---DKITDEDFV----------------IL-------------------------------------------------------------------------------------
111224948|Frankia_alni_ACN14a                                         ------GVP--------------LGNQDSVL-------------------------------------------------------------------------------------------------------
319790502|Thermovibrio_ammonificans_HB-1                              ------GTP-------------FTPRDIQIL-------------------------------------------------------------------------------------------------------
55376279|Haloarcula_marismortui_ATCC_43049                            ------GDD--------------PMDDQQFL-------------------------------------------------------------------------------------------------------
268317184|Rhodothermus_marinus_DSM_4252                               ------GTE--------------QGHRDRFL-------------------------------------------------------------------------------------------------------
150021525|Thermosipho_melanesiensis_BI429                             ------GKS---------------ENISNLI-------------------------------------------------------------------------------------------------------
385809791|Ignavibacterium_album_JCM_16511                             ------GVE--------KN------EITQFF-------------------------------------------------------------------------------------------------------
41614812|Nanoarchaeum_equitans_Kin4-M                                 ------GED----------------KDLQIY-------------------------------------------------------------------------------------------------------
389844442|Mesotoga_prima_MesG1-Ag-4-2                                 --I---GLD--------K----A--KETQIL-------------------------------------------------------------------------------------------------------
390962115|Thermococcus_sp-_CL1                                        --L---GRP--------P----D--EFDNVI-------------------------------------------------------------------------------------------------------
13540936|Thermoplasma_volcanium_GSS1                                  ------GND--------------KGNTSNMI-------------------------------------------------------------------------------------------------------
289596693|Aciduliprofundum_boonei_T469                                ------GIE--------------KGTTDMIF-------------------------------------------------------------------------------------------------------
33359262|Pyrococcus_horikoshii_OT3                                    ------GRD--------------KSPVEMII-------------------------------------------------------------------------------------------------------
150021527|Thermosipho_melanesiensis_BI429                             ------GIE--------KW---N--VSSNFI-------------------------------------------------------------------------------------------------------
84489237|Methanosphaera_stadtmanae_DSM_3091                           ------GIE--------K----N--PIEFIL-------------------------------------------------------------------------------------------------------
33359542|Pyrococcus_furiosus_DSM_3638                                 ------GIE--------------KNPIEEII-------------------------------------------------------------------------------------------------------
```

C

```
11499463|Archaeoglobus_fulgidus_DSM_4304                    LPMYLI-CSRYCPTS--RDVYL-TMVEKVRG----EESFPVSLGKAIHEAVAKAIREAKSKNFSAE--------------------------------PPNYDN--EMINVAVKAVWEYTMSLCQSSYLKVRSEQPYATEEGVLLTSMPF
383319613|Methanocella_conradii_HZ254                       LPMYSI-CGKYCPNS--RDVYL-RHVLKKPG----EVTYPVSLGSIVHSTVNGAYKHVRLRNFSAD----------------------FEKWYESQGFKPPTNP--DLVKKYSRIVWEHVLANAESRFREAASSQPYSEEEDMIMTAAPF
330509014|Methanosaeta_concilii_GP6                         MPVWEV-SDRYCETG--RDTYL-RHVQKVSV----PASEEMVAGRIYHQAMAELYPTAKRLIYKEGMSICAELPGILMERLDDAIVEANEQLSQTKFSPSAEAG--QSVSDNLRKLWSFEVSRIVSSIFAYLAKYKYLSDDSLINHALPF
408402999|Candidatus_Nitrososphaera_gargensis_Ga9-2         LGVYEV-ASQYCQSG--RDVFV-RRIDGIKG----VPSKDMVRGLVYHYAVADEISVAKAYMYKQGVLGAAGLYEHLKAKSEERFASYMLTHGEYFREAAFTEADIEEVKMNLDKLWGYEASQMAASIKAILSKQPRIGIDALVHTAIPV
167628645|Heliobacterium_modesticaldum_Ice1                 LGVFEV-ASGYCSSS--RDVYL-RRVEGARS----KPNGAMFQGSLLHGLAARFIVATKGLIYQEG----VDGYREILRKLQEMTLSELLDEGLVFELASLAEADREELERKSRLLLEFERGRLIARLQEALVKQPYIGEDALTALAVPV
170290047|Candidatus_Korarchaeum_cryptofilum_OPF8           LGVWEI-ASSICPTG--RDVWIRHKVLRRSV----PTTPQIAKGIVVHRLVSSMFLKAKKMVYMGK-------YELRDDLITESEGIVERELENMRKYVKLPDE--EGLRGFCRRIARWEATRIEGKVMEVRAKYPFLNEESLVQIAFPV
126465227|Staphylothermus_marinus_F1                        LPIWEI-AGSICPTY--RDVWIRRKVLRKKL----FTTSKQAEGVIIHKIVSHVFKEAKKLIYSGD------YSDLKNALLPGVKELVDNEVKYISELVENIDA--NKIKEFALQVADWEMLRIEGRINTVKAKYPYIDEEGLVSLAVPV
156937920|Ignicoccus_hospitalis_KIN4-I                      ARPSVSDLTSPCPTK--RDVYL-RKVLGVRL-----ESEAMSKGRAVHEAFMAPFKGVELETEDPL--------------------------------------------------------------LKRVYEEAKVRKAAAEEEGIPV
124027533|Hyperthermus_butylicus_DSM_5456                   YRPSVSEVASYCPSH--RDAYL-QHVRGLQP----PRSLRAQYGVLVHEFFLEPFRLVDRGADSLE------------------------------------------ALTAGKYRLARRLGVRADRFLERVYELGAILALQSRVDGDIP
118431381|Aeropyrum_pernix_K1                               ARPTLSDVTGVCPVG--RDTYL-RRVLNTRV-----NGGLLRLGALVHEAFLLPFKAHRVGDAYRM---------------------------------------------FNYLLRRVRARGNEIRVMNAVFEKAVEFSAAARVDGIPV
307595808|Vulcanisaeta_distributa_DSM_14429                 VRPTVSEVSSPCPTK--RDAYL-RRVAGVTM----EDSESLRLGRMVHEVFLTPFKESIPFSQLDN-------------------------------------------RFESIINEYGDLGNKYRAQLYEVYTKAVATSLNAREEGIPI
305663331|Ignisphaera_aggregans_DSM_17230                   ARPSISEVASPCPTK--RDVYL-RRVLHTYV-----DSQVLRLGRAVHEVFLLPFRYRGRDIEWLY-------------------------------------------GMFRRSLKDYADLRGYWKAFESVFRKSLALSMIAEEEQMPI
18311773|Pyrobaculum_aerophilum_str-_IM2                    LALSDF-AYGYCPTG--RNIYL-KYVLGERG----AAVKALVEGQALHQALFKALEDYKKYAYSGH----------------------------PMSPSFEGVP--EDVRPKAEALYKYVATRLLGEHSYVSAARLARSRDSAVFYTAPI
18311700|Pyrobaculum_aerophilum_str-_IM2                    LSVSDI-VGGGCESR--RDLYL-KRVLGVRA----EPNGGMKFGAYIHEVFKRSLAELRSLIEGGA---------------VKGWELVQSFNAEAIAREAAIAA--EAEADPRGVELARYLAIQVAARVDEIASRSSADSLSIAARAVPV
385805114|Fervidicoccus_fontis_Kam940                       LGVSEV-AYKYCETF--RDLWM-RRKEYVRGEV--LESSPLYMGKLIHSLMNYAFRDASEIVSRNL--------------------------GPLESIRRAQELSRKRASELTKSYYRKAMEFYMTMVLSAISTLWIQQSLTNDPLSYIS
15898287|Sulfolobus_solfataricus_P2                         LSISDV-AYKYCETK--RDVYL-RKVLKIQG----EQRTPLLLGQAIHEIVRNVSNEVVKLLSSGY--------------------------------KPWSVLETLFRNNKIVRSLCPPQFIKYCEEVYKAYALDLLSDFSDSTSFIPV
15922974|Sulfolobus_tokodaii_str-_7                         LTMGEI-AYDYCPTK--RDVYV-RRVLNQKG----SERAQLVFGQAIHKVFSKALNDVRISYVLGK----------------------------DPYQIVKESYLESEFCPQEISEYCRKVYGNLILFWSSWLAEAKAFYGGDSVGFLPW
347523054|Pyrolobus_fumarii_1A                              LGVSEV-AYRYCPTR--RDLWL-RRVRGLRG----QPTEAMLKGRIVHEAFRLAARDLSRLVLGEG------------------------KSTEDAVAELVGGARERASVILESAGIEGDGDGIAVFAARVYKKMVFAWAQWMEETGAPP
218884094|Desulfurococcus_kamchatkensis_1221n               LGVWEV-SSKYCGTR--RDLWL-RRKMGVVV----EPSESMAKGKMVHEAITTALREAGRYLSNGL---------------------------EVWSAYEAAKEKWRRIDTGGSRENARLVEKIYKSTLLSILGEAAYEEALYGSRRPPL
389861364|Thermogladius_cellulolyticus_1633                 LGVSEV-AGKYCPTR--RDVWL-RRKMGLAP----ETTEPLAKGRLVHEAVTAAVRCASRGLSRGW-------------------------------ETWDILTYCLGRFRPQGNGDTATALKAYKSTLVTLLGEVEYERLVSGGSGLPV
332296539|Thermodesulfobium_narugense_DSM_14796             LSPSSLNLLLTCP-----FAYKLKYIDRVQP----KFNINTIAGSAFHFALSKYFKTNRYDFESSW-----------------------------------------KEYRNYNPEVLSLSFNEEDKIKNNLFAMLKIYEPFARDLKVIS
167627879|Francisella_philomiragiaATCC_25017                ILITWLNDFIFCPYSIYLHNIY-SNASDATY-----YSSSQTKGRDAHKSIDKGIYSTKKDD----------------------------------------------------------------------------------------
56965355|Bacillus_clausii_KSM-K16                           MMLSGLQHFQFCK----RQWAL-IHIEQQWE-----ENVRTIEGQHLHKKADQPFIKEKRGSKLT-------------------------------------------------------------------------------------
76802275|Natronomonas_pharaonis_DSM_2160                    VHVSALNEYLYCP----RRFYY-QRYHDEMG-----TPYELVDGRSKHENAAQRGDW---------------------------------------------------------------------------------------------
189485222|uncultured_Termite_group_1_bacterium_Rs-D17       LHIRGVQQLSYCK----RRCAL-LFIEQQWS-----DNFFTAKGIVMHDKAHAEKIEHKKGVVI--------------------------------------------------------------------------------------
134101617|Saccharopolyspora_erythraea_NRRL_2338             IGARALSDFRFCP----RLFHL-EHAEGYRF-----DSEEMRLGRHVHASVDVTSGRERPAQRI----------------------------------------------------------------------------DPSN---WR-
134299482|Desulfotomaculum_reducens_MI-1                    FPISSVAEILYCP----RNFYY-RVVEGAED-----SNHHLLEGKLQEERRDERQRLVREGYRQ--------------------------------------------------------------------------------------
433631664|Mycobacterium_canettii_CIPT_140070010             VLLSEIEHWIYCP----RQWAI-IHHEQHFT-----DNDDTTRGHLEHQRVDAVGHESRHGVQV--------------------------------------------------------------------------------------
86742031|Frankia_sp-_CcI3                                   LPISALEHHAYCP----RQAAL-IHLDNVFA-----DNIETMRGNVAHAHVHEPSPAVPTPQGNRIRQ----------------------------------------------------------------------------------
344996860|Caldicellulosiruptor_lactoaceticus_6A             IPISSINTYVYCP----RRFYI-EYFLGLRE-----DNQYTLEGKNIQSELQFYGINGFY------------------------------------------------------------------------------------------
219683006|Bifidobacterium_animalis_subsp-_lactis_AD011      LPISLICDFIFCP----RRAWL--EIQGEKI-----ESLQMERGFHDHRAVDDANGGRGDTD----------------------------------------------------------------------------------------
347756303|Candidatus_Chloracidobacterium_thermophilum_B     TRVMALHALAYCE----RLFYL-EEVEEIRL-----ANDRVYEGRRVHLQDISEGIE---------------------------------------------------------------------------------------------
294828260|Leptospira_interrogans_serovar_Lai_str-_56601     ---MGIHSLLYCE----RLFYL-EEVEGILV-----ADDRVYAGRTLHEELEPNEDSSGR------------------------------------------------------------------------------------------
268318441|Rhodothermus_marinus_DSM_4252                     IPARMLNEFVYCP----RLFYL-EWVQGEWE-----DNADTLEGSSIHKRTDSASGTLPEIDEAFG---------------------------------------------------------------------------------EHA
```

```
378719041|Gordonia_polyisoprenivorans_VH2                  LPARMLNEYVYCP----RLFYF-EWVDQRWV-----DSHDTEQGHFTHRAVDKR-TGSLPEPES-------------------------------------------------------------------------------GDPAK
320102061|Isosphaera_pallida_ATCC_43644                    MPARMLNEYIYCP----RLFYY-EWVEGVFA-----ENLDVVEGRHRHTRVDGKTDDLPSVQQVEK----------------------------------------------------------------------------GEAERL
39995168|Geobacter_sulfurreducens_PCA                      IPVRMLNEHVYCP----RLAYL-MWVQGEFS-----HNEFTVDGVIRHRRVDAGGGVLPSETQEDS--------------------------------------------------------------------------------RI
83591519|Rhodospirillum_rubrum_ATCC_11170                  VPASMVNAWIYCP----RLAVL-EWGRGEKA-----RSVDLIAGLRAHQATESGPTPALPDPMVLR-----------------------------------------------------------------------------EDQSL
108757432|Myxococcus_xanthus_DK_1622                       VALSALQHLLYCE----RQAAL-IHVERQWR-----EDVNTASGRLLHERVDLPGHDARLGRRV------------------------------------------------------------------------------------
46255268|Thermus_thermophilus_HB27                         LPLSKVNTVVYCP----RRFYL-EYVLGEAH-----ANHHLIEGHYLHERAYTEPGE-------------------------------------------------------------------------------------------
159898754|Herpetosiphon_aurantiacus_DSM_785                IPISMLNALAYCP----RRFGY-EYLQAEML-----INEFVADGQILHSSVDAGGRTWIAGEVQ------------------------------------------------------------------------------------
38505683|Synechocystis_sp-_PCC_6803                        LPLAYLNAWEYCH----RRFYL-EYVLGEME-----INEHIIRGNHLHRNINEAGTSVDGDTRI------------------------------------------------------------------------------------
17229059|Nostoc_sp-_PCC_7120                               IPIAALNQYAYCS----HRCWR-MFCAGEFI-----DNQYTIEGTTLHDRVHTTSENQREETWQ------------------------------------------------------------------------------------
302391601|Acetohalobium_arabaticum_DSM_5501                VPLSAINAYNYCP----YRVYL-EYVLGEWK-----DNTHTIEGILKHDRAHSGERRYDSNRIQ------------------------------------------------------------------------------------
48477076|Picrophilus_torridus_DSM_9790                     YRGIDVSYVNICM----RRAWLSLHEIFYVR-----DSYYVSDGAYLADIN-------------------------------------------------------------------------------------------------
154174706|Campylobacter_curvus_525-92                      ITGTLVNYFMTCK----REAWLYAHHIHADQ-----DDENVLMGKALADIKEKD----------------------------------------------------------------------------------------------
111224946|Frankia_alni_ACN14a                              VGGVHIKYLLHCP----RQLWLYMRGYRPEG-----SSDLVSFGEIVDETTFTR----------------------------------------------------------------------------------------------
389844446|Mesotoga_prima_MesG1-Ag-4-2                      IGGNAINYLHYCK----TQLWFFLHNISMES-----TSESVQLGKLLQEEYSQGE---------------------------------------------------------------------------------------------
268317193|Rhodothermus_marinus_DSM_4252                    ITGTLVNYYVLCR----RKAWLAVHGLWMEQ-----ESEAVALGRLLDETTYAEELQ-------------------------------------------------------------------------------------------
28211140|Clostridium_tetani_E88                            VNGTLVNYYFHCK----RQCWLHGNRINLED-----NSQDVKIGKAIHEVKKEKGK--------------------------------------------------------------------------------------------
41614818|Nanoarchaeum_equitans_Kin4-M                      IGGAYIYYYFVCK----RKLWFYANKITMEH-----ESDLVQIGKYIELFYKEKHKY-------------------------------------------------------------------------------------------
289596695|Aciduliprofundum_boonei_T469                     FTGSLVHEYVVCK----RAAWY--SYRGVRL-----ENEYIILGKMLHEKSYKGE---------------------------------------------------------------------------------------------
18977491|Pyrococcus_furiosus_DSM_3638                      VRGIEINYLYVCP----TKLWFFTHGITMEQ-----ESEWVDLGKFLHEQRYENE---------------------------------------------------------------------------------------------
14590111|Pyrococcus_horikoshii_OT3                         VTGLMVQYYFTCK----RELWFFSRGINFDF-----ENEDMIIGRLIHEEAYEDS---------------------------------------------------------------------------------------------
55376277|Haloarcula_marismortui_ATCC_43049                 VTGVMMQYYEVCE----RELWFASRHLEIDR-----DNTAVVRGTQVDESAYDDKR--------------------------------------------------------------------------------------------
84489235|Methanosphaera_stadtmanae_DSM_3091                ITGVMIQYYKICK----RELWFYMNQINMNY-----DNNDIEIGRLIHENSYKRE---------------------------------------------------------------------------------------------
292492937|Nitrosococcus_halophilus_Nc_4                    LHGLHFQHVRLCE----RRAWMHLHGISFAQ-----WNRHVQIGMAQHDTSYARD---------------------------------------------------------------------------------------------
385805117|Fervidicoccus_fontis_Kam940                      ISLSLIRNLPYCE----RMVWF-KLNSDIEE----PPTPSMILSSERHRDLKRIAESLMLPQPM------------------------------------------------------------------------------------
347523047|Pyrolobus_fumarii_1A                             FGGGVVKEFVWCP----VAAWL-SVNLGVSP----PVTPRMLEGVEGHSRVLEVLHGLGFGDA-------------------------------------------------------------------------------------
15898285|Sulfolobus_solfataricus_P2                        VSVTDLKDYSLCK----VIPWI-RKKMGWKEPI--TNSQRIAKNVNLKEIVRDFPEPK------------------------------------------------------------------------------------------
15922977|Sulfolobus_tokodaii_str-_7                        VTVSDVKQYHFCK----AIPWI-NYVMGYRE----PSTFSMEEGKKISYEEVVKSLNLTT----------------------------------------------------------------------------------------
302347806|Acidilobus_saccharovorans_345-15                 VTGTAVKDYSYCP----ALPWL-KVRLGLRE----PETWGMRAGRSLRASDVAYKLGLERPW--------------------------------------------------------------------------------------
218884091|Desulfurococcus_kamchatkensis_1221n              VSPSLVKEYYYCP----VIPWI-QANLMVAE----PMTDSMALGSEAMKPPTG-----------------------------------------------------------------------------------------------
389861366|Thermogladius_cellulolyticus_1633                LSPSMVKEMVYCP----VIAWL-KARHLVQE----PATDSMLAGKQASGGGGLYARA-------------------------------------------------------------------------------------------
156937923|Ignicoccus_hospitalis_KIN4-I                     ----MVNAYALCP----AFAYF-KYHFPE------PQTPSMELGKELDVRELVEKFVRPKCEV-------------------------------------------------------------------------------------
15896973|Sulfolobus_solfataricus_P2                        ENTIYVTDLVRCP----RRVRY-ESEYKELAISQ-VYAPSAILGDILHLGLESV----------------------------------------------------------------------------------LKGNFNAETEVE
18311697|Pyrobaculum_aerophilum_str-_IM2                   PTPWDVLELEWCP----RYLWL-SKRHGV------PMTPNMAAGRGAEAGLRRFLAKALGAE--------------------------------------------------------------------------------------
170290049|Candidatus_Korarchaeum_cryptofilum_OPF8          LSVREIEEYFFCP----LLFYY-KNYLGIDT----QEGFWAYLGKRAQEEAEEEIRKRFEIL--------------------------------------------------------------------------------------
126465229|Staphylothermus_marinus_F1                       FTVRMVEEYIFCP----LVFYY-KYVLGIKK----PPSLWSNLGRQVEEELSRYVEEHYNVI--------------------------------------------------------------------------------------
257125860|Leptotrichia_buccalis_C-1013-b                   ISIFEVLDYIFCP----RIIYY-ENVLKMFGEK--KEDSNKKEEERVKGKNSINRKWIWDRLKLKK----------------------------------------------------------------QSIEKMMNSQENILYWYN
124027531|Hyperthermus_butylicus_DSM_5456                  IPIVLVKEYYWCP----VAAWH-KLVLWAER-----PTFSMEAGGLSGETRWRLLERLEEVFARWR-----------------------------------------------------------------------------VEARE
18311770|Pyrobaculum_aerophilum_str-_IM2                   INPSLIRQYLYCP----AAAYY-IMTGAAEP-----PTERMRAGKEVQREAVEAVAKALGAEKI------------------------------------------------------------------------------------
284040853|Spirosoma_linguale_DSM_74                        LTPSHIIEYLFCP----RFTYF-EYVLAIPQYE--EKNYKVMRGRNLHDERLERNKDYLRKRLGRP----------------------------------------------------------------------------GAPVTE
21226659|Methanosarcina_mazei_Go1                          IRISDVLEYLFCS----RFIYY-MYCLDIPQHE--EKRFKVIKGREVHETRKLTNRDYVRKKLNCIRK--------------------------------------------------------------------------------
218961575|Candidatus_Cloacamonas_acidaminovorans_str-_Evry ITPSEVMEYLFCP----RFVYY-MNILKIEQHE--HRRTLVNKGRDIHKLKMVQNKEYLRKKAGAIDK--------------------------------------------------------------------------------
150401499|Methanococcus_aeolicus_Nankai-3                  ITVSDVVEYMYCP----RFVYF-ERVLGIPQHE--EKREKVVIGRNIHELREKINKGYIRKSINAKSK--------------------------------------------------------------------------------
118497355|Francisella_novicida_U112                        IPVNLIRQWCFCP----RIVYY-QELLAIKP----NKPLWVAQGEEFHKKVEQLEKRRSFSRYGLE-------------------------------------------------------------------------------NAI
34557935|Wolinella_succinogenes_DSM_1740                   LPINLIRQYLFCP----RVVYF-LEVLNIPK----ASPVWVKEGSAHHRKQTELFKRRTLARFHLE-------------------------------------------------------------------------------NAK
54296141|Legionella_pneumophila_str-_Paris                 LPVNMIRQYCFCP----RIVYY-HMVVGLKV----AYPGWVQYGEEYQHKQQVLDKRRTLSKFKPI-----------------------------------------------------------------------------SEQAV
167628644|Heliobacterium_modesticaldum_Ice1                LRVTDVKQYVYCP----RIPYF-TYVQPLKP----RPTAKMDFGKDAHARLDKLEKRRTLAAYPEL------------------------------------------------------------------------------EAGE
386854795|Deinococcus_gobiensis_I-0                        ITPTELRQHHYCP----RVVFF-ERCTPVRR----RETILMTHGREKHQTELVRERRRTLSRYELT-------------------------------------------------------------------------------EGE
148656006|Roseiflexus_sp-_RS-1                             FEVTDLRQWAYCP----RVVYY-QYCLPDIR----PVTKLMEAGQEEHRAESGREERRSLRTYGLK-------------------------------------------------------------------------------TGE
118431378|Aeropyrum_pernix_K1                              RPVSMLKEYAYCP----RVAYY-MEVLRPSY----RPTEPMNLSREIYSVDHVRGILRSSGFRIVKE---------------------------------------------------------------------------------
307595805|Vulcanisaeta_distributa_DSM_14429                IPISMIKEYTYCP----RQAYL-KIMMMKEP-----PTESMKYAKQTQNTETITKTIKELGIQGEV----------------------------------------------------------------------------------
305663328|Ignisphaera_aggregans_DSM_17230                  IPIVLIKEYVYCP----RIAFY-KYFMYWEP-----PTESMKFARYSKIDILKIVRGYGIDGDI------------------------------------------------------------------------------------
408403002|Candidatus_Nitrososphaera_gargensis_Ga9-2        ISVTDIKHYFYCP----KIIYF-DKVMHADA----VLSSQQEDAKKTHKEKEKKDKRRKTMFYEEQ-----------------------------------------------------------------------------FPNCV
330509015|Methanosaeta_concilii_GP6                        -----MKQYLYCP----KIIYF-DHVLHIPK----PPDQKLQTGIEMHDSITAKEKRRKGAIFYDP----------------------------------------------------------------------------ELDQAE
11499461|Archaeoglobus_fulgidus_DSM_4304                   MDVSDIVQYLYCP----RKLYF-LKIAGIRI-----SKPKMEEGKAAQEEARGKLRSFARKLKGEL----------------------------------------------------------------------------------
383319611|Methanocella_conradii_HZ254                      LQVSDITQYFYCP----RKVYF-MKTMGIKV----KPRPKMDLGKEEHEREHRRVKERKTIYGFSE-----------------------------------------------------------------------------DEVKE
33359542|Pyrococcus_furiosus_DSM_3638                      --------------------MFHMYIVVVYDV-------------GVERVNKVKKFLRMH-LNWVQNSVFEG-EVTLAEFER---IKEGL------KK--IIDE----NS----DSVIIYKLR------SMP-----PR--ETL----
```

Continued

```
11499463|Archaeoglobus_fulgidus_DSM_4304                   -LVEHRMDGSL------LGCSGLIS-VDCYDYLR----------------NVVFDLKVGEFNE-------------------EMRLYTTGYALVLE-----SIHEIPVDVGCI---ISITFNGER---MRVSKDLHYI
383319613|Methanocella_conradii_HZ254                      -MIEHKISGNL------LGCSGILS-VDCFDYLH----------------NIIFDVKTATRKGENS----------------FYRLYATGYALVFE-----SVYEVPVDIGCT---VYLNFKDDR---LLIERDIFHI
330509014|Methanosaeta_concilii_GP6                        -VVEHKLDGKY------LGLSRNLS-ADAINMGG----------------MLICDLKTGQKRD-------------------FHRLSATGYALVYE-----SLYEVPINIGCT---VYLSFPKNHPT-PHIERDFFII
408402999|Candidatus_Nitrososphaera_gargensis_Ga9-2        -IVEQKIDGSA------LGLSTHLS-VDAFSFD-----------------SIILDLKTGEEKE-------------------FHKLSTTGYALVYE-----SIFEYPINVGCV---VYVNFSNNSPV-PIIKRVPHLI
167628645|Heliobacterium_modesticaldum_Ice1                -VVEQKLDGSF------LGLSAHLS-ADAYNIGQ----------------PMIVEVKFGEPRD-------------------FHRLATAGYALVME-----AVYEHPVDVGCL---VCVTFKGDR---VIIRRDIHLI
170290047|Candidatus_Korarchaeum_cryptofilum_OPF8          -ATELAIDGSL------LGLSQYLR-ADAAWIFG----------------GILFDVKTDRKED-------------------WHRIQLAGYALAFE-----SFFEMPVDIGCV---VYVSEVPEG---LRVFRDFFLI
126465227|Staphylothermus_marinus_F1                       -SLEMPVDGRL------LGLSSMLR-VDASWLPG----------------GLIYEIKTGYREK-------------------WHKLQVAGYALALE-----SIYERPVDIGVV---VYVNRAVEG---IRVNRDIFVV
156937920|Ignicoccus_hospitalis_KIN4-I                     -EVEPEIPGHS------VGLSSALK-PDFLVG------------------FIPVEVVYGNGLE-------------------RKKLAVTGYALALE-----SWIGHPVEVGIV---VSVTD-------KKTYWNVVRI
124027533|Hyperthermus_butylicus_DSM_5456                  LRVEPPIPGAP------VGLSDTVR-PDLLVG------------------FIPVEVTAAGPWGTYGS---------------KKELQLAGYALAIE-----AWTGIPVDYGVV---VYVTRRSDE---PRPEWRVVVI
118431381|Aeropyrum_pernix_K1                              -AVEPNVPGGP------VGLSDYVR-PDLLVG------------------IMPVDLVLAGHGDRWVA---------------RKELAIAGYALAVE-----AWTGNPVDYGVV---VGLRVNDG----LKIVWRLVRV
307595808|Vulcanisaeta_distributa_DSM_14429                -SVEPAIPGAP------IGLSDIVK-PDILVG------------------FIPVELVTSSQPTEELTS--------------RKDIAVTAYALAIE-----AWIGHPVDIGVV---LHISINGK----PRLTWRVVKI
305663331|Ignisphaera_aggregans_DSM_17230                  -SVEPYIPGGA------IGLSDFVR-PDLMVG------------------FMPIDITLSLGGNSLE----------------RKELALAGYALAIE-----AWSGNPVDIGIT---IGISINSD----VRFQWRVVRI
18311773|Pyrobaculum_aerophilum_str-_IM2                   -ATQVAVDGSP------LGMSYVV--ADGVAL------------------GAVIEFKFGPAQ--------------------NVDAALAGYAMAIE-----ADWGVPIDYGIH---VQITVNST----VEYKATAYPL
18311700|Pyrobaculum_aerophilum_str-_IM2                   -LAEYVIDGRP------LGLTLVR--ADALYY------------------NVVVEIKVGSYSE-------------------RHALALAGYALAIE-----ADEEVPVDAGLL---ISVAFNGG----VKLRAYPVAI
385805114|Fervidicoccus_fontis_Kam940                      IKTEFVIDGTP------VGLSNTLR-VDAYHEG-----------------GVIIEYKLGNPDPLLTE---------------RHSVGLAGYALAIE-----SSLDIPIDYGAI---VYISELSRR---PSIKVFPVYI
15898287|Sulfolobus_solfataricus_P2                        -ISEFKVDGSP------LGLSSRLS-VDAIMQL-----------------SLVVEMKVGSVQE-------------------FHRLALAGYGMALE-----SALELPIDFGAL---IYVNGINTS---SPEFKVEVFY
15922974|Sulfolobus_tokodaii_str-_7                        -ATEVRVDGSA------IGLSDRLT-VDALGE------------------FTVVEVKSGKREE-------------------FHKLALAGYALAIE-----SVLELPVDYGLL---VYINGFPND---VEIKFESYYI
347523054|Pyrolobus_fumarii_1A                             WITEYEVDGGL------LGLSRRLR-VDALAP------------------GLVVEVKYGRWSR-------------------DYPVALAGYAMALE-----ASLEIPFDYGLV---ILVNGDG-----SRLTLEPVYI
218884094|Desulfurococcus_kamchatkensis_1221n              AISEYRVDGSN------LGLAPNLS-ADLVAEG-----------------GVILDIKYGAPRE-------------------FHKLSLTGYALALE-----SEYEVPYDYGIL---IYIQGYLES---LKINIKPIYI
389861364|Thermogladius_cellulolyticus_1633                -VSEMKVDGTP------LGLSPQLS-IDVYAE------------------NIVVDFKTGAPRD-------------------FHKLSIAGYALALE-----AEYEVPVDYGIL---VYVNPSDSG---LRVLYRPVYV
332296539|Thermodesulfobium_narugense_DSM_14796            SEEPFNIDIND------IGFKGI---VDAITE------------------DSIIDFKTTNRIPSDITS--------------SHIIQTSIYSILTG-----------IDKVEL---HYISE-------KRVVIYKTQV
167627879|Francisella_philomiragiaATCC_25017               -LIGIDVINHK------YGLVGK---IDVFHKDK----------------GLLVERKRQIKTIYD-----------------GYKYQLYAQYFCLQ-----EMG-YDVKAIKF---YSMVDN------KSY---PIDI
56965355|Bacillus_clausii_KSM-K16                          -VRAMPIQSKQ------LQISGI---CDVVEFVEDPDGVELSGVSGTYQ-AFPVEYKRGKPKKGD-----------------EDIVQLVAQAICLE-----EMLVCRIDKGYL---FYNEIK------HRV---EVAI
76802275|Natronomonas_pharaonis_DSM_2160                   -INERYLSDSS------LGLHGK---IDIIESKDGV--------------LTPIERKRAESGGYYT----------------NDEVQLAGYCMLLS-----NAIGEPVNVGYI---YLYSTD------QRH---SIRI
189485222|uncultured_Termite_group_1_bacterium_Rs-D17      -ERDIYLKSYE------LGLIGK---SDVVEFHRSGNKLI----------PFPVEYKSGKAKSDN-----------------VDKVQLCAQALCLE-----EMMNVTIESGAI---FYGKTR------NRL---NVEF
134101617|Saccharopolyspora_erythraea_NRRL_2338            -ICALALSSDE------LGLVAV---CDVVEAVGGN--------------VRPVEYRRGAPQRDGQ----PWP---------NDRIQLLAQIVLLR------HHGYHCTHGYL---WYDSVR------RRV---PVPW
134299482|Desulfotomaculum_reducens_MI-1                   -DRSIHVSSEK------LNLYGI---VDIVEQGKE---------------IYPVEYKKGFRKESL-----------------NDDVQVCAQAMALE-----EKLGQDINRGYI---YYAGSK------ARR---EVIF
433631664|Mycobacterium_canettii_CIPT_140070010            -YWAVDVASDI------HRLRGR---CDRIVVEGGR--------------FVPVEHKSGRRSHH------------------SAMIQLVGQAICLE-----EMLGIPVARGRI---YLAASN------TFK---DVEV
86742031|Frankia_sp-_CcI3                                  -VTGLQVWSDR------LGLYGV---CDVVEISSTS--------------AIPIEHKVGPYVPGG-----------------PADLQAGAQALCLR-----DMLTLDVPYAEV---FSHTDR------RRH---RVDL
344996860|Caldicellulosiruptor_lactoaceticus_6A            -YKNLYVVSER------FKIEGI---IDIVIESSDR--------------VCIVEKKHGKRH--------AWE---------NDYFQVLAQILAFE-----ETTNIKVNEAYI---FYFESK------KRQ---KIEV
219683006|Bifidobacterium_animalis_subsp-_lactis_AD011     -YRAVNVNHQG------WGLSGR---LDAVRLNEDNG-------------VIIREYKATPVRRSMDVTH-------------AMRIQLALQAACME------DMGYRVDGTEI---FFTSHH------RIV---PVEL
347756303|Candidatus_Chloracidobacterium_thermophilum_B    -TLTYEMESER------WGIRGK---VDAVRYRDGS--------------LVAYEHKKGKSNRRA-----AWP---------SDRLQVIAYCALLE-----EATGKTILEGRI---RYHADN------VTV---RVPF
```

```
294828260|Leptospira_interrogans_serovar_Lai_str-_56601     -IESFHYTSEK------LEVSGK---VDRIQKRDGD-------------WIPYEHKRGRARIGTNGPE-AWE---------SDQCQVTVYALLLE-----EATGRNISEGKI---RYHGSK------DLV---KIEI
268318441|Rhodothermus_marinus_DSM_4252                     PVRSVLMSAPA------LGLIAR---MDVLEFSNGL-------------VYPVDYKKGSPSESD-----AWL---------TDKVQLCAQGLILR------ENGYRCEGGYV---FYVGTR------QRV---YVPF
378719041|Gordonia_polyisoprenivorans_VH2                   RVTSLRIESHA------LGLVAV---LDRVEDGGDGT------------VVPVDTKKGKPAANGL----PWS---------ADVVQLYAQAALLR------DAGYEVDRGVL---YYAEVH------KRV---VIPI
320102061|Isosphaera_pallida_ATCC_43644                     HARSVSLADDH------HGLIAK---LDLIECADGA-------------MIPVDYKKGKPRLADDGQPQVWE---------PERVQMAVQALVLR------ANGYRCERGVV---YYHQTR------QRV---DVPI
39995168|Geobacter_sulfurreducens_PCA                       HARSVSLSSER------LGITAK---IDLVEGEGAY-------------VSPVDYKRGKRPHVAGG---AYE---------PERVQLCAQGLLLR------EHGFASDGGAL---YFVASR------ERV---PVAF
83591519|Rhodospirillum_rubrum_ATCC_11170                   KTRRLSLSSER------LGLTAE---LDLLDVEEGM-------------VIPVEIKVGKRPSVDEG---AYL---------PERAQVCAQALLLR------EAGYTCLEGAL---WFAESR------ERV---TVDL
108757432|Myxococcus_xanthus_DK_1622                        -ERAVLLGSQR------LRVSGR---ADMVEYQRDAKGAGWR-------PFPVEVKRGRRKAGE-----------------ADRVQLCAQALCLE-----EMHGVDVPEGAL---FYGVEH------KRA---AVRF
46255268|Thermus_thermophilus_HB27                          -ESGLWVWSDR------LGLLGV---VDRLEWRRGE-------------ACPVEYKLGRAKEEAYL---------------SDAVQLAAQALCLR-----ESRGIKARRGFV---YYHKSH------TRR---EVVF
159898754|Herpetosiphon_aurantiacus_DSM_785                 -ERSVYVWSQR------WRITGI---VDLLESDRGE-------------FCPVEYKRGKLAQRR-----------------NDHVQLCAQALCLE-----EMFGCEVNQGFI---FSWANR------RRE---QVAF
38505683|Synechocystis_sp-_PCC_6803                         -HRQQWVWSDR------LKVKGI---IDAVEESDGL-------------LIPVEYKKGKMGKHL-----------------NDHFQLCAAALCLE-----EKTGKSITYGEI---FYYGNR------RRQ---PVEF
17229059|Nostoc_sp-_PCC_7120                                -VRAIWLKSDK------YQLIGK---SDLIESADGQ-------------FYPVEYKRGRKGEWD-----------------NDELQVCAQALCLE-----EMTGQTITTGYI---YYAHSH------QRQ---LVEI
302391601|Acetohalobium_arabaticum_DSM_5501                 -TTQVFVKSEQ------YRLVGK---IDVVEEKEGE-------------VYPVEYKKGRSGEWI-----------------NDHLQLCAQALCLE-----EQLGVKITKGYL---WYFSSR------DRE---EVEF
48477076|Picrophilus_torridus_DSM_9790                      -RKGLPISTGR----------SR---MDNVQIHKDF-------------IMVHEYKRGYKMIK------------------ADKMQILHYMYIIK-----LKENLDI-KGKL---HFLGSK------KVV---DIEL
154174706|Campylobacter_curvus_525-92                       -LQDFAFSNLK---------------FDKLSKQRGH-------------YCITEYKKSLKNEF------------------AGRMQLLFYIYILK-----TGLNLKEVKGRL-----ISGK------KVI---AVDD
111224946|Frankia_alni_ACN14a                               -RRDVDLGEAK---------------IDWVTTG----------------AVVHETKSSRAPSP------------------AHEAQVRHYCLLLE------RRGINVRSGVV---HYPLIR------RTV---TVAW
389844446|Mesotoga_prima_MesG1-Ag-4-2                       -DKDIVIDEIS---------------LDILRNT----------------REIVEIKYGKEVRE------------------GHIYQVKYYLYYLK-----HKLGLHEFTARL---TYPRAK------KMI---MLSL
268317193|Rhodothermus_marinus_DSM_4252                     -HVNVEADGPAG-----LRLAGR---IDWAELRD---------------GVLHETKRSPSGLE------------------AHRWQLRFYLWLLKLRGVTRPDGQPF-EGQL---NFPRQR------RTE---TVTL
28211140|Clostridium_tetani_E88                             -QTEISIDNIK---------------IDKITK-----------------DYLTEVKKSDSDIE------------------AAKWQLLLYLKVLK-----DK-GIE-RKGKL---EFIEKN------KSKSTIIIEL
41614818|Nanoarchaeum_equitans_Kin4-M                       -ERKVIINEIS---------------PDVIKRYKDF-------------VLVFEIKKSDKFKE------------------AALWQLKYYLWYLK-----RL-GLN-VKGKL---VIPEYN------KEE---YIEL
289596695|Aciduliprofundum_boonei_T469                      -IKNLFGEGVS---------------IDFTRKENGE-------------VWIYEVKRSSKMLN------------------AARMQLLFYLKWLK-----DR-GVE-VKGKI---AVPEEK------YTE---MVEL
18977491|Pyrococcus_furiosus_DSM_3638                       -EKEVLIGPIR---------------IDFISRGDV--------------IEVHEVKLGKRMER------------------AHEMQALYYLYYLK-----KH-GIK-AKAIL---HYPKLN------EKR---EVTL
14590111|Pyrococcus_horikoshii_OT3                          -WKEVLLEDIK---------------IDALKTKGG--------------LKVIEIKKSSKLEE------------------PAKWQLKYYLYYLK-----KA-GID-AIGVI---SYPKEG------KRE---IIEL
55376277|Haloarcula_marismortui_ATCC_43049                  -RNVSIDGTIA---------------IDVLDD-----------------GRVMEVKPSSSLIK------------------PAKLQLLYYLWYLD-----QVVGVE-REGVL---AHPTER------KRE---TVAL
84489235|Methanosphaera_stadtmanae_DSM_3091                 -KKEIRVDNLV---------------FDFIQTKDK--------------LTVFEVKKSSKITI------------------GALYQLYYYLYTLK-----NA-DIE-AKGML---VFPKEK------KRE---SIEL
292492937|Nitrosococcus_halophilus_Nc_4                     -RSTQGLFGLA---------------PDRIDWKH---------------NIVYESKGTGGAVE------------------AVSDQTAFYALLLS-----IATGKT-WRAVT---HVLSTR------KRR---EVVL
385805117|Fervidicoccus_fontis_Kam940                       -QFNKKLCHPQ------IPVCGS---PDIISGEER--------------KIIVEYKHFSRERGDFQ---------------DFRAQLLTYGWLVQ-----EVEGRVWKLVLM---V-----------GREVVWEQEY
347523047|Pyrolobus_fumarii_1A                              -RLLPPLASRR------WPIVVA---PDAYSPSR---------------GALLEVKWVARIDSR-----------------LYRVQALLYGVVAR-----EN-GLRVERVGL---VDAHRG------RVL---WFEL
15898285|Sulfolobus_solfataricus_P2                         -YYEVHLRDKD------SGLNGI---VDVLTE-----------------DSVIEIKAFYRRFYN-----------------HFRIQLLAYAYLAE-----RN-GFRVRNTVL----LMNKE------RKI---NIEV
15922977|Sulfolobus_tokodaii_str-_7                         -PVKYEVCLSY------NGLSGC---IDILAGDRK--------------YTVVEVKAFKRGNFS-----------------HFRYQLLAYAYLVR-----KRLGI-VERAIL----VMEGK------KVL---EIEP
302347806|Acidilobus_saccharovorans_345-15                  -LEEVPLRDRG------LGAAGV---VDLVSGSGR--------------LRVLEVKAFRAAGERLR---------------PFANQLKFYALLAN-----RLLG-RVREAYL---YHGG--------DLL---ALEV
218884091|Desulfurococcus_kamchatkensis_1221n               -RGQVRISTSR--------GVTI---IDEVIEEKNG-------------KIIIEKKAFKSHNFSR----------------YVSQAITSYVIARE-----KLPGVRKVRIAA-------GS------KIV---ELEV
389861366|Thermogladius_cellulolyticus_1633                 -----------------GGYSAL---VDEVREEGGG-------------VVVVERKAFKARDAQ-----------------RYMAQLVAGAYILS------LKGVRVRRLVL---EMGGVR------R-----ELEF
156937923|Ignicoccus_hospitalis_KIN4-I                      -RYQVPLRGK-------LG-SGR---ADAVVLCEGR-------------LEVVEAKLGGGFKR------------------QQLAQAAFYCLLAE-----EQFGLPCEALWL---CSPE--------GCE---RRPF
15896973|Sulfolobus_solfataricus_P2                         TLREINVGGKV------YKIKGR---ADAIIRNDNGK------------SIVIEIKTSRSDKGLPL--------------IHHKMQLQIYLWLFS-----------AEKGIL---VYITPD------------RIAE
18311697|Pyrobaculum_aerophilum_str-_IM2                    -PKPVYIDAG-------WA-HGV---VDILVKRVA--------------ASPVEIKTGLPRR-------------------EHKWQLYAEAYLVK------AAGCAVYTGYI---TYGE--------KIY---RLQI
170290049|Candidatus_Korarchaeum_cryptofilum_OPF8           -GKEVELRSDR------LGVVGK---VDFIVGRGRE-------------IMPLEVKFSGRLRP------------------WWKYSISLYAMLLE-----DSIGRPVKSGVV---LVTRGM------RFI---EVRV
126465229|Staphylothermus_marinus_F1                        -GKQVFLESKK------YSLAGI---VDYIVKYRGT-------------YTPLEIKYSRKLKP------------------WWKYTLTTYALLVE-----ETYYKPVKNAIL---IMPGPK------TTI----INI
257125860|Leptotrichia_buccalis_C-1013-b                    REFNKELSSEK------YHFYGK---LDDILLLEDKT------------IVPVYYYNAKYTAREEN---------------RYKYIVAMFLILIE-----EKYKIKSQKGYI---LFLNGL------SLK---KIEC
124027531|Hyperthermus_butylicus_DSM_5456                   LLWEHPVTSSR------LGLAGR---VDLAAVLSDDT------------LVVVEAKLSLPGKRRSL---------------HIRVQAAAYGMAAE-----ETLGYPLRMALI---YEAERN------RIV---EVEV
18311770|Pyrobaculum_aerophilum_str-_IM2                    -EHSPRLEG--------GGLCGV---VDAVLWLRGR-------------PAPLEVKLAKPPRKIPV---------------HHKAQAAAYAVAAQ-----HHYGKAVASAYI---YYTETG------EVK---YIAL
284040853|Spirosoma_linguale_DSM_74                         KHADQYLTNEL--------IRGV---VDEVLGFGDGT------------MAPLDYKFAEFKDKVYD---------------TYRTQLYCYAWLIE-----ANFGRPVDRGFL---VYTRSN------NRVI--EVPI
21226659|Methanosarcina_mazei_Go1                           -ESDVFIASKK------NHIKGI---VDEVLFLEDGT------------AAPLEYKFAEYKDKVFK---------------TYKFQLVLQALLIR-----ENYNIEVNRAYL---CFTRSN------SLVK--EMEI
218961575|Candidatus_Cloacamonas_acidaminovorans_str-_Evry  -LTDVYLSSEK------LKLVGK---VDEVLFLADGS------------AAPLDYKYAFWENKVYK---------------TLKYQQILYALLIM-----ENFQVPVYKAFI---VYTRSQ------NHLE--ELAI
150401499|Methanococcus_aeolicus_Nankai-3                   -YIDVYLSSKK------YRIVGV---VDEIVELRDGS------------YSPLDYKFAEFKNKLFK---------------THKYQSVLYGLLIK-----DNFDVDVKKGYV---VYVRSG------NLVK--EIEF
118497355|Francisella_novicida_U112                         RHFNLSIKSQK------YKLHGI---VDWVIETDTN-------------VYVVEYKTNPNPNSL-----------------GHKLQIAAYALLVQ-----EYFAKPCKTTFL-----TSDK------KSY---EIKI
34557935|Wolinella_succinogenes_DSM_1740                    LSTNVMLFHEN------FNFYGI---CDILIISEES-------------VYPVEIKLHGEKITK-----------------AQKAQIVAYGVLAE-----AIYRKKLEMGFV---LFEKNS------KTI---PIKI
54296141|Legionella_pneumophila_str-_Paris                  LRQNTMLEHPS------YPFYGI---CDGIIEDELE-------------IIPIELKSFASKISL-----------------GQKMQLTAYGMLSE-----LIFEKPCTRGYI---ITGEKA------KAK---LIKI
167628644|Heliobacterium_modesticaldum_Ice1                 RKFHTYIHSAR------LGLEGK---LDLHLVHRGE-------------IFPVEFKSTSRRPGL-----------------NHKYQLTAYAMLLE-----ELYQKPVRRGYF--KLVNPLADGARPEKSEITLPVEI
386854795|Deinococcus_gobiensis_I-0                         RRYEVRLTNHA------LGVTGE---LDLLIVDGAR-------------AFPVEFKHTSRLPDP-----------------GHKLQLCAYALLVE-----AELGLLCPHGYW---HSSRTR------QTH---MVLF
148656006|Roseiflexus_sp-_RS-1                              RFFDVPLRSET------LGLRGR---VDLVIAIPDRRQATE--------IVVVEYKSTEGKAGP-----------------HVKLQLTAYALMIE-----EAWGIPVKRGFI---YHMPER------KAE---AVTI
118431378|Aeropyrum_pernix_K1                               -EWAVPLRSKR------LGLQGV---ADGVVVEGSLG------------IIVVEAKLSVRSNRWLHTRGR-----------HVIFQAAAYALALE-----ETRGYSVDYLAI---VSLEDS------KTY---VVKM
307595805|Vulcanisaeta_distributa_DSM_14429                 -KLEAPVKSTS------LGIVGR---TDAVVINGNR-------------AVIIETKLNTGSKKKLYTKYN-----------HILAQATAYAIAVE-----ETMKVTVDKIII---INQEKQ------TTI---TIKP
305663328|Ignisphaera_aggregans_DSM_17230                   -YMEYPVKSRS------LGIYGK---IDIVIDNGKR-------------VSIAEIKLETSRARLWRKSY------------HHLIQLTAYTIATE-----ETFHKPIDRTFI---VTQIGG------EII---EINI
408403002|Candidatus_Nitrososphaera_gargensis_Ga9-2         KLFNIHMYSES------LRVEGV---VDCMIVNGNE-------------RIPVDYKRMFSKKGEAWT--------------DHRFQLTAYALLLE-----EQYSTIVRRGFV---YYLPED------KAV---EVKI
330509015|Methanosaeta_concilii_GP6                         KMFRVEMESST------LGLRGV---LDYLVRTNRE-------------FIPVDYKFGQSHKGSIHL--------------NHKYQLAAYALLVE-----ESFKTIVRRGFI---HYSRDQ------KNV---QIDF
11499461|Archaeoglobus_fulgidus_DSM_4304                    -LTSVYLESKR------YGLKGV---LDALVIVDGG-------------AFPVDVKLTRFQSITY----------------AWKMQLTAYALLVE-----ENYAVEVRKGYL---FLLNQG------KFV---EVSI
383319611|Methanocella_conradii_HZ254                       VIHDLAIENKT------IGLYGM---VDTVLIMKNGE------------VLPVDVKYSKSVSVKI----------------NWRKQLIAYSLLLE-----AYFGTTVRRGII---YLPAQH------KQI---QVEI
33359542|Pyrococcus_furiosus_DSM_3638
```

Continued

```
11499463|Archaeoglobus_fulgidus_DSM_4304                    -DANLRSWWIEERDRKAEIVFNEADPGVAKS-------CSSNCLFNAF--CL----------
383319613|Methanocella_conradii_HZ254                       -NDDLRSWWIEERDKKSEMVYEEKDPGVVQ--------CERKCMYANI--CKE---------
330509014|Methanosaeta_concilii_GP6                         -SDEVRQWFIEERDSKQDLIFYERDPGRCPN--------PGSCRYREF--CF----------
408402999|Candidatus_Nitrososphaera_gargensis_Ga9-2         -DETLRREFLEARDLRMKMIYDRRDPGLPRK-------CYAKCQYLRH--CGVGPEQIEVSA
167628645|Heliobacterium_modesticaldum_Ice1                 -DDELRQWFVEARDEKARMVYEEIDPGLAKD-------CPQSCPFANR--CS----------
170290047|Candidatus_Korarchaeum_cryptofilum_OPF8           -TDDLRSRFLERRDELQMMLIREDEPKRAER-------CPKYCLFSDM--CGVVT-------
126465227|Staphylothermus_marinus_F1                        -SDDLRSRFLEKRDEIQMLLLKGEEPRIPDK-------CPRKCLFRNI--CLGE--------
156937920|Ignicoccus_hospitalis_KIN4-I                      -HEGLRRRFLEVRDEVARILGEKEDPGAPEE-------CPKACPFAEV--CKR---------
124027533|Hyperthermus_butylicus_DSM_5456                   -DDALRRSFLRERDEVAMIVENRVDPGPAGE------GCPSWCPFRGVEGCPG---------
118431381|Aeropyrum_pernix_K1                               -DDHLRTRFLERRDAVARIIEYRSDPGLPEA-------CPPGCVFREV--CGV---------
307595808|Vulcanisaeta_distributa_DSM_14429                 -DDTLRRAFLDMRDNVARIIEHGDDPGPASS-------CPRTCPFYGV--CHG---------
305663331|Ignisphaera_aggregans_DSM_17230                   -DDSLRRRFIEARDDVARIIEYGEDPGKPSQ-------CLKMCPFREV--CGV---------
18311773|Pyrobaculum_aerophilum_str-_IM2                    -GDSARTKFLELRDEAIDIVIAGRDPGVAPD-------CPKTCPFYHI--CHADRSR-----
18311700|Pyrobaculum_aerophilum_str-_IM2                    -GEGLRREFLEERNRLMELVASGSDPGLSPK-------CDDKCPLWRH--CHGGSG------
385805114|Fervidicoccus_fontis_Kam940                       -SSLYRKKFLENRDEAIEVLMRATPPPRSGN-------CYQYCPFFKA--CNGGERS-----
15898287|Sulfolobus_solfataricus_P2                         VSPDMRKEFLDARDEVIDMISQEKDPGMPVS-------CSPSCPFLNY--CNNNKR------
15922974|Sulfolobus_tokodaii_str-_7                         -SPDMRREILDKRDEIIEMVLNGRDPGKPVK-------CPETCPFYEV--CWK---------
347523054|Pyrolobus_fumarii_1A                              -GNEERAEFLTARDEAAEIVVSQVDPGRPPR-------CPDQCPMRLV--CAP---------
218884094|Desulfurococcus_kamchatkensis_1221n               -NSYLRRYFIEERDEIIDMLIENREPPRDSA-------CNPSCPFYKV--CNK---------
389861364|Thermogladius_cellulolyticus_1633                 -SNELRRWFLEERDEIVDMLVSNREPPKDKN-------CPQTCPFYRV--CHQ---------
332296539|Thermodesulfobium_narugense_DSM_14796             -SSEMKEKTLMLIENVKDILENKIFLPNGLVH---PYACK-YCSVSNF--CYFVNGIS----
167627879|Francisella_philomiragiaATCC_25017                -PTSAKLEKFEKHIQTIKQYNPMDNSFRQNI-----EKCK-FCIYANL--CDKTDL------
56965355|Bacillus_clausii_KSM-K16                           -TDSLREKVVEMAKEMHHYYDNRYTPKVKTG-----PFCK-NCSLQHI--CLPKLMNKRSVK
76802275|Natronomonas_pharaonis_DSM_2160                    -TEDHRQAVTKIVDQIQSMSATNIPPLTDNP-----DKCD-ACSAREY--CMPAETAKLEPE
189485222|uncultured_Termite_group_1_bacterium_Rs-D17       -NKSLREETFALAQEFHSLVDSRETPKPEYS-----KKCY-NCSFKEL--CLPEIFDKQKSV
134101617|Saccharopolyspora_erythraea_NRRL_2338             -SATAEAELRHHLFRARHVASQGQPPPPLRH----SPKCP-RCALLPI--CMPDEINELSER
134299482|Desulfotomaculum_reducens_MI-1                    -DEDLRLMVENAVGLARNIALSGEIPPPLAD-----NRCE-GCALVNR--CLPFEVKGIKEN
433631664|Mycobacterium_canettii_CIPT_140070010             GNEELRAAVHTAASGVRVARADQSLPRPAND-----ARCP-ACSLNET--CLPRLVDDRHRQ
86742031|Frankia_sp-_CcI3                                   -TDTLATRIITAAERMRDILTTAALPEAVAD-----RRCR-RCSLHHD--CLPELANAAAGT
344996860|Caldicellulosiruptor_lactoaceticus_6A             -EENKITKLAEILDNMRKIAEGILKPAWEFS-----NKCP-KCSELLN--CFPYEIKQTNIF
```

```
219683006|Bifidobacterium_animalis_subsp-_lactis_AD011      -RKSDYEEAYGSVQEVRKLIECETAPLPFED----DPRCM-RCSHAGI--CLPEERAHNIPE
347756303|Candidatus_Chloracidobacterium_thermophilum_B     -DDAARQELRAAVARAAELRQHIERPPVTSD----VRLCR-HCALAPV--CLPEEDRLAVIT
294828260|Leptospira_interrogans_serovar_Lai_str-_56601     -DEELRSKALKTIDRAKGLSTSTNRPPVAQN----ENLCK-NCSLAPV--CLPEETRVITEN
268318441|Rhodothermus_marinus_DSM_4252                     -DEELITFTRESLEQLRQVARQPVPPPPLED----SPKCP-RCSLVGI--CLPDETHLLQQE
378719041|Gordonia_polyisoprenivorans_VH2                   -GNTEVAAAVEIADEAREVASRENAPLPLVG----SKRCP-RCSLLGL--CMPDETNALLRR
320102061|Isosphaera_pallida_ATCC_43644                     -DDSLIATTLRALDEARQLARSGTIPPPLVD----SPKCP-RCSLVGI--CLPDETNLLRNS
39995168|Geobacter_sulfurreducens_PCA                       -DDELIGRTLAAIDEMGRTALSGTMPPPLED----SPKCP-RCSLVGI--CLPDEVRFLSHL
83591519|Rhodospirillum_rubrum_ATCC_11170                   -TEALVTATLVATSDLRLTVASGRLPPPLDH----SAKCP-RCSLLPI--CLPDEIAWFRKG
108757432|Myxococcus_xanthus_DK_1622                        -DAELRQRTEDAIARMHEVLRRQEVPVVARA-----PRCD-KCSLEPL--CLPGVTAGQRSA
46255268|Thermus_thermophilus_HB27                          -TPELFRAVEAAVVRMRALLQSPRPPRVEVP----PSKCE-GCSVQGA--CQPELWRKGVAG
159898754|Herpetosiphon_aurantiacus_DSM_785                 -DPELRQKTIAIIEQAHAICQRGELPAPTSQ----KGKCR-DCSLKPL--CLPDEVAYLTTH
38505683|Synechocystis_sp-_PCC_6803                         -TQALRNMTEQAI-IAAHLACNGPMPRPIAN----TKKCA-DCSLERL--CLPKELKKLTK-
17229059|Nostoc_sp-_PCC_7120                                -NQELRKSAIATIQAVTNLLTTGIMPIPNYS-----KRCQ-GCSLNLQ--CLPKAIDRVKTY
302391601|Acetohalobium_arabaticum_DSM_5501                 -DLKLRDKTIETGESVLKIMEGEVIPENEYS-----RRCR-ACSIEEI--CLPKEVSILKKG
48477076|Picrophilus_torridus_DSM_9790                      -NDENLNELYKSYENIDKLKDMAKPPEPVRN-----YFCVHGCSYSDF--CWC---------
154174706|Campylobacter_curvus_525-92                       -SEENFNALETILREITALVNLAKPPKFTPQ-----KICK-NCAYNGY--CA----------
111224946|Frankia_alni_ACN14a                               -DDDARRSALETETQARAVISAPAVPPRLER-----RRCR-GCSYTDY--CWG---------
389844446|Mesotoga_prima_MesG1-Ag-4-2                       -TEEDEEQIRRDLDEIKTIFNLDNAPLPLFK-----EGCK-GCSYFDL--CFS---------
268317193|Rhodothermus_marinus_DSM_4252                     -TADDEARLRTLVQELRQTAARPTPPSRLEN----RRFCK-KCAYEEL--CFG---------
28211140|Clostridium_tetani_E88                             -DENNLSELEDVIKNIENLLIQENPPEVINE-----SKCK-KCAYFEY--CYI---------
41614818|Nanoarchaeum_equitans_Kin4-M                       -TEKDERKIKEILNEIKRIISMDKPPREVFK-----DYCK-ACAYYPL--CWEGY-------
289596695|Aciduliprofundum_boonei_T469                      -NEENEKKIEDLLKEMEVEIGKERPPKPVWR-----KICE-KCSYLEL--CWA---------
18977491|Pyrococcus_furiosus_DSM_3638                       --EGREREVEEAIVEVEVIKSSSKPPKPRKS-----KICR-KCAYKEL--CWG---------
14590111|Pyrococcus_horikoshii_OT3                          -KEENIKVIEEAIKGIEEVVSSPKPPKPVKK-----PYCR-KCSYRDL--CWV---------
55376277|Haloarcula_marismortui_ATCC_43049                  -TDENTRWVEEAIRGIYDVTTSDSPPPAEEK-----PFCE-SCAYHDF--CWSC--------
84489235|Methanosphaera_stadtmanae_DSM_3091                 -TEDIIKEIDDIIKEIKKISNETKPPHEKQK-----PYCQ-GCSFREL--CLI---------
292492937|Nitrosococcus_halophilus_Nc_4                     -DEARMQQLWESSERLERLAQQTQVPSAKRI-----PLCA-SCSLAAF--CGYD--------
385805117|Fervidicoccus_fontis_Kam940                       -SYELHISARHLVERLLKILDSPTPPSASYG-----GKCS-YCFFKSV--CSSSNQPAEIL-
347523047|Pyrolobus_fumarii_1A                              -DWGSFSIVERLVERVLETVSRPLPPPVTQP----PEKCL-ACRWRRW--CPERSV------
15898285|Sulfolobus_solfataricus_P2                         -RKEHVEYIGKVVNELVESLEDDSPPVVNPS----PLLCK-ACQYRRV--CSVSVLL-----
15922977|Sulfolobus_tokodaii_str-_7                         -TEEHYRYIENVVKKIEEIVNSEKLPAVNVE------KCD-FCQYRRI--CPVSVWS-----
302347806|Acidilobus_saccharovorans_345-15                  -DGHVLREAEELVARVRDVLESEEPPRAYRS-----ERCS-SCWFRRV--CPAWD-------
218884091|Desulfurococcus_kamchatkensis_1221n               -NDQLVRDVEASIRSVERMISESSLPKPGEQ-----GKCS-SCWYRRF--CPYA--------
389861366|Thermogladius_cellulolyticus_1633                 -TGELVEEARRYLGELERLASSDRPPQPTRP----GRRCS-SCWYRRF--CPYTE-------
156937923|Ignicoccus_hospitalis_KIN4-I                      -SY-SLREHVKGLLEALERDLDVLPRGRPS------RYCS-YCKYSSL--CPWRPP------
15896973|Sulfolobus_solfataricus_P2                         -YEINEPLDEATIVRLAEDTIMLQNSPRFN------WECK-YCIFSVI--CPAKLT------
18311697|Pyrobaculum_aerophilum_str-_IM2                    -SKAELKAAETLLVKAAEVIEGPPPPPVKTG------KCA-YCQYRAI--CTPT--------
170290049|Candidatus_Korarchaeum_cryptofilum_OPF8           -GDSERRFVLGAIEKCRRIME-GEIPRAYRS-----RSCE-NCDFKNR--CLEI--------
126465229|Staphylothermus_marinus_F1                        -TSNDRNYIIKTLENMKKTLEEQTTPKPTPS-----KSCI-NCDYKNI--CPYQTKNQ----
257125860|Leptotrichia_buccalis_C-1013-b                    -TNKNLEKVKESAAEALKIIETERYPIEIEG----GTKCR-DCYYKKI--CGR---------
124027531|Hyperthermus_butylicus_DSM_5456                   -TPMLRKMVEDAARELHEMLQRGEPLQHYTVS---PRRCR-VCSYRGV--CTYART------
18311770|Pyrobaculum_aerophilum_str-_IM2                    -TRDLVELVKHTAQQIRKIQNGWTPTPAYHP-----AKCP-SCWYRKY--CGFTSPHVEKI-
284040853|Spirosoma_linguale_DSM_74                         -AETDKVAVKRAAEAIFTIIERNFYPKATKA----KARCV-TCTYRNI--CIK---------
21226659|Methanosarcina_mazei_Go1                           -TSSDFKKAEKIIEEILDIVQKGLYPKTSRS----SRKCV-DCCYRNI--CV----------
218961575|Candidatus_Cloacamonas_acidaminovorans_str-_Evry  -TPKMLDKAKLILDEIFMIINLEKYPKPAKA----KRKCL-DCTYRNL--CSV---------
150401499|Methanococcus_aeolicus_Nankai-3                   -KKKDFDDAVAFVDLIFDIIENEEFPDGIKVN---KRKCI-DCCYRNV--CCK---------
118497355|Francisella_novicida_U112                         -TDELINKLIKTISDILSTLDSGNKPDSSAS----DHQCI-QCEYYNF--CNDR--------
34557935|Wolinella_succinogenes_DSM_1740                    -SDGDKAEINNFVKKIIEMIELGNLPHSSAD----GSKCT-QCEFLNY--CNDRF-------
54296141|Legionella_pneumophila_str-_Paris                  -DEKSKFKVIEICAKIKKIFGTETLPYSTAN----GLQCI-QCEFLNH--CNDRF-------
167628644|Heliobacterium_modesticaldum_Ice1                 -TVPMRDFVKDVLRRIRAMVKQESLPSPKRP----GNRCR-DCEYRVY--CLDLGDVVR---
386854795|Deinococcus_gobiensis_I-0                         -DTRLRNRTRAAIQVVRGFILAEQCPSPTAQ----TVKCL-ECELRNF--CGDTL-------
148656006|Roseiflexus_sp-_RS-1                              -SAALRTRVKETVAAIRATVDHERMPPPPAS----RRQCA-VCEFRRF--CNDVV-------
118431378|Aeropyrum_pernix_K1                               -SPSLRRDVIRLADDMNKTLDDGLEPPPKPG-----RKCV-ACRFRRV--CQPWVAERGSER
307595805|Vulcanisaeta_distributa_DSM_14429                 -TPNLRTWIKRTIQDMKKHIQNQEPPPKTTN----KAKCK-ACYFKDI--CPP---------
305663328|Ignisphaera_aggregans_DSM_17230                   -SPNLRRTTINIVEEFKKYIQQQKPPPKTPM----KTRCN-HCFYRKI--CV----------
408403002|Candidatus_Nitrososphaera_gargensis_Ga9-2         -TESMKTYLKKTVKAIQNVIESGEEPVARIP----KSRCTGGCGYLWI--CGGIWNRRH---
330509015|Methanosaeta_concilii_GP6                         -GDEIRRRTLKMIEKIAQIIEDEIEPAGTRN----SARCI-DCEYRKY--CVGNILPKPHTE
11499461|Archaeoglobus_fulgidus_DSM_4304                    -SPDDRKALLRIIKEIERLIEEEKYPRTSKS-----KKCG-YCEVREF--CV----------
383319611|Methanocella_conradii_HZ254                       -TCEGKEMVKQDIRKIKDLIKSEKMPNVNKG-----LQCN-YCEMAKF--CS----------
```

Supplementary Figure S2: Cas1, Cas2, Cas4 phylogenetic trees (please see Supplementary Materials and Methods for details and also see Makarova KS, Wolf YI, Koonin EV. **The basic building blocks and evolution of CRISPR-Cas systems.** Biochem Soc Trans. 2013 Dec 1;41(6):1392-400); Sequences are denoted by Genbank identifiers and organism genus and abbreviated species name; automatically identified CRISPR-Cas system subtype and gene in the Cas1 neighborhood are provided). The trees are provided in Newick format, could be open in any tree-viewing software.

Cas1 Tree:

```
(((347756303|CAS-I-C|Chloracidobacterium_the___6-3-8c-7-5-1+4-2:0.38914,(294828260|CAS-I-C|Leptospira_int___1+4-2-x-x-6-3-8c-7-
5:0.54801,162457462|CAS-I-C|Sorangium_cel___6-3-8c-7-5-1-2:0.50926)0.895:0.13172)1.000:0.61947,((15612904|CAS-I-
C|Bacillus_hal___cmr5SS-cmr6g7-x-x-x-3-5-8c-7-4-1-2:0.64423,108763792|CAS-I-C|Myxococcus_xan___2-1-4-7-8c-5-
3HD:0.49951)0.765:0.13333,(86742032|CAS-I-C|Frankia_CcI___3-5-8c-7-4-1-2:0.71514,433631663|CAS-I-C|Mycobacterium_can___2-1-
4-7-8c-5-3:0.60677)0.734:0.20305)0.990:0.32273)0.717:0.14616,(219683006|CAS-I-U|Bifidobacterium_ani___3-csb1-
1+4:1.07926,(((378719041|CAS-I-U|Gordonia_pol___2-1+4-csb2-csb1-8-3:0.51986,134101617|CAS-III-A|Saccharopolyspora_ery___1+4-
2-x-x-x-x-csm3g7-cmr3g5-csm3g7-10:0.62705)0.819:0.11364,(268318441|unknown|Rhodothermus_mar___2-1+4:0.43464,312114617|CAS-
III-A|Rhodomicrobium_van___RT-2-1-csm3g7-csx19-csm3g7-csm3g7-10-x-csm6-csx1-x-
csx16:1.15334)0.817:0.13822)0.378:0.07072,(320102061|CAS-I-U|Isosphaera_pal___1+4-2-csb2:0.57635,(39995168|CAS-I-
U|Geobacter_sul___3-8-csb1-csb2-x-x-1+4-2:0.33269,83591519|CAS-I-E|Rhodospirillum_rub___2-1-6e-5-7-cse2SS-8e-3-x-x-2-1+4-csb2-x-
x-csm3g7-csm3g7-csm3g7-10:0.61850)0.935:0.13614)0.960:0.17395)0.813:0.13734)0.850:0.11671,(((38505684|CAS-I-
D|Synechocystis_PCC___3-10d-csc2g7-csc1g5-6-4-1-2:0.83136,(344996859|CAS-I-D|Caldicellulosiruptor_lac___2-1-4-6-3-csc1g5-csc2g7-
10d:0.90289,46255267|CAS-I-A|Thermus_the___2-1-4-x-8a-7-5-6:0.82022)0.822:0.10403)0.844:0.08986,(((302391602|CAS-I-
D|Acetohalobium_ara___10d-csc2g7-csc1g5-3-4-1-2-6:0.44986,17229060|CAS-I-D|Nostoc_PCC___10d-csc2g7-csc1g5-3-2OG-6-4-1-x-
2:0.80140)0.816:0.13790,159898755|CAS-I-D|Herpetosiphon_aur___10d-csc2g7-csc1g5-3-6-4-1-2:0.73922)0.853:0.07489,134299482|CAS-
III-A|Desulfotomaculum_red___1+4-2-10-csm2SS-csm3g7-csm4g5-csx1-6:0.60078)0.965:0.13680)0.386:0.06285,((268318378|CAS-III-
U|Rhodothermus_mar___csx1-csx20-x-x-csx1-csx1-6-x-2-1+4:0.71035,(336116789|CAS-III-A|Microlunatus_pho___10-csm3g7-cmr3g5-
csm3g7-x-x-x-6-1+RT-2-x-10-csm3g7-cmr3g5-csm3g7:1.15764,(55978328|CAS-III-A|Thermus_the___csx1-1-x-10-csm2SS-csm3g7-
csm4g5-csm5g7-csx1-x-x-csx1:0.32915,297622675|CAS-III-U|Truepera_rad___2-1-csx1-csx1-
6:0.65702)0.999:0.43981)0.614:0.14144)0.987:0.24355,((76802276|CAS-I-D|Natronomonas_pha___R-6-10d-csc2g7-csc1g5-3-4-1-
2:1.07764,(390962116|CAS-III-A|Thermococcus_CL1___6-10-csm2SS-csm3g7-csm4g5-csm5g7-R2-csx3-2-1:1.08282,(154174048|CAS-I-
B|Campylobacter_cur___1-2-4-3-5-7-8-6:1.20432,(((((28211139|CAS-I-B|Clostridium_tet___1-4-3-5-7-8b-6:0.80992,(18977490|CAS-III-
B|Pyrococcus_fur___2-1-4-3-5-7-8b-cmr6g7-cmr5SS-cmr4g7-csm6-cmr3g5:0.60617,(((84489236|CAS-I-B|Methanosphaera_sta___6-8-7-
5-3-4-1-2:0.38030,(289596694|CAS-III-A|Aciduliprofundum_boo___2-1-4-6-csx1-csx1-csm5g7-csm4g5-csm3g7-csm2SS-
10:0.63757,13540937|CAS-III-A|Thermoplasma_vol___2-1-csx1-csm5g7-csm4g5-csm3g7-csm2SS-10-6-
csm6:0.86508)0.905:0.15392)0.906:0.14613,14590110|CAS-III-A|Pyrococcus_hor___10-csm2SS-csm3g7-csm4g5-csm5g7-csm6-8b-7-5-x-1-
2-4-3-csx1:0.63787)0.819:0.15012,41614813|CAS-I-B|Nanoarchaeum_equ___2-1-x-7-5-3-
4:0.77636)0.073:0.03836)0.175:0.14006)0.656:0.10760,268317185|CAS-III-A|Rhodothermus_mar___2-1-x-csx1-csm5g7-csm4g5-csm3g7-
csm2SS-10-4-3-5:0.62102)0.325:0.06368,(385809790|CAS-III-U|Ignavibacterium_alb___csm5g7-x-csx1-csx20-csx1-csx1-x-x-x-x-1-
2:0.72273,150021526|CAS-III-B|Thermosipho_mel___2-1-2-x-x-x-csm6-cmr1g7-10-cmr3g5-cmr4g7-
cmr5SS:0.71602)0.932:0.15479)0.396:0.05380,(389844445|CAS-III-B|Mesotoga_pri___cmr6g7-cmr5SS-cmr4g7-cmr3g5-10-cmr1g7-csm6-
2-x-x-1-4-3-5-7:0.68628,48477075|CAS-I-D|Picrophilus_tor___2-1-4-3-csc1g5-csc2g7-10d-
6:0.97627)0.470:0.06969)0.718:0.10249,(55376278|CAS-I-B|Haloarcula_mar___6-8-7-5-3-4-1-2:0.65921,(292492938|CAS-I-
D|Nitrosococcus_hal___6-10d-csc2g7-x-3-4-1-2:1.06977,111224947|CAS-I-B|Frankia_aln___6-x-7-3+5-4-1-
2:1.01974)0.382:0.12937)0.495:0.16546)0.929:0.19711)0.540:0.12245)0.997:0.40897)0.849:0.17499,(((((317051217|CAS-III-
A|Desulfurispirillum_ind___10-csm2SS-csm3g7-csm4g5-csm5g7-x-x-csx1-csx20-x-1-2-x-2:0.86471,17227877|CAS-III-
B|Nostoc_PCC___csx18-1-2:0.92025)0.768:0.14478,(((85858452|CAS-III-B|Syntrophus_aci___cmr5SS-cmr6g7-csm6-csx1-x-x-x-x-x-x-1-2-
2:0.74830,(328949009|CAS-III-A|Treponema_suc___csx1-x-x-2-2-1-x-x-x-csx20-csx1-x-csm5g7-csm4g5-csm3g7-
csm2SS:0.38670,333996012|CAS-III-B|Treponema_azo___csx20-csx1-cmr6g7-cmr5SS-cmr4g7-cmr3g5-10-cmr1g7-6-x-1-2-x-
csx1:0.43621)1.000:0.46620)0.996:0.33602,((328953422|CAS-III-A|Desulfobacca_ace___csm5g7-csm6-x-csx20-csx1-x-x-x-Prim2+x-x-1-
2:0.58824,(313673559|CAS-III-A|Calditerrivibrio_nit___2-Prim2-1-x-x-x-x-x-x-x-x-csx20-csm5g7:0.86969,218961167|CAS-III-
A|Cloacamonas_aci___csm5g7-csa3-csx1+x-x-x-x-x-x-Prim2-x-1-
2:0.68621)0.862:0.17223)0.840:0.15794,(154250301|unknown|Fervidobacterium_nod___1:0.90611,116748794|CAS-III-
A|Syntrophobacter_fum___csm5g7-csm4g5-csm3g7-csm2SS-10-6-csx20-csx1-x-Prim2-1-
2:0.93083)0.434:0.06593)0.080:0.06603)0.699:0.08120,(332296542|CAS-III-A|Thermodesulfobium_nar___4-x-2-1-Prim2-csm5g7-csm4g5-
csm3g7-csm2SS-10-csx1-csx1-6:0.88721,337286709|CAS-III-B|Thermodesulfatator_ind___cmr5SS-cmr4g7-x-x-cmr3g5-10-cmr1g7-x-2-x-
1-2-6:0.72067)0.539:0.11721)0.713:0.06506)0.646:0.10342,(((148657121|CAS-III-U|Roseiflexus_RS___2-x-1-RT-x-x-x-x-csx15-csx1-x-
csx1:0.61021,159898907|unknown|Herpetosiphon_aur___RT-1-2:0.55145)1.000:0.38100,(320161859|CAS-III-A|Anaerolinea_the___csx1-
10-csm2SS-csm3g7-csm4g5-csm5g7-csx1-csx1-csx15-6-1-2:0.68967,(156741961|CAS-III-A|Roseiflexus_cas___10-csm2SS-csm3g7-csm4g5-
csm5g7-csx1-6-1-2:0.47789,383762167|CAS-III-A|Caldilinea_aer___2-1-6-csx15-x-csx1-csm5g7-csm4g5-csm3g7-csm2SS-10-
csx1:0.57444)0.084:0.08868)0.989:0.24410)0.762:0.09771,37677204|CAS-III|Vibrio_vul___6-csm3g7-x-csm3g7-5g10-cmr4g7-10-x-x-
1+RT-2-2-csx1-csx16-csx16-csm6:1.23121)0.614:0.07387)0.702:0.04517,((((296133514|CAS-III-A|Thermincola_pot___2-1-csm6-6-
csm5g7-csm4g5-csm3g7-csm2SS-10:0.55968,(348026605|CAS-III-A|Megasphaera_els___1-2-10-csm2SS-csm3g7-csm4g5-csm5g7-6-
```

```
csm6:0.72836,114567264|CAS-III-A|Syntrophomonas_wol___csm6-x-2-1-x-6-csm5g7-csm4g5-csm3g7-csm2SS-
10:0.55211)0.097:0.09427)0.924:0.14860,(116627764|CAS-III-A|Streptococcus_the___1-2-6-csm2SS-csm3g7-csm4g5-csm5g7-x-
csm6:1.15855,15609954|CAS-III-A|Mycobacterium_tub___2-1-csm6-csm5g7-csm4g5-csm3g7-csm2SS-10-
6:0.80127)0.948:0.34739)0.991:0.30794,(218442809|CAS-III-A|Cyanothece_PCC___csm6-x-x-RT-1-2:0.82366,(38505760|CAS-III-
B|Synechocystis_PCC___cmr6g7-cmr5SS-cmr4g7-x-cmr3g5-10-csx18-1-2:1.05148,((119357853|CAS-III|Chlorobium_pha___2-1-
6:0.57832,119357846|CAS-III-A|Chlorobium_pha___6-csx1-csm5g7-csm4g5-csm3g7-csm2SS-10-x-csx15-csx1-1+RT-
2:0.71133)0.908:0.16639,(427708213|CAS-III|Nostoc_PCC___csx19-csm3g7-x-x-x-6-x-x-x-2-1-x-x-1+RT:1.09536,17228961|CAS-III-
A|Nostoc_PCC___1+RT-x-x-x-csm3g7-csx19-csm3g7-csm3g7-x-csm3g7-
csm3g7:0.37340)0.986:0.34526)0.475:0.09413)0.277:0.05347)0.828:0.10795)0.964:0.17610,(328953000|CAS-III|Desulfobacca_ace___6-x-x-
x-x-x-RT-x-x-1-2:0.79732,(261415948|CAS-III-A|Fibrobacter_suc___1-2-x-x-csx1-10-csm3g7-x-csm3g7-csx20:0.76621,269798874|CAS-
III-A|Veillonella_par___6-csm3g7-csx19-csm3g7-5g10-csm3g7-10-csx20-1-2-x-csx1-
csx1:1.17138)0.843:0.22745)0.760:0.16672)0.679:0.01909)0.789:0.04869,((((326793969|CAS-III-B|Marinomonas_med___2-6-x-x-cmr6g7-
cmr5SS-cmr4g7-cmr3g5-10-cmr1g7:0.97163,220934970|CAS-III-B|Thioalkalivibrio_sul___cmr1g7-10-cmr3g5-cmr4g7-cmr5SS-cmr6g7-
csx1-x-csx1-x-1-2-x-csx1:0.60767)0.999:0.58432,((436839745|CAS-III-B|Desulfovibrio_hyd___1+RT-2-x-x-cmr6g7-cmr5SS-cmr4g7-
cmr3g5-10-cmr1g7:1.37288,(283778924|CAS-III-A|Pirellula_sta___1+RT-2-x-10-csm3g7-
csm3g7:1.45956,118497969|unknown|Francisella_nov___2OG-x-x-x-x-csa3-2-1-
4:1.55280)0.320:0.28782)0.806:0.12534,(319790504|unknown|Thermovibrio_amm___2-2-2-1:1.30366,337288662|CAS-III-
A|Thermodesulfobacterium_OPB___10-csm2SS-csm3g7-csm4g5-csm5g7-6-csx1-x-2-1-
2:1.11781)0.980:0.53351)0.906:0.16162)0.600:0.09143,((((325980970|unknown|Nitrosomonas_AL2___1:1.73273,(407462288|unknown|Nitr
osopumilus_kor___1:1.50397,(21229351|unknown|Methanosarcina_maz___1:0.67446,147921327|unknown|Methanocella_arv___1:0.9953
7)0.976:0.45554)0.831:0.19198)0.904:0.19310,(((150398872|CAS-III-B|Methanococcus_van___10-cmr3g5-cmr4g7-cmr5SS-cmr6g7-x-x-6-
1-2:0.84562,(((((332668016|unknown|Haliscomenobacter_hyd___RT-1-PDEDxK-2:1.20570,(327314603|CAS-III|Prevotella_den___10-
cmr6g7-5g10-csm3g7-x-csm3g7-csx1-csx1-1-Prim-2:0.73847,(379728404|unknown|Saprospira_gra___2-x-
1:0.87455,392395717|unknown|Flexibacter_lit___2-x-1:0.68194)0.836:0.12742)0.831:0.12483)0.909:0.21557,284040851|CAS-I-
B|Spirosoma_lin___8-7-5-3-6-1-2-4:0.73413)0.213:0.10519,218961577|CAS-I-B|Cloacamonas_aci___4-2-1-3HD-5-7-8-
6:0.58565)0.358:0.09219,257125862|CAS-III|Leptotrichia_buc___csm3g7-6-x-x-x-x-x-x-4-2-1-x-x-csx1-
csx20:0.61093)0.324:0.10225,(21226661|CAS-I-B|Methanosarcina_maz___4-2-1-6-3-5-7-8-csa3:0.33047,150401497|CAS-III-
A|Methanococcus_aeo___6-3-5-7-8-1-2-4-x-csx1-csm5g7-csm4g5-csm3g7-csm2SS-
10:0.92232)0.484:0.09616)0.933:0.21764)1.000:0.55180,(((11499462|CAS-I-A|Archaeoglobus_ful___cmr1g7-csa3-csa5SS-7-5-csa8a2-3-
3HD-2-4-1-4:0.41184,383319612|CAS-I-A|Methanocella_con___4-1-4-2-csa3-x-x-x-x-x-7:0.53988)0.920:0.13193,(288560430|CAS-III-
A|Methanobrevibacter_rum___1-x-2-6-csx1-csm5g7-csm4g5-csm3g7-csm2SS-
10:0.47372,148642083|unknown|Methanobrevibacter_smi___1-2:0.88751)0.962:0.26391)0.999:0.31122,(((((((156937921|CAS-I-
A|Ignicoccus_hos___4-1-2-4-3HD-3-x-5-7-csa5SS-R:0.93227,124027534|CAS-I-A|Hyperthermus_but___csaXg8-5-7-csa5SS-R-x-x-4-2-4-
1-x-x-x-R:0.65197)0.613:0.22777,(305663330|CAS-I-A|Ignisphaera_agg___R-x-7-5-3-3HD-x-6-4-2-1-4:0.36006,307595807|CAS-I-
A|Vulcanisaeta_dis___3HD-3-x-5-7-x-R-x-4-2-1-4:0.64930)0.000:0.08743)0.820:0.16306,118431380|CAS-I-A|Aeropyrum_per___6-8a2-
3HD-3-5-7-csa5SS-R-4-2-1-4:0.62305)0.991:0.27322,(347523053|CAS-I-A|Pyrolobus_fum___4-x-x-x-x-2-1-4-R-x-x-3HD-3-8a2-5-7-
csa5SS:0.78895,((218884093|CAS-I-A|Desulfurococcus_kam___6-3HD-3-csaXg8-5-7-csa5SS-R-4-2-1-4-R-x-cmr3g5-10-cmr6g7-cmr1g7-
cmr5SS-cmr4g7-csm6:0.52821,389861365|CAS-I-A|Thermogladius_163___6-x-x-4-1-4-R:0.39368)0.988:0.22330,(302347808|CAS-III-
B|Acidilobus_sac___6-x-4-2-1-4-cmr3g5-10-cmr6g7-csm3g7-cmr5SS-cmr4g7-csm6-x-csm6:0.75489,(385805115|CAS-I-
A|Fervidicoccus_fon___6-R-3HD-3-8a2-5-7-x-x-4-1-2-4-R-R-x-x-csm6-cmr6g7-cmr1g7-cmr5SS:1.07762,(15898286|CAS-I-
A|Sulfolobus_sol___csaXg8-3HD-3-5-7-csa5SS-R-R-x-4-1-4:0.66104,15922975|CAS-I-A|Sulfolobus_tok___R-4-1-2-4-x-x-csa5SS-7-5-3-
3HD-csaXg8:0.46000)0.982:0.27868)0.852:0.19905)0.274:0.09946)0.000:0.06113)0.643:0.12690)0.943:0.18315,((18311772|CAS-I-
A|Pyrobaculum_aer___6-4-2-1-4-csa3-x-8a2-3HD-3-5-7-csa5SS-6:0.90508,18311699|CAS-I-A|Pyrobaculum_aer___3HD-3-x-x-x-6-csa3-x-
4-2-1-4:0.82264)0.635:0.11599,15898245|CAS-I-A|Sulfolobus_sol___csx1-csa5SS-7-5-csaXg8-3-3HD-2-1-
6:1.32874)0.273:0.10934)0.979:0.21717,(((34541577|CAS-III-B|Porphyromonas_gin___2-1+RT-cmr6g7-x-cmr4g7-cmr3g5-
10:1.17610,(386750121|CAS-III|Helicobacter_cet___csm5g7-csx1-x-x-x-x-x-x-x-2-1:1.23102,328954440|CAS-III-
A|Desulfobacca_ace___1+RT-x-x-csm3g7-x-csm6-csm3g7-x-csm3g7-csm3g7:0.77196)0.683:0.11126)0.716:0.14667,(54296139|CAS-II-
B|Legionella_pne___9-1-2-4:0.74021,(34557933|CAS-II-B|Wolinella_suc___9-1-2-4:0.46386,118497353|CAS-II-B|Francisella_nov___9-1-
2-4:0.52534)0.502:0.15130)1.000:0.52829)0.629:0.07561,(148656004|CAS-III-U|Roseiflexus_RS___csm3g7-10-csx1-csx1-csx3-8b-7-5-3-6-1-
2-4:0.77295,(386854797|CAS-I-A|Deinococcus_gob___4-2-1-5-7-8a2-3-6:0.75119,167628642|CAS-I-A|Heliobacterium_mod___4-4-2-1-x-6-
5-7-x-3:0.54240)0.076:0.07286)0.605:0.12022)0.964:0.15886)0.878:0.08588,((126465226|CAS-I-A|Staphylothermus_mar___csm3g7-
csm4g5-csm5g7-10-x-x-1-4-2-4-6-R-csa5SS-7-5-csa8a2-3:0.60120,170290046|CAS-I-A|Korarchaeum_cry___1-4-2-4-R-csa5SS-7-5-csa8a2-
x-3:0.38512)1.000:0.74406,(330509017|CAS-I-A|Methanosaeta_con___4-4-2-1-3-x-x-x-x-5:0.72091,408403000|CAS-
I|Nitrososphaera_gar___4-1-2-4-x-x-x-3-5-7-x-
6:0.85041)0.834:0.11037)0.627:0.05146)0.645:0.08920)0.700:0.09695)0.851:0.07534,(119720226|CAS-III-B|Thermofilum_pen___cmr3g5-
10-cmr6g7-cmr1g7-cmr5SS-cmr4g7-1-2-3-csc1g5-csc2g7-x-10d:1.21220,(156937861|CAS-III|Ignicoccus_hos___6-x-x-
1:0.82794,70607609|CAS-I-D|Sulfolobus_aci___3-csc1g5-csc2g7-10d-R-6-x-2-4-1-x-x-x-x-x-x-x-x-x-
csm3g7:0.96376)0.956:0.32568)1.000:0.59773)0.890:0.10459)0.784:0.05200,(((((187250661|CAS-II|Elusimicrobium_min___9-1-
2:0.61891,325972002|CAS-II|Spirochaeta_Bud___2-1-9:0.74940)0.971:0.28775,((((386318631|CAS-II|Staphylococcus_pse___9-1-
2:0.79215,(15675042|CAS-II-A|Streptococcus_pyo___9-1-2-csn2:0.79682,(352684362|CAS-II-A|Acidaminococcus_int___9-1-2-
csn2:0.51095,(169823756|CAS-II-A|Finegoldia_mag___csa3-x-x-x-x-9-1-2-csn2-PDEDxK:0.61846,42525844|CAS-II-
```

A|Treponema_den___9-1-2-csn2:0.40539)0.759:0.12239)0.928:0.19604)0.863:0.16984)0.000:0.05417,(302336019|CAS-II-A|Olsenella_uli___csn2-2-1-9:0.70415,(90961086|CAS-II-A|Lactobacillus_sal___9-x-x-1-2-csn2:0.80559,(310286727|CAS-II-A|Bifidobacterium_bif___csn2-1-9:0.67816,406600270|CAS-II-A|Leuconostoc_gel___csn2-2-1-9:0.48323)0.903:0.16893)0.956:0.22166)0.823:0.11869)0.980:0.22310,(((269123825|CAS-II|Streptobacillus_mon___2-1-9:1.01742,(47458867|CAS-II|Mycoplasma_mob___1-9:0.80008,(71894593|CAS-II-A|Mycoplasma_syn___9-1-2-csn2-1:0.77706,193216853|CAS-II|Mycoplasma_art___1-2:0.64612)0.976:0.30122)0.975:0.30654)0.473:0.11491,(238924076|CAS-II|Eubacterium_rec___9-1-2:0.79548,116627543|CAS-II|Streptococcus_the___9-1-2:0.73103)0.704:0.20755)0.000:0.04759,57865886|CAS-III-A|Staphylococcus_epi___6-csm6-csm5g7-csm4g5-csm3g7-csm2SS-10-2-1:1.23206)0.549:0.13652)0.933:0.18660,((187736488|CAS-II|Akkermansia_muc___2-1-9:1.39758,319957207|CAS-II|Nitratifractor_sal___9-1-2:0.78599)0.825:0.24500,(((189485226|CAS-I-C|uncultured_Ter___3-5-8c-7-4-1-2-9-1-2:0.72350,34557789|CAS-II|Wolinella_suc___2-1-9:0.70129)0.951:0.22060,((15602991|CAS-II|Pasteurella_mul___2-1-9:0.67060,(222109284|CAS-II|Acidovorax_ebr___2-1-9:0.79135,(310780383|CAS-II|Ilyobacter_pol___2-1-9:0.66637,220930481|CAS-II|Clostridium_cel___2-1-9:0.45815)0.495:0.12122)0.235:0.10067)0.527:0.07214,(57238545|CAS-II|Campylobacter_jej___2-1:0.68707,291276264|CAS-II|Helicobacter_mus___2-1-9:0.53967)0.991:0.31535)0.696:0.07398)0.467:0.12246,((90425960|CAS-II|Rhodopseudomonas_pal___2-1-9:0.59972,209542526|CAS-II|Gluconacetobacter_dia___9-x-1-2:0.34389)0.999:0.30076,((60683388|CAS-II|Bacteroides_fra___2-1-9:0.67044,154250554|CAS-II|Parvibaculum_lav___2-1-9:0.59613)0.616:0.18474,(294086112|CAS-II|Puniceispirillum_mar___9-1-2:0.75736,121608212|CAS-II|Verminephrobacter_eis___9-1-2:0.66201)0.970:0.28280)0.927:0.14692)0.886:0.15928)0.523:0.07411)0.916:0.17687)0.434:0.13234)0.871:0.16414,(117929157|CAS-II|Acidothermus_cel___2-1-9:0.43620,(283456134|CAS-II|Bifidobacterium_den___2-1-9:0.77563,38232679|CAS-II|Corynebacterium_dip___9-1-2:0.38942)0.984:0.34995)0.999:0.58714)1.000:0.76830,(((50085564|CAS-I-F|Acinetobacter_ADP___3f-x-8f-5f-7f-6f-1:0.87259,(77166205|CAS-I-F|Nitrosococcus_oce___3f-1-8f-5f-7f-6f:0.70997,15602176|CAS-I-F|Pasteurella_mul___8f-5f-7f-6f-x-x-1-3f:0.74246)0.432:0.16689)0.999:0.75717,(38234755|CAS-I-E|Corynebacterium_dip___5-7-cse2SS-8e-6e-3-1-2:1.23531,((53804737|CAS-I-E|Methylococcus_cap___2-1-6e-5-7-cse2SS-8e-3:0.57557,209966592|CAS-I-E|Rhodospirillum_cen___2-1-6e-5-7-cse2SS-8e-3:0.74976)0.998:0.72881,(((68535718|CAS-I-E|Corynebacterium_jei___1-6e-5-7-cse2SS-8e-3:0.84511,86738746|CAS-I-E|Frankia_CcI___3-8e-cse2SS-7-5-6e-1-2:0.59793)0.909:0.30308,((89885834|CAS-I-E|Rhodoferax_fer___1-6e-csf3g5-csf2g7-x-csf1g8-DinG:1.24990,16130662|CAS-I-E|Escherichia_col___2-1-6e-5-7-cse2SS-8e-3:0.63487)0.991:0.50550,(51891801|CAS-I-E|Symbiobacterium_the___1-3-8e-cse2SS-7-5-6e:0.60107,21674788|CAS-I-E|Chlorobium_tep___3-8e-cse2SS-6e-7-5-1-2:1.27314)0.707:0.10574)0.656:0.10724)0.799:0.10534,29834079|CAS-I-E|Streptomyces_ave___1-6e-5-7-cse2SS-8e-3:0.75713)0.915:0.20711)0.936:0.38681)0.855:0.25733)0.999:0.98797,(124485266|unknown|Methanocorpusculum_lab___1:0.82627,((154150254|unknown|Methanoregula_boo___1:0.58226,((126178701|unknown|Methanoculleus_mar___1:0.07443,397780289|unknown|Methanoculleus_bou___1:0.16232)0.999:0.36213,307352568|unknown|Methanoplanus_pet___1:0.81980)0.633:0.12886)0.049:0.09156,219852754|unknown|Methanosphaerula_pal___1:0.54240)0.915:0.20557)1.000:0.95417)0.185:0.13326)0.547:0.10037,332661941|CAS-III-B|Haliscomenobacter_hyd___2-1-x-RT-cmr6g7-x-cmr4g7-cmr3g5-10:2.03433)0.792:0.07389)0.993:0.24809,((288947693|CAS-III-U|Allochromatium_vin___csx1-x-csx16-x-2-2-1:0.63642,431929780|CAS-III-B|Thioflavicoccus_mob___cmr4g7-cmr3g5-10-cmr1g7-csx1-x-x-x-2-2-1-RT:0.71938)0.990:0.42353,((292492856|CAS-III|Nitrosococcus_hal___csm4g5-csm5g7-x-csx1-x-x-csx1-csx1-x-2-1-x-2:0.61028,((206890539|CAS-III-A|Thermodesulfovibrio_yel___csm2SS-csm3g7-csm4g5-csm5g7-6-x-x-x-x-2-1-1-2:0.94103,(52425690|CAS-III-U|Mannheimia_suc___2-1-2-x-csx1-csx1-csx16-x-6:0.83608,30248142|CAS-III|Nitrosomonas_eur___2-1-2-csx1-x-csx1-csx16-6-x-csm5g7-csm4g5-csm3g7:0.51689)0.754:0.15286)0.875:0.16834,430760780|CAS-III-B|Thioalkalivibrio_nit___cmr4g7-cmr5SS-cmr6g7-csx1-x-csx1-x-x-6-2-1-1-2:0.57693)0.171:0.10187)0.479:0.12125,121997419|CAS-III-B|Halorhodospira_hal___1-2-1-2-csx1-csx16-csx1-cmr1g7-10-cmr3g5-cmr4g7-cmr5SS-cmr6g7:0.80291)0.973:0.28121)0.994:0.39809)0.748:0.06442)0.802:0.09835,((88602029|CAS-III-A|Methanospirillum_hun___csx1-x-x-x-x-csx1-csx1-x-x-2-1-x-x-x-csm3g7-csm3g7-x-x-csm3g7-csm3g7-10:1.27757,83592490|CAS-III-A|Rhodospirillum_rub___csm3g7-csm3g7-csm3g7-10-6-2-1:0.90172)0.839:0.25935,((294676823|CAS-III-A|Rhodobacter_cap___RT-x-x-x-x-x-10-csm3g7-csm3g7-2-1+RT-RT-6:0.58881,(374998939|CAS-III-B|Azospirillum_lip___2-1+RT-x-csm3g7-x-x-x-cmr1g7-10-cmr3g5-cmr4g7-cmr5SS:0.39128,(389875622|CAS-III-A|Tistrella_mob___10-cmr4g7-csm3g7-csm3g7-6-x-1+RT-2:0.63303,(302343124|CAS-III-A|Desulfarculus_baa___csm3g7-csm3g7-csm3g7-10-1+RT-2-6:0.64546,288957883|CAS-III-A|Azospirillum_B51___csm3g7-csm3g7-csm3g7-6-2-1+RT:0.24663)0.814:0.09915)0.807:0.11976)0.877:0.14374)1.000:0.61254,20094748|CAS-III-A|Methanopyrus_kan___2-x-1-x-10-csm2SS-csm3g7-csm4g5-csx1-csm5g7:1.51613)0.666:0.10913)0.922:0.15239)0.946:0.10901)0.045:0.02893)0.725:0.07257)0.132:0.03771)0.879:0.12499);

Cas2 Tree:

((((((((269123824_Streptobacillus_mon____CAS-II_2-1-9:0.65435000,(90425959_Rhodopseudomonas_pal____CAS-II_2-1-9:0.22101000,209542527_Gluconacetobacter_dia____CAS-II_9-x-1-2:0.17402000)0.94:0.30411000)0.77:0.09933000,(((222109283_Acidovorax_ebr____CAS-II_2-1-9:0.42855000,310780382_Ilyobacter_pol____CAS-II_2-1-9:0.13912000)0.66:0.12085000,15602990_Pasteurella_mul____CAS-II_2-1-9:0.37916000)0.90:0.20861000,((291276263_Helicobacter_mus____CAS-II_2-1-9:0.31167000,57238544_Campylobacter_jej____CAS-II_2-1:0.28830000)0.97:0.34796000,189485227_uncultured_Ter____CAS-I-C_3-5-8c-7-4-1-2-9-1-2:0.13815000)0.81:0.07699000)0.63:0.13694000)0.85:0.07727000,(15675043_Streptococcus_pyo____CAS-II-A_9-1-2-

```
csn2:0.46351000,((374307736_Filifactor_alo____CAS-II-A_csn2-2-1-9:0.28271000,302336018_Olsenella_uli____CAS-II-A_csn2-2-1-
9:0.43920000)0.76:0.09571000,(42525845_Treponema_den____CAS-II-A_9-1-2-csn2:0.39033000,169823757_Finegoldia_mag____CAS-II-
A_csa3-x-x-x-x-9-1-2-csn2-
PDEDxK:0.34299000)0.77:0.11939000)0.84:0.10746000)0.96:0.38688000)0.79:0.08033000,220930480_Clostridium_cel____CAS-II_2-1-
9:0.29802000)0.88:0.09183000,((319957208_Nitratifractor_sal____CAS-II_9-1-2:0.70234000,187736487_Akkermansia_muc____CAS-
II_2-1-9:0.70126000)0.52:0.04877000,(34557788_Wolinella_suc____CAS-II_2-1-9:0.64671000,((116627544_Streptococcus_the____CAS-
II_9-1-2:0.56326000,238924077_Eubacterium_rec____CAS-II_9-1-2:0.13222000)0.94:0.21511000,(90961087_Lactobacillus_sal____CAS-
II-A_9-x-x-1-2-csn2:0.11325000,(406600269_Leuconostoc_gel____CAS-II-A_csn2-2-1-
9:0.31623000,331702226_Lactobacillus_buc____CAS-II_2-1-
9:0.07283000)0.66:0.08097000)0.97:0.21453000)0.73:0.09770000)0.09:0.07068000)0.16:0.13698000)0.85:0.12415000,(60683387_Bacteroide
s_fra____CAS-II_2-1-9:0.32820000,154250553_Parvibaculum_lav____CAS-II_2-1-
9:0.27515000)0.72:0.07487000)0.81:0.13684000,((294086113_Puniceispirillum_mar____CAS-II_9-1-
2:0.53000000,(189440766_Bifidobacterium_lon____CAS-II_9-1-2:0.31549000,(117929156_Acidothermus_cel____CAS-II_2-1-
9:1.13966000,38232680_Corynebacterium_dip____CAS-II_9-1-
2:0.36119000)0.34:0.23863000)0.96:0.35689000)0.60:0.03762000,(121608213_Verminephrobacter_eis____CAS-II_9-1-
2:0.53910000,(57865885_Staphylococcus_epi____CAS-III-A_6-csm6-csm5g7-csm4g5-csm3g7-csm2SS-10-2-
1:1.29537000,(325972001_Spirochaeta_Bud____CAS-II_2-1-9:0.39221000,187250662_Elusimicrobium_min____CAS-II_9-1-
2:0.87228000)0.61:0.25072000)0.68:0.09900000)0.79:0.12183000)0.88:0.15899000)0.82:0.22250000,(193216854_Mycoplasma_art____CAS
-II_1-2:0.60540000,71894594_Mycoplasma_syn____CAS-II-A_9-1-2-csn2-
1:0.22111000)0.93:0.55087000)0.97:0.66890000,((148642084_Methanobrevibacter_smi____unknown_1-
2:1.27532000,(((((((283778925_Pirellula_sta____CAS-III-A_1+RT-2-x-10-csm3g7-
csm3g7:0.89402000,319790502_Thermovibrio_amm____unknown_2-2-2-
1:0.93403000)0.80:0.21342000,((118497968_Francisella_nov____unknown_2OG-x-x-x-x-csa3-2-1-
4:1.02155000,((337288663_Thermodesulfobacterium_OPB____CAS-III-A_10-csm2SS-csm3g7-csm4g5-csm5g7-6-csx1-x-2-1-
2:1.15971000,337288661_Thermodesulfobacterium_OPB____CAS-III-A_10-csm2SS-csm3g7-csm4g5-csm5g7-6-csx1-x-2-1-
2:0.79480000)0.77:0.26073000,(319790501_Thermovibrio_amm____unknown_2-2-2-
1:0.48746000,319790503_Thermovibrio_amm____unknown_2-2-2-
1:0.90092000)0.74:0.15026000)0.72:0.06215000)0.82:0.16445000,((15609953_Mycobacterium_tub____CAS-III-A_2-1-csm6-csm5g7-
csm4g5-csm3g7-csm2SS-10-6:0.51178000,((296133513_Thermincola_pot____CAS-III-A_2-1-csm6-6-csm5g7-csm4g5-csm3g7-csm2SS-
10:0.29366000,114567263_Syntrophomonas_wol____CAS-III-A_csm6-x-2-1-x-6-csm5g7-csm4g5-csm3g7-csm2SS-
10:0.23322000)0.89:0.35153000,348026606_Megasphaera_els____CAS-III-A_1-2-10-csm2SS-csm3g7-csm4g5-csm5g7-6-
csm6:0.67555000)0.78:0.13760000)0.90:0.31648000,116627765_Streptococcus_the____CAS-III-A_1-2-6-csm2SS-csm3g7-csm4g5-csm5g7-
x-csm6:0.60642000)0.88:0.30205000)0.75:0.03587000)0.86:0.11603000,(((((269798875_Veillonella_par____CAS-III-A_6-csm3g7-csx19-
csm3g7-5g10-csm3g7-10-csx20-1-2-x-csx1-csx1:0.77156000,88602028_Methanospirillum_hun____CAS-III-A_csx1-x-x-x-x-csx1-csx1-x-x-
2-1-x-x-x-csm3g7-csm3g7-x-x-csm3g7-csm3g7:0.85308000)0.87:0.52835000,20094746_Methanopyrus_kan____CAS-III-A_2-x-1-x-10-
csm2SS-csm3g7-csm4g5-csx1-csm5g7:1.62011000)0.51:0.19062000,(((37677205_Vibrio_vul____CAS-III_6-csm3g7-x-csm3g7-5g10-
cmr4g7-10-x-x-1+RT-2-2-csx1-csx16-csx16-csm6:0.72100000,83592489_Rhodospirillum_rub____CAS-III-A_csm3g7-csm3g7-csm3g7-10-
6-2-1:0.71732000)0.78:0.22771000,436839746_Desulfovibrio_hyd____CAS-III-B_1+RT-2-x-x-cmr6g7-cmr5SS-cmr4g7-cmr3g5-10-
cmr1g7:0.74051000)0.17:0.14583000,317051220_Desulfurispirillum_ind____CAS-III-A_csm2SS-csm3g7-csm4g5-csm5g7-x-x-csx1-csx20-
x-1-2-x-2:1.75303000:0.05677000,(((386750120_Helicobacter_cet____CAS-III_csm5g7-csx1-x-x-x-x-x-x-x-2-
1:1.01565000,((((116748795_Syntrophobacter_fum____CAS-III-A_csm4g5-csm3g7-csm2SS-10-6-csx20-csx1-x-Prim2-1-
2:0.31037000,218961168_Cloacamonas_aci____unknown_csa3-csx1+x-x-x-x-x-x-Prim2-x-1-
2:0.54685000)0.85:0.15969000,(328953423_Desulfobacca_ace____CAS-III_csm6-x-csx20-csx1-x-x-x-Prim2+x-x-1-
2:0.41723000,(313673557_Calditerrivibrio_nit____unknown_2-Prim2-1:0.25389000,332296541_Thermodesulfobium_nar____CAS-III-
A_4-x-2-1-Prim2-csm5g7-csm4g5-csm3g7-csm2SS-10-csx1-csx1-
6:0.43566000)0.40:0.12454000)0.15:0.04882000)0.92:0.18403000,(383762166_Caldilinea_aer____CAS-III-A_2-1-6-csx15-x-csx1-csm5g7-
csm4g5-csm3g7-csm2SS-10:0.51565000,((156741962_Roseiflexus_cas____CAS-III-A_10-csm2SS-csm3g7-csm4g5-csm5g7-csx1-6-1-
2:0.24695000,320161860_Anaerolinea_the____CAS-III-A_10-csm2SS-csm3g7-csm4g5-csm5g7-csx1-csx1-csx15-6-1-
2:0.53996000)0.80:0.22964000,328953001_Desulfobacca_ace____CAS-III_6-x-x-x-x-x-RT-x-x-1-
2:0.37791000)0.38:0.04439000)0.63:0.03747000)0.84:0.09147000,(159898756_Herpetosiphon_aur____CAS-I-D_10d-csc2g7-csc1g5-3-6-4-
1-2:0.79678000,(((134299483_Desulfotomaculum_red____CAS-III-A_1+4-2-10-csm2SS-csm3g7-csm4g5-csx1-
6:0.63110000,119357847_Chlorobium_pha____CAS-III-A_csx1-csm5g7-csm4g5-csm3g7-csm2SS-10-x-csx15-csx1-1+RT-2-x-x-x-x-2-1-
6:0.38435000)0.57:0.12431000,(76802277_Natronomonas_pha____CAS-I-D_R-6-10d-csc2g7-csc1g5-3-4-1-
2:1.07211000,302391603_Acetohalobium_ara____CAS-I-D_10d-csc2g7-csc1g5-3-4-1-2-
6:0.31311000)0.70:0.11597000)0.69:0.09059000,75910375_Anabaena_var____CAS-I-D_10d-csc2g7-csc1g5-2OG-6-4-1-
2:0.30845000)0.71:0.06011000)0.92:0.23922000)0.82:0.08356000)0.60:0.01366000,(337286707_Thermodesulfatator_ind____CAS-III-
B_csx1-cmr6g7-cmr5SS-cmr4g7-x-x-cmr3g5-10-cmr1g7-x-2-x-1-2-6:0.54544000,(119357852_Chlorobium_pha____CAS-III-A_csx1-
csm5g7-csm4g5-csm3g7-csm2SS-10-x-csx15-csx1-1+RT-2-x-x-x-x-2-1-6:0.29854000,(((((317051218_Desulfurispirillum_ind____CAS-III-
A_csm2SS-csm3g7-csm4g5-csm5g7-x-x-csx1-csx20-x-1-2-x-2:0.47733000,85858453_Syntrophus_aci____CAS-III_cmr6g7-csm6-csx1-x-x-
x-x-x-x-1-2-2:0.66471000)0.86:0.26707000,85858454_Syntrophus_aci____CAS-III_cmr6g7-csm6-csx1-x-x-x-x-x-x-1-2-
2:0.90326000)0.42:0.18606000,(328949008_Treponema_suc____CAS-III-A_csx1-x-x-2-2-1-x-x-x-csx20-csx1-x-csm5g7-csm4g5-
```

```
csm3g7:0.26185000,333996013_Treponema_azo____CAS-III-B_csx1-cmr6g7-cmr5SS-cmr4g7-cmr3g5-10-cmr1g7-6-x-1-2-x-
csx1:0.75344000)0.10:0.23727000)0.72:0.19590000,((34541576_Porphyromonas_gin____CAS-III-B_2-1+RT-cmr6g7-x-cmr4g7-cmr3g5-
10:0.69930000,336116790_Microlunatus_pho____CAS-III-A_10-csm3g7-cmr3g5-csm3g7-x-x-x-6-1+RT-2-x-10-csm3g7-cmr3g5-
csm3g7:0.98282000)0.85:0.43712000,268318377_Rhodothermus_mar____CAS-III-U_csx1-csx20-x-x-csx1-csx1-6-x-2-
1+4:0.44664000)0.92:0.50562000)0.36:0.17225000,(((((((30248141_Nitrosomonas_eur____CAS-III-A_2-1-2-csx1-x-csx1-csx16-6-x-
csm5g7-csm4g5-csm3g7-csm2SS:0.44781000,52425689_Mannheimia_suc____CAS-III-U_2-1-2-x-csx1-csx1-csx16-x-
6:0.59489000)0.90:0.32569000,430760782_Thioalkalivibrio_nit____CAS-III-B_cmr3g5-cmr4g7-cmr5SS-cmr6g7-csx1-x-csx1-x-x-6-2-1-1-
2:0.51091000)0.80:0.19099000,(431929779_Thioflavicoccus_mob____CAS-III-B_cmr6g7-cmr5SS-cmr4g7-cmr3g5-10-cmr1g7-csx1-x-x-x-
2-2-1-RT:0.47668000,288947692_Allochromatium_vin____CAS-III-U_cmr6g7-x-x-x-x-x-csx1-x-csx16-x-2-2-
1:0.83407000)0.89:0.32391000)0.83:0.20682000,292492858_Nitrosococcus_hal____CAS-III-A_csm3g7-csm4g5-csm5g7-x-csx1-x-x-csx1-
csx1-x-2-1-x-2:0.56529000)0.85:0.21514000,(206890293_Thermodesulfovibrio_yel____CAS-III-A_10-csm2SS-csm3g7-csm4g5-csm5g7-6-
x-x-x-x-2-1-1-2:0.54995000,121997420_Halorhodospira_hal____CAS-III-B_1-2-1-2-csx1-csx16-csx1-cmr1g7-10-cmr3g5-cmr4g7-cmr5SS-
cmr6g7-6:0.80448000)0.82:0.18626000)0.92:0.35773000,((268318440_Rhodothermus_mar____CAS-I-C_2-
1+4:0.65320000,((320102062_Isosphaera_pal____CAS-I-Cu_1+4-2-csb2:0.62829000,(134101618_Saccharopolyspora_ery____CAS-III-
A_1+4-2-x-x-x-x-csm3g7-cmr3g5-csm3g7-10:0.50543000,((83591518_Rhodospirillum_rub____CAS-I-E_2-1-6e-5-7-cse2SS-8e-3-x-x-2-
1+4-csb2-x-x-csm3g7-csm3g7-csm3g7-10:0.49671000,39995169_Geobacter_sul____CAS-I-Cu_3-8-csb1-csb2-x-x-1+4-
2:0.23073000)0.92:0.28853000,312114616_Rhodomicrobium_van____CAS-III-A_RT-2-1-csm3g7-csx19-csm3g7-csm3g7-10-x-csm6-
csx1:0.80211000)0.90:0.26751000)0.34:0.06159000)0.41:0.13074000,378719040_Gordonia_pol____CAS-I-Cu_2-1+4-csb2-csb1-8-
3:0.70293000)0.27:0.09076000)0.90:0.28490000,((162457463_Sorangium_cel____CAS-I-C_6-3-8c-7-5-1-
2:0.58650000,347756304_Chloracidobacterium_the____CAS-I-C_6-3-8c-7-5-1+4-
2:0.47759000)0.36:0.08796000,24215881_Leptospira_int____CAS-I-C_1+4-2-x-x-6-3-8c-7-
5:0.48595000)0.92:0.30625000)0.86:0.18671000)0.74:0.07975000,337286710_Thermodesulfatator_ind____CAS-III-B_csx1-cmr6g7-
cmr5SS-cmr4g7-x-x-cmr3g5-10-cmr1g7-x-2-x-1-2-
6:0.72111000)0.83:0.20250000)0.69:0.06791000)0.69:0.05198000)0.91:0.17860000)0.78:0.06280000,((((427708212_Nostoc_PCC____CAS-
III-A_csm3g7-csx19-csm3g7-x-x-x-6-x-x-x-2-1-x-x-1+RT:0.48505000,(((433631662_Mycobacterium_can____CAS-I-C_2-1-4-7-8c-5-
3:0.46431000,86742033_Frankia_CcI____CAS-I-C_3-5-8c-7-4-1-2:0.59968000)0.25:0.10391000,((189485224_uncultured_Ter____CAS-I-
C_3-5-8c-7-4-1-2-9-1-2:0.23678000,108759816_Myxococcus_xan____CAS-I-C_2-1-4-7-8c-5-
3HD:0.66151000)0.76:0.28677000,15612905_Bacillus_hal____CAS-I-C_cmr5SS-cmr6g7-x-x-x-3-5-8c-7-4-1-
2:0.12115000)0.98:0.52307000)0.86:0.20392000,261415949_Fibrobacter_suc____CAS-III-A_1-2-x-x-csx1-10-csm3g7-x-csm3g7-
csx20:0.98191000)0.83:0.13763000)0.78:0.05780000,((297622674_Truepera_rad____CAS-III-U_2-1-csx1-csx1-
6:0.43915000,344996858_Caldicellulosiruptor_lac____CAS-I-D_2-1-4-6-3-csc1g5-csc2g7-
10d:0.72375000)0.71:0.35642000,(17227878_Nostoc_PCC____CAS-III-B_csx18-1-2:0.64378000,326793968_Marinomonas_med____CAS-
III-B_2-6-x-x-cmr6g7-cmr5SS-cmr4g7-cmr3g5-10-
cmr1g7:0.64153000)0.70:0.19291000)0.62:0.01818000)0.85:0.06567000,((38505761_Synechocystis_PCC____CAS-III-B_cmr6g7-cmr5SS-
cmr4g7-x-cmr3g5-10-csx18-1-2:0.33189000,(((389875623_Tistrella_mob____CAS-III-A_10-cmr4g7-csm3g7-csm3g7-6-x-1+RT-
2:0.59765000,((302343125_Desulfarculus_baa____CAS-III-A_csm3g7-csm3g7-csm3g7-10-1+RT-2-
6:0.43962000,294676822_Rhodobacter_cap____CAS-III-A_RT-x-x-x-x-x-10-csm3g7-csm3g7-2-1+RT-RT-
6:0.18051000)0.66:0.10955000,288957882_Azospirillum_B51____CAS-III-A_csm3g7-csm3g7-csm3g7-6-2-
1+RT:0.07997000)0.13:0.11882000)0.87:0.18367000,374998938_Azospirillum_lip____CAS-III-B_2-1+RT-x-csm3g7-x-x-x-cmr1g7-10-
cmr3g5-cmr4g7:0.10149000)0.95:0.45546000,220934971_Thioalkalivibrio_sul____CAS-III-B_10-cmr3g5-cmr4g7-cmr5SS-cmr6g7-csx1-
x-csx1-x-1-2-x-csx1:0.86008000)0.88:0.31193000)0.84:0.21401000,((148657119_Roseiflexus_RS____CAS-III-U_2-x-1-RT-x-x-x-x-csx15-
csx1:0.64747000,159898908_Herpetosiphon_aur____unknown_RT-1-
2:0.45464000)0.67:0.06560000,((38505685_Synechocystis_PCC____CAS-I-D_3-10d-csc2g7-csc1g5-6-4-1-
2:0.70582000,46255266_Thermus_the____CAS-I-A_2-1-4-x-8a-7-5-
6:0.80808000)0.34:0.16455000,218442810_Cyanothece_PCC____CAS-III-A_csm6-x-x-RT-1-
2:0.60968000)0.55:0.12524000)0.88:0.21577000)0.69:0.03635000)0.73:0.03037000,37677206_Vibrio_vul____CAS-III_6-csm3g7-x-csm3g7-
5g10-cmr4g7-10-x-x-1+RT-2-2-csx1-csx16-csx16-
csm6:0.81769000)0.32:0.05358000)0.72:0.05232000)0.82:0.11565000)0.15:0.12021000,((((30248143_Nitrosomonas_eur____CAS-III-A_2-
1-2-csx1-x-csx1-csx16-6-x-csm5g7-csm4g5-csm3g7-csm2SS:0.47226000,206889468_Thermodesulfovibrio_yel____CAS-III-A_10-csm2SS-
csm3g7-csm4g5-csm5g7-6-x-x-x-x-2-1-1-2:0.63151000)0.20:0.09160000,((430760779_Thioalkalivibrio_nit____CAS-III-B_cmr3g5-
cmr4g7-cmr5SS-cmr6g7-csx1-x-csx1-x-x-6-2-1-1-2:0.42728000,121997422_Halorhodospira_hal____CAS-III-B_1-2-1-2-csx1-csx16-csx1-
cmr1g7-10-cmr3g5-cmr4g7-cmr5SS-cmr6g7-6:0.48149000)0.23:0.15776000,292492855_Nitrosococcus_hal____CAS-III-A_csm3g7-
csm4g5-csm5g7-x-csx1-x-x-csx1-csx1-x-2-1-x-2:0.23880000)0.94:0.26331000)0.81:0.16973000,52425691_Mannheimia_suc____CAS-III-
U_2-1-2-x-csx1-csx1-csx16-x-6:1.01735000)0.74:0.20008000,(288947691_Allochromatium_vin____CAS-III-U_cmr6g7-x-x-x-x-x-csx1-x-
csx16-x-2-2-1:0.79768000,431929778_Thioflavicoccus_mob____CAS-III-B_cmr6g7-cmr5SS-cmr4g7-cmr3g5-10-cmr1g7-csx1-x-x-x-2-2-1-
RT:0.34180000)0.97:0.82237000)0.89:0.58176000)0.50:0.14925000,328949007_Treponema_suc____CAS-III-A_csx1-x-x-2-2-1-x-x-x-
csx20-csx1-x-csm5g7-csm4g5-csm3g7:0.94210000)0.64:0.06485000)0.83:0.09917000,(((((385809791_Ignavibacterium_alb____CAS-III-
U_csx1-csx20-csx1-csx1-x-x-x-x-1-2:0.40745000,150021525_Thermosipho_mel____CAS-III-B_2-1-2-x-x-x-csm6-cmr1g7-10-cmr3g5-
cmr4g7-cmr5SS-cmr6g7:0.72705000)0.35:0.23605000,268317184_Rhodothermus_mar____CAS-III-A_2-1-x-csx1-csm5g7-csm4g5-
csm3g7-csm2SS-10-4-3:0.47741000)0.76:0.10019000,((((289596693_Aciduliprofundum_boo____CAS-III-A_2-1-4-6-csx1-csx1-csm5g7-
csm4g5-csm3g7-csm2SS-10:0.27366000,33359262_Pyrococcus_hor____CAS-I-B_csm2SS-csm3g7-csm4g5-csm5g7-csm6-8b-7-5-x-1-2-4-3-
```

```
csx1:0.59521000)0.88:0.14079000,(33359542_Pyrococcus_fur____CAS-I-B_2-1-4-3-5-7-8b-cmr6g7-cmr5SS-cmr4g7-
csm6:0.30594000,(((41614812_Nanoarchaeum_equ____CAS-I-B_2-1-x-7-5-3-4:0.61326000,(48477074_Picrophilus_tor____CAS-I-D_2-1-
4-3-csc1g5-csc2g7-10d-6:0.64411000,154174506_Campylobacter_cur____CAS-I-B_1-2-4-3-5-7-8-
6:0.68988000)0.77:0.19975000)0.87:0.23894000,13540936_Thermoplasma_vol____CAS-III-A_2-1-csx1-csm5g7-csm4g5-csm3g7-csm2SS-
10-6-csm6:0.34232000)0.84:0.15228000,(84489237_Methanosphaera_sta____CAS-I-B_6-8-7-5-3-4-1-
2:0.35892000,150021527_Thermosipho_mel____CAS-III-B_2-1-2-x-x-x-csm6-cmr1g7-10-cmr3g5-cmr4g7-cmr5SS-
cmr6g7:0.44479000)0.80:0.08895000)0.93:0.15361000)0.11:0.00015000)0.85:0.11140000,(111224948_Frankia_aln____CAS-I-B_6-x-7-3+5-
4-1-2:0.86083000,55376279_Haloarcula_mar____CAS-I-B_6-8-7-5-3-4-1-
2:0.50539000)0.83:0.21709000)0.72:0.03878000,389844442_Mesotoga_pri____CAS-I-B_csm6-csm6+x-x-cmr6g7-cmr5SS-cmr4g7-
cmr3g5-10-cmr1g7-csm6-2-x-x-1-4-3-5-7:0.45847000)0.82:0.17195000)0.68:0.19571000,390962115_Thermococcus_CL1____CAS-III-
A_csx1-x-6-10-csm2SS-csm3g7-csm4g5-csm5g7-R2-csx3-2-1:0.80322000)0.79:0.16255000,292492939_Nitrosococcus_hal____CAS-I-D_6-
10d-csc2g7-x-3-4-1-2:1.14074000)0.86:0.32832000)0.90:0.17979000,(((408403001_Nitrososphaera_gar____CAS-I_4-1-2-4-x-x-x-3-5-7-x-
6:0.58554000,(18311771_Pyrobaculum_aer____CAS-I-A_6-4-2-1-4-csa3-x-8a2-3HD-3-5-7-
csa5SS:1.14033000,((126465228_Staphylothermus_mar____CAS-I-A_csm3g7-csm4g5-csm5g7-10-x-x-1-4-2-4-6-R-csa5SS-7-5-csa8a2-
3:0.26154000,170290048_Korarchaeum_cry____CAS-I-A_1-4-2-4-R-csa5SS-7-5-csa8a2-x-3-
3HD:0.40242000)0.86:0.17885000,(330509016_Methanosaeta_con____CAS-I-A_4-4-2-1-3-x-x-x-x-
5:0.80370000,((148656005_Roseiflexus_RS____CAS-III-U_10-csx1-csx1-csx3-8b-7-5-3-6-1-2-
4:1.02105000,386854796_Deinococcus_gob____CAS-I-A_4-2-1-5-7-8a2-3-
6:1.37160000)0.38:0.04011000,167628643_Heliobacterium_mod____CAS-I-A_4-4-2-1-x-6-5-7-x-
3:0.15876000)0.88:0.14551000)0.87:0.17843000)0.61:0.10408000)0.32:0.17182000)0.69:0.16217000,((((392395715_Flexibacter_lit____unkn
own_2-x-1:0.44069000,379728402_Saprospira_gra____unknown_2-x-1:0.77932000)0.43:0.26216000,327314605_Prevotella_den____CAS-
III_10-cmr6g7-5g10-csm3g7-x-csm3g7-csx1-csx1-1-Prim-2:0.60272000)0.91:0.33715000,15898244_Sulfolobus_sol____CAS-I-A_csx1-
csa5SS-7-5-csaXg8-3-3HD-2-1-6:0.58386000)0.60:0.05277000,332668018_Haliscomenobacter_hyd____unknown_RT-1-PDEDxK-
2:0.97255000)0.80:0.16408000)0.66:0.05231000,(((((284040852_Spirosoma_lin____CAS-I-B_8-7-5-3-6-1-2-
4:0.51641000,218961576_Cloacamonas_aci____CAS-I-B_4-2-1-3HD-5-7-8-
6:0.33190000)0.91:0.29536000,21226660_Methanosarcina_maz____CAS-I-B_4-2-1-6-3-5-7-8-
csa3:0.27037000)0.00:0.11790000,150401498_Methanococcus_aeo____CAS-III-A_6-3-5-7-8-1-2-4-x-csx1-csm5g7-csm4g5-csm3g7-
csm2SS-10:0.70927000)0.64:0.07037000,257125861_Leptotrichia_buc____CAS-III_csm3g7-6-x-x-x-x-x-x-4-2-1-x-x-csx1-
csx20:0.64294000)0.94:0.33287000,((((34557934_Wolinella_suc____CAS-II-B_9-1-2-4:0.68476000,54296140_Legionella_pne____CAS-II-
B_9-1-2-4:0.49844000)0.85:0.41634000,118497354_Francisella_nov____CAS-II-B_9-1-2-
4:0.44155000)0.97:0.63521000,332661940_Haliscomenobacter_hyd____CAS-III-B_2-1-x-RT-cmr6g7-x-cmr4g7-cmr3g5-
10:0.87623000)0.60:0.22102000,150398873_Methanococcus_van____CAS-III-B_10-cmr3g5-cmr4g7-cmr5SS-cmr6g7-x-x-6-1-
2:0.74887000)0.46:0.29206000)0.64:0.14281000)0.68:0.06174000)0.81:0.17631000,(((385805116_Fervidicoccus_fon____CAS-I-A_R-3HD-
3-8a2-5-7-x-x-4-1-2-4-R-R-x-x-csm6-cmr6g7-cmr1g7-cmr5SS-cmr4g7:0.87975000,((((119720227_Thermofilum_pen____CAS-III-
B_cmr3g5-10-cmr6g7-cmr1g7-cmr5SS-cmr4g7-1-2-3-csc1g5-csc2g7-x-10d:0.29880000,70607607_Sulfolobus_aci____CAS-I-D_3-csc1g5-
csc2g7-10d-R-6-x-2-4-1:0.72201000)0.98:0.47456000,(347523052_Pyrolobus_fum____CAS-I-A_4-x-x-x-x-2-1-4-R-x-x-3HD-3-8a2-5-
7:0.53973000,((305663329_Ignisphaera_agg____CAS-I-A_R-x-7-5-3-3HD-x-6-4-2-1-4:0.18600000,307595806_Vulcanisaeta_dis____CAS-
I-A_6-3HD-3-x-5-7-x-R-x-4-2-1-4:0.43362000)0.50:0.11061000,(118431379_Aeropyrum_per____CAS-I-A_6-8a2-3HD-3-5-7-csa5SS-R-4-
2-1-4:0.27746000,156937922_Ignicoccus_hos____CAS-I-A_4-1-2-4-3HD-3-x-5-7-csa5SS-
R:0.70425000)0.69:0.22932000)0.44:0.18671000)0.91:0.29882000)0.58:0.04873000,124027532_Hyperthermus_but____CAS-I-A_3-3-
csaXg8-5-7-csa5SS-R-x-x-4-2-4-1-x-x-x-R:0.62624000)0.91:0.28271000,302347807_Acidilobus_sac____CAS-I-A_6-x-4-2-1-4-cmr3g5-10-
cmr6g7-csm3g7-cmr5SS-cmr4g7-csm6-x-csm6:0.54554000)0.84:0.27918000)0.20:0.13179000,15922976_Sulfolobus_tok____CAS-I-A_R-4-
1-2-4-x-x-csa5SS-7-5-3-3HD-csaXg8:0.41222000)0.25:0.18822000,218884092_Desulfurococcus_kam____CAS-I-A_6-3HD-3-csaXg8-5-7-
csa5SS-R-4-2-1-4-R-x-cmr3g5-10-cmr6g7-cmr1g7-cmr5SS-
cmr4g7:0.81926000)0.92:0.64207000)0.74:0.09009000,(288560432_Methanobrevibacter_rum____CAS-III-A_1-x-2-6-csx1-csm5g7-
csm4g5-csm3g7-csm2SS-10:0.92672000,(11499460_Archaeoglobus_ful____CAS-I-A_cmr1g7-csa3-csa5SS-7-5-csa8a2-3-3HD-2-4-1-
4:0.35880000,383319614_Methanocella_con____CAS-I-A_4-1-4-2-csa3-x-x-x-x-x-7-x-x-
3:0.24133000)0.96:0.43829000)0.62:0.05416000)0.84:0.21065000)0.78:0.21994000,18311698_Pyrobaculum_aer____CAS-I-A_3HD-3-x-x-
x-6-csa3-x-4-2-1-4:1.11375000)0.67:0.07248000,((50085558_Acinetobacter_ADP____CAS-I-F_3f-x-8f-5f-7f-6f-
1:0.49611000,(77166204_Nitrosococcus_oce____CAS-I-F_3f-1-8f-5f-7f-6f:0.58084000,15602177_Pasteurella_mul____CAS-I-F_8f-5f-7f-
6f-x-x-1-3f:0.71605000)0.71:0.25381000)1.00:2.48456000,(((90111482_Escherichia_col____CAS-I-E_2-1-6e-5-7-cse2SS-8e-
3:0.19505000,77128445_Methylococcus_cap____CAS-I-E_2-1-6e-5-7-cse2SS-8e-
3:1.08019000)0.71:0.19962000,21674789_Chlorobium_tep____CAS-I-E_3-8e-cse2SS-6e-7-5-1-
2:1.33635000)0.81:0.25236000,(209966591_Rhodospirillum_cen____CAS-I-E_2-1-6e-5-7-cse2SS-8e-
3:1.03342000,(86738747_Frankia_CcI____CAS-I-E_3-8e-cse2SS-7-5-6e-1-2:0.29430000,38234756_Corynebacterium_dip____CAS-I-E_5-
7-cse2SS-8e-6e-3-1-2:1.16325000)0.85:0.31309000)0.72:0.13876000)0.81:1.00213000)0.76:0.43903000);
```

Cas4 Tree:

```
((170290047_Candidatus_Kor___CAS-I-A:0.32712000,126465227_Staphylothermus_mar___CAS-I-
A:0.33457000)0.99:0.59491000,(((408402999_Candidatus_Nit___CAS-I-A:0.46381000,(330509014_Methanosaeta_con___CAS-I-
A:0.57219000,(11499463_Archaeoglobus_ful___CAS-I-A:0.37486000,383319613_Methanocella_con___CAS-I-
A:0.37265000)0.99:0.34884000)0.65:0.14787000)0.92:0.15885000,167628645_Heliobacterium_mod___CAS-I-
A:0.38142000)0.98:0.30697000,(((385805114_Fervidicoccus_fon___CAS-I-A:0.87288000,(218884094_Desulfurococcus_kam___CAS-I-
A:0.32294000,389861364_Thermogladius_cel___CAS-I-
A:0.20434000)0.51:0.12318000)0.92:0.21199000,((15898287_Sulfolobus_sol___CAS-I-A:0.39623000,15922974_Sulfolobus_tok___CAS-I-
A:0.37770000)0.96:0.21913000,((18311700_Pyrobaculum_aer___CAS-I-A:0.81016000,347523054_Pyrolobus_fum___CAS-I-
A:0.43936000)0.20:0.06803000,18311773_Pyrobaculum_aer___CAS-I-
A:0.72574000)0.77:0.12556000)0.60:0.03377000)0.71:0.16243000,(((307595808_Vulcanisaeta_dis___CAS-I-
A:0.21634000,124027533_Hyperthermus_but___CAS-I-A:0.57735000)0.86:0.10869000,(118431381_Aeropyrum_per___CAS-I-
A:0.41512000,305663331_Ignisphaera_agg___CAS-I-A:0.21018000)0.90:0.10812000)0.92:0.19835000,156937920_Ignicoccus_hos___CAS-
I-A:0.23716000)1.00:0.68066000)0.23:0.03755000)0.78:0.09126000,((((111224946_Frankia_aln___CAS-I-
B:1.13081000,(((((84489235_Methanosphaera_sta___CAS-I-B:0.56652000,(14590111_Pyrococcus_hor___CAS-I-
B:0.27203000,55376277_Haloarcula_mar___CAS-I-B:0.96864000)0.88:0.20644000)0.80:0.17828000,(389844446_Mesotoga_pri___CAS-I-
B:0.83994000,41614818_Nanoarchaeum_equ___CAS-I-
B:0.35613000)0.97:0.39504000)0.09:0.06437000,(289596695_Aciduliprofundum_boo___CAS-III-
A:0.79798000,(268317193_Rhodothermus_mar___CAS-I-A:1.07613000,(154174706_Campylobacter_cur___CAS-I-
B:0.73180000,28211140_Clostridium_tet___CAS-I-
B:0.51649000)0.92:0.28583000)0.36:0.15318000)0.71:0.11810000)0.92:0.30744000,18977491_Pyrococcus_fur___CAS-I-
B:0.40489000)0.57:0.16692000,(48477076_Picrophilus_tor___CAS-I-D:0.57967000,292492937_Nitrosococcus_hal___CAS-I-
D:2.06459000)0.80:0.65148000)0.91:0.44422000)0.99:0.77832000,((((((((159898754_Herpetosiphon_aur___CAS-I-
D:0.42938000,(38505683_Synechocystis_sp___CAS-I-D:0.60781000,46255268_Thermus_the___CAS-I-
A:0.79859000)0.87:0.17858000)0.86:0.16668000,(((86742031_Frankia_CcI___CAS-I-
C:0.80886000,433631664_Mycobacterium_can___CAS-I-C:0.72421000)0.95:0.40152000,((189485222_uncultured_Ter___CAS-I-
C:0.69171000,56965355_Bacillus_hal___check:0.50337000)0.86:0.25583000,108757432_Myxococcus_xan___CAS-I-
C:0.55447000)0.02:0.08514000)0.98:0.37358000,(302391601_Acetohalobium_ara___CAS-I-D:0.50266000,17229059_Nostoc_sp___CAS-I-
D:0.66094000)0.17:0.14424000)0.77:0.10147000)0.49:0.10657000,344996860_Caldicellulosiruptor_lac___CAS-I-
D:1.17966000)0.54:0.03556000,134299482_Desulfotomaculum_red___CAS-III-
A:0.72503000)0.83:0.09076000,(((((39995168_Geobacter_sul___CAS-I-Cu:0.18797000,83591519_Rhodospirillum_rub___CAS-I-
Cu:0.67734000)0.91:0.14642000,320102061_Isosphaera_pal___CAS-I-Cu:0.29118000)0.59:0.09049000,378719041_Gordonia_pol___CAS-
I-Cu:0.76296000)0.86:0.12732000,268318441_Rhodothermus_mar___CAS-I-
C:0.35064000)0.86:0.14406000,134101617_Saccharopolyspora_ery___CAS-III-
A:0.92088000)0.97:0.30155000)0.76:0.10722000,(294828260_Leptospira_int___CAS-I-
C:0.34485000,347756303_Candidatus_Chl___CAS-I-
C:0.62198000)1.00:0.82703000)0.79:0.16940000,76802275_Natronomonas_pha___CAS-I-
D:1.19261000)0.84:0.23501000,167627879_Francisella_nov___check:2.53097000,219683006_Bifidobacterium_ani___CAS-I-
Cu:1.03586000:0.60263000)0.12:0.15433000)0.19:0.04486000,(332296539_Thermodesulfobium_nar___CAS-III-
A:2.10154000,(18311697_Pyrobaculum_aer___CAS-I-
A:1.34349000,15896973_Sulfolobus_sol___check:2.11406000)0.32:0.26671000)0.02:0.20274000)0.43:0.05955000,(((18311770_Pyrobaculu
m_aer___CAS-I-A:1.04920000,(124027531_Hyperthermus_but___CAS-I-A:1.08384000,((156937923_Ignicoccus_hos___CAS-I-
A:1.64041000,118431378_Aeropyrum_per___CAS-I-A:0.62763000)0.73:0.12480000,(305663328_Ignisphaera_agg___CAS-I-
A:0.64154000,307595805_Vulcanisaeta_dis___CAS-I-
A:0.34920000)0.96:0.35574000)0.14:0.05870000)0.24:0.21397000)0.80:0.16355000,(385805117_Fervidicoccus_fon___CAS-I-
A:1.48413000,(((218884091_Desulfurococcus_kam___CAS-I-A:0.89263000,389861366_Thermogladius_cel___CAS-I-
A:0.52075000)0.92:0.41883000,347523047_Pyrolobus_fum___CAS-I-A:1.69446000)0.60:0.04774000,(302347806_Acidilobus_sac___CAS-
I-A:1.01638000,(15898285_Sulfolobus_sol___CAS-I-A:0.78934000,15922977_Sulfolobus_tok___CAS-I-
A:0.61606000)0.94:0.36412000)0.85:0.25630000)0.84:0.12943000)0.90:0.22262000)0.75:0.19330000,(((((((284040853_Spirosoma_lin___CA
S-I-B:0.63299000,21226659_Methanosarcina_maz___CAS-I-B:0.36828000)0.01:0.10159000,(218961575_Candidatus_Clo___CAS-I-
B:0.52444000,150401499_Methanococcus_aeo___CAS-I-
B:0.46347000)0.16:0.01868000)0.89:0.24404000,257125860_Leptotrichia_buc___CAS-III-
A:1.09081000)0.97:0.41375000,(383319611_Methanocella_con___CAS-I-A:0.66135000,11499461_Archaeoglobus_ful___CAS-I-
A:0.51463000)0.97:0.34472000)0.89:0.14938000,(330509015_Methanosaeta_con___CAS-I-
A:0.50473000,408403002_Candidatus_Nit___CAS-I-B:0.92921000)0.89:0.21520000)0.72:0.06975000,(((148656006_Roseiflexus_sp-
___CAS-I-B:0.57761000,386854795_Deinococcus_gob___CAS-I-A:0.79274000)0.84:0.17086000,((54296141_Legionella_pne___CAS-II-
B:0.54352000,34557935_Wolinella_suc___CAS-II-B:0.66238000)0.97:0.37315000,118497355_Francisella_nov___CAS-II-
B:0.54478000)0.98:0.50460000)0.41:0.18627000,167628644_Heliobacterium_mod___CAS-I-
A:0.76117000)0.84:0.18847000)0.85:0.14480000,(170290049_Candidatus_Kor___CAS-I-
A:0.72124000,126465229_Staphylothermus_mar___CAS-I-
A:0.38343000)1.00:0.75378000)0.80:0.11271000)0.89:0.21765000)1.00:0.90558000);
```

Supplementary Figure S3. Multiple sequence alignments of type I and type III CRISPR repeats. See Figure 3. The alignments of short (24-26 nt) type I-A (**A**) and long (37 nt) type I and type III (**B**) CRISPR repeats used to create the logoplots of consensus repeat sequences are shown.

A

Pyrolobus fumarii 1A	GAAACAAC-CAAGAATGAATTGAAAG
Ignisphaera aggregans DSM 17230	GAAAC-TCCCAAAAGGGAGTAGAAAG
Hyperthermus butylicus DSM 5456	GAACA-ACTCAAAAGAGAATTGCAAG
Thermogladius cellulolyticus 1633	GAAGCAA-TCAAGAA-GAATTGAAAG
Desulfurococcus kamchatkensis 1221n	GAATA--CAATAGAAAGAATTGAAAG
Pyrobaculum aerophilum str. IM2	GAA-TCT-CAAAAAGAGGATTGAAAG
Fervidicoccus fontis Kam940	GAA-TCTGTT-AGATAGAATTGAAAG
Sulfolobus tokodaii str. 7	GATAA-TCCTTAATG-GAATTGAAAG
Sulfolobus solfataricus P2	GATAA-TCTCTTAT-AGAATTGAAAG
Vulcanisaeta distributa DSM 14429	GATAT-TCTCTAAAGAGAATAGAAGG
Sulfolobus solfataricus P2	GATTAATCCCAAAAG-GAATTGAAAG
Aeropyrum pernix K1	GCATA-TCCCTAAAGGGAATAGAAAG
Ignicoccus hospitalis KIN4/I	GTGAAGTCTCAAAAGAGGATAGAAAG
Acidilobus saccharovorans 345-15	GTTGCAACACCATCTCGTGGTTTCC-
Pyrobaculum aerophilum str. IM2	GTTTCAACTATCTTTTGATTTCTG-G

B

Desulfobacca acetoxidans DSM 11109	GTCGCAATCCCTTCGTAATCAGGTCAATGAAGCAGAG
Porphyromonas gingivalis W83	GTCTTAATAGCCTTACGGACTGTGTATGTATAGTGAG
Helicobacter cetorum MIT 00-7128	GTTTCAATCCCCTTGAACGGGTCAATCTGTAATAAAT
Heliobacterium modesticaldum Ice1	GTTTCACTGGATGACCCGCTATGTAGGGGATTGAAAG
Roseiflexus sp. RS-1	GTTTCAGTACTCTTCAACGAGTCGAAATTGCCGAAAT
Deinococcus gobiensis I-0	GTTTCAGTCCCCTATCGCCCGGGGAGTACGTTGCGGC

Supplementary Figure S4. Multiple alignment of Cas9 sequences divided into 5 groups according to tree topology and operon organization (Please see Supplementary Materials and Methods for details); Sequences are denoted by Genbank identifiers and complete organism name on the first page and further by abbreviated organism name. Color bars on the left group sequences according to the Cas9 tree topology (compare with Figure 4). *S. pyogenes* and *A. naeslundii* Cas9 strucutres with domain architecture are mapped on the alignment. r - RuvC motifs, l - loops, R - Arginine-rich motif, N - HNH domain, h - helical lobe, 1 - REC1 recognition lobe domain, 2 - REC2 recognition lobe domain, P - PAM recognition motifs. Functional and conserved amino acids are highlighted: yellow - catalytic, green - RNA-binding, red - target DNA binding, blue - conserved or semi-conserved (see Figure 2). Consensus is predicted using the "Consensus" online tool (http://coot.embl.de/Alignment//consensus.html). Map of conserved positions used for the tree reconstruction is shown below each alignment block.

```
Jnet 331001027                                                  EEEEEE------EEEEEE---------------E-----------------------------EEEEE---EEE----------------EEE-EEEE-E
331001027 Parasutterella excrementihominis YIT 11859            IIGVGLDLGGTYTGTFIT---------------SHPSDE--------------AEHRDHS-SAFTVVNSEKLSF----------------SSKSRTAV-R
 34557932 Wolinella succinogenes DSM 1740                       VSPISVDLGGKNTGFFSF---------------T-------------------DSLDNSQ-SGTVIYDES-FVL----------------SQVGRRSK-R
 54296138 Legionella pneumophila str. Paris                     LSPIGIDLGGKFTGVCLS---------------HLEAFAEL------------PNHANTK-YSVILIDHNNFQL----------------SQAQRRAT-R
319941583 Sutterella wadsworthensis 3 1 45B                     SCSIGIDMGAKYTGVFYA---------------LFDRE-EL------------PTNLNSK-AMTLVMPETGPRY----------------VQAQRTAV-R
254447899 gamma proteobacterium HTCC5015                        ISPIAIDLGAKFTGVALY---------------QYLEGA--------------DCTQEVA-KGLLVDDRGNVTW----------------SQEGRRGK-R
118497352 Francisella novicida U112                             ILPIAIDLGVKNTGVFSA---------------FYQKGTSL------------ERLDNKN-GKVYELSKDSYTL----------------LMNNRTAR-R
consensus/95%                                                   ...huhDhG...TG....               ........            ....... ............h                ....R.u. R
consensus/85%                                                   ...huhDhG...TG....               ........            ....... ............h                ....R.u. R

227494853 Actinomyces coleocanis DSM 15436#                     NYRIGIDVGLNSIGFCAV---------------EVDQHD--------------TPLGFLN-LSVYRHDAGIDPNGKKT----------NTTRLAMSGVAR
328956315 Coriobacterium glomerans PW2                          DYSIGLDMGTSSVGWAVT----D----------ERGTLA--------------HFKRKPT-WGSRLFREAQTAA-------------------VARMP-R
227824983 Acidaminococcus sp. D21                               MYYLGLDIGTNSVGYAVT----D----------PSYHLL--------------KFKGEPM-WGAHVFAAGNQSA-------------------ERRSF-R
303229466 Veillonella atypica ACS-134-V-Col7a                   DYFVGLDIGTNSVGWAVT----N----------TSYELL--------------KFHSHKM-WGSRLFEEGESAV-------------------TRRGF-R
 34762592 Fusobacterium nucleatum subsp. vincentii ATCC 49256   DYYLGFDIGTNSVGWCVT----D----------LDYNVL--------------RFNKKDM-WGSRLFDEAKTAA-------------------ERRVQ-R
374307738 Filifactor alocis ATCC 35896                          EYYLGLDVGTNSVGWAVT----D----------SQYNLC--------------KFKKKDM-WGIRLFESANTAK-------------------DRRLQ-R
320528778 Solobacterium moorei F0204                            NYYLGLDVGTSSVGWAVT----D----------TDYNVL--------------KFRGKSM-WGARLFDEASTAE-------------------ERRTH-R
291520705 Coprococcus catus GD-7                                EYFLGLDMGTGSLGWAVT----D----------STYQVM--------------RKHGKAL-WGTRLFESASTAE-------------------ERRMF-R
 42525843 Treponema denticola ATCC 35405                        DYFLGLDVGTGSVGWAVT----D----------TDYKLL--------------KANRKDL-WGMRCFETAETAE-------------------VRRLH-R
304438954 Peptoniphilus duerdenii ATCC BAA-1640                 EYYIGLDIGTASVGWAVT----D----------ESYNIP--------------KFNGKKM-WGVRLFDDAKTAE-------------------ERRTQ-R
224543312 Catenibacterium mitsuokai DSM 15897                   DYCIGLDLGTGSVGWAVV----D----------MNHRLM--------------KRNGKHL-WGSRLFSNAETAA-------------------NRRAS-R
116628213 Streptococcus thermophilus LMD-9                      PYSIGLDIGTNSVGWAVT--TDNYKVPSKKMKVLGNTSK--------------KYIKKNL-LGVLLFDSGITAE-------------------GRRLK-R
 24379809 Streptococcus mutans UA159                            PYSIGLDIGTNSVGWAVV--TDDYKVPAKKMKVLGNTDK--------------SHIEKNL-LGALLFDSGNTAE-------------------DRRLK-R
 13622193 Streptococcus pyogenes M1 GAS 4CMP                    KYSIGLDIGTNSVGWAVI--TDEYKVPSKKFKVLGNTDR--------------HSIKKNL-IGALLFDSGETAE-------------------ATRLK-R

                                                                --EEEEEE---EEEEEEE---------EEEEE----------------------EEEEE--EEEEE------HH-------------------HHH----
domains (Jinek et al.)                                          //rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr|RRRRRRRRRRRRRRRRRRRRRRRRRR
catalytic residues (Jinek et al.)                               //rrr*D*rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr|RRRRRRRRRRRRRRRRRRRRRRRRRR
domains (Nishimasu et al.)                                      //rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr|lllll|RRRRRRRRRRRRRRRRRRRRRRRRRR
catalytic residues (Nishimasu et al.)                           //rrr*D*rr*S*rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr|lllll|RRRRRRRRRRRRRRRRRRRRRRRRRR

310286728 Bifidobacterium bifidum S17                           DYAISLDIGNASVGWSAF----T----------PNYRLV--------------RAKGHEL-IGVRLFDPADTAE-------------------SRRMA-R
366983953 Oenococcus kitaharae DSM 17330                        DYSVGLDIGTSSVGWAAI----D----------NKYHLI--------------RAKSKNL-IGVRLFDSAVTAE-------------------KRRGY-R
422884106 Streptococcus sanguinis SK49                          NYSIGLDIGVSSVGYAVV----T----------EDYRVPAFKFKVLGNTEK--EKIKKNL-IGSTTFVSAQPAK-------------------GTRVF-R
339625081 Fructobacillus fructosus KCTC 3544                    VYDVGLDIGTGSVGWVAL----D----------ENGKLA--------------RAKGKNL-VGVRLFDTAQTAA-------------------DRRGF-R
306821691 Eubacterium yurii subsp. margaretiae ATCC 43715       QYYIGLDVGTNSVGWAVT----D----------TSYNLL--------------RAKGKDM-WGARLFEKANTAA-------------------ERRTK-R
336394882 Lactobacillus farciminis KCTC 3681                    PYNIGLDIGTSSVGWAVT----N----------DNYDLL--------------NIKKKNL-WGVRLFEEAQTAK-------------------ETRLN-R
323463801 Staphylococcus pseudintermedius ED99                  PYILSLDIGTGSVGYACM----D----------KGFNVL--------------KYHDKDA-LGVYLFDGALTAQ-------------------ERRQF-R
258509199 Lactobacillus rhamnosus GG                            PYGIGLDIGSNSIGFAVV----D----------ANSHLL--------------RLKGETA-IGARLFREGQSAA-------------------DRRGS-R
169823755 Finegoldia magna ATCC 29328                           KYYIGLDVGTNSVGWAVT----D----------EFYNIL--------------RAKGKDL-WGVRLFEKADTAA-------------------NTRIF-R
389815359 Planococcus antarcticus DSM 14505                     NYTIGLDIGVASVGWVCI----D----------ENYKIL--------------NYNNRHA-FGVHEFESAESAA-------------------GRRLK-R
consensus/95%                                                   .Y.huhDhG..ShGa...  ...............................  ....c.. hG...F..u..u.....          .......R...R
consensus/85%                                                   .Y.hGLDhG..SVGahh.  ................h..............  ....c.h hG...Fc.u.oA.....          .......R...R

Jnet 227501312                                                  --EEEEEE----EEEEEE-------------------EE--------------E-----E-EEEEEE----HHH-------------------HHH----
 47458868 Mycoplasma mobile 163K                                KVVLGLDLGIASVGWCLT----D----------ISQKED--------------NKFPIIL-HGVRLFETVDDSDDK-------------LLNETRRKK-R
284931710 Mycoplasma gallisepticum str. F                       EVTIGLDLGVGSVGWAIV-------------------DN--------------ETNIIHH-LGSRLFSQAKTAE-------------------DRRSF-R
 71894592 Mycoplasma synoviae 53                                KIVIGFDLGVASVGWSIV----NA-----------------------------ETKEVID-LGVRLFSEPEKAD-------------------YRRAK-R
363542550 Mycoplasma ovipneumoniae SC01                         NITIGFDLGIASIGWAII----DS-----------------------------TTSKILD-WGTRTFEERKTAN-------------------ERRAF-R
384393286 Mycoplasma canis PG 14                                KVTLGFDLGIASVGWAIV----DS-----------------------------ETNQVYK-LGSRLFDAPDTNL-------------------ERRTQ-R
consensus/95%                                                   .h.hGhDLGhuShGW.h.    ..         ......              ....hh. hG.R.F...c.....              ....RR.. R
consensus/85%                                                   .h.hGhDLGhuShGW.h.    ..         ......              ....hh. hG.R.F...c.....              ....RR.. R

238924075 Eubacterium rectale ATCC 33656                        KYILALDIGIASVGWAIL-------------------DK--------------ESETVIE-AGSNIFPEASAADNQ-----------------LRRDM-R
315149830 Enterococcus faecalis TX0012                          MYSIGLDLGISSVGWSVI-------------------DE--------------RTGNVID-LGVRLFSAKNSEKNL-----------------ERRTN-R
116627542 Streptococcus thermophilus LMD-9                      DLVLGLDIGIGSVGVGIL----NK-----------------------------VTGEIIH-KNSRIFPAAQAENNL-----------------VRRTN-R
315659848 Staphylococcus lugdunensis M23590                     KFILGLDIGITSVGYGLI-------------------DY--------------ETKNIID-AGVRLFPEANVENNE-----------------GRRSK-R
160915782 Eubacterium dolichum DSM 3991                         RLVLGLDIGITSVGFGII-------------------DL--------------DESEIVD-YGVRLFKEGTAAENE-----------------TRRTK-R
325677756 Ruminococcus albus 8                                  NYYLGLDVGIGSIGWAVI----NI-----------------------------EKKRIED-FNVRIFKSGEIQEKNRN----------SRASQQCRRS-R
225377804 Roseburia inulinivorans DSM 16841                     QYRIGLDIGITSVGWAVL---------------QNNSQD--------------EPVRITD-LGVRIFDVAENPKNG------------DALAAPRRDA-R
336393381 Lactobacillus coryniformis subsp. torquens KCTC 3535  GYRIGLDVGITSTGYAVL---------------KTDKNG--------------LPYKILT-LDSVIYPRAENPQTG------------ASLAEPRRIK-R
310780384 Ilyobacter polytropus DSM 2926                        KYSIGLDIGIASVGWSVI----NK-----------------------------KDERIED-MGVRIFQKAENPKDG------------SSLASSRREK-R
consensus/95%                                                   .h.huLDhGI.S.Ghuhh    ..         ......              ....h.. ....ha...........          .......R.. R
consensus/85%                                                   .h.hGLDhGI.ShGauhh    ..         ......              ....h.c h...hF..u........          ......RR.. R

301311869 Bacteroides sp. 20 3                                  KKIVGLDLGTNSIGWALI----N----------AYINKE--------------HLYGIEA-CGSRIIPMDAAILGNFDKGNS------ISQTADRTSY-R
385811609 Ignavibacterium album JCM 16511                       KKVLGLDIGTNSIGCALL----S----------LPKSIQDY-----------GKGGRLEW-LTSRVIPLDADYMKAFIDGKNGLPQV-ITPAGKRRQK-R
 60683389 Bacteroides fragilis NCTC 9343                        KRILGLDLGTNSIGWALV----N----------EAENKD--------------ERSSIVK-LGVRVNPLTVDELTNFEKGKS------ITTNADRTLK-R
319957206 Nitratifractor salsuginis DSM 16511                   KKILGVDLGITSFGYAIL--QET----------GK------------------DLYRCLD-NSVVMRNNPYDEKSG------------ESSQSIRSTQ-K
187250660 Elusimicrobium minutum Pei191                         PYRIGLDLGVGSIGFAIVS-MEE----------NDGNVL--------------LPKEIIM-VGSRIFKASAGAA-------------------DRKLS-R
325972003 Sphaerochaeta globus str. Buddy                       PYSIGLDLGVGSIGVAVA----A----------YDPIKK--------------QPSDLVF-VSSRIFIPSTGAA-------------------ERRQK-R
296446027 Methylosinus trichosporium OB3b                       MRVLGLDAGIASLGWALIE-IEE----------SNRGEL--------------SQGTIIG-AGTWMFDAPEEKTQAGA----------KLKSEQRRTF-R
347536497 Flavobacterium branchiophilum FL-15                   AKILGLDLGTNSIGWAV----------------VERENI---------------DFSLID-KGVRIFSEGV-KSEKGIE---------SSRAAERTGY-R
345885718 Prevotella sp. C561                                   QKVLGLDLGTNSIGSAV----------------RNLDLSD-------------DLQWQLEFFSSDIFRSSVNKESNGRE---------YSLAAQRSAH-R
282880052 Prevotella timonensis CRIS 5C-B1                      KRILGLDTGTNSLGWAV----------------VDWDEHA-------------QSYELIK-YGDVIFQEGV-KIEKGIE---------SSKAAERSGY-K
282880052 Prevotella timonensis CRIS 5C-B1                      KRILGLDTGTNSLGWAVV---------------DWDEHA--------------QSYELIK-YGDVIFQEGVKIEKGIE----------SSKAAERSGY-K
312879015 Aminomonas paucivorans DSM 12260                      ------MIGEHVRGGCLF----D----------DHWTPN--------------WGAFRLP-NTVRTFTKAENPKDG------------SSLAEPRRQA-R
294086111 Candidatus Puniceispirillum marinum IMCC1322          MRRLGLDLGTNSIGWCLL---------------DLGDDG--------------EPVSIFR-TGARIFSDGRDPKSL------------GSLKATRREA-R
330822845 Alicycliphilus denitrificans K601                     RYRLALDLGSTSLGWALF---------------RLDACN--------------RPTAVIK-AGVRIFSDGRNPKDG------------SSLAVTRRAA-R
344171927 Ralstonia syzygii R24                                 QHRWGLDIGTNSIGWAVI----------------ALIEG--------------RPAGLVA-TGSRIFSDGRNPKDG------------SSLAVERRGP-R
159042956 Dinoroseobacter shibae DFL 12                         -MRLGLDIGTSSIGWWLY--ETD----------GAGSDA--------------RITGVVD-GGVRIFSDGRDPKSG------------ASLAVDRRAA-R
 83591793 Rhodospirillum rubrum ATCC 11170                      PWILGLDIGTDSLGWAVF--SCE----------EKGPP---------------TAKELLG-GGVRLFDSGRDAKDH------------TSRQAERGAF-R
288957741 Azospirillum sp. B510                                 PYRLSFDLGTNSIGWGLL---------------NLDRQG--------------KPREIRA-LGSRIFSDGRDPQDK------------ASLAVARRLA-R
427429481 Caenispirillum salinarum AK4*                         RWSLALDIGEGSIGWAVA---------------EVDAEG--------------RVLQLTG-TGVTLFPSAWSNENG------------TYVAHGAA-D-R
 92109262 Nitrobacter hamburgensis X14                          LYRLGLDLGSNSLGWFVT--HLE----------KRGDRH--------------EPVALGP-GGVRIFPDGRDPQSG------------TSNAVDRRMA-R
148255343 Bradyrhizobium sp. BTAi1                              AYRLGVDLGANSLGWFVV--WLD----------DHG-----------------QPEGLGP-GGVRIFPDGRNPQSK------------QSNAAGRRLA-R
 34557790 Wolinella succinogenes DSM 1740                       ERILGVDLGISSLGWAIV----E----------YDKDDE--------------AANRIID-CGVRLFTAAETPKKK------------ESPNKARREA-R
218563121 Campylobacter jejuni subsp. jejuni NCTC 11168         ARILAFDIGISSIGWAFS-----------------------------------ENDELKD-CGVRIFTKVENPKTG------------ESLALPRRLA-R
Jnet 218563121                                                  -EEEEEE----EEEEEEE-----------------------------------E---EEE---EEEE-------------------------HHHHHH-H
291276265 Helicobacter mustelae 12198                           IRTLGIDIGIASIGWAVI--EGE----------YTDKGL--------------ENKEIVA-SGVRVFTKAENPKNK------------ESLALPRTLA-R
222109285 Acidovorax ebreus TPSY                                QHVFGLDIGIASVGWAIL-----------------------------------GEQRIID-LGVRCFDKAETAKEG------------DPLNLTRRQA-R
365156657 Bacillus smithii 7 3 47FAA                            NYKMGLDIGIASVGWAVI-------------------NL--------------DLKRIED-LGVRIFDKAEHPQNG------------ESLALPRRIA-R
220930482 Clostridium cellulolyticum H10                        KYTLGLDVGIASVGWAVI-------------------DK--------------DNNKIID-LGVRCFDKAEESKTG------------ESLATARRIA-R
297182908 uncult.delta proteobacterium HF0070 07E19             EYTLGLDLGIKSIGWAIL-----------------------------------SGERIAN-AGVYLFETAEELNSTGNKL--------ISKAAERGRK-R
154250555 Parvibaculum lavamentivorans DS-1                     ERIFGFDIGTTSIGFSVI---------------DYSSTQ--------------SAGNIQR-LGVRIFPEARDPDGT-------------PLNQQRRQK-R
218767588 Neisseria meningitidis Z2491                          NYILGLDIGIASVGWAMV---------------EID-ED-------------ENPICLID-LGVRVFERAEVPKTG------------DSLAMARRLA-R
 15602992 Pasteurella multocida str. Pm70                       SYILGLDLGIASVGWAVV---------------EIN-EN-------------EDPIGLID-VGVRIFERAEVPKTG------------ESLALSRRLA-R
187736489 Akkermansia muciniphila ATCC BAA-835                  SLTFSFDIGYASIGWAVI--------------ASASHDD--------------ADPSVCG-CGTVLFPKDDCQA------------------FKRREY-R
315605738 Actinomyces sp. oral taxon 180 str. F0310             RYRVGIDVGLRSIGFCAV---------------EVDDED--------------HPIRILN-SVVHVHDAGTGGPGETE------------SLRKRSGVAA
117929158 Acidothermus cellulolyticus 11B                       TWRLGVDVGERSIGLAAV----------------SYEED--------------KPKEILA-AVSWIHDGGVGDERS------------GASRLALRGMAR
189440764 Bifidobacterium longum DJO10A                         RYRIGIDVGLNSVGLAAV---------------EVSDEN--------------SPVRLLN-AQSVIHDGGVDPQKNKE----------AITRKNMSGVAR
584580150 Actinomyces naeslundii str. Howell 279                HLRVGIDVGTHSVGLATL---------------RVDDHG--------------TPIELLS-ALSHIHDSGVGKEGKKD----------HDTRKKLSGIAR

                                                                EEEEEEEE---EEEEEEE---------------EE--------------------EEEEE-EEEEE-----------------------------HHHHH
domains                                                         //rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr|RRRRRRRRRRRRRRRRRRRRRRRRRR
catalytic residues                                              //rrr*D*rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr|RRRRRRRRRRRRRRRRRRRRRRRRRR

403744858 Alicyclobacillus hesperidum URH17-3-68                AYRLGLDIGITSVGWAVV--------------ALEKDESG------------LKPVRIQD-LGVRIFDKAEDSKT-------------GASLALPRREAR
407803669 Alcanivorax sp. W11-5                                 RYRVGLDLGTASVGAAVF--------------SMDEQ---------------GNPMELIW-HYERLFSEPLVPDM----------GQLKPKKAARRLA-R
423317190 Bergeyella zoohelcum ATCC 43767                       KHILGLDLGTNSIGWALI--------------ERNIE----------------EKYGKIIG-MGSRIVPMGAELSK------FEQGQAQTKNADRRTN-R
402847315 Porphyromonas sp. oral taxon 279 str. F0450           KHVLGLDLGVGSIGWCLI--------------ALDAQ---------------GDPAEILG-MGSRVVPLNNATKAIEA----FNAGAAFTASQERTAR-R
404487228 Barnesiella intestinihominis YIT 11860                KNILGLDLGLSSIGWSVI--------------RENS-----------------EEQELVA-MGSRVVSLTAAELSS------FTQGNGVSINSQRTQK-R
374384763 Odoribacter laneus YIT 12061                          ETTLGIDLGTNSIGLALV--------------DQ-------------------EEHQILY-SGVRIFPEGINKDTI------GLGEKEESRNATRRAK-R
384109266 Treponema sp. JC4                                     KWRLGLDLGTNSIGWSVF--------------SLDKD---------------NSVQDLID-MGVRIFSDGRDPKT------------KEPLAVARRTA-R
402849997 Rhodovulum sp. PH10                                   GIRFAFDLGTNSIGWAVW--------------RTGPGVFGE-----------DTAASLDG-SGVLIFKDGRNPKD------------GQSLATMRRVP-R
consensus/95%                                                   ...huhDhG..ShG..h.. ...         .........           ..............h................................c
consensus/85%                                                   ...huhDhG..ShGh.hh. ...         .........           .....h....u..hh...........................R....R

#map                                                            ******************___________________________________*******_*********_______________________*******
```

```
Jnet 331001027         EHHH--HHHHHHHHHHHHHHHHHHHH-HH--------------------------------------------------------------------------------------------------------------------------
P. excrementihominis   HRVRSYKGFDLRRRLLLLVAEYQLLQ-KKQTL-----------------------------------------------------------------------------------------------------------------------
W. succinogenes        HSKRNNLRNKLVKRLFLLILQEHHGL-SI--------------------------------------------------------------------------------------------------------------------------
L. pneumophila         HRVRNKKRNQFVKRVALQLFQHILSR-DL--------------------------------------------------------------------------------------------------------------------------
S. wadsworthensis      HRLRGQKRYTLARKLAFLVVDDMIKK-QEKRL-----------------------------------------------------------------------------------------------------------------------
HTCC5015               HQVRGYKRRKMAKRLLWLILDSEYGI-KR--------------------------------------------------------------------------------------------------------------------------
F. novicida            HQRRGIDRKQLVKRLFKLIWTEQLNL-EW--------------------------------------------------------------------------------------------------------------------------
consensus/95%          H..R......h.cchh..hh...h.. .....
consensus/85%          H..R......h.cchh..hh...h.. .....

A. coleocanis          RTRRLFRKRKRRLAALDRFIEAQ----------------------------------------GWTLPDHADYKDP--------------------------------------------------------------------YTPWL--
C. glomerans           GQRRRYVRRRWRLDLLQKLFEQQMEQ-ADPDFFIRLRQS-RLLRDDRAEEHADY---------RWPLFNDCKFTER--------------------------------------------------------------DYYQRFPTIYH--
Acidaminococcus D21    TSRRRLDRRQQRVKLVQEIFAPVISP-IDPRFFIRLHES-ALWRDDVAETD------------KHIFFNDPTYTDK--------------------------------------------------------------EYYSDYPTIHH--
V. atypica             SMRRRLERRKLRLKLLEELFADAMAQ-VDSTFFIRLHES-KYHYEDKTTGHSS----------KHILFIDEDYTDQ--------------------------------------------------------------DYFTEYPTIYH--
F. nucleatum           NSRRRLKRRKWRLNLLEEIFSDEIMK-IDSNFFRRLKES-SLWLEDKNSKE------------KFTLFNDDNYKDY--------------------------------------------------------------DFYKQYPTIFH--
F. alocis              GNRRRLERKKQRIDLLQEIFSPEICK-IDPTFFIRLNES-RLHLEDKSNDF------------KYPLFIEKDYSDI--------------------------------------------------------------EYYKEFPTIFH--
S. moorei              GNRRRLARRKYRLLLLEQLFEKEIRK-IDDNFFVRLHES-NLWADDKSKPS------------KFLLFNDTNFTDK--------------------------------------------------------------DYLKKYPTIYH--
C. catus               TARRRLDRRNWRIQVLQEIFSEEISK-VDPGFFLRMKES-KYYPEDKRDAEGNCPEL------PYALFVDDNYTDK--------------------------------------------------------------NYHKDYPTIYH--
T. denticola           GARRRIERRKKRIKLLQELFSQEIAK-TDEGFFQRMKES-PFYAEDKTILQ------------ENTLFNDKDFADK--------------------------------------------------------------TYHKAYPTINH--
P. duerdenii           GSRRRLNRRKERINLLQDLFATEISK-VDPNFFLRLDNS-DLYREDKDEKLKS----------KYTLFNDKDFKDR--------------------------------------------------------------DYHKKYPTIHH--
C. mitsuokai           SIRRRYNKRRERIRLLRAILQDMVLE-KDPTFFIRLEHT-SFLDEEDKAKYLGTDYKD-----NYNLFIDEDFNDY--------------------------------------------------------------TYYHKYPTIYH--
S. thermophilus        TARRRYTRRRNRILYLQEIFSTEMAT-LDDAFFQRLDDS-FLVPDDKRDS-------------KYPIFGNLVEEKA---------------------------------------------------------------YHDEFPTIYH--
S. mutans              TARRRYTRRRNRILYLQEIFSEEMGK-VDDSFFHRLEDS-FLVTEDKRGE-------------RHPIFGNLEEEVK---------------------------------------------------------------YHENFPTIYH--
S. pyogenes            TARRRYTRRKNRICYLQEIFSNEMAK-VDDSFFHRLEES-FLVEEDKKHE-------------RHPIFGNIVDEVA---------------------------------------------------------------YHEKYPTIYH--

                       -HHHHHHHHHHHHHHHHHHHHHHHHH-H---HHHH-----------------------------------HHHHHH---------------------------------------------------------------HHHH---HHH--
domains                RRRRRRRRRRRR|hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
catalytic residues     RRRRRRRRRRRR|hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
domains                RRRRRRRRRRRRRRRRRRRRRRRRRRR|111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
catalytic residues     RRRRRRRRRRRRRRRRRRRRRRRRRRR|111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
.
B. bifidum             TTRRRYSRRRWRLRLLDALFDQALSE-IDPSFLARRKYS-WVHPDDENNADCWY---------GSVLFDSNEQDKR---------------------------------------------------------------FYEKYPTIYH--
O. kitaharae           TTRRRLSRRHWRLRLLNDIFAGPLTDFGDENFLARLKYS-WVHPQDQSNQAHFA---------AGLLFDSKEQDKD---------------------------------------------------------------FYRKYPTIYH--
S. sanguinis           VNRRRIDRRNHRITYLRDIFQKEIEK-VDKNFYRRLDESFRVLGDKSEDLQIK-----------QPFFGDKELETA---------------------------------------------------------------YHKKYPTIYH--
F. fructosus           TTRRRLSRRKWRLRLLDELFSAEINE-IDSSFFQRLKYS-YVHPKDEENKAHYY---------GGYLFPTEEETKK---------------------------------------------------------------FHRSYPTIYH--
E. yurii               TSRRRSEREKARKAMLKELFADEINR-VDPSFFIRLEESKFFLDDRSENNRQRY----------TLFNDATFTDKD---------------------------------------------------------------YYEKYKTIFH--
L. farciminis          STRRRYRRRKNRINWLNEIFSEELAK-TDPSFLIRLQNS-WVSKKDPDRKRDKY----------NLFIDGPYTDKE---------------------------------------------------------------YYREFPTIFH--
S. pseudintermedius    TSRRRKNRRIKRLGLLQELLAPLVQ---NPNFY-QFQRQFAWKNDNMDFKNKSL---------SEVLSFLGYES-------------------------------------------------------------------KKYPTIYH--
L. rhamnosus           TTRRRLSRTRWRLSFLRDFFAPHITK-IDPDFFLRQKYS-EISPKDKDRF-------------KYEKRLFNDRTDA--------------------------------------------------------------EFYEDYPSMYH--
F. magna               SGRRRNDRKGMRLQILREIFEDEIKK-VDKDFYDRLDES-KFWAEDKKVSG------------KYSLFNDKNFSDK--------------------------------------------------------------QYFEKFPTIFH--
P. antarcticus         GMRRRYNRRKKRLQLLQSLFDSYIT---DSGFFSKTDSQHFWKNNN-EFENRSL---------TEVLSSLRISS-------------------------------------------------------------------RKYPTIYH--
consensus/95%          ..RRR..cc..Rh.hL..hh...h.......Fh........h...............     ...h.........                                                              .....a.ThaH
consensus/85%          ..RRRh.Rc..Rh.hL..hF...h....D..Fh.R...o..h..cc............     ...h.........                                                              .ah..aPTIaH
.
Jnet 227501312         ---HHHHHHHHHHHHHHHHHHHHH-------HH------------------------------------------H--------------------------------------------------------------HHH-----HHH--
M. mobile              GQRRRNRRLFTRKRDFIKYLIDN-NI-IELEFDKNPKIL-VRNFIEKY---------------INPFSKNLELKYK--------------------------------------------------------------SVTNLPIGFHN--
M. gallisepticum       GVRRLIRRRKYKLKRFVNLIWKYNSY-F-----------------------------------GFKNKEDILNNYQ--------------------------------------------------------------EQQKLHNTVLN--
M. synoviae            TTRRLLRRKKFKREKFHKLILKNAEI-F-----------------------------------GLQSRNEILNVYK--------------------------------------------------------------DQSSKYRNILK--
M. ovipneumoniae        STRRNIRRKAYRNQRFINLILKYKDL-F-----------------------------------ELKNISDIQRANKKDTENYEKIISFFTEIY--------------------------------------------KKCAAKHSNILE--
M. canis               GTRRLLRRRKYRNQKFYNLV-KRTEV-F-----------------------------------GLSSREAIENRFR--------------------------------------------------------------ELSIKYPNIIE--
consensus/95%          ..RR..RR...c..cFh.hh.c...h h........... ........               .......h......................                                            .........hh.
consensus/85%          ..RR..RR...c..cFh.hh.c...h h........... ........               .......h......................                                            .........hh.

E. rectale             GAKRNNRRLKTRINDFIKLWENN----------------------------------------------NLSIPQF------------------------------------------------------------------KSTEIVG--
E. faecalis            GGRRLIRRKTNRLKDAKKILA------------------------------------------AVGFYEDKSLKNS---------------------------------------------------------------------CPYQ--
S. thermophilus        QGRRLARRKKHRRVRLNRLFEES-------------------------------------------------GLIT------------------------------------------------------------DFTKISINLNPYQ--
S. lugdunensis         GSRRLKRRRIHRLERVKKLLEDY------------------------------------------------NLLDQ--------------------------------------------------------------SQIPQSTNPYA--
E. dolichum            GGRRLKRRRVTRREDMLHLLKQA-------------------------------------------------GIIS--------------------------------------------------------------TSFHPLNNPYD--
R. albus               GLRRLYRRKSHRKLRLKNYLSII----------------------------------------GLTTSEKIDYYYE-----------------------------------------------------------------TADNNVIQ--
R. inulinivorans       TTRRRLRRRRHRLERIKFLLQENG---------------------------------------LIEMDSFMERYYK-----------------------------------------------------------------GNLPDVYQ--
L. coryniformis        GLRRRTRRTKFRKQRTQQLFIHS----------------------------------------GLLSKPEIEQILA---------------------------------------------------------------TPQAKYSVYE--
I. polytropus          GSRRRNRRKKHRLDRIKNILCES----------------------------------------GLVKKNEIEKIYK----------------------------------------------------------------NAYLKSPWE--
consensus/95%          ..cR..RR...R..c...hh....                                       .............                                                            ...........h.
consensus/85%          ..RR..RR...R..ch..hh....                                       .............                                                            ...........h.

Bacteroides sp.20      GIRRLRERHLLRRERLHRILDLL-GF-LPKHYSDSLNRY-GKFLNDIEC--------------KLPWVKDETGSYKFIFQESFKEMLANFTEHHPILIAN-------------------------------------NKKVPYDWTIYY--
I. album               GSRRLKHRYKLRRSRLIRVFKTL-NW-LPEDFPLDNPKRIKETISTEGKFSFRISD-------YVPISDESYREFYREFGYPENEIEQVIEEINFRRKTKGKNKNPMIKL------------------------------LPEDWVVYY--
B. fragilis            GMRRNLQRYKLRRETLTEVLKEH------------------------------------------KLITEDTILSE--------------------------------------------------------------NGNRTTFETYR--
N. salsuginis          SMRRLIEKRKKRIRCVAQTMERY----------------------------------------GILDYSETMKIND--------------------------------------------------------------PKNNPIKNRWQ--
E. minutum             GQRNNHRHTRERMRYLWKVLAEQ-KL-------------------------------------ALPVPADLDRKEN--------------SSEGETSAKR-------------------------------------FLGDVLQKDIYE--
S. globus              GQRNSLRHRANRLKFLWKLLAER-NL--------MLSYS-------------------------EQDVPDPARLRF--------------------------------------------------------------EDAVVRANPYE--
M. trichosporium       GQRRVVRRRRQRMNEVRRILHSH----------------------------------------GLLPSSDRDALKQ------------------------------------------------------------------PGLDPWR--
F. branchiophilum      SARKIKYRRKLRKYETLKVLSLN-RMCP-LSIEEVEEW-----KKSGFK--------------DYPLNP-EFLKWLS--------------------------------------------------------TDEESNVN-----PYF--
Prevotella C561        RSRGLNEVRRRRLWATLNLLIKH-GFCP-MSSESLMRWCTYDKRKGLFR--------------EYPIDDKDFNAWIL--------------------------------------------------------LDFNGDGRPDYSSPYQ--
P. timonensis          AIRKQYFRRRLRKIQVLKVLVKY-HLCPYLSDDDLRQW-------HLQK--------------QYPKSD-ELMLWQR--------------------------------------------------------TSDEEGKN-----PYY--
P. timonensis          AIRKQYFRRRLRKIQVLKVLVKY-HLCPYLSDDDLRQW-------HLQK--------------QYPKSDELMLWQR--------------------------------------------------------------TSDEEGKNPYY--
A. paucivorans         GLRRRLRRKTQRLEDLRRLLAKE----------------------------------------GVLSLSDLETLFR----------------------------------------------------------------ETPAKDPYQ--
P. marinum             LTRRRRDRFIQRQKNLINALVKY----------------------------------------GLMPADEIQRQAL------------------------------------------------------------------AYKDPYP--
A. denitrificans       AMRRRRDRLLKRKTRMQAKLVEH----------------------------------------GFFPADAGKRKAL------------------------------------------------------------------EQLNPYA--
R. syzygii             QMRRRRDRYLRRRDRFMQALINV----------------------------------------GLMPGDAAARKAL------------------------------------------------------------------VTENPYV--
D. shibae              AMRRRRDRYLRRRATLMKVLAET----------------------------------------GLMPADPAEAKAL------------------------------------------------------------------EALDPFA--
R. rubrum              RARRQTRTWPWRRDRLIALFQAA----------------------------------------GLTPPAAETRQIA-----------------------------------------------------------------------LA--
Azospirillum B510      QMRRRRDRYLTRRTRLMGALVRF----------------------------------------GLMPADPAARKRL-----------------------------------------------------------------EVAVDPYL--
C. salinarum           AVRGQQQRHDSRRRRLAG-LARL-----------------------------------CAPVLERSPEDLKDLTRT-----------------------------------------------------------------PPKADPRAIF
N. hamburgensis        GARKRRDRFVERRKELIAALIKY----------------------------------------NLLPDDARERRAL------------------------------------------------------------------EVLDPYA--
Bradyrhizobium BTAi1   SARRRRDRYLQRRGKLMGLLVKH----------------------------------------GLMPADEPARKRL------------------------------------------------------------------ECLDPYG--
W. succinogenes        GIRRVLNRRRVRMNMIKKLFLRA----------------------------------------GLIQDVDLDGEGGM------------------------------------------------------------FYSKANRADVWE--
C. jejuni              SARKRLARRKARLNHLKHLIANEFKL-------------------------------------NYEDYQSFDESLA-------------------------------------------------------------KAYKGSLISPYE--
Jnet 218563121         --HHHHHHHHHHHHHHHHHHHH----------------------------------------------------------------------------------------------------------------------------HHH--
H. mustelae            SARRRNARKKGRIQQVKHYLSKALGL-------------------------------------DLECFVQGEKLATL------------------------------------------------------------FQTSKDFLSPWE--
A. ebreus              LLRRRLYRRAWRLTQLSRLLKRK-------------------------------------------GLIADAKLFA--------------------------------------------------------------KAPSYGDSAWE--
B. smithii             SARRRLRRRKHRLERIRRLLVSE----------------------------------------NVLTKEEMNLLFK----------------------------------------------------------------QKKQIDVWQ--
C. cellulolyticum      GMRRRISRRSQRLRLVKKLFVQYE---------------------------------------IIKDSSEFNRIFD--------------------------------------------------------------TSRDGWKDPWE--
HF0070_07E19           RIRRMLDRKARRGRHIRYLLEREG---------------------------------------LPTDELEEVVVHQ------------------------------------------------------------------SNRTLWD--
P. lavamentivorans     MMRRQLRRRRIRRKALNETLHEAGF---------------------------------------LPAYGSADWPVV------------------------------------------------------------------MADEPYE--
N. meningitidis        SVRRLTRRRAHRLLRARRLLKR-----------------------------------------------EGVLQAA--------------------------------------------------------DFDENGLIKSLPNTPWQ--
P. multocida           STRRLIRRRAHRLLLAKRFLKR-----------------------------------------------EGILSTI-------------------------------------------------------------DLEKGLPNQAWE--
A. muciniphila         RLRRNIRSRRVRIERIGRLLVQA----------------------------------------------QIITPEM--------------------------------------------------------------KETSGHPAPFY--
Actinomyces F0310      RARRRGRAEKQRLKKLDVLLEEL----------------------------------------GWGVSSN-ELLDS--------------------------------------------------------------------HAPWH--
A. cellulolyticus      RARRLRRFRRARLRDLDMLLSEL----------------------------------------GWTPLPDKNVSPV---------------------------------------------------------------------DAWL--
B. longum              RTRRMRRRKRERLHKLDMLLGKF----------------------------------------GYPVIEPESLDKP--------------------------------------------------------------------FEEWH--
A. naeslundii          RARRLLHHRRTQLQQLDEVLRDL----------------------------------------GFPIPTPGEFLDL--------------------------------------------------------------NEQTDPYRVWR--

                       HHHHHHHHHHHHHHHHH--------------------------------------------------------------------------------------------------------------------------------------
domains                RRRRRRRRRRRRR|hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
catalytic residues     RRRRRRRRRRRRR|hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

A. hesperidum          SARRRTRRRRHRLWRVKRLLEQHGI---------------------------------------LSMEQIEALYAQRTSS----------------------------------------------------------------PDVYA--
Alcanivorax W11-5      QQRRQIDRRASRLRRIAIVSRRLGI---------------------------------------APGRNDSGVHG---------------------------------------------------------------------NDVPT--
B. zoohelcum           GARRLNKRYKQRRNKLIYILQKLDM---------------------------------------LPSQIKLKEDFSDPNK------IDKITILPISKKQEQLTA----------------------------------------FDLVS--
Porphyromonas F0450    TMRRGFARYQLRRYRLRRELEKVGM---------------------------------------LPDAALIQLPL---------------------------------------------------------------------LELWE--
B. intestinihominis    TQRKGYDRYQLRRTLLRNKLDTLGM---------------------------------------LPDDSLSYLPK---------------------------------------------------------------------LQLWG--
O. laneus              QMRRQYFRKKLRKAKLLELLIAYDM---------------------------------------CPLKPEDVRRWKNWDK------QQKSTVRQFPDTPAFREWLK--------------------------------------QNPYE--
Treponema JC4          SQRKLIYRRKLRRKQVFKFLQEQGL---------------------------------------FPK--TKEECMTLKS-----------------------------------------------------------------LNPYE--
Rhodovulum PH10        QSRKRRDRFVLRRRDLLAALRKAGL---------------------------------------FPV--DVEEGRRLAA-----------------------------------------------------------------TDPYH--
consensus/95%          ..R........R...h...h....................................  .....................................................                     ...............h...
consensus/85%          ..Rc...c...R...h..hh....................................  .....................................................                     ...............a...

#map                   **********************---------------------------------------------------------------------------------------------------------------------------------
```

```
Jnet 331001027          ----------------------HH-------HHHHHHHHHHHHHH----EEE--------------------H---------------------------HHHHHHHH----------------HHHHHH---------HH---------
P. excrementihominis    --------------------APEE-------RENLRIALSGYLKRRGYARTEAE---------------TDTSVLESLDPSVFSSAPSFTNFFN--DSEPLNIQWEAIANSPE------------TTKALN--------KELSGQK----
W. succinogenes         --------------------------------DVLPDEIRGLFNKRGYTYAGFE---------------LDEKKKDALESDTL--KEFLSEKLQSIDRDSDVEDFLNQIAS--------------NAESFK---------DYKKGF----
L. pneumophila          -------------------------------NAKEETALCHYLNNRGYTYVDTD---------------LDEYIKDETTINLL--KELLPS-----ESEHNFIDW--------------------FLQKMQ-------------------
S. wadsworthensis       --------------------TDEE-------WKRGREALSGLLKRRGYSRPNADG--------------EDLTPLENVRADVFAAHPAFSTYFS--EVRSLAEQWEEFTANIS------------NVEKFL--------GDPNIPA----
HTCC5015                ---------------------EEV-------TEPLLKFINGLLNRRGYTYI------------------SEEVDEESMNVSPLPFSEMMPDYFN--SSAPLLEQLAKLLSDKN------------KLVRFR--------AEGKIPS----
F. novicida             -------------------------------DKDTQQAISFLFNRRGFSFI------------------TDGYSPEYLNIVPEQVKAILMDIFDDYNGEDDLDSYLKLATEQESKISEIYNKLMQKILEFKLMKLCTDIKDDKVSTKTLK
consensus/95%                               ....       .......h..hh..RGa.......              .-....-............h...............h........................h....................
consensus/85%                               ....       .......h..hh..RGa.......              .-....-............h...............h........................h....................

A. coleocanis           -VRAELAQTP---IRDEN---DLH--------EKLAIAVRHIARHRGWRSPWVP------------------------------------------------------------------------------------------------
C. glomerans            -VR-SWLM-----ETDEQ---ADI--------RLIYLALHNIVKHRGNFLREGQ------------------------------------------------------------------------------------------------
Acidaminococcus D21     -LI-VDLM-----ESSEK---HDP--------RLVYLAVAWLVAHRGHFLNEVD------------------------------------------------------------------------------------------------
V. atypica              -LR--KDL-----MENGT---DDI--------RKLFLAVHHILKYRGNFLYEGA------------------------------------------------------------------------------------------------
F. nucleatum            -LR-DELI-----KNPEK---KDI--------RLIYLALHSIFKSRGHFLFEGQ------------------------------------------------------------------------------------------------
F. alocis               -LR-KHLI-----ESEEK---QDI--------RLIYLALHNIIKTRGHFLIDGD------------------------------------------------------------------------------------------------
S. moorei               -LR-SDLI-----HNSTE---HDI--------RLVFLALHHLIKYRGHFIYDNS------------------------------------------------------------------------------------------------
C. catus                -LR-KMLM-----ETTEI---PDI--------RLVYLVLHHMMKHRGHFLLSGD------------------------------------------------------------------------------------------------
T. denticola            -LI-KAWI-----ENKVK---PDP--------RLLYLACHNIIKKRGHFLFEGD------------------------------------------------------------------------------------------------
P. duerdenii            -LI-MDLI-----EDEGK---KDI--------RLLYLACHYLLKNRGHFIFEGQ------------------------------------------------------------------------------------------------
C. mitsuokai            -LR-KALC-----ESTEK---ADP--------RLIYLALHHIVKYRGNFLYEGQ------------------------------------------------------------------------------------------------
S. thermophilus         -LR-KYLA-----DSTKK---ADL--------RLVYLALAHMIKYRGHFLIEGE------------------------------------------------------------------------------------------------
S. mutans               -LR-QYLA-----DNPEK---VDL--------RLVYLALAHIIKFRGHFLIEGK------------------------------------------------------------------------------------------------
S. pyogenes             -LR-KKLV-----DSTDK---ADL--------RLIYLALAHMIKFRGHFLIEGD------------------------------------------------------------------------------------------------

                        -HH-HHHH-----H---------H--------HHHHHHHHHHHH----------------------------------------------------------------------------------------------------------
domains                 hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
catalytic residues      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
domains                 111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
catalytic residues      111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
.
B. bifidum              -LR-KALM-----EDDSQ---HDI--------REIYLAIHHMVKYRGNFLVEGT------------------------------------------------------------------------------------------------
O. kitaharae            -LR-LALM-----NDDQK---HDL--------REV-----HLVKYRGHFLIEGD------------------------------------------------------------------------------------------------
S. sanguinis            -LR-KHLA-----DADKNSPVADI--------REVYMAISHILKYRGHFLTLDK------------------------------------------------------------------------------------------------
F. fructosus            -LR-QELM-----AQPNKR--FDI--------REIYLAIHHLVKYRGHFLSSQE------------------------------------------------------------------------------------------------
E. yurii                -LR-SALI-----NSDEK---FDV--------RLVFLAILNLFSHRGHFLNASL------------------------------------------------------------------------------------------------
L. farciminis           -LR-KELI-----LNKDK---ADI--------RLIYLALHNILKYRGNFTYEHQ------------------------------------------------------------------------------------------------
S. pseudintermedius     -LQ-EALL-----LKDEK---FDP--------ELIYMALYHLVKYRGHFLFDHL------------------------------------------------------------------------------------------------
L. rhamnosus            -LR-LHLM-----THTHK---ADP--------REIFLAIHHILKSRGHFLTPGA------------------------------------------------------------------------------------------------
F. magna                -LR-KYLM-----EEHGK---VDI--------RYYFLAINQMMKRRGHFLIDGQ------------------------------------------------------------------------------------------------
P. antarcticus          -LR-SDLI-----ESNKK---MDL--------RLVYLALHNLVKYRGHFLQEGN------------------------------------------------------------------------------------------------
consensus/95%            h....h...   .........D.        c.hhhh...hh..RG.F.....
consensus/85%            L....Lh..   .........D.        R.hahAh..hhK.RG.Fh....
.
Jnet 227501312          -HH--HHH-----H---------H--------HHHHHHHHHHHHHH--------------------------------------------------------------------------------------------------------
M. mobile               -LR-KAAI-----NEKYK---LDK--------SELIVLLYFYLSLRGAFFDNPE------------------------------------------------------------------------------------------------
M. gallisepticum        -LK-SEAL-------NAK---IDP--------KALSWILHDYLKNRGHFYEDNR------------------------------------------------------------------------------------------------
M. synoviae             -LK-INAL-------KEE---IKP--------SELVWILRDYLQNRGYFYKN--------------------------------------------------------------------------------------------------
M. ovipneumoniae        -VK-VKAL-------DSK---IEK--------LDLIWILHDYLENRGFFYDLEEE-----------------------------------------------------------------------------------------------
M. canis                -LK-TKAL-------SQE---VCP--------DEIAWILHDYLKNRGYFYDEKE------------------------------------------------------------------------------------------------
consensus/95%            hc ..Ah     ....c   h..        ..h.hhL..YL..RGhFac....
consensus/85%            hc ..Ah     ....c   h..        ..h.hhL..YL..RGhFac....

E. rectale              -LK-VRAI-------TEE---ITL--------DELYLILYSYLKHRGISYLEDALDDT--------------------------------------------------------------------------------------------
E. faecalis             -LR-VKGL-------TEP---LSR--------GEIYKVTLHILKKRGISYLDEV------------------------------------------------------------------------------------------------
S. thermophilus         -LR-VKGL-------TDE---LSN--------EELFIALKNMVKHRGISYLDDAS-----------------------------------------------------------------------------------------------
S. lugdunensis          -IR-VKGL-------SEA---LSK--------DELVIALLHIAKRRGIHKIDVIDS----------------------------------------------------------------------------------------------
E. dolichum             -VR-VKGL-------NER---LNG--------EELATALLHLCKHRGSSVETIEDDEAK-------------------------------------------------------------------------------------------
R. albus                -LR-NKGL-------SEK---LTP--------EEIAACLIHICNNRGYKDFYEVNVEDIED------------------------------------------------------------------------------------------
R. inulinivorans        -LR-YEGL-------DRK---LKD--------EELAQVLIHIAKHRGFRSTRKAETK---------------------------------------------------------------------------------------------
L. coryniformis         -LR-VAGL-------DRR---LTN--------SELFRVLYFFIGHRGFKSNRKAELNP--------------------------------------------------------------------------------------------
I. polytropus           -LR-AKSL-------EAK---ISN--------KEIAQILLHIAKRRGFKSFRKTDRN---------------------------------------------------------------------------------------------
consensus/95%            hc ..uh       ...   h..        .Ehh.....h...RG.............
consensus/85%            hR hcuL       .c.   h..        .Ehh.hLh.h..cRGh............

Bacteroides sp.20       -LR-KKAL-------TQK---ISK--------EELAWILLNFNQKRGYYQLRGE-EEETPNKLVEYYSLKVEKVEDSG------------------------------------------------------------------------
I. album                -LR-KKAL-------IKP---TTK--------EELIRIIYLFNQRRGFKSSRKD-LTETA--ILDYDEFAKRLAEKEKYSAENYETKFVSITKVKEVVELK-------------------------------------------------
B. fragilis             -LR-AKAV-------TEE---ISL--------EEFARVLLMINKKRGYKSSRKAKGVEEGTLIDGMDIARELYNNNLTPGELCLQLLDAGKKFLPDFYRSDL------------------------------------------------
N. salsuginis           -LRAVDAW-------KRP---LSP--------QELFAIFAHMAKHRGYKSIATE----------------------DLIYELELELGLNDPEKE--------------------------------------------------------
E. minutum              -LR-VKSL-------DER---LSL--------QELGYVLYHIAGHRGSSAIRTFENDSE-------------------------------------------------------------------------------------------
S. globus               -LR-LKGL-------NEQ---LTL--------SELGYALYHIANHRGSSSVRTFLDEEK-------------------------------------------------------------------------------------------
M. trichosporium        -IR-AEAL-------DRL---LGP--------VELAVALGHIARHRGFKSNSKGAKTNDP------------------------------------------------------------------------------------------
F. branchiophilum       -FR-DRAS-------KHK---VSL--------FELGRAFYHIAQRRGFLSNRLDQSAEGILEEHCPKIEAIVEDLISIDEISTNITDYFFETGILD------------------------------------------------------
Prevotella C561         -LR-RELV-------TRQ---FDFEQPIER--YKLGRALYHIAQHRGFKSSKGETLSQQETNSKPSSTD---------------------------------------------------------------------------------
P. timonensis           -DR-HRCL-------HEK---LDLTVEADR--YTLGRALYHLTQRRGFLSNRLDTSAD--------------------------------------------------------------------------------------------
P. timonensis           -DR-HRCL-------HEK---LDLTVEADR--YTLGRALYHLTQRRGFLS----------------------------------------------------------------------------------------------------
A. paucivorans          -LR-AEGL-------DRP---LSF--------PEWVRVLYHITKHRGFQS----------------------------------------------------------------------------------------------------
P. marinum              -IR-KKAL-------DEA---IDP--------YEMGRAIFHINQRRGFKS----------------------------------------------------------------------------------------------------
A. denitrificans        -LR-AKGL-------QEA---LLP--------GEFARALFHINQRRGFKS----------------------------------------------------------------------------------------------------
R. syzygii              -LR-QRGL-------DQA---LTL--------PEFGRALFHLNQRRGFQS----------------------------------------------------------------------------------------------------
D. shibae               -LR-AAGL-------DEP---LPL--------PHLGRALFHLNQRRGFKS----------------------------------------------------------------------------------------------------
R. rubrum               -LR-REAV-------SRP---LAP--------DALWAALLHLAHHRGFRS----------------------------------------------------------------------------------------------------
Azospirillum B510       -AR-ERAT-------RER---LEP--------FEIGRALFHLNQRRGYKP----------------------------------------------------------------------------------------------------
C. salinarum            FLR-ADAA-------RRP---LDG--------PELFRVLHHMAAHRGIRLAELQ------------------------------------------------------------------------------------------------
N. hamburgensis         -LR-KTAL-------TDT---LPA--------HHVGRALFHLNQRRGFQS----------------------------------------------------------------------------------------------------
Bradyrhizobium BTAi1    -LR-AKAL-------DEV---LPL--------HHVGRALFHLNQRRGLFA----------------------------------------------------------------------------------------------------
W. succinogenes         -LR-HDGL-------YRL---LKG--------DELARVLIHIAKHRGYKF----------------------------------------------------------------------------------------------------
C. jejuni               -LR-FRAL-------NEL---LSK--------QDFARVILHIAKRRGYDD----------------------------------------------------------------------------------------------------
Jnet 218563121          -HH-HHHH-------HH---------------HHHHHHHHHHHHHH--------------------------------------------------------------------------------------------------------
H. mustelae             -LR-ERAL-------YRV---LDK--------EELARVILHIAKRRGYDD----------------------------------------------------------------------------------------------------
A. ebreus               -LR-RQGL-------DRL---LTP--------LEWARVIYHQCKHRGFHW----------------------------------------------------------------------------------------------------
B. smithii              -LR-VDAL-------ERK---LNN--------DELARVLLHLAKRRGFKS----------------------------------------------------------------------------------------------------
C. cellulolyticum       -LR-YNAL-------SRI---LKP--------YELVQVLTHITKRRGFKS----------------------------------------------------------------------------------------------------
HF0070_07E19            -VR-AEAV-------ERK---LTK--------QELAAVLFHLVRHRGYFP----------------------------------------------------------------------------------------------------
P. lavamentivorans      -LR-RRGL-------EEG---LSA--------YEFGRAIYHLAQHRHFKG----------------------------------------------------------------------------------------------------
N. meningitidis         -LR-AAAL-------DRK---LTP--------LEWSAVLLHLIKHRGYLS----------------------------------------------------------------------------------------------------
P. multocida            -LR-VAGL-------ERR---LSA--------IEWGAVLLHLIKHRGYLS----------------------------------------------------------------------------------------------------
A. muciniphila          -LA-SEAL-----KGHRT---LAP--------IELWHVLRWYAHNRGYDN----------------------------------------------------------------------------------------------------
Actinomyces F0310       -IRKRLVSEY---IEDET---ERR--------QCLSVAMAHIARHRGWRNSFSK------------------------------------------------------------------------------------------------
A. cellulolyticus       -ARKRLAEEY---VVDET---ERR--------RLLGYAVSHMARHRGWRNPWTT------------------------------------------------------------------------------------------------
B. longum               -VRAELATRY---IEDDE---LRR--------ESISIALRHMARHRGWRN----------------------------------------------------------------------------------------------------
A. naeslundii           -VRARLVEEK---LP-EE---LRG--------PAISMAVRHIARHRGWRN----------------------------------------------------------------------------------------------------

                        ---------------------------------HHHHHHHHHHHH---------------------------------------------------------------------------------------------------------
domains                 hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
catalytic residues      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

A. hesperidum           -LRVAGLD--------RC---LIA--------EEIARVLIHIAHRRGFQS----------------------------------------------------------------------------------------------------
Alcanivorax W11-5       -LRAMAVN--------ER---IEL--------GQLRAVLLRMGKKRGYGG----------------------------------------------------------------------------------------------------
B. zoohelcum            -LRVKALT--------EK---VGL--------EDLGKIIYKYNQLRGYAGGSLEPEKEDIFDEEQSKDKKNKSFIAFSKIVFLGEPQEEIFK----------------------------------------------------------
Porphyromonas F0450     -LRERAATAG------RR---LTL--------PELGRVLCHINQKRGYRH----------------------------------------------------------------------------------------------------
B. intestinihominis     -LRAKAVT--------QR---IEL--------NELGRVLLHLNQKRGYKS----------------------------------------------------------------------------------------------------
O. laneus               -LRKQAVT--------ED---VTR--------PELGRILYQMIQRRGFLS----------------------------------------------------------------------------------------------------
Treponema JC4           -LRIKALD--------EK---LEP--------YELGRALFNLAVRRGFKS----------------------------------------------------------------------------------------------------
Rhodovulum PH10         -LRAKALD--------ES---LTP--------HEMGRVIFHLNQRRGFRS----------------------------------------------------------------------------------------------------
consensus/95%           .hR.......   .....   .........  ..h..hh..h..cRGh......................................................
consensus/85%           .hR.......   ...c.   h........  .ch..hhhch..cRGa......................................................

#map                    --------------------------------******************----------------------------------------------------------------------------------------------------
```

```
Jnet 331001027          -----------------------------------------------HHHH--------HHH-HHE---------------------------HHHHHHHHHHHHHHHHHHHHHHHHHHH-----HHHHHHHHHHHHHHHHHHHHHHHH-
P. excrementihominis    -----------------------------------------------EADF--------KKY-IKTSFPEYS---------------------AKEILANYVEGRRAILDASKYIANLQSLGHKHRSKYLSDILQDMKRDSRITRLSEAF
W. succinogenes         -----------------------------------------------EAVFASATHSPNKKLELKDELKSEY----------------------GENAKELLAGLRVTKEILDEFDKQENQGNLPRAKYFEELGEYIATNEKVKSFFDSN
L. pneumophila          -----------------------------------------------SSEF--------RKI-LVSKVEE------------------------KKDDKELKNAVKNIKNFITGFEKNSVEGHRHRKVYFENIKSDITKDNQLDSIKKKI
S. wadsworthensis       -----------------------------------------------DKEF--------IEFAVAEGLID------------------------KTEKKAYQSALSTLRANANVLTGLRQMGHKPRSEYFKAIEADLKKDSRLAKINEAF
HTCC5015                ----------------------------------------------NKNEF--------KKL-LDTALDGKY----------------------KDEKKELSEAWGNILIASENVLKSTVDGHKSRSEYLANIKEDIKSNEELEKQISSK
F. novicida             EITSYEFELLADYLANYSESLKTQKFSYTDKQGNLKELSYYHHDKYNIQEFLKRHATINDRI-LDTLLTDDLDIWNFNFEKFDFDKNEEKL-QNQEDKDHIQAHLHHFVFAVNKIKSEMASGGRHRSQYFQEITNVLDENNHQEGYLKNF
consensus/95%           ..................................................F.........ch.h...h....................... ........h.............h......G...R..Yh..h...h.............
consensus/85%           ..................................................F.........ch.h...h....................... ........h.............h......G...R..Yh..h...h.............

A. coleocanis           ----------VRSLHVEQPP-------------------------------------------------------------------------------------------------------------------SDQYLALK-------
C. glomerans            ----------------SLSAKSARPDEALNHLRETLRVWSSERGFECSIADNGSILAMLTHPDLSPSDRRKKIAPLFDVKSDDAAADKKLGIALAGAVIGLKTEFKNIFGDFPCEDSSIYLS-------------NDEAVDAVRSACPDD
Acidaminococcus D21     ------------------KDNIGDVLSFDAFYPEFLAFLSDNGVSPWVCESKALQATLLSRNSVNDKYKALKSLIFGSQKPEDNFDANISEDGLIQLLAGKKVKVNKLFPQESNDASFTL---------------NDKEDAIEEI--LGT
V. atypica              ----------TFNSNAF-TFEDVLKQALVNITFNCFDTNSAISSISNILMESGKTKSDKAKAIERLVDTYTVFDEVNTPDKPQKEQVKEDKKTLKAFANLVLGLSANLIDLFGSVEDIDDDLKKLQ--IVGDTY-DEKRDELAKV--WGD
F. nucleatum            -----------------NLKEIKNFETLYNNLISFLEDNGINKSIDKDNIEKLEKIICDSGKGLKDKEKEFKGIFNSDKQLVAIFKLSVGSSVSLNDLFDTDEYKKEEVEKEKISFREQIY--------------EDDKPIYYSI--LGE
F. alocis               -----------------LQSAKQLRPILDTFLLSLQEEQNLSVSLSENQKDEYEEILKNRSIAKSEKVKKLKNLFEISDELEKEEKKAQSAVIENFCKFIVGNKGDVCKFLRVSKEELEIDSFSFSEGKYE----DDIVKNLEEK--VPE
S. moorei               -----------------ANGDVKTLDEAVSDFEEYLNENDIEFNIENKKEFINVLSDKHLTKKEKKISLKKLYGDITDSENINISVLIEMLSGSSISLSNLFKDIEFDGKQNLSLDSDI----------------EETLNDVVDI--LGD
C. catus                ------------------ISQIKEFKSTFEQLIQNIQDEELEWHISLDDAAIQFVEHVLKDRNLTRSTKKSRLIKQLNAKSACEKAILNLLSGGTVKLSDIFNNKELDESERPKVSFADSGY-------------DDYIGIVEAE--LAE
T. denticola            -----------------FDSENQFDTSIQALFEYLREDMEVDIDADSQKVKEILKDSSLKNSEKQSRLNKILGLKPSDKQKKAITNLISGNKINFADLYDNPDLKDAEKNSISFSKDDF----------------DALSDDLASI--LGD
P. duerdenii            ----------KFDTKNSFDKSINDLKIHLRDEYNIDLEFNNEDLIEIITDTTLNKTNKKKELKNIVGDTKFLKAISAIMIGSSQKLVDLFEDGEFEETTVKSVDFSTTAF-------------------------DDKYSEYEEA--LGD
C. mitsuokai            ----------KFNMDASNIEDKLSDIFTQFTSFNNIPYEDDEKKNLEILEILKKPLSKKAKVDEVMTLIAPEKDYKSAFKELVTGIAGNKMNVTKMILCEPIKQGDSEIKLKFSDSNY-----------------DDQFSEVEKD--LGE
S. thermophilus         -----------------FNSKNNDIQKNFQDFLDTYNAIFESDLSLENSKQLEEIVKDKISKLEKKDRILKLFPGEKNSGIFSEFLKLIVGNQADFRKCFNLDEKASLHFSKESY--------------------DEDLETLLGY--IGD
S. mutans               -----------------FDTRNNDVQRLFQEFLAVYDNTFENSSLQEQNVQVEEILTDKISKSAKKDRVLKLFPNEKSNGRFAEFLKLIVGNQADFKKHFELEEKAPLQFSKDTY--------------------EEELEVLLAQ--IGD
S. pyogenes             -----------------LNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTY--------------------DDDLDNLLAQ--IGD

                        ------------------------HHHHHHHHHHHHHHH------------HHHHHH----HHHHHHHHH----------HHHHHHHHHH------HHHH--------------H--------------------HHHHHHHHHH--H--
domains                 hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
catalytic residues      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
domains                 1111111111111111111111|2222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222
catalytic residues      1111111111111111111111|2222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222222

B. bifidum              ----------LESSNAFKEDELLKLLGRITRYEMSEGEQNSDIEQDDENKLVAPANGQLADALCATRGSRSMRVDNALEALSAVNDLSREQRAIVKAIFAGLEGNKLDLAKIFVSKEFSSENKKILGIYFNKSDYEEKCVQIVDSGLLDD
O. kitaharae            ----------VKADSAF---DVHTFADAIQRYAESNNSDENLLGKIDEKKLSAALTDKHGSKSQRAETAETAF--DILDLQSKKQIQAIL------KSVVGNQANLMAIFGLDSSAISKDEQKNYKFSFDDADI-DEKIADSEAL--LSD
S. sanguinis            --------INP--NNINMQNSWIDFI------ESCQEVFDLEISDES---------KNIADIFKSSE-NRQEKVKKILPYFQ--QELLKKDKSIFKQLLQLLFGLKTKFKDCFELEEEPD--------LNFSKENYDENLENFL--GSLE
F. fructosus            --------KIT-IGSTYNPEDLANAI------EVYADEKGLSWELNNPEQL-----TEIISGEAGYGLNKSMKADEALKLFE---FDNNQDKVAIKTLLAGLTGNQIDFAKLFG-KDISDKDEAKLWKLKLDDEALEEKSQTIL--SQLT
E. yurii                --------KGD--GDIQGMDVFYN--------DLVESCEYFEIELPRITNI-----DNFEKILSQKGKSRTKILEELSEELS----ISKKDKSKYN-LIKLISGLEASVVELYNIEDIQDENKKI--KIGFRESDYEESSLKVK--EIIG
L. farciminis           --------KFN--ISNLNNNLSKELI------ELNQQLIKYDISFPDDCDW-----NHISDILIGRG-NATQKSSNILKDFT----LDKETKKLLKEVINLILGNVAHLNTIFKTSLTKDEEK-----LNFSGKDIESKLDDLD--SILD
S. pseudintermedius     --------KIENLTNNDNMHDFVELI------ETYENLNNIKLNLDYEKT------KVIYEILKDNEMTKNDRAKRV-----------KNMEKKLEQFSIMLLGLKFNEGKLFNHADNAEELKGANQSHTFAD-NYEENLT-----PFLT
L. rhamnosus            ----------AKDFNTDKVDLEDIFPALTEAYAQVYPDLELTFDLAKADDFKAKLLDEQATPSDTQKALVNLLLSSDGEKEIVKKRKQVLTEFAKAITGLKTKFNLALGTEVDEADASNWQFSMGQL--------DDKWSNIETS--MTD
F. magna                ----------ISHVTDDKPLKEQLILLINDLLKIELEEELMDSIFEILADVNEKRTDKKNNLKELIKGQDFNKQEGNILNSIFESIVTGKAKIKNIISDEDILEKIKEDNKEDFVLTGDSY--------------EENLQYFEEV--LQE
P. antarcticus          --------WSEAASAEGMDDQLLELV------TRYAELENLSPLDLSESQW-----KAAETLLLNRNLTKTDQSKELTAMFG----------KEYEPFCKLVAGLGVSLHQLFPSSEQALAYKETKTKVQLSNENVEEVM------ELLL
consensus/95%                   ................................................................................................................................-.............
consensus/85%                   ................................................................................................................................-.............
Jnet 227501312          ----------------HHHHHHHHHHHHHH---------HHHHH-HHHHH----HHHHHHHHHHHHHH-----HHHHHHH----HHH----E----------------------------H--------------HHHHHHHHH------
M. mobile               ---------------------------------------------------------------------------------------------------------------------------------------DTKSKEMNKN-----
M. gallisepticum        -------------------------------------------------------------------------------------------------------------------------------------------DF---------
M. synoviae             ---------------------------------------------------------------------------------------------------------------------------------------EKLTDEFVSNSF---
M. ovipneumoniae        ----------------------------------------------------------------------------------------------------------------------------------------NVADKYEGIEH---
M. canis                -----------------------------------------------------------------------------------------------------------------------------------------TKEDFDQQTVESM
consensus/95%                                                                                                                                                  ....c..........
consensus/85%                                                                                                                                                  ....c..........

E. rectale              ---------------------------------------------------------------------------------------------------------------------------------------VSGSSAYANG-----
E. faecalis             ---------------------------------------------------------------------------------------------------------------------------------------DTEAAKESQD-----
S. thermophilus         ---------------------------------------------------------------------------------------------------------------------------------------DDGNSSVGD------
S. lugdunensis          ---------------------------------------------------------------------------------------------------------------------------------------NDDVGNELS------
E. dolichum             ---------------------------------------------------------------------------------------------------------------------------------------AKEAGETKKV--LSM
R. albus                ---------------------------------------------------------------------------------------------------------------------------------------PDERNEYKEE---HD
R. inulinivorans        ---------------------------------------------------------------------------------------------------------------------------------------EKEGGAVLKA--TTE
L. coryniformis         ---------------------------------------------------------------------------------------------------------------------------------------ENEADKKQMGQLLN-
I. polytropus           ---------------------------------------------------------------------------------------------------------------------------------------ADDTGKLLSG--IQE
consensus/95%                                                                                                                                                  ...............
consensus/85%                                                                                                                                                  ...............

Bacteroides sp.20       ---------------------------------------------------------------------------------------------------------------------------------------ERKGKDTWYNVHLEN
I. album                -------------------------------------------------------------------------------------------------------------------------------------TDGRKGKKR-FKVILED
B. fragilis             ---------------------------------------------------------------------------------------------------------------------------------------QNELDRIWEK--QKE
N. salsuginis           ---------------------------------------------------------------------------------------------------------------------------------------SEKKADERRQ--VYN
E. minutum              ---------------------------------------------------------------------------------------------------------------------------------------EAQKENTENKKIAGN
S. globus               ---------------------------------------------------------------------------------------------------------------------------------------SSDDKKLEE------
M. trichosporium        ---------------------------------------------------------------------------------------------------------------------------------------ADDTSKMKRA--VNE
F. branchiophilum       ---------------------------------------------------------------------------------------------------------------------------------------SNEKNGYAKD--LDE
Prevotella C561         ------------------------------------------------------------------------------------------------------------------------------------------------------
P. timonensis           ------------------------------------------------------------------------------------------------------------------------------------------------------
P. timonensis           -------------NRLDTSAD------------------------------------------------------------------------------------------------------------------NKEDGVVKSG--ISQ
A. paucivorans          -------------NRRNPVEDGQERSR------------------------------------------------------------------------------------------------------------QEEEGKLLSG--VGE
P. marinum              -------------NRKSA---------------------------------------------------------------------------------------------------------------------DNEAGVVKQ------
A. denitrificans        -------------NRKTDKK-------------------------------------------------------------------------------------------------------------------DNDSGVLKKA--IGQ
R. syzygii              -------------NRKT-DRAT-----------------------------------------------------------------------------------------------------------------AKESGKVKNA--IAA
D. shibae               -------------NRKTDRG-------------------------------------------------------------------------------------------------------------------DNESGKIKD------
R. rubrum               -------------NRIDKRERAAAKALAKAKPAKATAKATAPAKE------------------------------------------------------------------------------------------ADDEAGFWEG-----
Azospirillum B510       -------------VRTATKP-------------------------------------------------------------------------------------------------------------------DEEAGKVKE------
C. salinarum            --------------EVDPPP-------------------------------------------------------------------------------------------------------------------ESDADDAAP------
N. hamburgensis         -------------NRKT-DSK------------------------------------------------------------------------------------------------------------------QSEDGAIKQA-----
Bradyrhizobium BTAi1    -------------NRAI-EQG------------------------------------------------------------------------------------------------------------------DKDASAIKAA--AGR
W. succinogenes         -------------IGDDEA--------------------------------------------------------------------------------------------------------------------DEESGKVKKA-----
C. jejuni               -------------IKNS-D--------------------------------------------------------------------------------------------------------------------DKEKGAILKA--IKQ
Jnet 218563121          -----------------------------------------------------------------------------------------------------------------------------------------HHHHHHHH--HHH
H. mustelae             -------------ITYGVE--------------------------------------------------------------------------------------------------------------------DNDSGKIKKA--IAE
A. ebreus               -------------TSKAEEAKADS---------------------------------------------------------------------------------------------------------------DAEGGRVKQG--LAH
B. smithii              -------------NRKSE-------------------------------------------------------------------------------------------------------------------RNSKESSEFLKN--IEE
C. cellulolyticum       -------------NRKE-DLS------------------------------------------------------------------------------------------------------------------TTKEGVVITS--IKN
HF0070_07E19            -------------NTKK-LPP-----------------------------------------------------------------------------------------------------------DDESDSADEEQGKINRA--TSR
P. lavamentivorans      --------------RELEESDTPD---------------------------------------------------------------------------------------------------------------PDVDDEKEAA-----
N. meningitidis         -------------QRKNE-GETAD---------------------------------------------------------------------------------------------------------------K-ELGALLKG--VAD
P. multocida            -------------KRKNE-SQTNN---------------------------------------------------------------------------------------------------------------K-ELGALLSG--VAQ
A. muciniphila          -------------NASWSNSL------------------------------------------------------------------------------------------------------------------SEDGGNGEDT-----
Actinomyces F0310       ----------------VDTLLLEQAP-------------------------------------------------------------------------------------------------------------SDRMQGLK-------
A. cellulolyticus       ----------------IKDLKNLPQP-------------------------------------------------------------------------------------------------------------SDSWERTRESLEARY
B. longum               -------------PYRQVDSLISDNPY------------------------------------------------------------------------------------------------------------SKQYGELKEK-----
A. naeslundii           -------------PYSKVESLLSPAEE------------------------------------------------------------------------------------------------------------SPFMKALRER-----

                        ------------------------------------------------------------------------------------------------------------------------------------------------------
domains                 hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
catalytic residues      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

A. hesperidum           -------------NRK-------SEIK------------------------------------------------------------------------------------------------------------DSDAGKLLKA-----
Alcanivorax W11-5       -------------TFKAVRKVGEAGEV------------------------------------------------------------------------------------------------------------ASGASRLEEE-----
B. zoohelcum            -------------NKKLNRRAIIVETE------------------------------------------------------------------------------------------------------------EGNFEGSTFL----E
Porphyromonas F0450     -------------VKSDAAAIVGDEGE------------------------------------------------------------------------------------------------------------KKKDSNSAYL----A
B. intestinihominis     -------------IKS-------DFSG------------------------------------------------------------------------------------------------------------DKKITDYVKT-----
O. laneus               -------------SRK-----------------------------------------------------------------------------------------------------------------------GKEEGKIFTG-----
Treponema JC4           -------------NRKDGSREEVSEKK------------------------------------------------------------------------------------------------------------SPDEIKTQAD-----
Rhodovulum PH10         -------------NRK-------ADRQ------------------------------------------------------------------------------------------------------------DREKGKIAEG-----
consensus/95%                        ................................                                                                                   ......................
consensus/85%                        ................................                                                                                   ......................

#map                    ------------------------------------------------------------------------------------------------------------------------------------------------------
```

```
Jnet 331001027          ---------------------HHHHHH------------------HHHHHHH--------------HHHHHHHHHHHHHH---------HHHHHHHHHHHHH------------------------------------------------
P. excrementihominis    G----------------STDNLWRIIGNIS------------NLQERAVRWYFNDAKFEQGQEQLDAVKLKNVLVRALKY-LRSDDKEWSASQKQIIQSLEQ------------------------------------------------
W. succinogenes         SL---------------KLTDMTKLIGNIS------------NYQLKELRRYFNDKEMEKG-DIWIPNKLHKITERFVRS-WHPKNDADRQRRAELMKDLKS------------------------------------------------
L. pneumophila          -----------------PSVCLSNLLGHLS------------NLQWKNLHRYLAKNP-----KQFDEQTFGNEFLRMLKNFRHLKGSQESLAVRNLIQQLEQ------------------------------------------------
S. wadsworthensis       G----------------GAERLARLLGNLS------------NLQLRAERWYFNAPDIMKD-RGWEPDRFKKTLVRAFKF-FHP-AKDQNKQHLELIKQIEN------------------------------------------------
HTCC5015                -----------------EIDGFYNLVGHLS------------NFQLRLLRKYFNDPNMSGV-SYWDEKRLEKYFYQWVQG-WHTKGGTDEAEKKNIILKTKG------------------------------------------------
F. novicida             CENLHN-----KKYSNLSVKNLVNLIGNLS------------NLELKPLRKYFNDKIHAKA-DHWDEQKFTETYCHWILG-EWRVGVKDQDKKDGAKYSYKDLCNELKQK----------------------------------------
consensus/95%           ......     ..........h..hhG.hS            Nh..c..c.Yh...........h....h......hh.................h....c.........
consensus/85%           ......     ..........h..hhG.hS            Nh..c..c.Yh...........h....h......hh.................h....c.........

A. coleocanis           ---------ERVEAKTLLQMPEGATPAEMV----------VALDLSVDVNLRPKNREKTDTRPENKKPGF--------------------------------------------------------------------------------
C. glomerans            ---------CAELFDRLCEVYSAYVLQGLL-----SYAPGQTISANMVEKYRRYGEDLALLKKLVKIYAPDQYRMFFSGATYPGTGIYDAAQARGYTKYNL-------------------------------------------------
Acidaminococcus D21     LTPD-----ECEWIAHIRRLFDWAIMKHAL-------KDGRTISESKVKLYEQHHHDLTQLKYFVKTYLAKEYDDIFRNVDSETTKNYVAYSYHVKEVKGTLP-----------------------------------------------
V. atypica              ---------EIHIIDDCKSVYDAIILMSIK-------EPGLTISQSKVKAFDKHKEDLVILKSLLKLDRNVYNEMFKSDKKGLHNYVHYIKQGRTE------------------------------------------------------
F. nucleatum            ---------KIELLDIAKSFYDFMVLNNIL-------SDSNYISEAKVKLYEEHKKDLKNLKYIIRKYNKENYDKLFKDKNENNYPAYIGLNKEKD------------------------------------------------------
F. alocis               ---------KVYLFEQMKAMYDWNILVDIL-------ETEEYISFAKVKQYEKHKTNLRLLRDIILKYCTKDEYNRMFNDEKEAGSYTAYVGKLKK------------------------------------------------------
S. moorei               ---------NIDLLIHAKEVYDIAVLTSSL-------GKHKYLCDAKVELFEKNKKDLMILKKYIKKNHPEDYKKIFSSPTEKKNYAAYSQTNS--------------------------------------------------------
C. catus                ---------QYYIIASAKAVYDWSVLVEIL-------GNSVSISEAKIKVYQKHQADLKTLKKIVRQYMTKEDYKRVFVDTEEKLNNYSAYIGMTKKNGKKVDLK---------------------------------------------
T. denticola            ---------SFELLLKAKAVYNCSVLSKVI-------GDEQYLSFAKVKIYEKHKTDLTKLKNVIKKHFPKDYKKVFGYNKNEKNNNNYSGYVG--------------------------------------------------------
P. duerdenii            ---------TISLLNILKSIYDSSILENLLKDADKSKDGNKYISKAFVKKFNKHGKDLKTLKRIIKKYLPSEYANIFRNKSINDNYVAYTKSNITSNKRTK-------------------------------------------------
C. mitsuokai            ---------YVEFVDALHNVYSWVELQTIM---GATHTDNASISEAMVSRYNKHHDDLKLLKDCIKNNVPNKYFDMFRNDSEKSKGYYNYINR---------------------------------------------------------
S. thermophilus         D--------YSDVFLKAKKLYDAILLSGFL--TVTDNETEAPLSSAMIKRYNEHKEDLALLKEYIRNISLKTYNEVFKDDTKNGYAGYIDGKTNQEDFYVYL------------------------------------------------
S. mutans               N--------YAELFLSAKKLYDSILLSGIL--TVTDVGTKAPLSASMIQRYNEHQMDLAQLKQFIRQKLSDKYNEVFSDVSKDGYAGYIDGKTNQEAFYKYL------------------------------------------------
S. pyogenes             Q--------YADLFLAAKNLSDAILLSDIL--RVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFI------------------------------------------------

                        ----------HHHHHHHHHHHHHHHHHHHH-----------HHHHHHHHHHHHHHHHHHHHHHHHHHH--HHHHHHHH------HHHHH-----HHHHHHHH------------------------------------------------
domains                 hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
catalytic residues      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
domains                 222222222222222222222222222222222|1111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
catalytic residues      222222222222222222222222222222222|1111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
.
B. bifidum              E--------EREFLDRMQGQYNAIALKQLL-------GRSTSVSDSKCASYDAHRANWNLIKLQLRTKENEKDINENYGILVGWKIDS--------------------------------------------------------------
O. kitaharae            T--------EFEFLCDLKAAFDGLTLKMLL-------GDDKTVSAAMVRRFNEHQKDWEYIKSHIRNAKN----AGNGLYEKSKKFDG--------------------------------------------------------------
S. sanguinis            ED-------FSDVFAKLKVLRDTILLSGMLTY---TGATHARFSATMVERYEEHRKDLQRFK-FFIKQNLSEQDYLDIF---GRKTQNGFDVDKETKGY---------------------------------------------------
F. fructosus            DE-------EIELFHAVVQAYDGFVLIGLL-------NGADSVSAAMVQLYDQHREDRKLLKSLAQKAGLKHKRFSEIY-----------------------------------------------------------------------
E. yurii                DE-------YFDLVERAKSVHDMGLLSNII-------GNSKYLCEARVEAYENHHKDLLKIKELLKKYD--KKAYNDMFRKMTDKNYSAYVGSVNSNIAKE-------------------------------------------------
L. farciminis           DD-------QFTVLDAANRIYSTITLNEIL-------NGESYFSMAKVNQYENHAIDLCKLRDMWHTTKNEEAVEQ------SRQAYDDYINKPKYGT----------------------------------------------------
S. pseudintermedius     VE-------QSEFIERANKIYLSLTLQDIL-------KGKKSMAMSKVAAYDKFRNELKQVKDIVYKADSTRTQFKKIFVS-SKKSLKQYDATPNDQTFSS-------------------------------------------------
L. rhamnosus            Q--------GTEIFEQIQELYRARLLNGIV-------PAGMSLSQAKVADYGQHKEDLELFKTYLKKLNDHELAKTIRGLYDRYINGDDAKPFLREDFVKALTK----------------------------------------------
F. magna                ---------NITLFNTLKSTYDFLILQSIL-------KGKSTLSDAQVERYDEHKKDLEILKKVIKKYDEDGKLFKQVFKEDNGNGYVSYIGYYLNKNKKITA-----------------------------------------------
P. antarcticus          EE-------ESALLEAVQPFYQQVVLYELL-------KGETYVAKAKVSAFKQYQKDMASLKNLLDKT-FGEKVYRSYFIS-DKNSQREYQKSHKVEVLCK-------------------------------------------------
consensus/95%           ....     ...hh..h........h..hh............h.......a.........hc..h........................................
consensus/85%           ....     ...hh..h..ha....L..hh............hu.u.h..a..a..-h..hK.hh........................................
.
Jnet 227501312          ---------HHHHHHHHHHHHHHHHH---------------HHHHHHHHHHHHHHHHHHHHHHHHH----HHH-HHHHHHHHHH------------------------------------------------------------------
M. mobile               ---------EIEIFDKNESIKNAEFPIDKI-----------------IEFYKISGKIRST------------------------------------------------------------------------------------------
M. gallisepticum        ------------------NVYPTKELAKYF---------------DKYGYYKGIIDSKEDNDNKLEEEL---------------------------------------------------------------------------------
M. synoviae             ---------PSKKLHEHYEKYGFFRGSVKL-----------------DNKLDNKKDKAKEKDEEEESDAKKES-----------------------------------------------------------------------------
M. ovipneumoniae        ---------PSILLYDFFKKNGFFKSNSSI---------------------------------------PKDL-----------------------------------------------------------------------------
M. canis                ---------PSYKLNEFYKKYGYFKGALSQ----------------------------PTESEMKDNKDLKEA-----------------------------------------------------------------------------
consensus/95%                    .....................               ............................
consensus/85%                    .....................               ............................

E. rectale              ---------LKLNAKELETHYPCEIQQERL---------------NTIGKYRGQSQIINENGEVLD------------------------------------------------------------------------------------
E. faecalis             ---------YKEQVRKNAQLLTKYTPGQIQ-----------------LQRLKENNRVKTGINAQGNY-----------------------------------------------------------------------------------
S. thermophilus         ---------YAQIVKENSKQLETKTPGQIQ-----------------LERYQTYGQLRGDFTV-EKDGKKHR------------------------------------------------------------------------------
S. lugdunensis          ---------TKEQLNKNSKLLKDKFVCQIQ-----------------LERMNEGQVRG--------------------------------------------------------------------------------------------
E. dolichum             ---------NDQLL------KSGKYVCEIQ-----------------KERLRTNGHIRG-------------------------------------------------------------------------------------------
R. albus                ---------SIVLISNLMNEGGYCTPAEMI---------------CNCREFDEPNSVYRKFHNSAASK----------------------------------------------------------------------------------
R. inulinivorans        ---------NQKIM----QEKGYRTVGEML--------------YLDEAFHTECLWNEKGY-----------------------------------------------------------------------------------------
L. coryniformis         ---------SIEEIRKAIAEKGYRTVGELY---------------LKDPKYNDHKRNKGYIDG---------------------------------------------------------------------------------------
I. polytropus           ---------NKKIM----EEKGYLTIGDMV---------------AKDPKFNTHVRNKAGS-----------------------------------------------------------------------------------------
consensus/95%                    ....h................              ......h.....................
consensus/85%                    ....h..............h.              .....ch.....................

Bacteroides sp.20       GM-------IYRRTSNIPLDWEGKTKEFIV-----------------TTDLEADGSPKKDKEGNIKRSFRAPKDDDWTLIKKKTEADIDKI-----------------------------------------------------------
I. album                SR-----IEPYEIERKEKPDWEGKEYTFLV-----------------TQKLE-KGKFKQNKPD-------LPKEEDWALCTTALDNRMG-------------------------------------------------------------
B. fragilis             YYPEILTDVLKEELRGKKRDAVWAICAKYF---------------VWKENYTEWNKEKGKTEQQEREHKLEGIYSKRKRDEAKRENLQWRVNGLKEKL----------------------------------------------------
N. salsuginis           ---------ALRHLEELRKKYGGETIAQTI---------HRAVEAGDLRSYRNHDD----------------------------------------------------------------------------------------------
E. minutum              ---------IKRLM----AKKNYRTYGEYL-----------------YKEFFENKEKHKREKISNAANN---------------------------------------------------------------------------------
S. globus               ---------QQAMTEQLAKEKGISTFIEVL------------------TAFNTNGLIGYRNSESVKSK----------------------------------------------------------------------------------
M. trichosporium        ---------TREKLARFGSAAKMLVEDESF-----------------VLRQTPTKNGASEIVRRFRNREGD-------------------------------------------------------------------------------
F. branchiophilum       GDKKLVSLYKSLLAILKKNESDFENCKSEI--------------IERLNKKDVLGKVKGKIKDISQA-----------------------------------------------------------------------------------
Prevotella C561         ------------------------------------------------EIPDVAGAMKASEEKLSKGLSTY-------------------------------------------------------------------------------
P. timonensis           ------------------------------------------------NKED--GVVKSGISQLSTE-----------------------------------------------------------------------------------
P. timonensis           ---------LSTEM----EEAGCEYLGDYF-----------------YKLYDAQGNKVRIR-----------------------------------------------------------------------------------------
A. paucivorans          ---------NERLL----REGGYRTAGEML---------------ARDPKFQDHRRNRAGD-----------------------------------------------------------------------------------------
P. marinum              ---------SIADLEMKLGEAGARTIGEFL-----------------ADRQATNDTVRARRLSGTNAL----------------------------------------------------------------------------------
A. denitrificans        ---------LRQQMAEQGS----RTVGEYL-----------------WTRLQ------QGQGVRARYREKPYTTEEGKKRIDKS------------------------------------------------------------------
R. syzygii              ---------FRAGM------GNARTVGEAL-----------------ARRLEDGRPVRARMVGQGKDEH---------------------------------------------------------------------------------
D. shibae               ---------ATARLDMEMMANGARTYGEFL-----------------HKRRQKATDPRHVPSVRTRLSIANRGGPDGKEEAG--------------------------------------------------------------------
R. rubrum               ---------AEAALRQRMAASGAPTVGALL-----------------ADDLDRGQPVRMRYNQSDRD-----------------------------------------------------------------------------------
Azospirillum B510       ---------AVERLEAAIAAAGAPTLGAWF-----------------AWRKTRGETLRARLAGKGKEAA---------------------------------------------------------------------------------
C. salinarum            ---------AATEDEDGTRRAAADER---------------------AFRRLMAEHMHRHGTQPTCGEIMAGRLRETPAGAQPVTRARDGLRVGG-------------------------------------------------------
N. hamburgensis         ---------ASRLATDKGNETLGVFFADMH----------------LRKSYEDRQTAIRAELVRLGKDHLTGNARKKIWAKVRK------------------------------------------------------------------
Bradyrhizobium BTAi1    ---------LQTSM----QACGARTLGEFL---------------NRRHQLRATVRARSPVGGDVQAR----------------------------------------------------------------------------------
W. succinogenes         ----------GVVLRQNFEAAGCRTVGEWL-----------------WRERGANGKKRNKHGD---------------------------------------------------------------------------------------
C. jejuni               ---------NEEKL------ANYQSVGEYL-------------YKEYFQKFKENSKEFTNVRNKKES-----------------------------------------------------------------------------------
Jnet 218563121          ---------HHHHH------H----HHHHH-------------HHHHHHH----------------------------------------------------------------------------------------------------
H. mustelae             ---------NSKRI----KEEQCKTIGEMM-----------------YKLYFQKSLNVRNKKES--------------------------------------------------------------------------------------
A. ebreus               ---------TKALM----QAKNYRSAAEMV-----------------LAEFPDAQRNKRGQ-----------------------------------------------------------------------------------------
B. smithii              ---------NQSIL------AQYRSVGEMI---------------VKDSKFAYHKRNKLDS-----------------------------------------------------------------------------------------
C. cellulolyticum       ---------NSEML----RTKNYRTIGEMI-----------------FMETPENSNKRNKVDE---------------------------------------------------------------------------------------
HF0070_07E19            ---------LREEL----KASDCKTIGQFL--------------------AQNRDRQRNREGD---------------------------------------------------------------------------------------
P. lavamentivorans      ---------NERAATLKALKNEQTTLGAWL-----------------ARRPPSDRK----------------------------------------------------------------------------------------------
N. meningitidis         ---------NAHAL----QTGDFRTPAELA--------------LNKFEKESGHIRNQRGD-----------------------------------------------------------------------------------------
P. multocida            ---------NHQLL----QSDDYRTPAELA--------------LKKFAKEEGHIRNQRGA-----------------------------------------------------------------------------------------
A. muciniphila          ----ERVKHAQDLMDKHGTATMAETICREL-------------------KLEEGKADAPMEVSTPAYKN---------------------------------------------------------------------------------
Actinomyces F0310       ---------ERVEDRTGLQFSEEVTQGELV---------ATLLEHDGDVTIRGFVRKGGKA---TKVHGV--------------------------------------------------------------------------------
A. cellulolyticus       ---------SVSLEPGTVGQWAGYLLQRAP-----------------GIRLNPTQQSAGRRAELSNATA---------------------------------------------------------------------------------
B. longum               ---------AKAYNDDATAAEEESTPAQLV-----------------VAMLDAGYAEAPRLRWRTGSKKPDAEGY---------------------------------------------------------------------------
A. naeslundii           ---------ILATTGEVL--DDGITPGQ--------------------AMAQVALTHNISMR--------GPEGI---------------------------------------------------------------------------

                        ------------------------------------------------------------------------------------------------------------------------------------------------------
domains                 hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
catalytic residues      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

A. hesperidum           -VQENE-NLMQ--------SKGYRTVAEML--------------VSEATKTDAEGKLVHGKKHGYVSNVRNKAGEY--------------------------------------------------------------------------
Alcanivorax W11-5       -MVAL--ASVQ--------NKDSVTVGEYLAARVEHGLPSKLKVAANNEYYAPEYALFRQYLGL--PAIKGRPDCLP-------------------------------------------------------------------------
B. zoohelcum            NIKVG--DSLEL-----LINISASKSGDTITIKLPNKTNWRKKMENIENQLKEKSKEMGREFYISEFLLELLKENRWAKIR---------------------------------------------------------------------
Porphyromonas F0450     GIRAND-EKLQ---------AEHKTVGQYF------------------------AEQLRQNQSES-----PTGGISYR-IK---------------------------------------------------------------------
B. intestinihominis     -VKTRY-DELK---------EMRLTIGELF------------------------FRRLTENAFF-----RCKEQVY--------------------------------------------------------------------------
O. laneus               -KDRMV-GIDET-----RKNLQKQTLGAYL------------------------YDIAPKNGEK----YRFRTERVR-------------------------------------------------------------------------
Treponema JC4           -MQTHLEKAIK--------ENGCRTITEFL------------------------YKNQGENGGI-----RFAPGRM--------------------------------------------------------------------------
Rhodovulum PH10         -SKRLA-ETLA--------ATNCRTLGEFL-----------------------WSRHRGTPRTRSPTRIRMEGEGAKA-LY---------------------------------------------------------------------
consensus/95%           ..................................................................................................
consensus/85%           .............................h....................................................................

#map                    ------------------------------------------------------------------------------------------------------------------------------------------------------
```

```
Jnet 331001027          -----------------------------------------------------HHHHHH-----------------------EEE--HHHH-----------------------------------------------------------
P. excrementihominis    --------------------------------------------------SGDVLDVLAGLDPDRTIPPYEDQNNRRPPEDQTLYLNPKALSSEYG------------------------------------------------------
W. succinogenes         --------------------------------------------------KE-IMELLTTTEPVMTIPPYDDMNNRGAVKCQTLRLNEEYLDKHL-------------------------------------------------------
L. pneumophila          --------------------------------------------------SQDYISILEKTPPEITIPPYEARTNTGMEKDQSLLLNPEKLNNLY-------------------------------------------------------
S. wadsworthensis       --------------------------------------------------SEDIIETLCTLDPNRTIPPYEDQNNRRPPLDQTLLLSPEKLTRQYG------------------------------------------------------
HTCC5015                --------------------------------------------------AP-LLKTLKSLSADLTIPPYEDQNNRRPPKCQSVLLSDEKLTMHY-------------------------------------------------------
F. novicida             ------------------------------------------------VTKAGLVDFLLELDPCRTIPPYLDNNNRKPPKCQSLILNPKFLDNQY-------------------------------------------------------
consensus/95%                                                           .....hh..L.......TIPPY....N......Qoh.L..c.L...h.
consensus/85%                                                           .....hh..L.......TIPPY....N......Qoh.L..c.L...h.

A. coleocanis           -------------------------------------------------------------LGGKLMQSDNANELRKIAKIQGL---------DDALLREL-------------------------------------------------
C. glomerans            ---GPKKSEYKPSESMQYDDFRKAVEKLFA--KTDARADERYRMMMDRFDKQQFLRRLKTSDNGSIYHQLHLEELKAIVENQGRFYPFL----KRDADKLV-------------------------------------------------
Acidaminococcus D21     ---KNKATQEEFCKYVLGKVKNI---------ECSEADKVDFDEMIQRLTDNSFMPKQVSGENRVIPYQLYYYELKTILNKAASYLPFL----TQCGKDAI-------------------------------------------------
V. atypica              ---ETSCSREDFYKYTKKIVEGLAD-------SKDKEY------ILNEIELQTLLPLQRIKDNGVIPYQLHLEELKVILDKCGPKFPFL-----HTVSDGF-------------------------------------------------
F. nucleatum            ---KKEVVEKSRLKIDDLIKVIKGYLPKPE--RIEEKDKTIFNEILNKIELKTILPKQRISDNGTLPYQIHEVELEKILENQSKYYDFL--NYEENGVSTK-------------------------------------------------
F. alocis               ---NNKKYWIEKKRNPEEFYKSLGKLLDKI--EPLKEDLEVLTMMIEECKNHTLLPIQKNKDNGVIPHQVHEVELKKILENAKKYYSFL----TETDKDGY-------------------------------------------------
S. moorei               ---KNVCSQEEFCLFIKPYIR-----------DMVKSENEDEVRIAKEVEDKSFLTKLKGTNNSVVPYQIHERELNQILKNIVAYLPFM----NDEQEDIS-------------------------------------------------
C. catus                ---SKQCTQADFYDFLKKNV------------IKVIDHKEITQEIESEIEKENFLPKQVTKDNGVIPYQVHDYELKKILDNLGTRMPFI-----KENAEKI-------------------------------------------------
T. denticola            ---VCKTKSKKLIINNSVNQEDFYKFLKTI--LSAKSEIKEVNDILTEIETGTFLPKQISKSNAEIPYQLRKMELEKILSNAEKHFSFL----KQKDEKGL-------------------------------------------------
P. duerdenii            ---ASKFTKQEDFYKFIKKHLDTIKETKLN--SSENEDLKLIDEMLTDIEFKTFIPKLKSSDNGVIPYQLKLMELKKILDNQSKYYDFL--NESDEYGTVK-------------------------------------------------
C. mitsuokai            ---PSKAPVDEFYKYVKK--------------CIEKVDTPEAKQILNDIELENFLLKQNSRTNGSVPYQMQLDEMIKIIDNQAEYYPIL-----KEKREQL-------------------------------------------------
S. thermophilus         --------------------------------KKLLAEFEGADYFLEKIDREDFLRKQRTFDNGSIPYQIHLQEMRAILDKQAKFYPFL-----AKNKERI-------------------------------------------------
S. mutans               --------------------------------KGLLNKIEGSGYFLDKIEREDFLRKQRTFDNGSIPHQIHLQEMRAIIRRQAEFYPFL-----ADNQDRI-------------------------------------------------
S. pyogenes             --------------------------------KPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFL-----KDNREKI-------------------------------------------------

                        --------------------------------HHHHHHH----HHHHHHHH----------------HHHHHHHHHHHHHHH----HHH-----HHHHHHH-------------------------------------------------
domains                 hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
catalytic residues      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
domains                 111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
catalytic residues      111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
.
B. bifidum              ---GQRKSVRGESAYENMRKKANVFFKKMI--ETSDLSETDKNRLIHDIEEDKLFPIQRDSDNGVIPHQLHQNELKQIIKKQGKYYPFL----LDAFEKDG-------------------------------------------------
O. kitaharae            ---INAAYLALQSDNEDDRKKAKKIFQDEI--SSADIPDDVKADFLKKIDDDQFLPIQRTKNNGTIPHQLHRNELEQIIEKQGIYYPFL----KDTYQENSHE-----------------------------------------------
S. sanguinis            ---VGYITNKMVLTNPQKQKTIQQNFYDYI--SGKITGIEGAEYFLNKISDGTFLRKLRTSDNGAIPNQIHAYELEKIIERQGKDYPFL----LENK-----------------------------------------------------
F. fructosus            -------EQLALATDEATIKNGISTARELV--EESNLSKEVKEDTLRRLDENEFLPKQRTKANSVIPHQLHLAELQKILQNQGQYYPFL----LDTFEKED-------------------------------------------------
E. yurii                ---RRSVDKRKI---EDLYKYIEDTALKNIPDD-----NKDKIEILEKIKLGEFLKKQLTASNGVIPNQLQSRELRAILKKAENYLPFL----KEKGEKNL-------------------------------------------------
L. farciminis           ---------------KELYTSLKKFLKVALPTN-------LAKEAEEKISKGTYLVKPRNSENGVVPYQLNKIEMEKIIDNQSQYYPFL----KENK-----------------------------------------------------
S. pseudintermedius     ---LCLFDQYLIRP-KKQYSLLIKELKKIIPQD---------SELYFEAENDTLLKVLNTTDNASIPMQINLYEAETILRNQQKYHAEI----TDEM-----------------------------------------------------
L. rhamnosus            ------------------------------EVT--AHPNEVSEQLLNRMGQANFMLKQRTKANGAIPIQLQQRELDQIIANQSKYYDWLA--APNPVEAHR-------------------------------------------------
F. magna                ---KKKISNIEFTKYVKGILE---------KQC--DCEDEDVKYLLGKIEQENFLLKQISSINSVIPHQIHLFELDKILENLAKNYPSFN--NKKEEFTKI-------------------------------------------------
P. antarcticus          ---LDQFNKEAKFA-ETFYKDLKKLLEDKSKTSIGTTEKDEMLRIIKAIDSNQFLQKQKGIQNAAIPHQNSLYEAEKILRNQQAHYPFI----TTEW-----------------------------------------------------
consensus/95%              ..................................................hh.......Nu.h..Q....Eh..Ih......h..h. ............
consensus/85%              .........................................hh.ch....hh.......Nu.hPhQh...Ehc.Ih......h.hh. ............
.
Jnet 227501312          --------HHHHHHHHHHHH-------------------HHHHHHHHHHHH------E--------HHHHHHHHHHHHHHHHH-----------------H-------------------------------------------------
M. mobile               -------------------------------------------------------------INLKFGHQDYLKEIKQVFEKQNIDFMNY-----EKFAMEE-------------------------------------------------
M. gallisepticum        -------------------------------------------------------------TKYKFSNKHWLEEVKKVLSNQTGL--------PEKFKEEY-------------------------------------------------
M. synoviae             -------------------------------------------------------------EELIFSNKQWINEIVKVFENQSYL--------TESFKEEY-------------------------------------------------
M. ovipneumoniae         -------------------------------------------------------------GGYSFSNLQWVNEIKKLFEVQEIN---------PEFSEKF-------------------------------------------------
M. canis                -------------------------------------------------------------FFFDFSNKEWLKEINYFFNVQKNILS-------ETFIEEF-------------------------------------------------
consensus/95%                                                                        ..h.Fu...ah.Eh..hh..Q.......    ...F..c.
consensus/85%                                                                        ..h.Fu...ah.Eh..hh..Q.......    ...F..c.

E. rectale              -------------------------------------------------------------LSNVFTIGAYRKEIQRVFEIQKKYHPEL----TDEFCDGY-------------------------------------------------
E. faecalis             -------------------------------------------------------------QLNVFKVSAYANELATILKTQQAFYPNEL---TDDWIALF-------------------------------------------------
S. thermophilus         -------------------------------------------------------------LINVFPTSAYRSEALRILQTQQEFNPQI----TDEFINRY-------------------------------------------------
S. lugdunensis          -------------------------------------------------------------EKNRFKTADIIKEIIQLLNVQKNFHQL-----DENFINKY-------------------------------------------------
E. dolichum             -------------------------------------------------------------HENNFKTRAYVDEAFQILSHQDL---------SNELKSAI-------------------------------------------------
R. albus                -------------------------------------------------------------NHYLITRHMLVKEVDLILENQSKYYGIL----DDKTIAKI-------------------------------------------------
R. inulinivorans        ---------------------------------------------------VLTPRNRPDDYKHTILRSMLVEEVHAIFAAQRAHGNQKA---TEGLEEAY-------------------------------------------------
L. coryniformis         -------------------------------------------------------------YLSTPNRQMLVDEIKQILDKQRELGNEKL---TDEFYATY-------------------------------------------------
I. polytropus           -------------------------------------------------------------YLFSFSRKLLEDEVRKIQAKQKELGNTHF---TDDVLEKY-------------------------------------------------
consensus/95%                                                              ...................h..Eh..h...Q........   .......h
consensus/85%                                                              ..............h...hh..Eh..hh..Q..h.....   .-.h...h

Bacteroides sp.20       -----------------------------------KMTVGAYIYDTLLQKPDQ-KIRGK--LVRTIERKYYKNELYQILKTQSEFHEELR---DKQLYIACL---NE----LYPNNE---------------------------------
I. album                -----------------------------------SKHPGEFFFDELLKAFKE-K-RGYKIRQYPVNRWRYKKELEFIWTKQCQLNPELN---NLNINKEIL---RKLATVLYPSQSKF-------------------------------
B. fragilis             ---SLEQLVIVFQEMNTQINNSSGYLGAISDRS--KELYFNKQTVGQYQMEMLDKNPNASLRNMVFYRQDYLDEFNMLWEKQAVYHKEL----TEELKKEI-------------------------------------------------
N. salsuginis           -------------------------------------------------------------YEKMIRREDIEEEIEKVLLRQAELGALGL--PEEQVSELI-------------------------------------------------
E. minutum              -------------------------------------------------------------HKFSPTRDLVIKEAEAILKKQAGKDGFHKE-LTEEYIEKL-------------------------------------------------
S. globus               -------------------------------------------------------------GVPVPTRDIISNEIDVLLQTQKQFYQEIL---SDEYCDRI-------------------------------------------------
M. trichosporium        -------------------------------------------------------------YSRSLLRDDLAAEMRALFTAQARFQSAIA---TADLQTAF-------------------------------------------------
F. branchiophilum       --------------------------------MLDGNYKTLGQYFYSLY----SKEKIR--NQYTSREEHYLSEFITICKVQGIDQINE----EEKINEKK-------------------------------------------------
Prevotella C561         --------------------------------MKEHNLLTVGAAFAQLED---EGVRVRNNNDYRAIRSQFQHEIETIFKFQ--QGLSV----ESELYERL-------------------------------------------------
P. timonensis           --------------------------------MEEAGCEYLGDYFYKLYDAQGNKVRIR--QRYTDRNKHYQHEFDAICEKQ-----------------EL-------------------------------------------------
P. timonensis           -------------------------------------------------------------QRYTDRNKHYQHEFDAICEKQEL---------SSELIEDL-------------------------------------------------
A. paucivorans          -------------------------------------------------------------YSHTLSRSLLLEEARRLFQSQRTLGNPHA---SSNLEEAF-------------------------------------------------
P. marinum              -------------------------------------------------------------YEFYPDRYMLEQEFDTLWAKQAAFNPSLY---IEAARERL-------------------------------------------------
A. denitrificans        -------------------------------------------------------------YDLYIDRAMIEQEFDALWAAQAAFNPTLF---HEAARADL-------------------------------------------------
R. syzygii              -------------------------------------------------------------YELYIAREWIAQEFDALWASQQRFHAEVL---ADAARDRL-------------------------------------------------
D. shibae               -------------------------------------------------------------YDFYPDRRHLEEEFHKLWAAQGAHHPEL----TETLRDLL-------------------------------------------------
R. rubrum               -------------------------------------------------------------GVVAPTRALIAEELAEIVARQSSAYPGL------DWPAVT-------------------------------------------------
Azospirillum B510       -------------------------------------------------------------YPFYPARRMLEAEFDTLWAEQARHHPDLL---TAEAREIL-------------------------------------------------
C. salinarum            -------------------------------------------------------------GVAVPTRALIEQEFDAIRAIQAPRHPDL-------PWDSL-------------------------------------------------
N. hamburgensis         ------------------------------RLF--GDEVLPRADAPHGVRARATITGTKASYDYYPTRDMLRDEFNAIWAGQSAHHATI----TDEARTEI-------------------------------------------------
Bradyrhizobium BTAi1    -------------------------------------------------------------YEFYPTRAMVDAEFEAIWAAQAPHHPTM----TAEAHDTI-------------------------------------------------
W. succinogenes         -------------------------------------------------------------YEISIHRDLLVEEVEAIFVAQQEMRSTIA---TDALKAAY-------------------------------------------------
C. jejuni               -------------------------------------------------------------YERCIAQSFLKDELKLIFKKQREFGFSF----SKKFEEEV-------------------------------------------------
Jnet 218563121          -------------------------------------------------------------------HHHHHHHHHHHH--------------HHHHHHHH-------------------------------------------------
H. mustelae             -------------------------------------------------------------YNRCVGRSELREELKTIFQIQQELKSPWV---NEELIYKLLGNP---------------------------------------------
A. ebreus               -------------------------------------------------------------YDKALSRVLLGEELALLFATQRRLGNPH----ASDFFEKL-------------------------------------------------
B. smithii              -------------------------------------------------------------YSNMIARDDLEREIKLIFEKQREFNNPVC---TERLEEKY-------------------------------------------------
C. cellulolyticum       -------------------------------------------------------------YIHTIAREDLLNEIKYIFSIQRKLGSPFV---TEKLEHDF-------------------------------------------------
HF0070_07E19            -------------------------------------------------------------YSNLMARKLVFEEALQILAFQRKQGHE-L---SKDFEKTY-------------------------------------------------
P. lavamentivorans      -------------------------------------------------------------RGIHAHRNVVAEEFERLWEVQSKFHPALK---SEEMRARI-------------------------------------------------
N. meningitidis         -------------------------------------------------------------YSHTFSRKDLQAELILLFEKQKEFGNPHV---SGGLKEGI-------------------------------------------------
P. multocida            -------------------------------------------------------------YTHTFNRLDLLAELNLLFAQQHQFGNPHC---KEHIQQYM-------------------------------------------------
A. muciniphila          -------------------------------------------------------------LNTAFPRLIVEKEVRRILELSAPLIPGL----TAEIIELI-------------------------------------------------
Actinomyces F0310       -------------------------------------------------------------LEGKYMQSDLVAELRQICRTQRV---------SETTFEKL-------------------------------------------------
A. cellulolyticus       -------------------------------------------------------------FETRLRQEDVLWELRCIADVQGL---------PEDVVSNV-------------------------------------------------
B. longum               -------------------------------------------------------------LPVRLMQEDNANELKQIFRVQRVP--------ADEWKPLF-------------------------------------------------
A. naeslundii           -------------------------------------------------------------LG-KLHQSDNANEIRKICARQGVS--------PDVCKQLL-------------------------------------------------

                        ------------------------------------------------------------------------------------------------HHHHH-------------------------------------------------
domains                 hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh|PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
catalytic residues      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh|PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP

A. hesperidum           --------------------------------------------------------------RHTVSRQAIVDEVRKIFAAQRAL---GNDVMSEELEDSY-------------------------------------------------
Alcanivorax W11-5       --------------------------------------------------------------NMYALRHQIEHEFERIWATQSQF---HDVMKDHGVKEEI-------------------------------------------------
B. zoohelcum            --------------------------------------------------------------NNTILRARYESEFEAIWNEQVKH---YPFLENLDKKTL------------------------------------IEIVSFIFPGEKES
Porphyromonas F0450     --------------------------------------------------------------DQIFSRQCYIDEYDQIMAVQRVH---YPDILTDEFIRML-------------------------------------------------
B. intestinihominis     ------------------------------------------------------------------PRQAYVEEFDCIMNCQRKF---YPDILTDETIRCI-------------------------------------------------
O. laneus               --------------------------------------------------------------ARYTLRDMYIREFEIIWQRQAGHLGLAHEQATRKKNIFL-------EGSATNVRNSKLITHLQAKYGRGHVLIEDTRITVTFQLPLKE
Treponema JC4           --------------------------------------------------------------TYYPTRKMYEEEFNLIRSKQEKY---YPQVDWDDIYKAI-------------------------------------------------
Rhodovulum PH10         --------------------------------------------------------------AFYPTREMVRAEFERLWTAQSRF---APDLLTPERHEEI-------------------------------------------------
consensus/95%              ...................................................................h..Eh..h...Q....................................................................
consensus/85%              ................................................................R..h..Eh..hh..Q..................h.................................................

#map                    ------------------------------------------------------------------------------------------------------------------------------------------------------
```

```
Jnet 331001027          ---------------------HHHHHHHHHH-----------------------------------------------------HHH----HHHHHHHH-------E--------------------------EEE----------HHHH
P. excrementihominis    ---------------------EKWKSWANKF-----------------------------------------------------AGAYPLLTEDLTEILKNTDRKSR--------------------------IKIRSDVLPDSDYRLAY
W. succinogenes         ---------------------PNWRDIAKRL-------------------------------------------------------NHGKFNDDLADSTVKGYS------------------------------------------EDST
L. pneumophila          ---------------------PNWRNLIPGI-----------------------------------------------------IDAHPFLEKDLEHTKLRDRKR-------------------------------IISPSKQDEKRDSY
S. wadsworthensis       ---------------------EIWKTWSARL-----------------------------------------------------TSAEPTLAPAAEILERSTDRKSR-------------------------VAVNGHEPLPTLAYQLSY
HTCC5015                ---------------------PKWKEWVGQL--------------------------------------------------VK-QNDNAYLNENVT---------------------------------------------------LAN
F. novicida             ---------------------PNWQQYLQEL--------------------------------------------------KKLQSIQNYLDSFETDLKVLKSSKDQPYF--V---------------EYKSSNQQIASGQRDYKDLDAR
consensus/95%                                ..W..h...h                                                  .........h...................  .               ....................u.
consensus/85%                                ..W..h...h                                                  .........h...................  .               ....................u.

A. coleocanis           -----------------IELVFAADSPKGA------------------------------------------------------------SGELVGYDV------LPGQHGK------------------------RR------AEKAHP
C. glomerans            -------------------SLVSFRIPYYVGPL---------STRNARTDQHGEN-RFAWSERKPGMQDEPIFPWNWESIIDRSKSAEKFILRMTGMCT------YLQQE-P------------------------VL------P-KSSL
Acidaminococcus D21     ------------SNQDKLLSIMTFRIPYFVGPL---------RKDNSEHAWL-------------ERKAGKIYPWNFNDKVDLDKSEEAFIRRMTNTCT------YYPGE-D------------------------VL------P-LDSL
V. atypica              ------------SVTEKLIKMLEFRIPYYVGPL---------NTHHNIDNGG--------FSWAVRKQAGRVTPWNFEEKIDREKSAAAFIKNLTNKCT------YLFGE-D------------------------VL------P-KSSL
F. nucleatum            ---------------DKLLKTFKFRIPYYVGPL---------NSYHKDKGGN---------SWIVRKEEGKILPWNFEQKVDIEKSAEEFIKRMTNKCT------YLNGE-D------------------------VI------P-KDSF
F. alocis               ------------SVVQKIESIFRFRIPYYVGPL---------STRHQEKGSN-------VWMVRKPGREDRIYPWNMEEIIDFEKSNENFITRMTNKCT------YLIGE-D------------------------VL------P-KHSL
S. moorei               -------------VVDKIKLIFKFKIPYYVGPL---------NTKSTRSWVY--------------RSDEKIYPWNFSNVIDLDKTAHEFMNRLIGRCT------YTNDP--------------------------VL------P-MDSL
C. catus                ------------------QQLFEFRIPYYVGPL---------NRVDDGKDGK--------FTWSVRKSDARIYPWNFTEVIDVEASAEKFIRRMTNKCT------YLVGE-D------------------------VL------P-KDSL
T. denticola            ------------SHSEKIIMLLTFKIPYYIGPI---------NDNHKKFFPD-----RCWVVKKEKSPSGKTTPWNFFDHIDKEKTAEAFITSRTNFCT------YLVGE-S------------------------VL------P-KSSL
P. duerdenii            ---------------DKVESIMEFRIPYYVGPL---------NPDSKYAWIK--------------RENTKITPWNFKDIVDLDSSREEFIDRLIGRCT------YLKEE-K------------------------VL------P-KASL
C. mitsuokai            ------------------LSILTFRIPYYFGPL---------NETSEHAWIK----------RLEGKENQRILPWNYQDIVDVDATAEGFIKRMRSYCT------YFPDE-E------------------------VL------P-KNSL
S. thermophilus         ------------------EKILTFRIPYYVGPL--------ARGNSDFAWSI-------------RKRNEKITPWNFEDVIDKESSAEAFINRMTSFDL------YLPEE-K------------------------VL------P-KHSL
S. mutans               ------------------EKLLTFRIPYYVGPL--------ARGKSDFAWLS-------------RKSADKITPWNFDEIVDKESSAEAFINRMTNYDL------YLPNQ-K------------------------VL------P-KHSL
S. pyogenes             ------------------EKILTFRIPYYVGPL--------ARGNSRFAWMT-------------RKSEETITPWNFEEVVDKGASAQSFIERMTNFDK------NLPNE-K------------------------VL------P-KHSL

                        ------------------HHHHH-----------------------------------------------------------------------------------------------------------------EE------E----H
domains                 hhhhhhhhhhhhhhhhhhhhhhhhh|PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP*W*PPPPPPPPPPPPPPPPPPPPPPPPPPPP|hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
catalytic residues      hhhhhhhhhhhhhhhhhhhhhhhhh|PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP*W*PPPPPPPPPPPPPPPPPPPPPPPPPPPP|hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
domains                 111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
catalytic residues      111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
.
B. bifidum              ------------KQINKIEGLLTFRVPYFVGPL-VVPEDLQKSDNSENHWMV-------------RKKKGEITPWNFDEMVDKDASGRKFIERLVGTDS------YLLGE-P------------------------TL------P-KNSL
O. kitaharae            --------------LNKITALINFRVPYYVGPL-VEEEQKIADDGKNIPDPTNHWMV--------RKSNDTITPWNLSQVVDLDKSGRRFIERLTGTDT------YLIGE-P------------------------TL------P-KNSL
S. sanguinis            ---------------DKLLSILTFKIPYYVGPL-------AKGSNSRFAWIKRATSSDILDDNDEDTRNGKIRPWNYQKLINMDETRDAFITNLIGNDI------ILLNE-K------------------------VL------P-KRSL
F. fructosus            ------------GQDNKIEELLRFRIPYYVGPLVTKKDVEHAGGDADNHWVER----------NEGFEKSRVTPWNFDKVFNRDKAARDFIERLTGNDT------YLIGE-K------------------------TL------P-QNSL
E. yurii                ------------TVSEMIIQLFEFQIPYYVGPL----DKNPKKDKNKANSWAK------------IKQGGRILPWNFEDKVDVKGSRKEFIEKMVRKCT------YISDE-H------------------------TL------P-KQSL
L. farciminis           ---------------EKLLSILSFRIPYYVGPL-------QSAEKNPFAWME-------------RKSNGHARPWNFDEIVDREKSSNKFIRRMTVTDS------YLVGE-P------------------------VL------P-KNSL
S. pseudintermedius     --------------IEKVLSLIQFRIPYYVGPL------VNDHTASKFGWME-------------RKSNESIKPWNFDEVVDRSKSATQFIRRMTNKCS------YLINE-D------------------------VL------P-KNSL
L. rhamnosus            -----------WKMPYQLDELLNFHIPYYVGPLITPKQQAESG-ENVFAWMV------------RKDPSGNITPYNFDEKVDREASANTFIQRMKTTDT------YLIGE-D------------------------VL------P-KQSL
F. magna                ---------------EKIRKTFTFRIPYYVGPLNDYHKNNGG------NAWI------------FRNKGEKIRPWNFEKIVDLHKSEEEFIKRMLNQCT------YLPEE-T------------------------VL------P-KSSI
P. antarcticus          --------------IEKVKQILAFRIPYYIGPL------VKDTTQSPFSWVE-------------RKGDAPITPWNFDEQIDKAASAEAFISRMRKTCT------YLKGQ-E------------------------VL------P-KSSL
consensus/95%                      ..........h.FchPYahGPh........................................PaNh...h........Fh..h.....      h......                        .h      P...Sh
consensus/85%                      .........hh.FcIPYYhGPL......................................h.PWNa...hD...o...FI.ch.....      Yh..c..                        .L      P.K.SL
.
Jnet 227501312          ---------------HHHHHHHH-----EE-------------------EEE------------EEE----------------HHHHHHHHHHHH-------------------------------------------------------
M. mobile               ------------------KSFFSRIRNYSEGPGNEKSFSKYGL----------YANENGN----------PELIINEKGQKIYTKIFKTLWESKIGKCS------YDKKL-Y------------------------RA------P-KNSF
M. gallisepticum        ------------------ESLFSYVRNYSEGPGSINSVSPYGI------------------------------YHLDEKEGKVVQKYNNIWDKTIGKCN------IFPDE-Y------------------------RA------P-KNSP
M. synoviae             ------------------LKLFNYVRPFNKGPGSKNSRTAYGV-----------------------------FSTDIDPETNKFKDYSNIWDKTIGKCS------LFEEE-I------------------------RA------P-KNLP
M. ovipneumoniae         ------------------LNLFTSVRDYAKGPGSEHSASEYGI-----------------------------FQK--DEKGKVFKKYDNIWDKTIGKCS------FFVEE-N------------------------RS------P-VNYP
M. canis                ------------------KKIFSFTRDISKGPGSDNMPSPYGI-----------------------------FGEF-GDNGQG-GRYEHIWDKNIGKCS------IFTNE-Q------------------------RA------P-KYLP
consensus/95%                             ..hF...R.h.cGPG.....o.YGh          .......          ................a..hW-..IGKC.      h.... .                        Ru      P ....
consensus/85%                             ..hF...R.h.cGPG.....o.YGh          .......          ................a..hW-..IGKC.      h.... .                        Ru      P ....

E. rectale              ------------------MLIFNRKRKYYEGPGNEKSRTDYGR------------------------------FTTKLDANGNYITEDNIFEKLIGKCS------VYPDE-L------------------------RA------A-AASY
E. faecalis             ----------VQPGIAEEAGLIYRKRPYYHGPGNEANNSPYGR---------------------------------WSDFQKTGEPATNIFDKLIGKD-------FQGEL--------------------------RA------S-GLSL
S. thermophilus         ------------------LEILTGKRKYYHGPGNEKSRTDYGR------------------------------------YRTSGETLDNIFGILIGKCT------FYPDE-F------------------------RA------A-KASY
S. lugdunensis          ------------------IELVEMRREYFEGPGKGSPYGWEGD-------------------------------------------PKAWYETLMGHCT------YFPDELR------------------------SV--------KYAY
E. dolichum             ------------------ITIISRKRMYYDGPGGPLSPTPYGR-----------------------------------YTYFGQKEPIDLIEKMRGKCS------LFPNE-P------------------------RA------P-KLAY
R. albus                ------------------KDIIFAQRDFEIGPGKNERFRRFTG-----------------------------------------------YLDSIGKCQ------FFKDQ-E------------------------RG------S-RFTV
R. inulinivorans        ------------------VEIMTSQRSFDMGPGLQPDGKPS-------------------------------------------PYAMEGFGDRVGKCT------FEKDE-Y------------------------RA------P-KATY
L. coryniformis         -----------LLGDENRAGIFQAQRDFDEGPGAGPYAGDQ-------------------------------------------------IKKMVGKDI------FEPTE-D------------------------RA------A-KATY
I. polytropus           ------------------IEVFNSQRNFDEGPSKPSPYYSEIGQ----------------------------------------------IAKMIGNCT------FESSE-K------------------------RT------A-KNTW
consensus/95%                     ..........hh...R.a..GPu...........                             .................h....G...      h......                        ..      . ...h
consensus/85%                     ........h.hh...R.a..GPG...........                             .................h...hGc..      h......                        R.      . .h.h

Bacteroides sp.20       --PRRNSISTRDFCHLFIEDIIFYQRPLKS------------------------------------------------------------KKSLIDNCP------YEENR-YI---------------DKESGEIKHA-SIKCIA-KSHP
I. album                FGPKIKEFENSDVLHIISEDIIYYQRDLKS------------------------------------------------------------QKSLISECR------YEKRK-GI---------------D---GEI--Y-GLKCIP-KSSP
B. fragilis             -----------------RDIIIFYQRRLKS------------------------------------------------------------QKGLIGFCE------FESRQIEV---------------DIDGKKKIKTVGNRVIS-RSSP
N. salsuginis           ---------------DELKACITDQEMPTID------------------------------------------------------------ESLFGKCT------FYKDE-L------------------------AA------P-AYSY
E. minutum              ------------------TKAIGYESEK--------------------------------------------------------------LIPESGFCP------YLKDE-K------------------------RL------P-ASHK
S. globus               ------------------VSAILFENEK--------------------------------------------------------------IVPEAGCCP------YFPDE-K------------------------KL------P-RCHF
M. trichosporium        ------------------TKAAFFQRPLQD------------------------------------------------------------SEKLVGPCP------FEVDE-K------------------------RA------P-KRGY
F. branchiophilum       ----------FDGLAKDLYKAIFFQRPLKS------------------------------------------------------------QKGLIGKCS------FEKSK-S------------------------RC------A-ISHP
Prevotella C561         ----------ISE--KKNVGTIFYKRPLRS------------------------------------------------------------QRGNVGKCT------LERSK-P------------------------RC------A-IGHP
P. timonensis           ----------SSELIEDLQRAIFFQLPLKS------------------------------------------------------------QRHGVGRCT------FERGK-P------------------------RC------A-DSHP
P. timonensis           ------------------QRAIFFQLPLKS------------------------------------------------------------QRHGVGRCT------FERGK-P------------------------RC------A-DSHP
A. paucivorans          ------------------LHLVAFQNPFASGED---------------------------------------------------------IRNKAGHCS------LEPDQ-I------------------------RA------P-RRSA
P. marinum              ------------------KEIVFFQRKL--------------------------------------------------------------KPQEVGRCI------FLSDE-D------------------------RI------S-KALP
A. denitrificans        ------------------KDTLLHQRPL--------------------------------------------------------------RPVKPGRCT------LLPEE-E------------------------RA------P-LALP
R. syzygii              ------------------RAILLFQRKL--------------------------------------------------------------LPVPVGKCF------LEPNQ-P------------------------RV------A-AALP
D. shibae               ------------------FEKIFFQRPL--------------------------------------------------------------KEPEVGLCLFSGHHGVPPKD-P------------------------RL------P-KAHP
R. rubrum               -------------------RLVLDQRPL--------------------------------------------------------------RSKGAGPCA------FLPGE-D------------------------RA------L-RALP
Azospirillum B510       ------------------RHRIFHQRPL--------------------------------------------------------------KPPPVGRCT------LYPDD-G------------------------RA------P-RALP
C. salinarum            ------------------RRLVLDQAPI--------------------------------------------------------------AVPPATPCL------FLEELRR--------RGETFQGRTITREAIDRG------L-TVDP
N. hamburgensis         ------------------EHIIFYQRPL--------------------------------------------------------------KPAIVGKCT------LDPAT-RP---------------FKEDPEGYRA------P-WSHP
Bradyrhizobium BTAi1    ------------------REAIFSQRAM--------------------------------------------------------------KRPSIGKCS------LDPATSQD---------------DVDGF---RC------A-WSHP
W. succinogenes         ------------------REIAFFVRPMQR------------------------------------------------------------IEKMVGHCT------YFPEE-R------------------------RA------P-KSAP
C. jejuni               ------------------LSVAFYKRALKD------------------------------------------------------------FSHLVGNCS------FFTDE-K------------------------RA------P-KNSP
Jnet 218563121          ------------------HHHHHHHH---------------------------------------------------------------------------------------------------------------------------H
H. mustelae             ----------DAQSKQEREGLIFYQRPLKG------------------------------------------------------------FGDKIGKCS------HIKKGENS---------------PY------RA------C-KHAP
A. ebreus               ----------ILGDGDRKSGLFWQQKPALSGAD---------------------------------------------------------LLKMLGKCT------FEKGE-Y------------------------RA------P-KASF
B. smithii              ------------------LNIWSSQRPFASKED---------------------------------------------------------IEKKVGFCT------FEPKE-K------------------------RA------P-KATY
C. cellulolyticum       ------------------LNIWEFQRPFASGDS---------------------------------------------------------ILSKVGKCT------LLKEE-L------------------------RA------P-TSCY
HF0070_07E19            ------------------LDVLMGQRSGR--------------------------------------------------------------SPKLGNCS------LIPSE-L------------------------RA------P-SSAP
P. lavamentivorans      ------------------SDTIFAQRPVFW------------------------------------------------------------RKNTLGECR------FMPGE-P------------------------LC------P-KGSW
N. meningitidis         ------------------ETLLMTQRPALSGDA---------------------------------------------------------VQKMLGHCT------FEPAE-P------------------------KA------A-KNTY
P. multocida            ------------------TELLMWQKPALSGEA---------------------------------------------------------ILKMLGKCT------HEKNE-F------------------------KA------A-KHTY
A. muciniphila          ---------------AQHHPLTTEQRGVLLQHGIKLARRYRGS-------------------------------------LLFGQLIPRFDNRIISRCPVTWAQVYEAELKKG---------------NSEQSARERAEKLSKVPTANCP
Actinomyces F0310       -----------------VLSIFHSKEPAPS---------------------------------------------------------AARQRERVGLDELQ--LALDPAAKQ------------------------PR------AERAHP
A. cellulolyticus       -----------------IDAVFCQKRPSV-------------------------------------------------------------PAERIGRDP------LDPSQ-L------------------------RA------S-RACL
B. longum               ------------------RSVFYAVSPKGS------------------------------------------------------------AEQRVGQDP------LAPEQA-------------------------RA------L-KASL
A. naeslundii           ------------------RAVFKADSPRGS------------------------------------------------------------AVSRVAPDP------LPGQGSFR-----------------------RA------P-KCDP

                        ------------------HHHHHH----------------------------------------------------------------------------------------------------------------EE------E----H
domains                 PPPPPPPPPPPPPPPPPPPPPPPPP|PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP*W*PPPPPPPPPPPPPPPPPPPPPPPPPPPP|hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
catalytic residues      PPPPPPPPPPPPPPPPPPPPPPPPP|PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP*W*PPPPPPPPPPPPPPPPPPPPPPPPPPPP|hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

A. hesperidum           LKILCSQRNFDDGPGGDSPYGHGSVSPDGVR--------------------------------------------------------QSIYERMVGSCT------FETGE-------------------------------KRAP-RSSY
Alcanivorax W11-5       ------------------RNAIFFQRPLKS------------------------------------------------------------PADKVGRCS------LQTNL-------------------------------PRAP-RAQI
B. zoohelcum            QKKYRELGLEKGLKYIIKNQVVFYQRELKD------------------------------------------------------------QSHLISDCR------YEPNE-------------------------------KAIA-KSHP
Porphyromonas F0450     -----------------RDEVIFMQRPLKS------------------------------------------------------------CKHLVSLCE------FEKQERVM------RVQQDDGKGGWQLVERRVKFGPKVAP-KSSP
B. intestinihominis     -----------------RDEIIYYQRPLKS------------------------------------------------------------CKYLVSRCE------FEKRF------------------YLNAAGKKTEAGPKVSP-RTSP
O. laneus               VLGGKIEIEEEQLKFKSNESVLFWQRPLRS------------------------------------------------------------QKSLLSKCV------FEGRN-----------------FYDPVHQKWIIAGPTPAP-LSHP
Treponema JC4           ----------------------FYQRPLK--------------------------------------------------------------PQQRGYCI------YENDK-------------------------------ERTF-KAMP
Rhodovulum PH10         ------------------AGILFRQRDLA--------------------------------------------------------------PPKIGCCT------FEPSE-------------------------------RRLP-RALP
consensus/95%           .....................h.....................                                     ...............u.........h.......      ...............................
consensus/85%           .....................h.....................                                     ..............hu.C.......h.......      ...............................

#map                    ---------------------------------------------------------------------------------------------------------------------------------------------------***
```

```
Jnet 331001027          HH-HHHHHHH-----HHHHHH-----------------------------HHHHHHHHHHH---HHH------HHHHHHHHHHHH-HHHHH-----------------------------HHHHHHH--------HHHHHHHHHHHH---
P. excrementihominis    IL-QRAFDRS-IALDECSIRR---------------------TAEDFENGVVIKNEKLEDVLSGHQLE------EFLEFANRYYQE-TAKAKNG-LW---------------------FP-ENALLERADLHPPMKNKILNVIVGQALGV-
W. succinogenes         LL-HRLLDTS-KEIDIYELR-----------------------GKKPNELLVK------TLGQSDAN------RLYGFAQNYYELIRQKVRAG-IWV--------PVKNKDDSLN--LEDNSNMLKRCNHNPPHKKNQIHNLVAGILGV-
L. pneumophila          IL-QRYLDLN-KKIDKFKIKKQLSFLG--------------QGKQLPANLIETQKEMETHFNSSLVS------VLIQIASAYNKE-REDAAQG-IW---------------------FDNAFSLCELSNINPPRKQKILPLLVGAILSED
S. wadsworthensis       AL-QRAFDRS-KALDPYALRAL-------------------AAGSKSNKLTSARTALENCIGGQNVK------TFLDCARRYYRE-ADDAKVG-LW---------------------FDNADGLLERSDLHPPMKKKILPLLVANILQT-
HTCC5015                AL-HRIVERS-RSIDPYQLRLLIS-----------------ITDAEKRNDLAGYKRLKLSLGSEVDE-------FLLLVKNIVDE-TKEAREG-LW---------------------FETENKLFFKCGKTPPRKEKLKSTLLSAVLGK-
F. novicida             ILKQFIFDRV-KASDELLLNEIYFQAKKL---QKASSELEKLESSKKLDEVIANSQLSQILKSQHTNGIFEQGTFLHLVCKYYKQ-RQRARDSRLYIMPEYRYDKKLHKYNNTGR--FDDDNQLLTYCNHKPRQKRYQLLNDLAGVLQVS
consensus/95%           hL...hh-.. ...D...h..........   ............................h.............hh..h..h......ch..u.ha...................  h.....h.......P..K.......hu.hL...
consensus/85%           hL...hh-.. ...D...h..........   ............................h.............hh..h..h......ch..u.ha...................  h.....h.......P..K.......hu.hL...

A. coleocanis           AF-QRYRIAS-IVSNLRI--------------------------RHLGSGADERLDVETQKRVFEYLLN----AKPTAD-------IT---WSD----VA--------EEIGVER---NLLMGTATQT--------------ADGERASA
C. glomerans            LY-EEFCVLN-ELNGAHW---------------------------SIDGDDEHRFDAADREGIIEELFR-----R-KRT-------VS---YGD----VA--------GWMERERNQIGAHVC-------------------GGQGEKGF
Acidaminococcus D21     IY-EKFMILN-EINNIRI--------------------------------DGYPISVDVKQQVFGLFEK-----K--RR-------VT---VKD----IQ--------NLLLSLG---ALDKHGK-----------------LTGIDTTI
V. atypica              LY-SEFMLLN-ELNNVRI--------------------------------DGKALAQGVKQHLIDSIFK-----QDHKK-------MT---KNR----IE--------LFLKDNN---YITKKHKPE---------------ITGLDGEI
F. nucleatum            LY-SEYIILN-ELNKVQV--------------------------------NDEFLNEENKRKIIDELFK-----E-NKK-------VS---EKK----FK--------EYLLVNQI--ANRTVE------------------LKGIKDSF
F. alocis               LY-SKYMVLN-ELNNVKV--------------------------------RGKKLPTSLKQKVFEDLFE-----N-KSK-------VT---GKN----LL--------EYLQIQDK--DIQIDD------------------LSGFDKDF
S. moorei               LY-SKYNVLN-EINPIKV--------------------------------NGKAIPVEVKQAIYTDLFE-----NSKKK-------VT---RKS----IY--------IYLLKNG---YIEKEDI-----------------VSGIDIEI
C. catus                VY-SKFMVLN-ELNNLRL--------------------------------NGEKISVELKQRIYEELFC-----K-YRK-------VT---RKK----LE--------RYLVIE----GIAKKGVE----------------ITGIDGDF
T. denticola            LY-SEYTVLN-EINNLQI-----------------------------IIDGKNICDIKLKQKIYEDLFK-----K-YKK-------IT---QKQ----IS--------TFIKHEG---ICNKTDEVI---------------ILGIDKEC
P. duerdenii            IY-NEFMVLN-ELNNLKL--------------------------------NEFLITEEMKKAIFEELFK-----T-KKK-------VT---LKA----VS--------NLLKK-----EFNLTGDIL---------------LSGTDGDF
C. mitsuokai            IV-SKYEVYN-ELNKIRV--------------------------------DDKLLEVDVKNDIYNELFM-----K-NKT-------VT---EKK----LK--------NWLVNNQ---CCSKDAEIK---------------GFQKENQF
S. thermophilus         LY-ETFNVYN-ELTKVRF--------------------------IAESMRDYQFLDSKQKKDIVRLYFK-----D-KRK-------VT---DKD----II-------E-YLHAIY---GYDGIE------------------LKGIEKQF
S. mutans               LY-EKFTVYN-ELTKVKY--------------------------KTEQ-GKTAFFDANMKQEIFDGVFK-----V-YRK-------VT---KDK----LM-------DFLEKEFD---EFRIVD---------------LTGLDKENKVF
S. pyogenes             LY-EYFTVYN-ELTKVKY--------------------------VTEGMRKPAFLSGEQKKAIVDLLFK-----T-NRK-------VT---VKQ----LK-------EDYFKKIE---CFDSVE------------------ISGVEDRF

                        HH-HHHHHHH-HH-------------------------------------------HHHHHHHHHH-------------------------HHH----HH-------H------------------------------------------
domains                 hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
catalytic residues      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
domains                 111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
catalytic residues      111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
.
B. bifidum              LY-QEYEVLN-ELNNVRL---------------------SVRTGNHWNDKRRMRLGREEKTLLCQRLFM-----K-GQT-------VT---KRT----AE--------NLLRKEY---GRTYELS-----------------GLSDESKF
O. kitaharae            LY-QKFDVLQ-ELNNIRV--------------------------------SGRRLDIRAKQDAFEHLFK-----V-QKT-------VS---ATN----LK--------DFLVQAGYISEDTQIE---------------GL-ADVNGKNF
S. sanguinis            IY-EEVMLQN-ELTRVKY---------------------KDKYG------KAHFFDSELRQNIINGLFK----NN-SKR-------VN---AKS----LI--------KYLSDNHKDL-------NAIEIV----------SGVEKGKSF
F. fructosus            RY-QLFTVLN-ELNNVRV--------------------------------NGKKFDSKTKADLINDLFK-----A-RKT-------VS---LSA----LK--------DYLKAQG----------KGDVTI----------TGLADESKF
E. yurii                LY-EKFMVLN-EINNIKI--------------------------------DGEKISVEAKQKIYNDLFV-----K-GKK-------VS---QKD----IK--------KELISLNIM--------DKDSVL----------SGT------
L. farciminis           IY-QRYEVLN-ELNNIRI---------------------TENLK---TNPIGSRLTVETKQRIYNELFK-----K-YKK-------VT---VKK----LT--------KWLIAQGYY---------KNPIL----------IGLSQKDEF
S. pseudintermedius     LY-QEMEVLN-ELNATQI---------------------RLQTD---PKNRKYRMMPQIKLFAVEHIFK-----K-YKT-------VS---HSK----FL--------EIMLNSNHRENFMNH--GEKLSI----------FGTQDDKKF
L. rhamnosus            LY-QKYEVLN-ELNNVRI--------------------------------NNECLGTDQKQRLIREVFE-----R-HSS-------VT---IKQ----VA-------DNLVAHGD---FARRPEIR----------------GLADEKRF
F. magna                LY-SEYMVLN-ELNNLRI--------------------------------NGKPLDTDVKLKLIEELFK-----K-KTK-------VT---LKS----IR--------DYMVRNN---FADKEDFD----------------NSEKNLEI
P. antarcticus          TY-ERFEVLN-ELNGIQL---------------------RTTGA---ESDFRHRLSYEMKCWIIDNVFK-----Q-YKT-------VS---TKR----LL--------QELKKSPYADELYDEHTGEIKEV----------FGTQKENAF
consensus/95%           .a ..h.hh. Eh..h.h                     ...............h....c..hh..hh.    ......       ho   ...    h.       ...h....................        ...........
consensus/85%           hY .ca.hhN Eh..h.h                     ...............h....K..hh..hF.    ......       Vo   ...    h.       ..hh....................        .........h
.
Jnet 227501312          HH-HHHHHHH-H---EEE--------------------------------------HHHHHHHHHHHHH----------------------HHH----HH-------HHHHH------------EE----------------EEE-----
M. mobile               SA-KVFDITN-KLTDWKH-------------------------------KNEYISERLKRKILLSRFLN-----KDSKS-------AVEKILKEENIKFE--------NLSEIAY---NKDDNKINLPIINAYHSLTT----IFKKHLIN
M. gallisepticum        IA-MIFNEIN-ELSTIRS---------------------------YSIYLTGWFINQEFKKAYLNKLLDLLIKTNGEKP-------IDARQFKKLREETIAESI----GKETLKDVENEEKLEKEDHKWKLKGLKLNT----NGKIQYND
M. synoviae             SA-LIFNLQN-EICTIKN---------------------------EFTEFKNWWLNAEQKSEILKFVFTELFNWKDKKY-------SDKKFNKNLQDKIK--------KYLLNFA---LENFNLNEEILKNRDLENDTVLGLKGVKYYEK
M. ovipneumoniae        SY-EIFNLLN-QLINLST------------------DLKT-------TNKKIWQLSSNDRNELLDELLKVKEKAKI----------ISISLKKNEIKKII------------------LKDFGFEKSDIDDQD-------TIEGRKIIKE
M. canis                SA-LIFNFLN-ELANIRL------------------YSTD-----KKNIQPLWKLSSVDKLNILLNLFNLPISEK-KKKL------TSTNINDIVKKESI--------KSIM-IS---VEDIDMIKDEWAGKE-PNVYGVGLSGLNI--E
consensus/95%           .h .hF...N .h..h..                  ....     .......a......c..hL..hh...........       ......c....c......    ...........c........h.....................
consensus/85%           .h .hF...N .h..h..                  ....     .......a......c..hL..hh...........       ......c....c......    ...........c........h.....................

E. rectale              TA-QEYNVLN-DLNNLTI--------------------------------NGRKLEENEKHEIVERIKS-----S--NT-------IN----------MR--------KIISDCMGENIDDFAGAR----------------IDKSGKEI
E. faecalis             SA-QQYNLLN-DLTNLKI-------------------------------DGEVPLSSEQKEYILTELMT-----KEFTR-------FG---VND----VV--------KLLGVK----KERLSGWR----------------LDKKGKPE
S. thermophilus         TA-QEFNLLN-DLNNLTV------------------------------PTETKKLSKEQKNQIINYVKN-----E--KA-------MG---PAKLFKYIA--------KLLSCD----VADIKGYR----------------IDKSGKAE
S. lugdunensis          SA-DLFNALN-DLNNLVI-----------------------------QRDGLSKLEYHEKYHIIENVFK-----QKKKP--------T---LKQ----IA--------NEINVN----PEDIKGYR----------------ITKSGKPQ
E. dolichum             SA-ELFNLLN-DLNNLSI--------------------------------EGEKLTSEQKAMILKIVHE-----K--GK-------IT---PKQ----LA--------KEVGVS----LEQIRGFR----------------IDTKGSPL
R. albus                IA-DIYAFVN-VLSQYTY----------------------------TNNRGESVFDTSFANDLINSALK-----N--GS-------MD---KRE----LK--------AI--------AKSYHIDIS---------------DKNSDTSL
R. inulinivorans        TA-ELFVALQ-KINHTKL---------------------------IDEFGTGRFFSEEERKTIIGLLLS-----S--KE-------LK---YGT----IR--------KKLNIDP---SLKFNSLNYS---------AKKEGETEEERVL
L. coryniformis         TF-QYFNLLQ-KMTSLNY--------------------------QNTTGDTWHTLNGLDRQAIIDAVFA-----KAEKP----TKTYKPTDFGE----LR--------KLLKLPD---DARFNLVNYGSLQTQKEI------ETVEKKTR
I. polytropus           SG-ERFVFLQ-KLNNFRI----------------------------VGLSGKRPLTEEERDIVEKEVYL-----K--KE-------VR---YEK----LR--------KILYLKE---EERFGDLNYS---------------KDEKQDK
consensus/95%           .. ..a.hh. .h....h                          ..........h.......h...h..     .....    ...............h.        .............h.............. .............
consensus/85%           ou ..a.hL. ch..h.h                          ..........h....c..hh..h..     .....    ...h...........h.        ..h..........h..h........... .............

Bacteroides sp.20       LY-QEFRLWQ-FIVNLRI------------------------------------YRKETDVDVTQELLP-----T-EAD------------YVT----LF--------EWLNEKKEIDQKAFFKYPP------FGFKK----TTSNYRWN
I. album                LY-QEFRIWQ-DIHNIKV------------------------------------IRKESEVNGKKKINI-----D-ETQL---------YINENIKEKLF--------ELFNSKDSLSEKDILELISLNIINSGIKISK-KEEETTHRIN
B. fragilis             LF-QEFKIWQ-ILNNIEVTVVGKKRKRRKLKENYSALFEELNDAEQLELNGSRRLCQEEKELLAQELFI-----R--DK-------MT---KSE----VL--------KLLFDNPQELDLNFKTIDGNKTGY----------ALFQAYSK
N. salsuginis           LY-DLYRLYK-KLADLNI--------------------------------DGYEVTQEDREKVIEWVEKKIAQGKNLKK-------IT---HKD----LR--------KILGLAP---EQKIFGVED------ERIV-----KGKKEPRT
E. minutum              LN-EERRLWE-TLNNARY-----------------SDPIVDIVTGEITGYYEKQFTKEQKQKLFDYLLT-----G--SE-------LT---PAQ----TK--------KLLGLK----NTNFEDI-----------------ILQGRDKK
S. globus               LN-EERRLWE-AINNARI---------------------------------KMPMQEGAAKRYQSASFS-----DEQRH-------IL---FHI----AR--------SGTDITPKLVQKEFPALKTSII------------VLQGKEKA
M. trichosporium        SF-ELFRFLS-RLNHVTL----------------------------RDGKQERTLTRDELALAAADFGA-----A--AK-------VS---FTA----LR--------KKLKLPE---TTVFVGVK----------------ADEESKLD
F. branchiophilum       DF-EEYRMWT-YLNTIKI--------------------------GTQSDKKLRFLTQDEKLKLVPK-FYR----K--ND-------FN---FDV----LA--------KELIEK----GSSFGFYKSSKKNDFF-------YWFNYKPTD
Prevotella C561         LF-EKFRAWT-LINNIKV--------------------------RMSVDTLDEQLPMKLRLDLYNECFLAFV--R--TE-------FK---FED----I---------RKYLEK----RLGIHFSYNDKT-------------INYKDST
P. timonensis           DY-EEFRMLC-FVNNIQV--------------------------KGPHDLELRPLTYEEREKIEPL-FFRKS--K--PN-------FD---FED----IA--------KALAGK---K---NYAWIHDKEERA--------YKFNYRMTQ
P. timonensis           DY-EEFRMLC-FVNNIQV--------------------------KGPHDLELRPLTYEEREKIEPLFFR-----KSKPN-------FD---FED----IA--------KALAGKK---NYAWIHDKE----------ERAY-KFNYRMTQ
A. paucivorans          SA-ETFMLLQ-KTGNLRL--------------------------IHRRTGEERPLTDKEREQIHLLAWK-----Q--EK-------VT---HKT----LR--------RHLEIPE---EWLFTGLPYHRS------------GDKAEEKL
P. marinum              SF-QRFRIYQ-ELSNLAW-------------------------IDHDGVAHRITASLALRDHLFDELEH-----K--KK-------LT---FKA----MR--------AILRKQGV--VDYPVGFNL---------------ESDNRDHL
A. denitrificans        ST-QRFRIHQ-EVNHLRL---------------------------LDENLREVALTLAQRDAVVTALET-----K--AK-------LS---FEQ----IR--------KLLKLSG---SVQFN-------------------LEDAKRTE
R. syzygii              SA-QRFRLMQ-ELNHLRV--------------------------MTLADKRERPLSFQERNDLLAQLVA-----R--PK-------CG---FDM----LR--------KIVF------GANKEAYRF---------------TIESERRK
D. shibae               LT-QRRVLYE-TVNQLRV---------------------------TADGREARPLTREERDQVIHALDN-----KKPTK-SLSSMVLK---LPA----LA--------KVLKLRDGERFTLETGVRD---------------AIACDPLR
R. rubrum               TV-QDFIIRQ-TLANLRL--------------------------PSTSADEPRPLTDEEHAKALALLST-----ARFVE------------WPA----LR--------RALGLKR---GVKFTAETE---------------RNGAKQAA
Azospirillum B510       SA-QRLRLFQ-ELASLRV---------------------------IHLDLSERPLTPAERDRIVAFVQG-----RPPKAGRKPGKVQKSVPFEK----LR--------GLLELPP---GTGFSL------------------ESDKRPEL
C. salinarum            LI-QALRIRE-TVGNLRLHE---------RITEPDGRQRYVPRAMPELGLSHGELTAPERDTLVRALMH-----DPDGLAAKDGR----IPYTR----LR--------KLIGYDNS--PVCFAQ------------------ERDTSGGG
N. hamburgensis         LA-QRFRILS-EARNLEI---------------------------RDTGKGSRRLTKEQSDLVVAALLA-----N--RE-------VK---FDK----LR--------TLLKLPA---EARFNL------------------ESDRRAAL
Bradyrhizobium BTAi1    LA-QRFRIWQ-DVRNLAV---------------------------VETGPTSSRLGKEDQDKVARALLQ-----T--DQ-------LS---FDE----IR--------GLLGLPS---DARFNL------------------ESDRRDHL
W. succinogenes         TA-EKFIAISKFFSTVII---------------------------DNEGWEQKIIERKTLEELLDFAVS-----R--EK-------VE---FRH----LR--------KFLDLSD---NEIFKGLHYKGKPKTAK-------KREATLFD
C. jejuni               LA-FMFVALTRIINLLNN----------------------------LKNTEGILYTKDDLNALLNEVLK-----N--GT-------LT---YKQ----TK--------KLLGLSD---DYEFKG------------------EKGTYFIE
Jnet 218563121          HH-HHHHHHHHHH--EEE--------------------------------------HHHHHHHHHHHHH----------------------HHH----HH--------HHH---------------------------------------
H. mustelae             SA-EEFVALTKSINFLKN----------------------------LTNRHGLCFSQEDMCVYLGKILQ-----EAQKN----EKGLT---YSK----LK--------LLLDLPS---DFEFLGLDYS--------------GKNPEKAV
A. ebreus               SV-ERHVWLT-RLNNLRI----------------------------VVDGRSRPLNEAERQAALLLPYQ-----TETSK------------YKT----LK-------NAFIKAGLWGDGVRFGGLAYPSQAQIDAEK-----TKDPEDQF
B. smithii              TF-QSFIVWE-HINKLRL----------------------------VSPDETRALTEIERNLLYKQAFS-----K--NK-------MT---YYD----IR--------KLLNLSD---DIHFKGLLYDPKSS----------LKQIENIR
C. cellulolyticum       TS-EYFGLLQ-SINNLVL----------------------------VEDNNTLTLNNDQRAKIIEYAHF-----K--NE-------IK---YSE----IR--------KLLDIEP---EILFKAHNLTHKNP----------SGNNESKK
HF0070_07E19            ST-EWFKFLQ-NLGNLQI---------------------------SNAYREEWSIDAPRRAQIIDACSQ-----R--ST-------SS---YWQ----IR--------RDFQIPD---EYRFNLVNYERRDP-------------DVDLQ
P. lavamentivorans      LS-QQRRMLE-KLNNLAI-----------------------------AGGNARPLDAEERDAILSKLQQ-----Q--AS-------MS---WPGVRSALK--------ALYKQRGEPGAEKSLKFNL---------------ELGGESKL
N. meningitidis         TA-ERFIWLT-KLNNLRI----------------------------LEQGSERPLTDTERATLMDEPYR-----K--SK-------LT---YAQ----AR--------KLLGLED---TAFFKGLRYGKDN--------------AEAST
P. multocida            SA-ERFVWLT-KLNNLRI----------------------------LEDGAERALNEEERQLLINHPYE-----K--SK-------LT---YAQ----VR--------KLLGLSE---QAIFKHLRYSKEN--------------AESAT
A. muciniphila          EF-YEYRMAR-ILCNIRA--------------------------------DGEPLSAEIRRELMNQARQ-----E--GK-------LT---KAS----LE--------KAISSRLGKETETNVSNYF---------------TLHPDSEE
Actinomyces F0310       AF-QKFKVVA-TLANMRI--------------------------REQSAG-ERSLTSEELNRVARYLLN----HTESES-------PT---WDD----VA--------RKLEVPR---HRLRGSSRAS--------------LETGGGLT
A. cellulolyticus       EF-QEYRIVA-AVANLRI----------------------------RDGSGSRPLSLEERNAVIEALLA-----QTERS-------LT---WSD----IA-------LEILKLPN---ESDLTSVPE---------------EDGPSSLA
B. longum               AF-QEYRIAN-VITNLRI---------------------------KDASAELRKLTVDEKQSIYDQLVS-----PSSED-------IT---WSD----LC--------DFLGFK----RSQLKGVGS---------------LTEDGEER
A. naeslundii           EF-QRFRIIS-IVANLRI---------------------------SETKGENRPLTADERRHVVTFLTE-----DSQAD-------LT---WVD----VA--------EKLGVH----RRDLRGTAV---------------HTDDG-ER

                        HH-HHHHHHH-HH---EE----------------------------------EE--HHHHHHHHHHH-------------------------HH----HH--------HHH-------HHHEE---------------------------
domains                 hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
catalytic residues      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

A. hesperidum           SF-ERFQLLT-KVVNLRI--YRQQ---EDGGRYPC------------------ELTQTERARVI-DCAY------EQTK-------IT---YGK----LR--------KLLDMK----DTESFAGLTYGLN----------------RSR
Alcanivorax W11-5       AA-QNFRIEK-QMADLRW---------GMGRRAE--------------------MLNDHQKAVIRELLN------QQKE-------LS---FRK----IY--------KEL--E----RAGCPGPEGKGLNMDRAAL---------GGRD
B. zoohelcum            VF-QEYKVWE-QINKLIV--NTKIEAGTNRKGEKKYKYIDR------------PIPTALKEWIFEELQN-------KKE-------IT---FSA----IFKKL----KAEFDLR----EGIDFLNGMSPKD-------------------
Porphyromonas F0450     LF-QLCCIYE-AVNNIRL---------TRPNGSPC------------------DITPEERAKIV-AHLQ------SSAS-------LS---FAA----LKKLL----KEKALIA----DQLTSKSGLKGNSTRVALA---------SALQ
B. intestinihominis     LF-QVCRLWE-SINNIVV--K------DRRNEIVF-------------------ISAEQRAALF-DFLN------THEK-------LK---GSD----LLKLL--GLSKTYGYR----LGEQFKTGIQGNKTRVEIE---------RALG
O. laneus               EF-EEFRAYQ-FINNIIY------------GKNEH-------------------LTAIQREAVFELMCT------ESKD-------FN---FEK----IPKHL----------K----LFEKFNFDDTTKV-------------------
Treponema JC4           CS-QKLRILQ-DIGNLAY--YE-----GGSKKRVE-------------------LNDNQDKVLY-ELLN------SKDK-------VT---FDQ----MRKAL--------CLA----DSNSFNLEENRDF-------------------
Rhodovulum PH10         SV-EARGIYE-RLAHLRI--T------TGPVSDRG-------------------LTRPERDVLASALLA-------GKS-------LT---FKA----VRKTL--------KIL----PHALVNFEEAGEK-----------------GL
consensus/95%           .. ....h....h..h.h....................................h.......h...................................h....  ................................. ...........
consensus/85%           .. ....hh...h..h.h....................................h.......hh...........................a......h....  .....h........................... ...........

#map                    **-*******-*******------------------------------------------------------------------------------------------------------------------------------------
```

```
Jnet 331001027          -----HHHHHHHHHHH---------HHHHHHHHHHHHHHH---------HHHH----HHHH---------HHHHHHH------HHHHHH------HHHHHHHHHH------------------HHHHHH---H----HHHHHHHHHH---
P. excrementihominis    --SPAEGTDFIEEIWNS--KVKGRSTVRSICNAIENERKTYGPYFSEDYKFVK----TALKEGKTEKELSKKFAAVI------KVLKMV------SEVVPFIGKELRLS--------------DEAQSK---FDNLYSLAQLYNLIETER
W. succinogenes         KLDEAKFAEFEKELWSA---KVGNKKLSAYCKNIEELRKTHGNTFKIDIEELR---------KKDPAELSKEEKAKL------RLTDDVI-----LNEWSQKIANFFDID-------------DKHRQR---FNNLFSMAQLHTVIDTPR
L. pneumophila          FINNKDKWAKFKIFWNT--HKIGRTSLKSKCKEIEEARKNSGNAFKIDYEEAL-----NHPEH-------SNNKALI------KIIQTI------PDIIQAIQSHLGHN--------------DSQALI---YHNPFSLSQLYTILETKR
S. wadsworthensis       --DETTGQKFLDEIWRK--QIKGRETVASRCARIETVRKSFGGGFNIAYNTAQYREVNKLPRN-------AQDKELL------TIRDRV------AETADFIAANLGLS--------------DEQKRK---FANPFSLAQFYTLIETEV
HTCC5015                NLSDDEQSSFIEEFWKSGTPKIERRNVRGWCRLASQVQKTYGVYLK-EYGLQQ---LHKLEAGKK-----LDDKPLA------LLYKNS------GLIASKIGEALNIE--------------PDEVSR---FASPHSLAQIFNIIEGDV
F. novicida             PNFLKDKIGSDDDLFISKWLVEHIRGFKKACEDSLKIQKDNRGLLNHKINIAR------NTKGKCEKEIFNLICKIE------GSEDKK------GNYKHGLAYELGVLLFGEPNEASKPE-FDRKIKK---FNSIYSFAQIQQIAFAER
consensus/95%           ...........c.ha...........h...C.......K.....h...h............................      .......     .......h...h.h............. .......   a...aShuQh..hh....
consensus/85%           ...........c.ha...........h...C.......K.....h...h............................      .......     .......h...h.h............. .......   a...aShuQh..hh....

A. coleocanis           KPPVDVT----------NVAFATCKIKPLKEWWLNADYEARCVMVSALSH-----------AEK----LTEGTAAEV------EVAEFLQNL---------------SDEDNEK-LDSF---SLPIG------RAAYSVDSLERLTKRMI
C. glomerans            ESKLGSY-IF--FCK--DVFKVERLEQS----------------DYPMIER----------IILW----NTLF-EDRK------ILSQRLKEEYGSR----LS-----AEQIKT--ICKK----RFTG------WGRLSEKFLTGITVQVD
Acidaminococcus D21     HSNYNTY-HH--FK---SLMERGVL------------------TRDDVER----------IVER----MTYS-DDTK------RVRLWLNNNYGT-----LT-----ADDVKH--ISRL----RKHD------FGRLSKMFLTGLKGVHK
V. atypica              KNDLTSY-RDMV-----RILGNNF-------------------DVSMAED----------IITD----ITIFGESKK------MLRQTLRNKFGSQ----LN-----DETIKK--LSKL----RYRD------WGRLSKKLLKGIDGCDK
F. nucleatum            NSNYVSYIKFKD------IFGEKLNLDI---------------YKEISEK----------SILW----KCLYGDDKK------IFEKKIKNEYGDI----LN-----KDEIKK--INSF----KFNT------WGRLSEKLLTGIEFINL
F. alocis               KTSLKSY-----LDFKKQIFGEEIEKES---------------IQNMIED----------IIKW----ITIYGNDKE------MLKRVIRANYSNQ----LT-----EEQMKK--ITGF----QYSG------WGNFSKMFLKGISGSDV
S. moorei               KSKLKSH-HD--FT---QIVQENKC------------------TPEEIER----------IIKG----ILVYSDDKS------MLRRWLKNNIKG-----LS-----ENDVKY--LAKL----NYKE------WGRLSKTLLTDIYTINP
C. catus                KASLTAY-HD--FKE--RLTDVQL-------------------SQRAKEA----------IVLN----VVLFGDDKK------LLKQRLSKMYPN-----LT-----TGQLKG--ICSL----SYQG------WGRLSKTFLEEITVPAP
T. denticola            TSSLKSY------IELKNIFGKQVDEIS---------------TKNMLEE----------IIRWAT--IYDEGEGKT------ILKTKIKAEYGKY----CS-----DEQIKK--ILNL----KFSG------WGRLSRKFLETVTSEMP
P. duerdenii            KQGLNSY------IDFKNIIGDKVDRDD---------------YRIKIEE----------IIKL----IVLYEDDKT------YLKKKIKSAYKND----FT-----DDEIKK--IAAL----NYKD------WGRLSKRFLTGIEGVDK
C. mitsuokai            STSLTPW-IDFT-----NIFGKIDQS-----------------NFDLIEN----------IIYD----LTVF-EDKK------IMKRRLKKKYA------LP-----DDKVKQ--ILKL----KYKD------WSRLSKKLLDGIVADNR
S. thermophilus         NSSLSTYHDLLNIINDKEFLDDSS-------------------NEAIIEE----------IIHT----LTIF-EDRE------MIKQRLSKFEN------IF-----DKSVLKK-LSRR----HYTG------WGKLSAKLINGIRDEKS
S. mutans               NASYGTYHDLCKIL-DKDFLDNSK-------------------NEKILED----------IVLT----LTLF-EDRE------MIRKRLENYSD------LL-----TKEQVKK-LERR----HYTG------WGRLSAELIHGIRNKES
S. pyogenes             NASLGTYHDLLKIIKDKDFLDNEE-------------------NEDILED----------IVLT----LTLF-EDRE------MIEERLKTYAH------LF-----DDKVMKQ-LKRR----RYTG------WGRLSRKLINGIRDKQS

                        ----HHHHHHHHHH--HHHHH-----------------------HHHHHH----------HHHH----HHH----HH------HHHHHH-------------------HHHHHH-HHH---------------EEEEEHHHHH-------
domains                 hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
catalytic residues      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
domains                 111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
catalytic residues      111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
.
B. bifidum              TSSLSTY------GKMCRIFGEKYVNE---------------HRDLMEK----------IVEL----QTVF-EDKE------TLLHQLRQLEG------IS-----EADCAL--LVNT----HYTG------WGRLSRKLLTTKAGECK
O. kitaharae            NNALTTY------NYLVSVLGREFVENPS--------------NEELLEE----------ITEL----QTVF-EDKK------VLRRQLDQLDG------LS-----DHNREK--LSRK----HYTG------WGRISKKLLTTKIVQNA
S. sanguinis            NSTLKTY------NDLKTIFSEELLDSEI--------------YQKELEE----------IIKV----ITVF-DDKK------SIKNYLTKFFG--HLEILD-----EEKINQ--LSKL----RYSG------WGRYSAKLLLDIRDED-
F. fructosus            NSSLSSY------NDLKKTFDAEYLENED--------------NQETLEK----------IIEI----QTVF-EDSK------IASREL-SKLP------LD-----DDQVKK--LSQT----HYTG------WGRLSEKLLDSKIIDER
E. yurii                DTVCNAY------LSSIGKFTGVFKEEINKQS-----------IVDMIED----------IIFL----KTVYGDEKR------FVKEEIVEKYG----DEID-----KDKIKR--ILGF----KFSN------WGNLSKSFLELEGADVG
L. farciminis           NSTLTTY------LDMKKIFGSSFMEDNK--------------NYDQIEE----------LIEW----LTIF-EDKQ------ILNEKLHSSKY-----SYT-----PDQIKK--ISNM----RYKG------WGRLSKKILMDITTETN
S. pseudintermedius     ASKLSSY------QDMTKIFGDIEG------------------KRAQIEE----------IIQW----ITIF-EDKK------ILVQKLKECYP-----ELT-----SKQINQ--LKKL----NYSG------WGRLSEKLLTHAY----
L. rhamnosus            LSSLSTY------HQLKEILHEAIDDPT---------------KLLDIEN----------IITW----STVF-EDHT------IFETKLAEIEW------LD-----PKKINE--LSGI----RYRG------WGQFSRKLLDGLKLGN-
F. magna                ASNMKSY------IDFNNILEDKF-------------------DVEMVED----------LIEK----ITIHTGNKK------LLKKYIEETYPD-----LS-----SSQIQK--IINL----KYKD------WGRLSRKLLDGIKGTKK
P. antarcticus          ATSLSGY------ISMKSILGAVVDD-----------------NPAMTEE----------LIYW----IAVF-EDRE------ILHLKIQEKYP-----SIT-----DVQRQK--LALV----KLPG------WGRFSRLLIDGLPLDE-
consensus/95%           ...h..a...........h..........................E.          .h....  .........      .h...h............      ....... h...    .....      au.hS..hh........
consensus/85%           ...h..Y...........hh.........................hEc         hh....  ..ha.-.c.      .h...h..........h.      ....... h...    ..a..      WG.hS..hh........
.
Jnet 227501312          ----HHH-HH--HH---HHHHHHH------------------------H---------HHHH----HEE----HH------HHHHHHHH---------------HHHHHH--HHH---------------HHHHHHHHHHHHHH---
M. mobile               FENYLIS-NENDLSKLMSFYKQQSEKLFVPNEKGSYEINQNNNVLHIFDA--------I--SNI----LNKF-S---------TIQDRIRILEGYFEFSNLK-----KDVKSSEIYSEIAKLREFSG------TSSLSFGAYYKFIPNLI
M. gallisepticum        LSSLAKFVHKLKQHLKLDFLLEDQYATLDKINFLQSLFVYLGKHLRYSNR--------VDSANL----KEFS-DSNK------LFERILQKQKDGLFKLFEQ-----TDKDDEK-ILAQ-------------THSLSTKAMLLAITRMT
M. synoviae             SNATADAALEFSSLKPLYVFIKFLKEKKLDLNYLLGLE-----NTEILYF--------LDSIYLAISYSSDLKERNE------WFKKLLKELYPKIKNNNLEIIENVEDIFE---ITDQEKFESFSK------THSLSREAFNHIIPLLL
M. ovipneumoniae        EPTT---KLEVTKHLLATIYSHSSDS-----NWININN-----ILEFLPY--------LDAICIILDREKS-RGQDE-----VLKKLTEKNIFEVLKIDREKQLDFVKSIFS---NTKF----NFKK------IGNFSLKAIREFLPKMF
M. canis                ESAK-ENKFKFQDLKILNVLINLLDNVGIKFEFKDRND-----IIKNLEL--------LDNLYLFLIYQKESNNKDS-----SIDLFIAKNESLNIENLKLKLKEFLLGAGN---EFENHN----SK------THSLSKKAIDEILPKLL
consensus/95%           .........c......h.hh.........................c....        h....h............     ............h.............................      ...hS..Ah..hh..h.
consensus/85%           .........c......h.hh.........................c....        h....h............     ............h.............................      ...hS..Ah..hh..h.

E. rectale              FHKFEVY----------NKMRKALLEIGIDISNY---------SREELDE----------IGYI----MTIN-TDKE------AMMEAFQKSWID-----LS-----DDVKQCLINMRKTNGALFNK------WQSFSLKIMNELIPEMY
E. faecalis             IHTLKGY----------RNWRKIFAEAGIDLATL---------PTETIDC----------LAKV----LTLN-TERE------GIENTLAFE--------LP-----ELSESVK-LLVLDRYKELSQSISTQSWHRFSLKTLHLLIPELM
S. thermophilus         IHTFEAY----------RKMKTLETLDIEQMD------------RETLDK----------LAYV----LTLN-TERE------GIQEALEHE--------FA-----DGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELY
S. lugdunensis          FTEFKLY----------HDLKSVLFDQSILE-----------NEDVLDQ----------IAEI----LTIY-QDKD------SIKSKLTELDI------LL-----NEEDKEN-IAQL---TGYTG------THRLSLKCIRLVLEEQW
E. dolichum             LSELTGY----------KMIREVLEKSNDEHLE----------DHVFYDE----------IAEI----LTKT-KDIE------GRKKQISE---------LS-----SDLNEES-VHQLAGLTKFTA------YHSLSFKALRLINEEML
R. albus                TKCFKYI----------KVVKPLFEKY--GYDWDKLIENYTDTDNNVLNR----------IGIV----LSQA-QTPK------RRREKLKALNIG-----LD-----DGLIN-E-LTKL----KLSG------TANVSYKYMQGSIEAFC
R. inulinivorans        DTEKAKF------------ASMFWTYEYSKCLKDRTEEMPVGEKADLFDR----------IGEI----LTAY-KNDD------SRSSRLKELG-------LS-----GEEIDGL-LDLS-----PAK------YQRVSLKAMRKMQPYLE
L. coryniformis         FVDFKAY----------HDLVKVLPEEMWQ-------------SRQLLDH----------IGTA----LTLY-SSDK------RRRRYFAEELN------LP-----AELIE-K-LLPL----NFSK------FGHLSIKSMQNIIPYLE
I. polytropus           KTEKTKF------------ISLIGNYTIKKLNLSEKLKSEIEEDKSKLDK----------IIEI----LTFN-KSDK------TIESNLKKLE-------LS-----REDIEIL-LSEE-----FSG------TLNLSLKAIKKILPYLE
consensus/95%           ......h          ..h..h......................h..        h..h    ho.. ...c      .....h......     h.     ...........................hShK.h........
consensus/85%           ......a          ..h..h......................hD.        hu.h    LT.. ...c      .....h......     L.     ........h........h..........hShK.h..h...h.

Bacteroides sp.20       YVEDKPY-PCNETHA--QIIARLGKAHIPKAFL----------SKEKEET----------LWHI----LYSI-EDKQ------EIEKALHSFANKNN---LS-----EEFIE-Q-FKNF---PPFKKE-----YGSYSAKAIKKLLPLMR
I. album                LFANRKELKGNETKS--RYRKVFKKLGFDGEYILN--------HPSKLNR----------LWHS----DYSNDYADKE-----KTEKSILSSLGWKNRNGKWEKSKNYDVFNLPLEVAKAIANLPPLKKE---YGSYSALAIRKMLVVMR
B. fragilis             MIEMSGH-EPVDFKKPVEKVVEYIKAVFDLLNWNTDILGFN--SNEELDNQPYYK-----LWHL----LYSF-EGDNTPTGNGRLIQKMTELYGF------------EKEYATI-LANV----SFQDD-----YGSLSAKAIHKILPHLK
N. salsuginis           FVPFFFL-AD--IAKFKELFASIQKHPD---------------ALQIFRE----------LAEI----LQRS-KTPQ------EALDRLRALMAGKG---ID-----TDDRE---LLEL-FKNKRSG------TRELSHRYILEALPLFL
E. minutum              AQKIKGY----------KLIKLES-----MPFWARL-------SEAQQDS----------FLYD----WNSC-PDEK------LLTEKLSNEYH------LT-----EEEIDNA-FNEI---VLSSS------YAPLGKSAMLIILEKIK
S. globus               IQKIAGF----------------RFRRLEEKSFWKRL------SEEQKDD----------FFSA----WTNT-PDDK------RLSKYLMKHLL------LT-----ENEVVDA-LKTV---SLIGD------YGPIGKTATQLLMKHLE
M. trichosporium        VVARSGK-----AAEGTARLRSVIVDALGELAWGALLC-----SPEKLDK----------IAEV----ISFR-SDIG------RISEGLAQAGCNAP---LV-----DALTA---AASDGRFDPFTG------AGHISSKAARNILSGLR
F. branchiophilum       TVAACQVAASLKNAIGEDWK--------TKSFKYQTINSNKEQVSRTVDY--------KDLWHL----LTVA-TSDV------YLYEFAIDKLG------LD-----EKNAK-A-FSKT---KLKKD------FASLSLSAINKILPYLK
Prevotella C561         SVAGCPITARFRKMLGEEWE-----SFRVEGQKERQAHSKNNISFHRVSY------SIEDIWHF----CYDA-EEPE------AVLAFAQETLR------LE-----RKKAE-E-LVRIWS-AMPQG------YAMLSQKAIRNINKILM
P. timonensis           GVPGCPTIAQLKSIFGDDWK-----TGIAETYTLIQKKNGSKSLQEMVD----------DVWNV----LYSF-SSVE------KLKEFAHHKLQ------LD-----EESAE-K-FAKI---KLSHS------FAALSLKAIRKFLPFLR
P. timonensis           GVPGCPTIAQLKSIFGDDWKTGIAETYTLIQKKNGSKSL----QEMVDD----------VWNV----LYSF-SSVE------KLKEFAHHKLQ------LD-----EESAE-K-FAKI---KLSHS------FAALSLKAIRKFLPFLR
A. paucivorans          FVHLAGI----------HEIRKALDKGPDPAVWDTLRS-----RRDLLDS----------IADT----LTFY-KNED------EILPRLESLG-------LS-----PENAR-A-LAPL----SFSG------TAHLSLSALGKLLPHLE
P. marinum              IGNLTSC-------------IMRDAKKMIGSAWDRL-------DEEEQDS----------FILM----LQDDQKGDD------EVRSILTQQYG------LS-----DDVAEDC-LDVR----LPDG------HGSLSKKAIDRILPVLR
A. denitrificans        LKGNATS-------------AALARKELFGAAWSGF-------DEALQDE----------IVWQ----LVTE-EGEG------ALIAWLQTHTG------VD-----EARAQAIVDVSL-----PEG------YGNLSRKALARIVPALR
R. syzygii              --ELKGC----------DTAAKLAKVNALGTRWQAL-------SLDEQDR----------LVCL----LLDG-ENDA------VLADALREHYG------LT-----DAQID-T-LLGL---SFEDG------HMRLGRSALLRVLDALE
D. shibae               ASPAHP-----------DRFGPRWSILDADAQW-----------EVISR-----------IRR----VQSD-AEHA------ALVDWLTEAHG------LD-----RAHAEATAHAPL-----PDG------YGRLGLTATTRILYQLT
R. rubrum               RGTAGNL--------------TEAILAPLIPGWSGW-------DLDRKDR----------VFSD----LWAARQDRS------ALLALIGDPRGPTR---VT-----EDETAEA-VADAIQIVLPTG------RASLSAKAARAIAQAMA
Azospirillum B510       LGDETGA----------------RIAPAFGPGWTAL-------PLEEQDA----------LVEL----LLTE-AEPE------RAIAALTARWA------LD-----EAT-----AAKLAGATLPDF------HGRYGRRAVAELLPVLE
C. salinarum            ITVNPTD----------------PLMARWIDGWVDL-------PLKARSL----------YVRD----VVARGADSA------ALARLLAEGAHGVP-PVA------AAAVPAATAAILESDIMQPG------RYSVCPWAAEAILDAWA
N. hamburgensis         DGDQTA-----------ARLSDKKGF---NKAWRGF-------PPERQIA----------IVAR----LEET-EDEN------ELIAWLEKECA------L------DGAAAAR-VANT---TLPDG------HCRLGLRAIKKIVPIMQ
Bradyrhizobium BTAi1    KGDATG-----------AILSARRHF---GPAWHDR-------SLDRQID----------IVAL----LESA-LDEA------AIIASLGTTHS------LD-----EAAAQRA-LSAL----LPDG------YCRLGLRAIKRVLPLME
W. succinogenes         PNEPTEL-EFDKVEAEKKAWISLRGAAKLREALGNEFYGRFVALGKHADE----------ATKI----LTYY-KDEG------QKRRELTKLP-------LE-----AEMVE-R-LVKI----GFSD------FLKLSLKAIRDILPAME
C. jejuni               FKKYKEF-IK-------ALGEHNL-------------------SQDDLNE----------IAKD----ITLI-KDEI------KLKKALAKYD-------LN-----QNQID-S-LSKL----EFKD------HLNISFKALKLVTPLML
Jnet 218563121          ----HHH-HH-------H-----------------------HHHHHH----------HHHH----HH-----HH------HHHHHHHH---------------HHHHH-H-HH----------------HHHHHHHHHHHHHHHHH
H. mustelae             FLSLPST----------FKLNKITQ------------------DRKTQDK----------IANI----LGAN-KDWE------AILKELESLQ-------LS-----KEQIQTIKDAKL----NFSK------HINLSLEALYHLLPLMR
A. ebreus               LVKLPAW----------HELRKAFKAAGHEALWQQISTPALDGDPTLLDQ----------IATV----LSVY-KDGA------EVVQQLRQLA-------LP-----EPAASIAVLEKI----SFDK------FSSLSLKALRRIVPLMQ
B. smithii              FLELDSY----------HKIRKCIENVYGKDGIRMF-------NETDIDT----------FGYA----LTIF-KDDE------DIVAYLQNEYITKNGKRVSNLA--NKVYDKS-LIDELLNLSFSK------FAHLSMKAIRNILPYME
C. cellulolyticum       FYEMKSY----------HKLKSTLPTDIWGKLHS---------NKESLDN----------LFYC----LTVY-KNDN------EIKDYLQANN--------------LDYLIEY-IAKL---PTFNK------FKHLSLVAMKRIIPFME
HF0070_07E19            EYLQQQERKTLA-----NFRNWKQLEKIIGTGHP----------IQTLDE----------AARL----ITLI-KDDE------KLSDQLADL--------LP---EASDKAITQ-LCEL----DFTT------AAKISLEAMYRILPHMN
P. lavamentivorans      LGNALEA-KLADMFG--PDWPAHPRKQEIRHAVHERLWAADYGETPDKKR----------VIIL------SE-KDRK------AHREAAANSF-------VA-----DFGITGEQAAQLQALKLPTG------WEPYSIPALNLFLAELE
N. meningitidis         LMEMKAY----------HAISRALEKEGLKDKKSPLNL-----SPELQDE----------IGTA----FSLF-KTDE------DITGRLKD-R-------IQ-----PEILEAL-LKHI----SFDK------FVQISLKALRRIVPLME
P. multocida            FMELKAW----------HAIRKALENQGLKDTWQDLAK-----KPDLLDE----------IGTA----FSLY-KTDE------DIQQYLTN-K-------VP-----NSVINAL-LVSL----NFDK------FIELSLKSLRKILPLME
A. muciniphila          ALYLNPAVEVLQRSGIGQILSPSVYRI----------------AANRLRR--GKSVTPNYLLNL----LKSRGESGE------ALEKKIEKESKK-----KE-----ADYADTP-LKPK----YATG------RAPYARTVLKKVVEEIL
Actinomyces F0310       YPPVDDT----------TVRVMSAEVDWLADWWDCANDESRGHMIDAISN-----------GCG----SEPDDVEDE------EVNELISSA--------------TAEDMLK-LELLAKK-LPSG------RVAYSLKTLREVTAAIL
A. cellulolyticus       YSQFAPF-DETSARIAEFIAKNRRKIPTFAQWWQEQ-------DRTSRSD----------LVAA----LADNSIAGE------EEQELLVH---------LP-----DAELE-A-LEGL---ALPSG------RVAYSRLTLSGLTRVMR
B. longum               ISSRPPR-----LTSVQRIYESDNKIRKPLVAWWKSA------SDNEHEA----------MIRL----LSNT-VDID------KVREDVAYASAIEFIDGLD-----DDA-----LTKLDSVDLPSG------RAAYSVETLQKLTRQML
A. naeslundii           SAARPP------IDATDRIM-RQTKI-SSLKTWWEEA------DSEQRGA----------MIRY----LY---------------EDPTDSECAEIIAELP-----EED-----QAKLDSLHLPAG------RAAYSRESLTALSDHML

                        -------------HHHHHH-HH----HHHHHHHH-----------HHH----------HHHH----HH-----------------------------------HHHHH----HHHH------------EEEEEHHHHHHHHHHHHH
domains                 hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
catalytic residues      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

A. hesperidum           NKTEDTV-----FVEMKFYH-EVRKALQRAGVFIQD-----------LSIETLDQ-----IGWI----LSV---WKS---------DDNRRKKLSTL--GLS-----DNV-----IEELLPL----N------GSKFGHLSLKAIRKILP
Alcanivorax W11-5       DLSGNTT-----LAAWRKLGLEDRWQELDEVTQIQVINFLA---DLGSPEQLDTD-----DWSC----RFM---GKN------------GRPRNFSDEFVAFM-----NELRMTDGFDRLSKMGFEGG------RSSYSIKALKALTEWMI
B. zoohelcum            KLKGNET-----KLQLQKSL----------GELWDVLGL--------DSINRQIE-----LWNI----LYNE-KGNE------YDLTSDRTSKVLEFINKYG-----NNIVDDNAEETAIRISKIKF------ARAYSSLSLKAVERILP
Porphyromonas F0450     PYPQYHH-----LLDMELETRMMTVQLTDEETGEVTEREVAV-VTDSYVRKPLYR-----LWHI----LYS---IEE---------REAMRRALITQLGMKE-----EDL-DGGLLDQLYRLDFV--------KPGYGNKSAKFICKLLP
B. intestinihominis     NYPDKKR-----LLQFNLQEESSSMVNTETGEIIPM-------ISLSFEQEPLYR-----LWHV----LYS---IDD---------REQLQSVLRQKFGIDD-----DEV-----LERLSAIDLV--------KAGFGNKSSKAIRRILP
O. laneus               --PACTT-----ISQLRKLF---------PHPVWEE---------------KREE-----IWHC----FYF---YDD---------NTLLFEKLQKDYALQT-----NDL------EKIKKIRL---------SESYGNVSLKAIRRINP
Treponema JC4           -LIGNPT-----AVKMRSKN--------RFGKLWDE-----------IPLEEQDL-----IIET----IIT---ADE---------DDAVYEVIKKYD--LT-----QEQ----RDFIVKNTILQSG------TSMLCKEVSEKLVKRLE
Rhodovulum PH10         DGALTAK-----LLSKPDHY----------GAAWHG-----------LSFAEKDT-----FVGK----LLD---EAD---------EERLIRRLVT-ENRLS-----EDA-----ARRCASIPLADG------YGRLGRTANTEILAALV
consensus/95%           ..............................................................    .......................................................................   ...hu......h...h.
consensus/85%           ............................................................hh..    h...................h..........................h.............   ...hu..uh..h...h.

#map                    -------------------------------------------------------------------------------------------------------------------------------------*****************
```

```
Jnet 331001027          ------------------EEEEE--E----------------------------------------------------------------------------------------------------------------------------
P. excrementihominis    -------NGF----SKVSLAAHLENAWRMT---MTDG-----------------------------------------------------------------------------------------------------------------
W. succinogenes         -------SGF----SSTCKRCTAENRFRSETAFYNDETGE--------------------------------------------------------------------------------------------------------------
L. pneumophila          -------DGF----HKNCVAVTCENYWRSQKTEIDPEISY--------------------------------------------------------------------------------------------------------------
S. wadsworthensis       -------SGF----SATTLAVHLENAWRMT---IKDAVIN--------------------------------------------------------------------------------------------------------------
HTCC5015                -------AGF----NKTCRACTYENIWRMQEEKVESLLTNQLLSE---------------------------------------------------------------------------------------------------------
F. novicida             -------KGN----ANTCAVCSADNAHRMQQIKITEPVED--------------------------------------------------------------------------------------------------------------
consensus/95%                  .G.    .........-N.aR.....h...........
consensus/85%                  .G.    .........-N.aR.....h...........

A. coleocanis           ------ENGE----DLFEAR----------------------------------------------------------------------------------------------------------------------------------
C. glomerans            ------EDSV----SIMDVL-----------------------------------------------------------------------------------REGCPVS----------------------------------------
Acidaminococcus D21     ------ETGE----RASILD------------------------------------------------------------------F---------------------------------------------------------------
V. atypica              -------AGN----GAPKTI------------------------------------------------------------------I---------------------------------------------------------------
F. nucleatum            ------ETGE----CYSSVM----------------------------------------------------------------------------------------------------------------------------------
F. alocis               ------STGE----TFDIIT----------------------------------------------------------------------------------------------------------------------------------
S. moorei               ------EDGE----ACSILD------------------------------------------------------------------I---------------------------------------------------------------
C. catus                ------GTGE----VWNIMT------------------------------------------------------------------A---------------------------------------------------------------
T. denticola            --------------GFSEPV----------------------------------------------------------------------------------------------------------------------------------
P. duerdenii            ------TTGE----KGSIIY----------------------------------------------------------------------------------------------------------------------------------
C. mitsuokai            -------FGS----SVTVLD----------------------------------------------------------------------------------------------------------------------------------
S. thermophilus         -----GNTIL----DYLIDD----------------------------------------------------------------------------------------------------------------------------------
S. mutans               -----RKTIL----DYLIDD----------------------------------------------------------------------------------------------------------------------------------
S. pyogenes             -----GKTIL----DFLKSD----------------------------------------------------------------------------------------------------------------------------------

                        --------HH----HHH-------------------------------------------------------------------------------------------------------------------------------------
domains                 hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
catalytic residues      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
domains                 111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
catalytic residues      111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111111
.
B. bifidum              I--SDDFAPR----KHSIIE----------------------------------------------------------------------------------------------------------------------------------
O. kitaharae            D-KIDNQTFDVPRMNQSIID----------------------------------------------------------------------------------------------------------------------------------
S. sanguinis            ---------T----GFNLLQ----------------------------------------------------------------------------------------------------------------------------------
F. fructosus            G--------Q----KVSILD----------------------------------------------------------------------------------------------------------------------------------
E. yurii                T--G-----E----VRSIIQ----------------------------------------------------------------------------------------------------------------------------------
L. farciminis           T--PQLLQLS----NYSILD----------------------------------------------------------------------------------------------------------------------------------
S. pseudintermedius     ---------Q----GHSIIE----------------------------------------------------------------------------------------------------------------------------------
L. rhamnosus            --------------GHTVIQ----------------------------------------------------------------------------------------------------------------------------------
F. magna                -------ETE----KTDTVI------------------------------------------------------------------N---------------------------------------------------------------
P. antarcticus          ---------Q----GQSVLD----------------------------------------------------------------------------------------------------------------------------------
consensus/95%           . ..................                                                                  .                .......
consensus/85%           . ..................                                                                  .                .......
.
Jnet 227501312          ---------H----HHHHHH----------------------------------------------------------------------------------------------------------------------------------
M. mobile               ------SEGS----KNYSTI----------------------------------------------------------------------------------------------------------------------------------
M. gallisepticum        --------NL----DNDEDN----------------------------------------------------------------------------------------------------------------------------------
M. synoviae             --------------SNNEGK----------------------------------------------------------------------------------------------------------------------------------
M. ovipneumoniae        ----------------EQNK----------------------------------------------------------------------------------------------------------------------------------
M. canis                --------------DNNEGW----------------------------------------------------------------------------------------------------------------------------------
consensus/95%                 ....    ......
consensus/85%                 ....    ......

E. rectale              -----AQPKE----QMTLLT----------------------------------------------------------------------------------------------------------------------------------
E. faecalis             -------NAT----SEQNTL----------------------------------------------------------------------------------------------------------------------------------
S. thermophilus         -----ETSEE----QMTILT----------------------------------------------------------------------------------------------------------------------------------
S. lugdunensis          ---------------YSSRN----------------------------------------------------------------------------------------------------------------------------------
E. dolichum             -----KTELN----QMQSIT----------------------------------------------------------------------------------------------------------------------------------
R. albus                -------EGD----LYGKYQ----------------------------------------------------------------------------------------------------------------------------------
R. inulinivorans        -------DGL----IYDKAC----------------------------------------------------------------------------------------------------------------------------------
L. coryniformis         -------MGQ----VYSEAT----------------------------------------------------------------------------------------------------------------------------------
I. polytropus           -------KGL----SYNEAC----------------------------------------------------------------------------------------------------------------------------------
consensus/95%                .....    ......
consensus/85%                .....    .h....

Bacteroides sp.20       -------MGK----YWSIEN------------------------------------------------------------------IDNGTRIRINKIIDGEYDENIR------------------------------------------
I. album                -------DGKYWQHPDQIAKDQENTSLMLFDKNLIQLTNNQRKVLNKYLLTLAEVQKRSTLIKQKLNEIEHNPYKLELVS----------------------------------------------------------------------
B. fragilis             -------EGN----RYDVAC----------------------------------------------------------------------------------------------------------------------------------
N. salsuginis           -------EGY----DEKEVQ------------------------------------------------------------------R------IL-------------------------------------------------------
E. minutum              -------NDL----SYTEAV----------------------------------------------------------------------------------------------------------------------------------
S. globus               -------DGL----TYTEAL----------------------------------------------------------------------------------------------------------------------------------
M. trichosporium        -------QGM----TYDKAC----------------------------------------------------------------------------------------------------------------------------------
F. branchiophilum       -------EGL----LYSHAVFVANIENIV---DENIWKDEKQRDYIKTQISEIIENYTLEKSRFEIINGLLKEYKSENEDGK--RVYYSKEAEQSFENDLKKKLVLFYKSN--EIENKEQQETIFNELLPIFIQQLKDYE--FIKIQRLD
Prevotella C561         -------LGL----KYSDAVILAKVPELVDVSDEELLSIAKDYYLVEAQVNY-------DKRINSIVNGLIAKYKSVSEEYRFADHNYEYLLDESDEKDIIRQIENSLGARRWSLMDANEQTDILQKVRDRYQDFFRSHERKFVESPKLG
P. timonensis           -------KGM----YYTHASFFANIPTIV---GKEIWNKEQNRKYIMENVGE-----------------LVFNYQPKHREVQ----------------------------------------------------------------GTIE
P. timonensis           -------KGM----YYTHASFFANIPTIVGKEIWNK--------------------------------------------------EQNRKYIMENVGELVFNYQPKHREVQGTIEMLIKDFL---------------------------
A. paucivorans          -------EGK----SYTQAR----------------------------------------------------------------------------------------------------------------------------------
P. marinum              ------DQGL----IYYDAV----------------------------------------------------------------------------------------------------------------------------------
A. denitrificans        ------AAVI----TYDKAV-------------------------------------------------------------------------------------QAAGFDHHSQLGFEYDAS---------------------------
R. syzygii              SGRDEQGLPL----SYDKAV------------------------------------------------------------------V---------------------------------------------------------------
D. shibae               ------ADVV----TYADAV----------------------------------------------------------------------------------------------------------------------------------
R. rubrum               -------PGI----GYDEAV----------------------------------------------------------------------------------------------------------------------------------
Azospirillum B510       RETRGDPDGRVRPIRLDEAV----------------------------------------------------------------------------------------------------------------------------------
C. salinarum            NA-----PTE---GFYDVTRGLFGFAPGEIVLEDLRRARGALLA----------------------------------------------------------------------------------------------------------
N. hamburgensis         -------DGL----DEDGVA-----------------------------------------------------------------------------------GAGYHI-----------------------------------------
Bradyrhizobium BTAi1    -------AGR----TYAEAA----------------------------------------------------------------------------------------------------------------------------------
W. succinogenes         -------SGA----RYDEAV----------------------------------------------------------------------------------------------------------------------------------
C. jejuni               -------EGK----KYDEAC------------------------------------------------------------------N---------------------------------------------------------------
Jnet 218563121          ---------------HHHHH----------------------------------------------------------------------------------------------------------------------------------
H. mustelae             -------EGK----RYDEGV----------------------------------------------------------------------------------------------------------------------------------
A. ebreus               -------SGL----RYDEAV----------------------------------------------------------------------------------------------------------------------------------
B. smithii              -------QGE----IYSKAC----------------------------------------------------------------------------------------------------------------------------------
C. cellulolyticum       -------KGY----KYSDAC----------------------------------------------------------------------------------------------------------------------------------
HF0070_07E19            -------QGM----GFFDAC----------------------------------------------------------------------------------------------------------------------------------
P. lavamentivorans      -------KGE----RFGALV----------------------------------------------------------------------------------------------------------------------------------
N. meningitidis         -------QGK----RYDEAC----------------------------------------------------------------------------------------------------------------------------------
P. multocida            -------QGK----RYDQAC----------------------------------------------------------------------------------------------------------------------------------
A. muciniphila          -------DGE----DPTRPA----------------------------------------------------------------------------------------------------------------------------------
Actinomyces F0310       ------ETGD----DLSQAI----------------------------------------------------------------------------------------------------------------------------------
A. cellulolyticus       ------DDGV----DVHNAR----------------------------------------------------------------------------------------------------------------------------------
B. longum               ------TTDD----DLHEAR----------------------------------------------------------------------------------------------------------------------------------
A. naeslundii           ------ATTD----DLHEAR----------------------------------------------------------------------------------------------------------------------------------

                        ------H--------HHHHH----------------------------------------------------------------------------------------------------------------------------------
domains                 hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
catalytic residues      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh

A. hesperidum           ----FLEDGY----SYDVACE-------LAGYQFQ-------------------------------------------------------------------------------------------------------------------
Alcanivorax W11-5       AP--HWRETP----ETHRVDE-------EAAIREC--------------------------------------------------------------------------------------YP---------------------------
B. zoohelcum            ----LVRAGK----YFNNDFSQQLQSKILKLLNENVEDPFAKAAQ-------------------------------------------------------------------------TYLDN---------------------------
Porphyromonas F0450     ----QLQQGL----GYSEACA-------AVGYRHS--------------------------------------------------------------------------------------NS---------------------------
B. intestinihominis     ----FLQLGM----NYAEACE-------AAGYNHS--------------------------------------------------------------------------------------NN---------------------------
O. laneus               ----YLKKGY----AYSTAVLLGGIRN-SFGKRFEYFKEYEPEIE-------------------------------------------------------------------------KAVCR---------------------------
Treponema JC4           -----EIADL----KYHEAVE-------SLGYKFA-------------------------------------------------------------------------------------------------------------------
Rhodovulum PH10         EETDETGTVV----TYAEAVR-------RAGERTG--------------------------------------------------------------------------------------RN---------------------------
consensus/95%           ......................................................................................................................................................
consensus/85%           ...............h..h...................................................................................................................................

#map                    ------------------------------------------------------------------------------------------------------------------------------------------------------
```

```
Jnet 331001027          ----------------------------------------------------------------------------EE---------HHHHHHHHHHHHHHHHHHHHHHHHH----------------------EEE--EEE-EE-----
P. excrementihominis    ------------------------------------------------------------------------SAQCCRLPADCVRPFDGFIRKAIDRNSWEVAKRIAEEVKKSVDFT-------------NG-TVKI--PVA-IEANSFN
W. succinogenes         -------------------------------------------------------------------FHKKATATCQRLPADTQRPFSGKIERYIDKLGYELAKIKAKELEGMEA----------------K-EIKV--PII-LEQNAFE
L. pneumophila          ---------------------------------------------------------------------------ASRLPADSVRPFDGVLARMMQRLAYEIAMAKWEQIKHIPD---------------NS-SLLI--PIY-LEQNRFE
S. wadsworthensis       -------------------------------------------------------------------GETVRAAQCSRLPAETARPFDGLVRRLVDRQAWEIAKRVSTDIQSKVDFS-------------NG-IVDV--SIF-VEENKFE
HTCC5015                --------------------------------------------------------------IHGERKVPLKSAMCTRLSADSTRPFDGQMASIIEHIARKIAQHKIAQINDVPK---------------EF-SIDI--PII-IESNQFS
F. novicida             -----------------------------------------------------------------NKDKIILSAKAQRLPAIPTRIVDGAVKKMATILAKNIVDDNWQNIKQVLS-A-------------KH-QLHI--PII-TESNAFE
consensus/95%                                                                         ...............RL.A...R.h.G.h..hh...u..hh......h.......             .. .h.h  .hh .E.N.F.
consensus/85%                                                                         ...............RL.A...R.h.G.h..hh...u..hh......h.......             .. .h.h  .hh .E.N.F.

A. coleocanis           ------------VNEFGVSEDWRPPA-----------------------------------------------------EPIGARVGNPAVDRVLKAVNRYLMAAEAE----------------------WG--APL--SVN-IEHVREG
C. glomerans            ------------GKRGRAMVMMEILRDEELGFQKKVDDFNRAFF----------------------------AENAQALG-VNELPGSPAVRRSLNQSIRIVDEIASI----------------------AGK-APA--NIF-IEVTRDE
Acidaminococcus D21     ------------MWNTNDNLMQLLSECYT--FSDEITKLQEAYY--------------------------AKAQLSLNDF-LDSMYISNAVKRPIYRTLAVVNDIRKA----------------------CGT-APK--RIF-IEMARDG
V. atypica              ------------ELMRNDSYNLMEILGDKFSFMECIEEENAKLA--------------------------QGQVVNPHDI-IDELALSPAVKRAVWQALRIVDEVAHI----------------------KKA-LPS--RIF-VEVARTN
F. nucleatum            ------------EALRRTNYNLMELLSSKFTLQESIDNENKEMN-----------------------------EVSYRDL-IEESYVSPSLKRAILQTLKIYEEIKKI----------------------TGR-VPK--KVF-IEMARGG
F. alocis               -------------AMWETDNNLMQILSKKFTFMDNVEDFNSGKV-------------------------GKIDKITYDST-VKEMFLSPENKRAVWQTIQVAEEIKKV----------------------MGC-EPK--KIF-IEMARGG
S. moorei               ------------MWNTNATLMEILSNEKYQ-FKQNIENYKAENY---------------------------DEKQNLHEE-LDDMYISPAARRSIWQALRIVDEIVDI----------------------KKS-APK--KIF-IEMAREK
C. catus                ------------LWQTNDNLMQLLSRNYG--FTNEVEEFNTLKK----------------------------ETDLSYKT-VDELYVSPAVKRQIWQTLKVVKEIQKV----------------------MGN-APK--RVF-VEMAREK
T. denticola            ---------NIITAMRETQNNLMELLSSEFTFTENIKKINSGFE-------------------------DAEKQFSYDGL-VKPLFLSPSVKKMLWQTLKLVKEISHI----------------------TQA-PPK--KIF-IEMAKGA
P. duerdenii            ------------FMREYNLNLMELMSGHYT-FTEEVEKLNPVEN-----------------------------RELCYEM-VDELYLSPSVKRMLWQSLRVVDEIKRI----------------------IGK-DPK--KIF-IEMARAK
C. mitsuokai            ------------VLEMSRLNLMEIINDKDLGYAQMIEEATSCPE----------------------------DGKFTYEE-VERLAGSPALKRGIWQSLQIVEEITKV----------------------MKC-RPK--YIY-IEFERSE
S. thermophilus         ------------GISNRNFMQLIHDD-----ALSFKKKIQKAQI------------------------IGDEDKGNIKEV-VKSLPGSPAIKKGILQSIKIVDELVKV----------------------MGGRKPE--SIV-VEMAREN
S. mutans               ------------GNSNRNFMQLINDD-----ALSFKEEIAKAQV------------------------IGETD--NLNQV-VSDIAGSPAIKKGILQSLKIVDELVKI----------------------MG-HQPE--NIV-VEMAREN
S. pyogenes             ------------GFANRNFMQLIHDD-----SLTFKEDIQKAQV------------------------SGQGD--SLHEH-IANLAGSPAIKKGILQTVKVVDELVKV----------------------MGRHKPE--NIV-IEMAREN

                        ------------------HHHHH----------HHHHHHHHHHH-------------------------------HHHHH-HH-----HHHHHHHHHHHHHHHHHHHH-------------------------------EEE-EEE----
domains                 hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh|llllll|rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr|ll
catalytic residues      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh|llllll|rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr*E*r|ll
domains                 1111111111111111111111111111111111111111111111111111111111111111111|llllll|rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
catalytic residues      1111111111111111111111111111111111111111111111111111111111111111111|llllll|rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr*E*rrrr
.
B. bifidum              ------------IMRAEDRNLMEIITDKQLGFSDWIEQENLGAE----------------------------NGSSLMEV-VDDLRVSPKVKRGIIQSIRLIDDISKA----------------------VGK-RPS--RIF-LELADDI
O. kitaharae            ------------TLYNTKMNLMEIINNAEDDFGVRAWIDKQNTT--------------------------DGDEQDVYSL-IDELAGPKEIKRGIVQSFRILDDITKA----------------------VGY-APK--RVY-LEFARKT
S. sanguinis            ----------FLRNDEENRNLTKLISDNTLSFEPKIKDI-QSKST---------------------------IEDDIFDE-IKKLAGSPAIKRGILNSIKIVDELVQI----------------------IGY-PPH--NIV-IEMAREN
F. fructosus            ------------KLKSTSQNFMSIINNDKYGVQAWITEQ---NTG---------------------------SSKLTFDEKVNELTTSPANKRGIKQSFAVLNDIKKA----------------------MKE-EPR--RVY-LEFARED
E. yurii                ------------SLWETNFNLMELLSSRFTYMDELEKRVKKLEKP---------------------------LSEWTIED-LDDMYLSSPVKRMIWQSMKIVDEIQTV----------------------IGY-APK--RIF-VEMTRSE
L. farciminis           ------------LMWATNNNFISIMSNDKYDFKNYIENH-NLNKN---------------------------EDQNISDL-VNDIHVSPALKRGITQSIKIVQEIVKF----------------------MGH-APK--HIF-IEVTRET
S. pseudintermedius     ------------LLRHSDENFMEILTNDVYGFQNFIKEE--NQVQ---------------------------SNKIQHQD-IANLTTSPALKKGIWSTIKLVRELTSI----------------------FGE--PE--KII-MEFATED
L. rhamnosus            ------------ELMLSNHNLMQILADET--LKETMTELNQDKL-----------------------------KTDDIEDVINDAYTSPSNKKALRQVLRVVEDIKHA----------------------ANGQDPS--WLF-IETADGT
F. magna                ------------FLRNSSDNLMQIIGSQNYSFNEYIDKLRKKYI-----------------------------PQEISYEVVENLYVSPSVKKMIWQVIRVTEEITKV----------------------MGY-DPD--KIF-IEMAKSE
P. antarcticus          ------------HMEQYSSVFMEVLKNKGFGLEKKIQKMNQHQVDG--------------------------TKKIRYED-IEELAGSPALKRGIWRSVKIVEELVSI----------------------FGE--PA--NIV-LEVARED
consensus/95%                    .....................................                      .............h.........cc.h...h.hh.-h..h                      .....P.  .hh hEh.c..
consensus/85%                    ......................h..............                      .............h..h..SP.hKc.h...h.hh.-h..h                      .....P.  .hh hEh.c..
.
Jnet 227501312          -----------H------HHHHHHHH-----HHHHHHHHHH------------------------------------------------HHHHHHHHHHHHHHHHHHH----------------------H--------EEE-EEEEE--
M. mobile               ------------SYEEKALQNQKN-------NFSHSNLFEKTWV--------------------------------------EDLIASPTVKRSLRQTMNLLKEIFKY--------------------SEKNNLEIE--KIV-VEVTRSS
M. gallisepticum        ------------QKNNDKGWNFEAIKNFDQKFIDITKKNNNLSL--------------------------KQNKRYLDDRFINDAILSPGVKRILREATKVFNAILKQ----------------------FSE-EYDVTKVV-IELAREL
M. synoviae             ------------NYESLKHSNEELKKRTEKAELKAQQN-----------------------------------QKYLKDNFLKEALVPLSVKTSVLQAIKIFNQIIKN----------------------FGK-KYEISQVV-IEMAREL
M. ovipneumoniae         ------------NSEYLKWKDEEIRRKWEEQKSKLGKTDKKT--------------------------------KYLNPRIFQDEIISPGTKNTFEQAVLVLNQIIKK----------------------YSK-ENIIDAII-IESPREK
M. canis                ------------NLEAIKNYDEEIKSQIEDNSSLMAKQDK----------------------------------KYLNDNFLKDAILPPNVKVTFQQAILIFNKIIQK----------------------FSK-DFEIDKVV-IELAREM
consensus/95%                       ..........c...........h.........                          .............-.hh....K..h.....hh..Ih..                    ......c.....hh hE..R..
consensus/85%                       ..........c...........h.........                          .............-.hh....K..h.....hh..Ih..                    ......c.....hh hE..R..

E. rectale              ------------EMGVTKGTQEEFAG------LKYIPV-----------------------------------------DVVSEDIFNPVVRRSVRISFKILNAVLKK----------------------YK--ALD--TIV-IEMPRDR
E. faecalis             -------------LEQFQLKSDVRKRYSEYKKLPT--------------------------------------------KDVLAEIYNPTVNKTVSQAFKVIDALLVK----------------------YGKEQIR--YIT-IEMPRDD
S. thermophilus         --------RLGKQK------TTSSSNK----------------------------------------------TKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKE----------------------YG--DFD--NIV-IEMARET
S. lugdunensis          ------------QMEIFTHLNIKPKKI----NLTAANKIPKAMI--------------------------------------DEFILSPVVKRTFGQAINLINKIIEK----------------------YG--VPE--DII-IELAREN
E. dolichum             ------------LFGLKQNNELSVKG-------------------------------------------------MKNIQADDTAILSPVAKRAQRETFKVVNRLREI----------------------YG--EFD--SIV-VEMAREK
R. albus                ------------AKFNKEIPDIDENAKPQ--KLPPFKN-----------------------------------------EDDCEFFKNPVVFRSINETRKLINAIIDK----------------------YG--YPA--AVN-IETADEL
R. inulinivorans        ------------EAAGYDFRALNDGNKKH--LLKGEEINAIV-----------------------------------------NDITNPVVKRSVSQTIKVINAIIQK----------------------YG--SPQ--AVN-IELAREM
L. coryniformis         ------------TNTGYDFRKKQISK-----------------------------------------------------DTIREEITNPVVRRAVTKTIKIVEQIIRR----------------------YG--KPD--GIN-IELAREL
I. polytropus           ------------EKADYDYKNNGIKFKRGE-LLPVVDK---------------------------------------------DLIANPVVLRAISQTRKVVNAIIRK----------------------YG--TPH--TIH-VEVARDL
consensus/95%                   ....................................                             ............h..P.h.c.......hh..h...                      Y......  .h. hE..c-.
consensus/85%                   ....................................                             ............I..PVV.c.h....Khh..hh.c                      YG.....  .h. hEh.R-.

Bacteroides sp.20       ------------ERVRQKAINLTDIT-----HFRALPLWLACYL----VYDRHSEVKDIVKWKTPK------DIDLYLKSFKQHSLRNPIVEQVITETLRTVRDIWQQ----------------------VG--HID--EIH-IELGREM
I. album                -------DQDLEKQVLKSFLEKKNESD----YLKGLKTYQAGYL----IYGKHSE-KDVPIVNSPDELGEYIR-----KKLPNNSLRNPIVEQVIRETIFIVRDVWKS----------------------FG--IID--EIH-IELGREL
B. fragilis             ------------VYAGYRHSESSLT------REEIANKVLKDRL----------------------------------MLLPKNSLHNPVVEKILNQMVNVINVIIDI----------------------YG--KPD--EIR-VELAREL
N. salsuginis           ------------GFDDRE-------------DYSRYPKSLRHLH----------------------------LREGNLFEKEENPINNHAVKSLASWALGLIADLSWR----------------------YG--PFD--EII-LETTRDA
E. minutum              ------------EEALKEGKLTKEKQAIKD-RLPYYGAVLQEST----------QKIIAKGFSPQFKDKGYKTPHTNKYELEYGRIANPVVHQTLNELRKLVNEIIDI----------------------LGK-KPC--EIG-LETAREL
S. globus               ------------ERGMETGEFQELSVWEQQSLLPYYGQILTGST------QALMGKYWHSAFKEKRDSEGFFKPNTNSDEEKYGRIANPVVHQTLNELRKLMNELITI----------------------LGA-KPQ--EIT-VELAREL
M. trichosporium        ------------CAADYDHTASRERG-----AFDVGGHGREALK-----------------------------RILQEERISRELVGSPTARKALIESIKQVKAIVER----------------------YG--VPD--RIH-VELARDV
F. branchiophilum       QKVLIFLKGKNETGQIFCTEEKGTAEEKEKKI--------------KNRLKKLYHPSDIEKFKKKIIKDEFGNEKIVLGSPLTPSIKNPMAMRALHQLRKVLNALILE--------------------GQID--EKT--IIH-IEMAREL
Prevotella C561         ESFENYLTKKFPMVE-------------------------------REQWKKLYHPSQITIYRPVSVGKDRS--VLRLGNPDIGAIKNPTVLRVLNTLRRRVNQLLDD--------------------GVISP-DET--RVV-VETAREL
P. timonensis           MLIKDFLANNFELP--------------------------------AGATDKLYHPSMIETY-PNAQRNEFG--ILQLGSPRTNAIRNPMAMRSLHILRRVVNQLLKE--------------------SIID--ENT--EVH-VEYAREL
P. timonensis           --------ANNFELPAGATDKLYHPS-----MIETYPNAQRNEF-----------------------------GILQLGSPRTNAIRNPMAMRSLHILRRVVNQLLKE--------------------SIID--ENT--EVH-VEYAREL
A. paucivorans          ------------ADAGYAAPPPDRHP-----KLPPLEE---------------------------------------------ADWRNPVVFRALTQTRKVVNALVRR----------------------YG--PPW--CIH-LETAREL
P. marinum              ------------KEAGLGEANLYDPYAALSDKLDYYGKALAGHV------------------------MGASGKFEDSDEKRYGTISNPTVHIALNQVRAVVNELIRL----------------------HG--KPD--EVV-IEIGRDL
A. denitrificans        --------EVEDL---VHPETGEIRSV--FKQLPYYGKALQRHV------------------------AFGSGKPEDPDEKRYGKIANPTVHIGLNQVRMVVNALIRR----------------------YG--RPT--EVV-IELARDL
R. syzygii              ------------AAGYPAHTADLENGERD--ALPYYGELLWRYT-------------------------QDAPTAKNDAERKFGKIANPTVHIGLNQLRKLVNALIQR----------------------YG--KPA--QIV-VELARNL
D. shibae               ------------KACGWHHSDGRTGECFD--RLPYYGEVLERHV------------------------IPGSYHPDDDDITRFGRITNPTVHIGLNQLRRLVNRIIET----------------------HG--KPH--QIV-VELARDL
R. rubrum               ------------TLALGLHHSHRPRQERLA-RLPYYAAALPDVG-----------------LDGDPVGPPPAEDDGAAAEAYYGRIGNISVHIALNETRKIVNALLHR----------------------HG--PIL--RLVMVETTREL
Azospirillum B510       -----------KLLRGGKDHSDFSREGALLDALPYYGAVLERHV------------------------AFGTGNPADPEEKRVGRVANPTVHIALNQLRHLVNAILAR----------------------HG--RPE--EIV-IELARDL
C. salinarum            -----------HLPRTMAAARTPNRAAQQRGPLPAYESVIPSQL-----------ITSLRRAHKGRAADWSAADPEERNPFLRTWTGNAATDHILNQVRKTANEVITK-----------------YGNRRGWDPLPS--RIT-VELAREA
N. hamburgensis         ------------AAKRAGYDHAKLPTGEQLGRLPYYGQWLQDAV-------------------------VGSGDARDQKEKQYGQFPNPTVHIGLGQLRRVVNDLIDK----------------------YG--PPT--EIS-IEFTRAL
Bradyrhizobium BTAi1    ------------SAAGYDHALLPGGKLSPTGYLPYYGQWLQNDV-------------------------VGSDDERDTNERRWGRLPNPTVHIGIGQLRRVVNELIRW----------------------HG--PPA--EIT-VELTRDL
W. succinogenes         -------------------LMLGVPHKEKSAILPPLNK-------------------------------------------TDIDILNPTVIRAFAQFRKVANALVRK----------------------YG--AFD--RVH-FELAREI
C. jejuni               ------------ELNLKVAINEDKKD-----FLPAFNE-----------------------------------------TYYKDEVTNPVVLRAIKEYRKVLNALLKK----------------------YG--KVH--KIN-IELAREV
Jnet 218563121          -----------HHHHHHHHHHHH------------------------------------------------------------------HHHHHHHHHHHHHHHHHHH----------------------H--------EEE-EEEEE--
H. mustelae             ------------EILQERGIFSKPQPKNRQ-LLPPLSELAK--------------------------------------EESYFDIPNPVLRRALSEFRKVVNALLEK----------------------YG--GFH--YFH-IELTRDV
A. ebreus               ----------AQIPEYGHHSQRIEPGAAKHLYLPPFYEAQRKYA------------------------GKGDHIGSMQFRDDADIPRNPVVLRALNQARKVVNALIRE----------------------YG--SPI--AVN-IEMARDL
B. smithii              ------------ELAGYNFTGPKKKEKAL--LLPVIPNI-----------------------------------------------ANPVVMRALTQSRKVVNAIIKK----------------------YG--SPV--SIH-IELARDL
C. cellulolyticum       ------------NMAELDFTGSS--------KLEKCNKLTV--------------------------------------EPIIENVTNPVVIRALTQARKVINAIIQK----------------------YG--LPY--MVN-IELAREA
HF0070_07E19            ------------QQESLPEIGVP--------PAGDRVPP-------------------------------------------FDEMYNPVVNRVLSQSRKLINAVIDE----------------------YG--MPA--KIR-VELARDL
P. lavamentivorans      ---------NGPDWEGWRRTNFPHRNQPTGEILDKLPSPASKEE-----------------------------------RERISQLRNPTVVRTQNELRKVVNNLIGL----------------------YG--KPD--RIR-IEVGRDV
N. meningitidis         ------------AEIYGDHY-GKKNTEEKI-YLPPIPA---------------------------------------------DEIRNPVVLRALSQARKVINGVVRR----------------------YG--SPA--RIH-IETAREV
P. multocida            ------------REIYGHHY-GEANQKTSQ-LLPAIPA---------------------------------------------QEIRNPVVLRTLSQARKVINAIIRQ----------------------YG--SPA--RVH-IETGREL
A. muciniphila          ------------RGEAHPDGELKAHDGCLYCLLDTDSSVNQHQK----------------------------------ERRLDTMTNNHLVRHRMLILDRLLKDLIQD----------------------FADGQKDRISRVCVEVGKEL
Actinomyces F0310       ------------TRLYGVDPGWVPTP-----------------------------------------------------APIEAPVGNPSVDRVLKQVARWLKFASKR----------------------WG--VPQ--TVN-IEHTREG
A. cellulolyticus       ------------KTCFGVDDNWRPPL-----------------------------------------------------PALHEATGHPVVDRNLAILRKFLSSATMR----------------------WG--PPQ--SIV-VELARGA
B. longum               ------------KTLFNVTDSWRPPA-----------------------------------------------------DPIGEPLGNPSVDRVLKNVNRYLMNCQQR----------------------WG--NPV--SVN-IEHVRSS
A. naeslundii           ------------KRLFGVDDSWAPPA-----------------------------------------------------EAINAPVGNPSVDRTLKIVGRYLSAVESM----------------------WG--TPE--VIH-VEHVRDG

                        ------------HHHH------------------------------------------------------------------------HHHHHHHHHHHHHHHHHHHH----------------------H--------EEE-EEE----
domains                 hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh|llllll|rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
catalytic residues      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh|llllll|rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr*E*rrrr

A. hesperidum           ------------GKTEYVKQRLLP-------------------------------------------------PLGEGE------VTNPVVRRALSQAIKVVNAVIRK----------------------HG---SPE-SIH-IELAREL
Alcanivorax W11-5       --------ESLATPAQGGRQSKLE-------------------------------------------------PPPL--------TGNEVVDVALRQVRHTINMMIDD----------------------LG--SVPA-QIV-VEMAREM
B. zoohelcum            --------NQSVLSEGGVGNSIAT--ILVYDKHTAKE----------------------------YSHDELYKSYKEINLLKQGDLRNPLVEQIINEALVLIRDIWKN----------------------YG--IKPN-EIR-VELARDL
Porphyromonas F0450     --------PTSEEITERTLLEKIP-------------------------------------------------LLQRNE------LRQPLVEKILNQMINLVNALKAE----------------------YG----ID-EVR-VELAREL
B. intestinihominis     --------YTKAENEARALLDRLP-------------------------------------------------AIKKNE------LRQPVVEKILNQMVNVVNALMEK----------------------YG---RFD-EIR-VELAREL
O. laneus               --------ILKEKNAEGEVIRKIKDYLVHNRFGFAKNDRAFQKLY--------------------HHSQAITTQAQKERLPETGNLRNPIVQQGLNELRRTVNKLLATCREK------------------YGPSFKFD-HIH-VEMGREL
Treponema JC4           -------------DQTVEKYDLLP-----YYGKVLPG----------------------------STMEIDLSAPETNPEKHYGKISNPTVHVALNQTRVVVNALIKE----------------------YG---KPS-QIA-IELSRDL
Rhodovulum PH10         --------WHHSDERDGVILDRLP-----YYGEILQR----------------------------HVVPGSGEPEEKNEAARWGRLANPTVHIGLNQLRKVVNRLIAA----------------------HG---RPD-QIV-VELAREL
consensus/95%           ............................................. ............................................h...h......h..h.......             .....h.........h..hE..R.h
consensus/85%           ............................................. .......................................h.NP.V...h.....hh..hh......             .....hG........h..hE..R-h

#map                    -------------------------------------------------------------------------------------***********************-------------------------------***********
```

```
Jnet 331001027          ---------------HHHHHH------------------HHHHHHH-----------------HHHHH-----HHHHH---------------------------------------HHH-HHH--------H-----------------
P. excrementihominis    F-------TASLTDLKYIQLK--------------EQKLKKKLEDI-----------------QRNEENQEKRWLSKE---------------------------------------ERI-RAD--------SHGI----------CAY-
W. succinogenes         Y-------EESLRKSKTGSND-----------RVINSKKDRDGKKL-----------------AKAKENAEDRLKDKD---------------------------------------KRI-KAF--------SSGI----------CPY-
L. pneumophila          F-------EESFKKIK-GSSS--------------DKTLEQ----------------------AIEKQNIQ--WEEKF---------------------------------------QRI-INA--------SMNI----------CPY-
S. wadsworthensis       F-------SASVADLK----K--------------NKRVKDKM--------------------LSEAEKLETRWLIKN---------------------------------------ERI-KKA--------SRGT----------CPY-
HTCC5015                F-------TAELEEIKRGRGS--------------AKAKKA----------------------KELGEKSKAGWVSKT---------------------------------------ERI-KTS--------SEGI----------CPY-
F. novicida             F-------EPALADVKGKSLK--------------DRRKK-----------------------ALERISPENIFKDKN---------------------------------------NRI-KEF--------AKGI----------SAY-
consensus/95%           a       ...h.c.K.....           .......c......                 ..........h..K.                                       .RI ...        u...          ..Y
consensus/85%           a       ...h.c.K.....           .......c......                 ..........h..K.                                       .RI ...        u...          ..Y

A. coleocanis           F-------ISKRQAVEIDRENQKRY--RQNQAVRSQIADHINATS-------------------------GVRGSDV----------------------------------------TRY-LAIQR------QNGE----------CLY-
C. glomerans            DPK-KKGRRTKRRYNDLKDALEAFK--KEDPELWRELC--------------------------------ETAPNDM---------------------------------------DERL-SLYFM------QRGK----------CLY-
Acidaminococcus D21     E---SKKKRSVTRREQIKNLYRSIR--KDFQQEVDFLEKILENKSDG-----------------------QLQSDAL--------------------------------------------YLYFA------QLGR----------DMY-
V. atypica              K---SEKKKKDSRQKRLSDLYSAI---KKDDVLQSGLQDKEFGALKSG----------------L-----ANYDDAALR-------------------------------------SKKL-YLYYT------QMGR----------CAY-
F. nucleatum            DES-MKNKKIPARQEQLKKLYDSCG--NDIANFSIDIKEMKNSLS-------------------------SYDNNSLRQ--------------------------------------KKL-YLYYL------QFGK----------CMY-
F. alocis               E---KVKKRTKSRKAQLLELYAACE--EDCRELIKEIE--------------------------------DRDERDFNS--------------------------------------MKL-FLYYT------QFGK----------CMY-
S. moorei               KSA-MKKKRTESRKDTLLELYKSCK--SQADGFYDEELFEKLSNESNS----------------------RLRRDQL--------------------------------------------YLYYT------QMGR----------SMY-
C. catus                Q----EGKRSDSRKKQLVELYRACK--NEERDWITELN--------------------------------AQSDQQLRS--------------------------------------DKL-FLYYI------QKGR----------CMY-
T. denticola            E---LEPARTKTRLKILQDLYNNCK--NDADAFSSEIKDLSGKIE-------------------------NEDNLRLRS--------------------------------------DKL-YLYYT------QLGK----------CMY-
P. duerdenii            E---AKNSRKESRKNKLLEFYKFGKKAFINEIGEERYNYLLNEIN-------------------------SEEESKFRW--------------------------------------DNL-YLYYT------QLGR----------CMY-
C. mitsuokai            E----AKERTESKIKKLENVYKDLD-EQTKKEYKSVLEELKGFDNTK-----------------------KISSDSL--------------------------------------------FLYFT------QLGK----------CMY-
S. thermophilus         Q-------YT-------NQGKSNSQ--QRLKRLEKSLKELGSKILKENIPAKLS----------------KIDNNALQN--------------------------------------DRL-YLYYL------QNGK----------DMY-
S. mutans               Q-------FT-------NQGRRNSQ--QRLKGLTDSIKEFGSQILKEH----------------------PVENSQLQN--------------------------------------DRL-FLYYL------QNGR----------DMY-
S. pyogenes             Q-------TT-------QKGQKNSR--ERMKRIEEGIKELGSQILKEH----------------------PVENTQLQN--------------------------------------EKL-YLYYL------QNGR----------DMY-

                        ----------------------HHH--HHHHHHHHHH--------------------------------------------------------------------------------HHH-HHHHH------------------------
domains                 llllllllllllllllll|NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
catalytic residues      llllllllllllllllll|NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
domains                 rrrrrrrrrrrrrrrrrrrr|NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
catalytic residues      rrrrrrrrrrrrrrrrrrrr|NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
.
B. bifidum              Q----PSGRTISRKSRLQDLYRNANLGKEFKGIADELN--------------------------------ACSDKDLQD--------------------------------------DRL-FLYYT------QLGK----------DMY-
O. kitaharae            Q----ESHLTNSRKNQLSTLLKNA--GLSELVTQVSQY----------------------------------DAAALQN--------------------------------------DRL-YLYFL------QQGK----------DMY-
S. sanguinis            MTTEEGQKKAKTRKTKLESAL--KNIEN--SLLENGKV--------------------------------PHSDEQLQS--------------------------------------EKL-YLYYL------QNGK----------DMY-
F. fructosus            Q----TSVRSVPRYNQLKEKYQSKSLSEEAKVLKKTLD--------------------------------G-NKNKMSD--------------------------------------DRY-FLYFQ------QQGK----------DMY-
E. yurii                G----EKVRTKSRKDRLKELY--NGIKEDSKQWVKELD--------------------------------SKDESYFRS--------------------------------------KKM-YLYYL------QKGR----------CMY-
L. farciminis           K----KSEITTSREKRIKRLQS-KLLNKANDFKPQLRE-YLVPNKKIQE-------------------ELKKHKNDLSS--------------------------------------ERI-MLYFL------QNGK----------SLY-
S. pseudintermedius     Q---QKGKKQKSRKQLWDDNIKKNKLKSVDEYKYIIDV------------------------------ANKLNNEQLQQ--------------------------------------EKL-WLYLS------QNGK----------CMY-
L. rhamnosus            G---TAGKRTQSRQKQIQTVYANAAQELIDSAVRGELEDKI-----------------------------ADKASFT----------------------------------------DRL-VLYFM------QGGR----------DIY-
F. magna                E----EKKTTISRKNKLLDLYKAIKKDERDSQYEKLLTGLN-----------------------------KLDDSDLRS--------------------------------------RKL-YLYYT------QMGR----------DMY-
P. antarcticus          G----EKKRTKSRKDQWEELTK-TTLKNDPDLKSFIGE------------------------------IKSQGDQRFNE--------------------------------------QRF-WLYVT------QQGK----------CLY-
consensus/95%           ......................................................          .   ........h..                                     .... hLYh.      Q.Gc          .hY
consensus/85%           ................h.....................................          .   ........h..                                     ...h hLYa.      Q.Gc          .hY
.
Jnet 227501312          -------HHHHHHHHHHHHHHHHHHHHHHH---HHHH--------------------------------------HHHH--------------------------------------HHH-HHHHH------H-----------------
M. mobile               N--------NKHERKKIEGINKYRK--EKYEELKKVYDLP------------------------------NENTTLL----------------------------------------KKL-WLLRQ------QQGY----------DAY-
M. gallisepticum        S--------EEKELENTKNYKKLIK--KNGDKISEGLKALGISEDEIK----------------------DILKSPTKS--------------------------------------YKF-LLWLQ------QDHI----------DPY-
M. synoviae             TKP-NLEKLLNNATNSNIKILKEKL--DQTEKFDDFTKKKFID---------------------------KIENSVVFR--------------------------------------NKL-FLWFE------QDRK----------DPY-
M. ovipneumoniae         NDK-KTIEEIKK---RNKKGKGKTL--EKLFQILNLENKGYKLS-------------------------DLETKPAKLL--------------------------------------DRL-RFYHQ------QDGI----------DLY-
M. canis                TQD-QENDALKG-IAKAQKSKKSLV--EERLEANNIDKSVFN---------------------------DKYEKLI-----------------------------------------YKI-FLWIS------QDFK----------DPY-
consensus/95%           ... .....................  c....h...............                     ..........                                      .ch .hh..      Q...          D.Y
consensus/85%           ... .....................  c....h...............                     ..........                                      .ch .hh..      Q...          D.Y

E. rectale              N--------SEEQKKRINDSQKLNE--KEMEYIEKKLAVTYGIKLSPS----------------------DFSSQKQLS--------------------------------------LKL-KLWNE------QDGI----------CLY-
E. faecalis             N--------EEDEKKRIKELHAKNS--QRKNDSQSYFMQKSGWSQEKF----------------QTTI--QKNRRFL----------------------------------------AKL-LYYYE------QDGI----------CAY-
S. thermophilus         N--------EDDEKKAIQKIQKANK--DEKDAAMLKAANQYNGKAELPHSVF------------------HGHKQLA----------------------------------------TKI-RLWHQ------QGER----------CLY-
S. lugdunensis          N--------SKDKQKFINEMQKKNE--NTRKRINEIIGKYGN----------------------------QNAKRLV----------------------------------------EKI-RLHDE------QEGK----------CLY-
E. dolichum             N--------SEEQRKAIRERQKFFE--MRNKQVADIIGDDR-----------------------------KINAKLR----------------------------------------EKL-VLYQE------QDGK----------TAY-
R. albus                N-------KTFEDRAIDTKRNNDNQ--KENDRIVKEIIECIKCDE-------------------------VHARHLI----------------------------------------EKY-KLWEA------QEGK----------CLY-
R. inulinivorans        S-------KNFQDRTNLEKEMKKRQ--QENERAKQQIIELGKQ---------------------------NPTGQDI----------------------------------------LKY-RLWND------QGGY----------CLY-
L. coryniformis         G-------RNFKERGDIQKRQDKNR--QTNDKIAAELTELGI----------------------------PVNGQNI----------------------------------------IRY-KLHKE------QNGV----------DPY-
I. polytropus           A-------KSYDDRQTIIKENKKRE--LENEKTKKFISEEFGIK--------------------------NVKGKLL----------------------------------------LKY-RLYQE------QEGR----------CAY-
consensus/95%           .       ..........c......  .........h...............            ....  .........                                      .ch .ha..      Q...          ..Y
consensus/85%           .       ...c.c..h.c......  ...c.....h...............            ....  .........                                      .Kh .La..      Q.G.          .hY

Bacteroides sp.20       K-------NPADKRARMSQQMIKNE--NTNLRIKALLTEFLNPEFGIENVRPY-SPSQQDLLRIYEEGVLNSILELPEDIGIILGKFNQTDTLKRPTRSEI----------------LRY-KLWLE------QKYR----------SPY-
I. album                K-------NNSEERKKTSESQEKNF--QEKERARKLLKELLNS----SNFEHY-DENGN---KIFSSFTVNPNPDSPLDIEKFRIWKNQSGLTDEELNKKLKDEKIPTEIEV-----KKY-ILWLT------QKCR----------SPY-
B. fragilis             K-------KNAKEREELTKSIAQTT--KAHEEYKTLLQTEFGLT--------------------------NVSRTDI----------------------------------------LRY-KLYKE---LESCGYK----------TLY-
N. salsuginis           L--------PEKIRKEIDKAMRERE--KALDKIIGKYKKEF-----------------------------PSIDKRL---------------------------------------ARKI-QLWER------QKGL----------DLY-
E. minutum              K-------KSAEDRSKLSREQNDNE--SNRNRIYEIYIRPQQQVIITR----------------R-----ENPRNYI----------------------------------------LKF-ELLEE------QKSQ----------CPF-
S. globus               K-------VGAEKREDIIKQQTKQE--KEAVLAYSKYCEPN-----------------------------NLDKRYI----------------------------------------ERF-RLLED------QAFV----------CPY-
M. trichosporium        G-------KSIEEREEITRGIEKRN--RQKDKLRGLFEKEVGRPPQDG----------------------ARGKEEL----------------------------------------LRF-ELWSE------QMGR----------CLY-
F. branchiophilum       N--------DANKRKGIQDYQNDNK--KFREDAIKEIKKLYFEDCKKEVEP---------------------TEDDI----------------------------------------LRY-QLWME------QNRS----------EIY-
Prevotella C561         N--------DANRKWALDTYNRIRH--DENEKIKKILEEFYPKRDGISTDDIDKARYVIDQREVDYFTGSKTYNKDI----------------------------------------KKY-KFWLE------QGGQ----------CMY-
P. timonensis           N--------DANKRRAIADRQKEQD--KQHKKYGDEIRKLYKEETGKDIEP---------------------TQTDV----------------------------------------LKF-QLWEE------QNHH----------CLY-
P. timonensis           N--------DANKRRAIADRQKEQD--KQHKKYGDEIRKLYKEETGKD---------------------IEPTQTDV----------------------------------------LKF-QLWEE------QNHH----------CLY-
A. paucivorans          S-------QPAKVRRRIETEQQANE--KKKQQAEREFLDIVGT---------------------------APGPGDL----------------------------------------LKM-RLWRE------QGGF----------CPY-
P. marinum              P-------MGADGKRELERFQKEGR--AKNERARDELKKLGH----------------------------IDSRESR----------------------------------------QKF-QLWEQLAKEP-VDRC----------CPF-
A. denitrificans        K-------QSREQKVEAQRRQADNQ--RRNARIRRSIAEVLGIGEE------------------------RVRGSDI----------------------------------------QKW-ICWEELSFDA-ADRR----------CPY-
R. syzygii              K-------AGLEEKERIKKQQTANL--ERNERIRQKLQDAGV----------------------------PDNRENR----------------------------------------LRM-RLFEELGQ---GNGLGTP-------CIY-
D. shibae               K-------KSEEQKRADIKRIRDTT--EAAKKRSEKLEELEI----------------------------EDNGRNR----------------------------------------MLL-RLWEDLNPDDAMRRF----------CPY-
R. rubrum               K-------AGADERKRMIAEQAERE--RENAEIDVELRKSDRWM--------------------------ANARERR----------------------------------------QRV-RLARR------QNNL----------CPY-
Azospirillum B510       K-------RSAEDRRREDKRQADNQ--KRNEERKRLILSLGE----------------------------RPTPRNL----------------------------------------LKL-RLWEEQGPV--ENRR----------CPY-
C. salinarum            K-----HGVIRRNEIAKENRENEGR--RKKESAALDTFCQDNTVSWQAGG--------------------LPKERAA----------------------------------------LRL-RLAQRQEFF----------------CPY-
N. hamburgensis         K-------LSEQQKAERQREQRRNQ--DKNKARAEELAKFGR----------------------------PANPRNL----------------------------------------LKM-RLWEELAHDP-LDRK----------CVY-
Bradyrhizobium BTAi1    K-------LSPRRLAELEREQAENQ--RKNDKRTSLLRKLGL----------------------------PASTHNL----------------------------------------LKL-RLWDE------QGDVASE-------CPY-
W. succinogenes         N--------TKGEIEDIKESQRKNE--KERKEAADWIAETSFQV--------------------------PLTRKNI----------------------------------------LKK-RLYIQ------QDGR----------CAY-
C. jejuni               G-------KNHSQRAKIEKEQNENY--KAKKDAELECEKLGL----------------------------KINSKNI----------------------------------------LKL-RLFKE------QKEF----------CAY-
Jnet 218563121          ----------HHHHHHHHHHHHHHH--HHHHHHHHHHHHH----------------------------------HHH----------------------------------------HHH-HHHHH------------------------
H. mustelae             C-------KAKSARMQLEKINKKNK--SENDAASQLLEVLGL----------------------------PNTYNNR----------------------------------------LKC-KLWKQ------QEEY----------CLY-
A. ebreus               S-------RPLDERNKVKRAQEEFR--DRNDRARSEFERDFGY---------------------------KPKAAAF----------------------------------------EKW-MLYRE------QLGQ----------CAY-
B. smithii              S-------HSFDERKKIQKDQTENR--KKNETAIKQLIEYELTK--------------------------NPTGLDI----------------------------------------VKF-KLWSE------QQGR----------CMY-
C. cellulolyticum       G-------MTRQDRDNLKKEHENNR--KAREKISDLIRQNGR----------------------------VASGLDI----------------------------------------LKW-RLWED------QGGR----------CAY-
HF0070_07E19            G-------KGRELRERIKLDQL------DKSKQNDQRAEDFRAEFQQ-----------------------APRGDQS----------------------------------------LRY-RLWKE------QNCT----------CPY-
P. lavamentivorans      G-------KSKREREEIQSGIRRNE--KQRKKATEDLIKNGIA---------------------------NPSRDDV----------------------------------------EKW-ILWKE------GQER----------CPY-
N. meningitidis         G-------KSFKDRKEIEKRQEENR--KDREKAAAKFREYFPNFVG------------------------EPKSKDI----------------------------------------LKL-RLYEQ------QHGK----------CLY-
P. multocida            G-------KSFKERREIQKQQEDNR--TKRESAVQKFKELFSDFSS------------------------EPKSKDI----------------------------------------LKF-RLYEQ------QHGK----------CLY-
A. muciniphila          T------TFSAMDSKKIQRELTLRQ--KSHTDAVNRLKRKLPGK--------------------------ALSANLI----------------------------------------RKC-RIAMD------MNWT----------CPF-
Actinomyces F0310       L-------KSA-SLLEEERERWERF-EARREIRQKEMYKRLGISG-------------------------PFRRSDQ----------------------------------------VRY-EILDL------QDCA----------CLY-
A. cellulolyticus       S-------ESRERQAEEEAARRAHR--KANDRIRAELRASGLS---------------------------DPSPADL----------------------------------------VRA-RLLEL------YDCH----------CMY-
B. longum               F-------SSVAFARKDKREYEKNN--EKRSIFRSSLSEQLRADEQME----------------------KVRESDL----------------------------------------RRL-EAIQR------QNGQ----------CLY-
A. naeslundii           F---TSERMADERDKANRRRYNDNQ--EAMKKIQRDYGKEG-----------------------------YISRGDI----------------------------------------VRL-DALEL------QGCA----------CLY-

                        -----------HHHHHHHHHHHHHH--HHHHHHHHH-------------------------------------HHHH----------------------------------------HHH-HHHHH------------------------
domains                 rrrr|NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
catalytic residues      rrrr|NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

A. hesperidum           S-------KNLDERRKIEKAQKENQ--KN----NEQIKDEIRE-ILGS---------------------AHVTGRDI----------------------------------------VKY-KLFKQ------QQEF----------CMY-
Alcanivorax W11-5       K-------GGVTRRNDIEKQNKRFA--SE----RKKAAQSIEE--NGK----------------------TPTPARI----------------------------------------LRY-QLWIE------QGHQ----------CPY-
B. zoohelcum            K-------NSAKERATIHKRNKDNQ--TI----NNKIKETLVK--NKK----------------------ELSLANI----------------------------------------EKV-KLWEA------QRHL----------SPY-
Porphyromonas F0450     K-------MSREERERMARNNKDRE--ER----NKGVAAKIRE--CGL----------------------YPTKPRI----------------------------------------QKY-MLWKE------AGRQ----------CLY-
B. intestinihominis     K-------QSKEERSNTYKSINKNQ--RE----NEQIAKRIVE--YGV-----------------------PTRSRI----------------------------------------QKY-KMWEE------SKHC----------CIY-
O. laneus               R-------SSKTEREKQSRQIRENE--KK----NEAAKVKLAE--YGL----------------------KAYRDNI----------------------------------------QKY-LLYKEIE----EKGGTVC-------CPY-
Treponema JC4           K-------NNVEKKAEIARKQNQRA--KENIAINDTISALYHTAFPGK--------------------SFYPNRNDR----------------------------------------MKY-RLWSELG----LGNK----------CIY-
Rhodovulum PH10         K-------LNREQKERLDRENRKNR--EE----NERRTAILAE--HGQ----------------------RDTAENK----------------------------------------IRL-RLFEE------QARANAGIAL----CPY-
consensus/95%           .     ................... ......................................................................................    ..c. .hh..................    ..a
consensus/85%           .     .......c........... ..........h...........................................................................    ..ch .Lh..................    C.Y

#map                    *-----------------------------------------------------------------------------------------------------------------------------------****----------***-
```

```
Jnet 331001027          --------------------------------------------EEE---------EEEEE----EEEEEE-----------EEEE------------------HHHH-------HHHHHHHHHHHH-------HHH-H-----------
P. excrementihominis    ----------------------TGRPLDDV----------GEIDHIIPRSL-TLKKSESIYNSEVNLIFVSAQGNQ-EKKNNIYLLS-------------NLAKNYLAAVFGTSDLSQITNEIESTV-------LQL-KAAGRL------
W. succinogenes         ----------------------CGDTIGDD----------GEIDHILPRSH-TLKIYGTVFNPEGNLIYVHQKCNQ-AKADSIYKLS-------------DIKA------------GVSAQWIEEQV-------ANI-KGY---------
L. pneumophila          ----------------------KGASIGGQ----------GEIDHIYPRSL-SKKHFGVIFNSEVNLIYCSSQGNR-EKKEEHYLLE-------------HLSPLYLKHQFGTDNVSDIKNFISQNV-------ANI-KKY---------
S. wadsworthensis       ----------------------TGDRLAEG----------GEIDHILPRSL-IKDARGIVFNAEPNLIYASSRGNQ-LKKNQRYSLS-------------DLKANYRNEIFKTSNIAAITAEIEDVV-------TKL-QQTHRL------
HTCC5015                ----------------------TGAPLGGS----------GEIDHIIPRSL-TGRTKKTVFNSEANLIYCSSKGNH-DKGNRVYVIE-------------QLNDKYLKKQFSTSDVNLIKKKIKTTI-------QRFTEGGEKL------
F. novicida             ----------------------SGANLTDGDFDGAK----EELDHIIPRS----HKKYGTLNDEANLICVTRGDNK-NKGNRIFCLR-------------DLADNYKLKQFETTDDLEIEKKIADTIWDANKKDFKF-GNY---------
consensus/95%                                 .G..h.........    .EhDHIhPRS. ..c.....hN.E.NLI......N. .K....a.h.             .h....................I...h.........h.......
consensus/85%                                 .G..h.........    .EhDHIhPRS. ..c.....hN.E.NLI......N. .K....a.h.             .h....................I...h.........h.......

A. coleocanis           ----------------------CGTAITFVN---------SEMDHIVPRAGLGST------NTRDNLVATCERCNK-SKSN-KPFAV-------WAAECGIP-GVSVAEALK-------RVDF----------------WIADG-FASSK
C. glomerans            ----------------------SGRAIDIHQLSNAGI---YEVDHIIPRTY-VKD------DSLENKALVYREENQ-RKTDMLLIDP--------------EIRRRMSGYWR-------MLHE----------------AKLIG------
Acidaminococcus D21     ----------------------TGDPIKLEHIKDQSF---YNIDHIYPQSM-VKD------DSLDNKVLVQSEING-EKSSRYPLDA--------------AIRNKMKPLWD-------AYYN----------------HGLIS------
V. atypica              ----------------------TGNIIDLNQLNTDN----YDIDHIYPRSL-TKD------DSFDNLVLCERTANA-KKSDIYPIDN--------------RIQTKQKPFWA-------FLKH----------------QGLIS------
F. nucleatum            ----------------------TGREIDLDRLLQNNDT--YDIDHIYPRSKVIKD------DSFDNLVLVLKNENA-EKSNEYPVKK--------------EIQEKMKSFWR-------FLKE----------------KNFIS------
F. alocis               ----------------------SGDDIDINELIRGNSK--WDRDHIYPQSK-IKD------DSIDNLVLVNKTYNA-KKSNELLSED---------------IQKKMHSFWL-------SLLN----------------KKLIT------
S. moorei               ----------------------TGKRIDFDKLINDKNT--YDIDHIYPRSK-IKD------DSITNRVLVEKDING-EKTDIYPISE--------------DIRQKMQPFWK-------ILKE----------------KGLIN------
C. catus                ----------------------SGETIQLDELWDNTK---YDIDHIYPQSK-TMD------DSLNNRVLVKKNYNA-IKSDTYPLSL--------------DIQKKMMSFWK-------MLQQ----------------QGFIT------
T. denticola            ----------------------CGKPIEIGHVFDTSN---YDIDHIYPQSK-IKD------DSISNRVLVCSSCNK-NKEDKYPLKS--------------EIQSKQRGFWN-------FLQR----------------NNFIS------
P. duerdenii            ----------------------SLEPIDLADLKSNNI---YDQDHIYPKSK-IYD------DSLENRVLVKKNLNH-EKGNQYPIPE-------------KVLNKNAYGFWK-------ILFD----------------KGLIG------
C. mitsuokai            ----------------------SGKKLDIDSLDK------YQIDHIVPQSL-VKD------DSFDNRVLVVPSENQ-RKLDDLVVPF--------------DIRDKMYRFWK-------LLFD----------------HELIS------
S. thermophilus         ----------------------TGDDLDIDRLSN------YDIDHIIPQAF-LKD------NSIDNKVLVSSASNR-GKSDDVPSLE---------------VVKKRKTFWY-------QLLK----------------SKLIS------
S. mutans               ----------------------TGEELDIDYLSQ------YDIDHIIPQAF-IKD------NSIDNRVLTSSKENR-GKSDDVPSKD---------------VVRKMKSYWS-------KLLS----------------AKLIT------
S. pyogenes             ----------------------VDQELDINRLSD------YDVDHIVPQSF-LKD------DSIDNKVLTRSDKNR-GKSDNVPSEE---------------VVKKMKNYWR-------QLLN----------------AKLIT------

                        ----------------------------HHHHHH------H-EEE---------------------EEE--------------------------------------HHHHH-------HHHH---------------------------
domains                 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
catalytic residues      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN*DH*NNNNNNNNNNNNNNNNNN*N*NNNNNNNNNNNNN*N*NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
domains                 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
catalytic residues      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN*H*NNNNNNNNNNNNNNNNNN*N*NNNNNNNNNNNNN*N*NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
.
B. bifidum              ----------------------TGEELDLDRLSSA-----YDIDHIIPQAV-TQN------DSIDNRVLVARAENA-RKTDSFTYMP--------------QIADRMRNFWQ-------ILLD----------------NGLIS------
O. kitaharae            ----------------------SGEKLNLDNLSN------YDIDHIIPQAY-TKD------NSLDNRVLVSNITNR-RKSDSSNYLP--------------ALIDKMRPFWS-------VLSK----------------QGLLS------
S. sanguinis            ----------------TLDKTGSPAPLYLDQLDQ------YEVDHIIPYSF-LPI------DSIDNKVLTHRENNQ-QKLNNIP-DK--------------ETVANMKPFWE-------KLYN----------------AKLIS------
F. fructosus            ----------------------TGRPINFERLSQD-----YDIDHIIPQAF-TKD------DSLDNRVLVSRPENA-RKSDSFAYTD--------------EVQKQDGSLWT-------SLLK----------------SGFIN------
E. yurii                ----------------------SGEVIELDKLMDDN---LYDIDHIYPRSF-VKD------DSLDNLVLVKKEINN-RKQND-PITP--------------QIQASCQGFWK-------ILHD----------------QGFMS------
L. farciminis           ----------------------SEESLNINKLSD------YQVDHILPRTY-IPD------DSLENKALVLAKENQ-RKADDLLLNS--------------NVIDRNLERWT-------YMLN----------------NNMIG------
S. pseudintermedius     ----------------------SGQSIDLDALLSPNATKHYEVDHIFPRSF-IKD------DSIDNKVLVIKKMNQ-TKGDQVPLQFI-------------QQPYERIAYWK-------SLNK----------------AGLIS------
L. rhamnosus            ----------------------TGAPLNIDQLSH------YDIDHILPQSL-IKD------DSLDNRVLVNATINR-EKNNVFASTL---------------FAGKMKATWR-------KWHE----------------AGLIS------
F. magna                ----------------------TGEKIDLDKLFDSTH---YDKDHIIPQSMKKDD------SIINNLVLVNKNANQTTKGNIYPVPS------------SIRNNPKIYNYWK-------YLME----------------KEFIS------
P. antarcticus          ----------------------TGKALDIQNLSM------YEVDHILPQNF-VKD------DSLDNLALVMPEANQ-RKNQVGQNKM-------PLEIIEANQQYAMRTLWE-------RLHE----------------LKLIS------
consensus/95%                           ..........h.h..h........a..DHIhP.......      .oh.N.hL.....N...K.........      ................W.       .h..                ..hh. .....
consensus/85%                           ......o...h.h..L........Y-hDHIhP.u....D      .Sh.N.VL.....N...K.........      ...............hW.       .h..                ..hI. .....
.
Jnet 227501312          -----------------------------------------EEEEEE-------------------HHHHHHHHH-HH-------HH------------H-----HHHHHHH-------HHHH---------------------------
M. mobile               ----------------------SLRKIEANDVINKPWN--YDIDHIVPRSI-SFD------DSFSNLVIVNKLDNA-KKSNDLSAKQ------FIEKIYGIEKLKEAKENWG-------NWYL----------------RNANG------
M. gallisepticum        ----------------------SLKEIAFDDIFTKTEK--FEIDHIIPYSI-SFD------DSSSNKLLVLAESNQ-AKSNQTPYEF--------------ISSGNAGIKWE-------DYEA----------------YCRKFKDGDSS
M. synoviae             ----------------------TQLDIKINEIEDE-----TEIDHVIPYSK-SAD------DSWFNKLLVKKSTNQ-LKKNKTVWEY-------------YQNESDPEAKWN-------KFVA----------------WAKRIYLVQKS
M. ovipneumoniae         ----------------------TLDKINIDQLINGSQK--YEIEHIIPYSM-SYD------NSQANKILTEKAENL-KKGKLIASEY-------------IKRNGDE--FYN-------KYYE----------------KAKELFI----
M. canis                ----------------------TGAQISVNEIVN--NK--VEIDHIIPYSL-CFD------DSSANKVLVHKQSNQ-EKSNSLPYEY-------------IK-QGHSGWNWD-------EFTK----------------YVKRVFV----
consensus/95%                                 o...I.h..h......  .-I-HhhP.S. .hD      .S..N.hh.....N. .K......c.      .................a.       .a..                ...........
consensus/85%                                 o...I.h..h......  .-I-HhhP.S. .hD      .S..N.hh.....N. .K......c.      .................a.       .a..                ...........

E. rectale              ----------------------SGKTIDPNDIINNPQL--FEIDHIIPRSI-SFD------DARSNKVLVYRSENQ-KKGNQTPYYY---------------LTHSHSEWS-----FEQYKA-----------------TVMNL----SK
E. faecalis             ----------------------TGLPISPELLVSDS----TEIDHIIPISI-SLD------DSINNKVLVLSKANQ-VKGQQTPYDA--------WMDGSFKKINGKFSNWD-------DYQK----------------WVESR------
S. thermophilus         ----------------------TGKTISIHDLINNSNQ--FEVDHILPLSI-TFD------DSLANKVLVYATANQ-EKGQRTPYQA---------------LDSMDDAWS-----FRELKA-----------------FVRES------
S. lugdunensis          ----------------------SLESIPLEDLLNNPNH--YEVDHIIPRSV-SFD------NSYHNKVLVKQSENS-KKSNLTPYQY--------------FNSGKSKLSYN-------QFKQ----------------HILNL-----S
E. dolichum             ----------------------SLEPIDLKLLIDDPNA--YEVDHIIPISI-SLD------DSITNKVLVTHRENQ-EKGNLTPISA-------------FVKGRFTKGSLA-------QYKA----------------YCLKL-KEKNI
R. albus                ----------------------SGETITKEDMLRDKDKL-FEVDHIVPYSL-ILD------NTINNKALVYAEENQ-KKGQRTPLMY---------------MNEAQAADYR-------VRVN----------------TMFKS------
R. inulinivorans        ----------------------SGKKIPLEELFDGG----YDIDHILPYSI-TFD------DSYRNKVLVTAQENR-QKGNRTPYEY----------------FGADEKRWE-------DYEA----------------SVRLL------
L. coryniformis         ----------------------TGDQIPFERAFSEG----YEVDHIIPYSI-SWD------DSYTNKVLTSAKCNR-EKGNRIPMVY-------------LANNEQRLNALT-------NIAD----------------NIIRN------
I. polytropus           ----------------------SRKELSLSEVILDESM--TDIDHIIPYSR-SMD------DSYSNKVLVLSGENR-KKSNLLPKEY----------------FDRQGRDWD-------TFVL----------------NVKAM------
consensus/95%                                 o...h....hh...... .-hDHIhP.S. .hD      ....NKhL.....N. .Ku...P..h         ..............h.     ......                ..... .....
consensus/85%                                 o.c.I....hh...... .-hDHIhP.Sh ohD      .oh.NKVLV....N. .Ku...Ph.h         ..............h.     ...h..                .h... .....

Bacteroides sp.20       ----------------------TGEMIPLSKLFTPA----YEIEHIIPQSR-YFD------DSLSNKVICESEINK-LKDRSLGYEF-----------IKNHHGEKVELAFD-------KPVE----------------VLSVEAYEKLV
I. album                ----------------------TGKIIPLSKLFDSNV---YEIEHIIPRSK-MKN------DSTNNLVICELGVNK-AKGDRLAANF-----------ISESNG-KCKFG-E-------VEYT----------------LLKYGDYLQYC
B. fragilis             ----------------------SNTYISREKLFSKE----FDIEHIIPQAR-LFD------DSFSNKTLEARSVNI-EKGNKTAYDF----------VKEKFGESGADNSLE-------HYLN----------------NIEDL------
N. salsuginis           ----------------------SGKVINLSQLLDGS----ADIEHIVPQSL-GGL------STDYNTIVTLKSVNA-AKGNRLPGDW-------------LAGNPDYRERIG-------MLSE----------------KGLID------
E. minutum              ----------------------CGGQISPNDIINNQ----ADIEHLFPIAE-SED------NGRNNLVISHSACNA-DKAKRSPWAA--------------FASAAKDSKYD---------YN----------------RILSNVKENIP
S. globus               ----------------------CLEHISVADIAAGR----ADVDHIFPRDD-TAD------NSYGNKVVAHRQCND-IKGKRTPYAA-----------------FSNTSAWG-------PIMH----------------YLDETPGMWRK
M. trichosporium        ----------------------TDDYISPSQLVATDDA--VQVDHILPWSR-FAD------DSYANKTLCMAKANQ-DKKGRTPYEW--------------FKAEKTDTEWD-------AFIV----------------RVEAL------
F. branchiophilum       ---------------------EEGKNISICDIIGSNPA--YDIEHTIPRSR-SQD------NSQMNKTLCSQRFNREVKKQSMPIELNNHLE-----------ILPRIAHWKEEADNLTREIEIISRSIKAAATKEI--KDKKIRRRHYL
Prevotella C561         ----------------------TGRTINLSNLFDPN-A--FDIEHTIPESL-SFD------SSDMNLTLCDAHYNRFIKKNHIPTDMPNYDKAITIDGKEYPAITSQLQRWVERVERLNRNVEYWKGQARRAQNKDR--KDQCMREMHLW
P. timonensis           ----------------------TGEQIGITDFIGSNPK--FDIEHTIPQSV-GGD------STQMNLTLCDNRFNREVKKAKLPTELANHEE-----------ILTRIEPWKNKYEQLVKERD--KQRTFAGMDKAV--KDIRIQKRHKL
P. timonensis           ----------------------TGEQIGITDFIGSNPK--FDIEHTIPQSV-GGD------STQMNLTLCDNRFNREVKKAKLPTEL-----------ANHEEILTRIEPWKNKYEQLVKERDKQRTFAGMDKAV----KDIRIQKRHKL
A. paucivorans          ----------------------CEEYLNPTRLAEPGY---AEMDHILPYSR-SLD------NGWHNRVLVHGKDNR-DKGNRTPFEA----------------FGGDTARWD-------RLVA----------------WVQAS------
P. marinum              ----------------------TGKMMSISDLFSDK----VEIEHLLPFSL-TLD------DSMANKTVCFRQANR-DKGNRAPFDA--------------FGNSPAGYDWQ-------EILG----------------RSQNL------
A. denitrificans        ----------------------SGVQISAAMLLSDE----VEVEHILPFSK-TLD------DSLNNRTVAMRQANR-IKRNRTPWDA-------------RAEFEAQGWSYE-------DILQ----------------RAERM----PL
R. syzygii              ----------------------SGRQISLQRLFSND----VQVDHILPFSK-TLD------DSFANKVLAQHDANR-YKGNRGPFEA--------------FGANRDGYAWD-------DIRA----------------RAAVL------
D. shibae               ----------------------TGTRISAAMIFDGS----CDVDHILPYSR-TLD------DSFPNRTLCLREANR-QKRNQTPWQA-----------------WGDTPHWH-------AIAA----------------NLKNL------
R. rubrum               ----------------------TSTPIGHADLLGDA----YDIDHVIPLAR-GGR------DSLDNMVLCQSDANK-TKGDKTPWEA-----------------FHDKPGWI------AQRDD----------------FLARLDPQTAK
Azospirillum B510       ----------------------SGETISMRMLLSEQ----VDIDHILPFSV-SLD------DSAANKVVCLREANR-IKRNRSPWEA----------------FGHDSERWA-------GILA----------------RAEAL------
C. salinarum            -------------------CAERPKLRATDLFSPAE----TEIDHVIERRM-G-G------DGPDNLVLAHKDCNN-AKGKKTPHEH----------AGDLLDSPALAALWQ-------GWRKENADRLKGKGHKARTPREDKD------
N. hamburgensis         ----------------------TGEQISIERLLSDE----VDIDHILPVAM-TLD------DSPANKIICMRYANR-HKRKQTPSEA----------FGSSPTLQGHRYNWD-------DIAA----------------RATGL------
Bradyrhizobium BTAi1    ----------------------TGEAIGLERLVSDD----VDIDHLIPFSI-SWD------DSAANKVVCMRYANR-EKGNRTPFEA-------------FGHRQGRPYDWA-------DIAE----------------RAARL------
W. succinogenes         ----------------------TGDVIELERLFDEGY---CEIDHILPRSR-SAD------DSFANKVLCLARANQ-QKTDRTPYEW----------------FGHDAARWN-------AFET----------------RTSAP------
C. jejuni               ----------------------SGEKIKISDLQDEKM---LEIDHIYPYSR-SFD------DSYMNKVLVFTKQNQ-EKLNQTPFEA----------------FGNDSAKWQ---------------------------KIEVL------
Jnet 218563121          ----------------------------------------EEEEEEEEE-----------------EEHHHHHHHH-H-----HHHH----------------HH---HHHH---------------------------HHHHH------
H. mustelae             ----------------------SGEKITIDHLKDQRA---LQIDHAFPLSR-SLD------DSQSNKVLCLTSSNQ-EKSNKTPYEW----------------LGSDEKKWD-------MYVG----------------RVYSS------
A. ebreus               ----------------------SQQPLDIQRVLDDHNY--AQVDHALPYSR-SYD------DSKNNKVLVLTHENQ-NKGNRTAFEY-----------LTSFPDGEDGERWR-------TFVA----------------WVQGN------
B. smithii              ----------------------SLKPIELERLLEPGY---VEVDHILPYSR-SLD------DSYANKVLVLTKENR-EKGNHTPVEY----------------LGLGSERWK-------KFEK----------------FVLAN------
C. cellulolyticum       ----------------------SGKPIPVCDLLNDSL---TQIDHIYPYSR-SMD------DSYMNKVLVLTDENQ-NKRSYTPYEV-----------------WGSTEKWE-------DFEA----------------RIYSM------
HF0070_07E19            ----------------------SGRMIPVNSVLSED----TQIDHILPISQ-SFD------NSLSNKVLCFTEENA-QKSNRTPFEY---------------------LDAA------DFQRL----------------EAISG------
P. lavamentivorans      ----------------------TGDQIGFNALFREGR---YEVEHIWPRSR-SFD------NSPRNKTLCRKDVNI-EKGNRMPFEA----------------FGHDEDRWS-------AIQI----------------RLQGM---VSA
N. meningitidis         ----------------------SGKEINLGRLNEKGY---VEIDHALPFSR-TWD------DSFNNKVLVLGSENQ-NKGNQTPYEY--------------FNGKDNSREWQ-------EFKA----------------RVETS------
P. multocida            ----------------------SGKEINIHRLNEKGY---VEIDHALPFSR-TWD------DSFNNKVLVLASENQ-NKGNQTPYEW--------------LQGKINSERWK-------NFVA----------------LVLGS------
A. muciniphila          ----------------------TGATYGDHELEN------LELEHIVPHSF-RQS------NALSSLVLTWPGVNR-MKGQRTGYDF-----------VEQEQENPVPDKPNLHICSLNNYRE----------------LVEKLDDKKGH
Actinomyces F0310       ----------------------CGNEINFQT---------FEVDHIIPRVDASSD------SRRTNLAAVCHSCNS-AKGG-LAFGQ-------WVKRGDCPSGVSLENAIK-------RVRS----------------WSKDR-LGLTE
A. cellulolyticus       ----------------------CGAPISWEN---------SELDHIVPRTD-GGS------NRHENLAITCGACNK-EKGR-RPFAS--------------WAETSNRVQLR-------DVID----------------RVQKL----KY
B. longum               ----------------------CGRTITFRT---------CEMDHIVPRKGVGST------NTRTNFAAVCAECNR-MKSN-TPFAI----------WARSEDAQTRGVSLA-------EAKK----------------RVTMFTFNPKS
A. naeslundii           ----------------------CGTTIGYHT---------CQLDHIVPQAGPGSN------NRRGNLVAVCERCNR-SKSN-TPFAV----------WAQKCGIPHVGVK---------EAIG----------------RVRGWRKQTPN

                        -----------------------------------------EEEE------------------HHHEEEE-HHHHH-HH-----HHH----------HHHHH------HH---------HHHH----------------HHH--------
domains                 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
catalytic residues      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN*DH*NNNNNNNNNNNNNNNNNN*N*NNNNNN*N*NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN

A. hesperidum           ----------------------SGEKLDVTRLFEP---GYAEVDHIIPYGI-SFD------DSYDNKVLVKTEQNR-QKGNRTPLEY----------LRDKPE-------QKA------KFIA----------------LVES--IPLSQ
Alcanivorax W11-5       ----------------------CESNISLEQALSG---AYTNFEHILPRTLTQIG------RKRSELVLAHRECND-EKGNRTPYQA----------FGHDDR-------RWR------IVEQ----------------RANA--LP-KK
B. zoohelcum            ----------------------TGQPIPLSDLFDK---EKYDVDHIIPISR-YFD------DSFTNKVISEKSVNQ-EKANRTAMEY----------FEVGSL-------KYS------IFTK----------------EQFI--AHVNE
Porphyromonas F0450     ----------------------CGRSIEEEQCLRE---GGMEVEHIIPKSV-LYD------DSYGNKTCACRRCNK-EKGNRTALEY----------IRAKGR--------EA------EYMK----------------RIND--LLKEK
B. intestinihominis     ----------------------CGQPVDVGDFLRG---FDVEVEHIIPKSL-YFD------DSFANKVCSCRSCNK-EKNNRTAYDY----------MKSKGE------KALS------DYVE----------------RVNT--MYTNN
O. laneus               ----------------------TGKTLNISHTLGSD--NSVQIEHIIPYSI-SLD------DSLANKTLCDATFNR-EKGELTPYDF----------YQKDPSPEKWGASSWE------EIED----------------RAFR--LLPYA
Treponema JC4           ----------------------CGKGISGAELFTK----EIEIEHILPFSR-TLL------DAESNLTVAHSSCNA-FKAERSPFEA----------FGTNPS-----GYSWQ------EIIQ----------------RANQ--LKNTS
Rhodovulum PH10         ----------------------TGRAIGIAELFTS----EVEIDHILPVSL-TLD------DSLANRVLCRREANR-EKRRQTPFQA----------FGATPA--------WN------DIVA----------------RAAK--LPPNK
consensus/95%                              .......h...............h-H.hP.......      ....N..h.....N...K.......h...............................................................
consensus/85%                              .......h....h..........h-HhhP.u.....      .u..N..h.....N...K.....h.h...............................................................

#map                    ----------------------*****************-***********************************************---------------------------------------------------------------
```

```
Jnet 331001027          ----------------------------------EEEE---HHHHHHHHHHHH-----HHHHHHHHHHH-----------------------EEE----HHHHHHHHH--HHHHHHHHHHHHH----EEEEEE-----------------
P. excrementihominis    ----------------------------------GYFDLLSEKERACARHALFLNSDSEARRAVIDVLGSR--------------------RKASVNGTQAWFVRSIF--SKVRQALAAWTQETGNELIFDAISVPA-------------
W. succinogenes         ----------------------------------KTFSVLSAEQQKAFRYALFLQNDNEAYKKVVDWLRTD--------------------QSARVNGTQKYLAKKIQ--EKLTKML------PNKHLSFEFILADA-------------
L. pneumophila          ----------------------------------ISFHLLTPEQQKAARHALFLDYDDEAFKTITKFLMSQ--------------------QKARVNGTQKFLGKQIM--EFLSTLA----DSKQLQLEFSIKQITA-------------
S. wadsworthensis       ----------------------------------KFFDLLNEHEQDCVRHALFLDDGSEARDAVLELLATQ--------------------RRTRVNGTQIWMIKNLA--NKIREELQNWCKTTNNRLHFQAAATNV-------------
HTCC5015                ----------------------------------RSFSELSREDQKAFRHALFV---PELKSEVTSLLAVK--------------------NITRVNGTQAWLAKKIA--SLLAEHL----DKQGRDYTLSAHQIDP-------------
F. novicida             ----------------------------------RSFINLTPQEQKAFRHALFLADENPIKQAVIRAINNR--------------------NRTFVNGTQRYFAEVLA--NNIYLRA-KKENLNTDKISFDYFGIPTIGNG---------
consensus/95%                                             ..F..L.......hRaALFh.....h...h..hh...                    ....VNGTQ.ah.c.h.  ..h...h..........h.h.h.........
consensus/85%                                             ..F..L.......hRaALFh.....h...h..hh...                    ....VNGTQ.ah.c.h.  ..h...h..........h.h.h.........

A. coleocanis           EHREL-QKGVKDRLKRKV-----------------------SDPEIDNRSMESVAWMARELAHRVQYYFDEKHTGT---------------KVRVFRGSLTSAARKAS--------------------GFE-------------------
C. glomerans            -------DKKFRNLLRSR-------------------IDDKALKGFIARQLVETGQMVKLVRSLLEARYPET-------------------NIISVKASISHDLRTAA--------------------ELV-------------------
Acidaminococcus D21     -------LKKYQRLTRST------------------PFTDDEKWDFINRQLVETRQSTKALAILLKRKFPDT-------------------EIVYSKAGLSSDFRHEF--------------------GLV-------------------
V. atypica              -------ERKYERLTRIA------------------PLTADDLSGFIARQLVETNQSVKATTTLLRRLYPDI-------------------DVVFVKAENVSDFRHNN--------------------NFI-------------------
F. nucleatum            -------DEKYKRLTGKD---------------------DFELRGFMARQLVNVRQTTKEVGKILQQIEPEI-------------------KIVYSKAEIASSFREMF--------------------DFI-------------------
F. alocis               -------KSKYDRLTRKG------------------DFTDEELSGFIARQLVETRQSTKAIADIFKQIYSS--------------------EVVYVKSSLVSDFRKKPL-------------------NYL-------------------
S. moorei               -------EEKYKRLTRNY------------------ELTDEELSSFVARQLVETQQSTKALATLLKKEYPSA-------------------KIVYSKAGNVSEFRNRKD------------------KELP-------------------
C. catus                -------KEKYVRLVRSD------------------ELSADELAGFIERQIVETRQSTKAVATILKEALPDT-------------------EIVYVKAGNVSNFRQTY--------------------ELL-------------------
T. denticola            -------LEKLNRLTRAT------------------PISDDETAKFIARQLVETRQATKVAAKVLEKMFPET-------------------KIVYSKAETVSMFRNKF--------------------DIV-------------------
P. duerdenii            -------QKKYTRLTRRT------------------PFEERELAEFIERQIVETRQATKETANLLKNICQDS-------------------EIVYSKAENASRFRQEF--------------------DII-------------------
C. mitsuokai            -------PKKFYSLIKTE-------------------YTERDEERFINRQLVETRQITKNVTQIIEDHYSTT-------------------KVAAIRANLSHEFRVKN--------------------HIY-------------------
S. thermophilus         -------QRKFDNLTKAE----------------RGGLSPEDKAGFIQRQLVETRQITKHVARLLDEKFNNKKDENNRAVRTV--------KIITLKSTLVSQFRKDF--------------------ELY-------------------
S. mutans               -------QRKFDNLTKAE----------------RGGLTDDDKAGFIKRQLVETRQITKHVARILDERFNTETDENNKKIRQV--------KIVTLKSNLVSNFRKEF--------------------ELY-------------------
S. pyogenes             -------QRKFDNLTKAE----------------RGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREV--------KVITLKSKLVSDFRKDF--------------------QFY-------------------

                        -------HHHHHHHH------------------------HHHHHHHHHHHH----HHHHHHHHHHHHHH----------------------EEEE--HHHHHHHHHH-------------------------------------------
domains                 NNNNNNNNNNNNNNNN|llllllllllllllllll|rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
catalytic residues      NNNNNNNNNNNNNNNN|llllllllllllllllll|rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
domains                 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN|rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
catalytic residues      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN|rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
.
B. bifidum              -------RVKFERLTRQN------------------EFSEREKERFVQRSLVETRQIMKNVATLMRQRYGNSA------------------AVIGLNAELTKEMHRYL--------------------GFS-------------------
O. kitaharae            -------KHKFANLTRTR------------------DFDDMEKERFIARSLVETRQIIKNVASLIDSHFGGET------------------KAVAIRSSLTADMRRYV--------------------DIP-------------------
S. sanguinis            -------QTKYQRLTTSERTPDGV-------------LTESMKAGFIERQLVETRQIIKHVARILDNRF--SDT-----------------KIITLKSQLITNFRNTF--------------------HIA-------------------
F. fructosus            -------RKKYERLTKAGK------------------YLDGQKTGFIARQLVETRQIIKNVASLIEGEY--ENS-----------------KAVAIRSEITADMRLLV--------------------GIK-------------------
E. yurii                -------NEKYSRLTRKTQE-----------------FSDEEKLSFINRQIVETGQATKCMAQILQKSM-GEDV-----------------DVVFSKARLVSEFRHKF--------------------ELF-------------------
L. farciminis           -------LKKFKNLTRRV-------------------ITDKDKLGFIHRQLVQTSQMVKGVANILDNMYKNQGT-----------------TCIQARANLSTAFRKALSGQDDTYHFKHP--------ELV-------------------
S. pseudintermedius     -------DSKLHKLMKPE-------------------FTAMDKEGFIQRQLVETRQISVHVRDFLKEEY--PNT-----------------KVIPMKAKMVSEFRKKF--------------------DIP-------------------
L. rhamnosus            -------GRKLRNLMLRP------------------DEIDKFAKGFVARQLVETRQIIKLTEQIAAAQYPNT-------------------KIIAVKAGLSHQLREEL--------------------DFP-------------------
F. magna                -------KEKYNRLIRNT------------------PLTNEELGGFINRQLVETRQSTKAIKELFEKFYQKS-------------------KIIPVKASLASDLRKDM--------------------NTL-------------------
P. antarcticus          -------SGKLGRLKKPS-------------------FDEVDKDKFIARQLVETRQIIKHVRDLLDERF--SKS-----------------DIHLVKAGIVSKFRRFS--------------------EIP-------------------
consensus/95%           ..... ...Kh..L..........          ...........Fh.R.hV...Q..c....hh..................        .hh..cu.....hR...............       ..h.
consensus/85%           ..... ...Kh..L.c........          ...h.......Fh.RQhVET.Q..K.h..hh...h..............        chh..cu.....hR...............       ..h.
.
Jnet 227501312          -------HHHHHHH-------------------------HHHHHHHHHH----HHHHHHHHHHHHHHH-----------------------EEEEEEEEEHHHHHHH-------------------------------------------
M. mobile               --KAFNDKGKFIKLYTID------------------NLDEFDNSDFINRNLSDTSYITNALVNHLTFSNSKYKY-----------------SVVSVNGKQTSNLRNQIAFVGIKNNKETEREW-KRPEGFK-------------------
M. gallisepticum        LLDSTQRSKKFAKMMKTD-------------------TSSKYDIGFLARNLNDTRYATIVFRDALEDYANNHLVEDKPMF-----------KVVCINGSVTSFLRKNFD-----------------DSSYA-------------------
M. synoviae             DKESKDNSEKNSIFKNKK-------PNLKFKNITKKLFDPYKDLGFLARNLNDTRYATKVFRDQLNNYSKHHSKDDENKLF----------KVVCMNGSITSFLRKSMW-------------------RKN-------------------
M. ovipneumoniae        -NKYKKNKKLDSYVDLDEDS-----AKNRFRFLTLQDYDEF-QVEFLARNLNDTRYSTKLFYHALVEHFENNEFFTYIDENSSKHKV----KISTIKGHVTKYFRAKPVQKNNGPNEN-------------------LNNNK--------
M. canis                -------NNVDSI--LSKKE-----RLKKSENLLTASYDGYDKLGFLARNLNDTRYATILFRDQLNNYAEHHLIDNK-KMF----------KVIAMNGAVTSFIRKNM--------------------SYDN------------------
consensus/95%           .................c..     ....................Fh.RNL.DT.Y.T.hh...L......................    .h..h.G..T..hR.................. ........     .....
consensus/85%           .................c..     ....................Fh.RNL.DT.Y.T.hh...L......................    .h..h.G..T..hR.................. ........     .....

E. rectale              KKEYAISRKKIQNLLYSE-----------------DITKMDVLKGFINRNINDTSYASRLVLNTIQNFFMANEADT---------------KVKVIKGSYTHQMRCNL--------------------KLD-------------------
E. faecalis             ----HFSHKKENNLLETR-----------------NIFDSEQVEKFLARNLNDTRYASRLVLNTLQSFFTNQET-----------------KVRVVNGSFTHTLRKKW--------------------GAD-------------------
S. thermophilus         ---KTLSNKKKEYLLTEE-----------------DISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDT---------------KVSVVRGQFTSQLRRHW--------------------GIE-------------------
S. lugdunensis          KSQDRISKKKKEYLLEER-----------------DINKFEVQKEFINRNLVDTRYATRELTNYLKAYFSANNMNV---------------KVKTINGSFTDYLRKVW--------------------KFK-------------------
E. dolichum             KTNKGYRKKVEQYLLNEN-----------------DIYKYDIQKEFINRNLVDTSYASRVVLNTLTTYFKQNEIPT---------------KVFTVKGSLTNAFRRKI--------------------NLK-------------------
R. albus                ---KKCSKKKYQYLMLPD------------------LNDQELLGGWRSRNLNDTRYICKYLVNYLRKNLRFDRSYESSDEDDLKIRDHY--RVFPVKSRFTSMFRRWW---------------------LN-------------------
R. inulinivorans        ----VRDYKKQQKLLKKN-------------------FTEEERKEFKERNLNDTKYITRVVYNMIRQNLELEPFNHPEKKK----------QVWAVNGAVTSYLRKRW--------------------GLM-------------------
L. coryniformis         -------SRKRQKLLKQK-------------------LSDEELKDWKQRNINDTRFITRVLYNYFRQAIEFNPELEKKQ------------RVLPLNGEVTSKIRSRW--------------------GFL-------------------
I. polytropus           ----KIHPRKKSNLLKEK-------------------FTREDNKDWKSRALNDTRYISRFVANYLENALEYRDDSPKK-------------RVFMIPGQLTAQLRARW--------------------RLN-------------------
consensus/95%           ........c....Lh...                 .....-....a..R.h.DT.ah.c.h.N.h...h....................  .V..h.u.hT..hR..h                    .h.
consensus/85%           ........cK...LL...                 .....-..c.a..RNh.DT.YhoRhhhN.h...h....................  cV..h.G.hT..hR..h                    .h.

Bacteroides sp.20       HESYSHNRSKMKKLLMED-----------------------IPDQFIERQLNDSRYISKVVKSLLSNIVREENEQEAISK-----------NVIPCTGGITDRLKKDW--------------------GIN-------------------
I. album                KDTFKYQKAKYKNLLATE-----------------------PPEDFIERQINDTRYIGRKLAELLTPVVKD-------SK-----------NIIFTIGSITSELKITW--------------------GLN-------------------
B. fragilis             FKSGKISKTKYNKLKMAE---------------------QDIPDGFIERDLRNTQYIAKKALSMLNEISH---------------------RVVATSGSVTDKLREDW--------------------QLI-------------------
N. salsuginis           -------WKKRKNLLAQS-------------------LDEIYTENTHSKGIRATSYLEALVAQVLKRYYPFPDPELRKNGI----------GVRMIPGKVTSKTRSLL--------------------GIK-------------------
E. minutum              HKAWRFNQGAFEKFIENK--------------------------PMAARFKTDNSYISKVAHKYLACLFEKP-------------------NIICVKGSLTAQLRMAW--------------------GLQ-------------------
S. globus               RRKFETNEEEYAKYLQSK--------------------------GFVSRFESDNSYIAKAAKEYLRCLFNPNNVT----------------AVGSLKGMETSILRKAW--------------------NLQ-------------------
M. trichosporium        --ADMKGFKKRNYKLRNA---------------------EEAAAKFRNRNLNDTRWACRLLAEALKQLYPKGEKDKDGKERR---------RVFSRPGALTDRLRRAW--------------------GLQ-------------------
F. branchiophilum       TLKRDYLQGKYDRFIWEE-----------------------PKVGFKNSQIPDTGIITKYAQAYLKSYFKKVE---------------------SVKGGMVAEFRKIW--------------------GIQ-------------------
Prevotella C561         KMELEYWKKKLERFTVTE-----------------------VTDGFKNSQLVDTRVITRHAVLYLKSIFPHVD---------------------VQRGDVTAKFRKIL--------------------GIQ-------------------
P. timonensis           QMEIDYWRGKYERFTMTE-----------------------VPEGFSRRQGTGIGLISRYAGLYLKSLFHQADSRNKSNVY-------------VVKGVATAEFRKMW--------------------GLQ-------------------
P. timonensis           QMEIDYWRGKYERFTMTE-----------------------VPEGFSRRQGTGIGLISRYAGLYLKSLFHQADSRNKS-------------NVYVVKGVATAEFRKMW--------------------GLQ-------------------
A. paucivorans          ----HLSAPKKRNLLRED-------------------FGEEAERELKDRNLTDTRFITKTAATLLRDRLTFHPEAPKD-------------PVMTLNGRLTAFLRKQW--------------------GLH-------------------
P. marinum              ------PYAKRWRFLPDA------------------MKRFEADGGFLERQLNDTRYISRYTTEYISTIIPKN-------------------KIWVVTGRLTSLLRGFW--------------------GLN-------------------
A. denitrificans        RKRYRFAPDGYERWLGDD-------------------------KDFLARALNDTRYLSRVAAEYLRLVCP---------------------GTRVIPGQLTALLRGKF--------------------GLN-------------------
R. syzygii              ------PRNKRNRFAETA-----------------MQDWLHNETDFLARQLTDTAYLSRVARQYLTAICSKD-------------------DVYVSPGRLTAMLRAKW--------------------GLN-------------------
D. shibae               ------PENKRWRFAPDA------------------MTRFEGENGFLDRALKDTQYLARISRSYLDTLFTKGG------------------HVWVVPGRFTEMLRRHW--------------------GLN-------------------
R. rubrum               ALAWRFADDAGERVARKS--------------------AEDEDQGFLPRQLTDTGYIARVALRYLSLVTNEPN------------------AVVATNGRLTGLLRLAW--------------------DITPGPAPRDLLPTPRDALRDD
Azospirillum B510       ------PKNKRWRFAPDA------------------LEKLEGEGGLRARHLNDTRHLSRLAVEYLRCVCP---------------------KVRVSPGRLTALLRRRW--------------------GID-------------------
C. salinarum            -----FMDRVGWRFEEDA-----------------RAKAEENQERRGRRMLHDTARATRLARLYLAAAVMPEDPAEIGAPPVETPPSPEDPTGYTAIYRTISRVQPVN--------------------GSVTHMLRQRLLQRDKNRDYQT
N. hamburgensis         ------PRNKRWRFDANA------------------REEFDKRGGFLARQLNETGWLARLAKQYLGAVTDPN-------------------QIWVVPGRLTSMLRGKW--------------------GLN-------------------
Bradyrhizobium BTAi1    ------PRGKRWRFGPGA------------------RAQFEELGDFQARLLNETSWLARVAKQYLAAVTHPH-------------------RIHVLPGRLTALLRATW--------------------ELN-------------------
W. succinogenes         SNRVRTGKGKIDRLLKKN-------------------FDENSEMAFKDRNLNDTRYMARAIKTYCEQYWVFKNSHTKA-------------PVQVRSGKLTSVLRYQW--------------------GLE-------------------
C. jejuni               --AKNLPTKKQKRILDKN-------------------YKDKEQKNFKDRNLNDTRYIARLVLNYTKDYLDFLPLSDDENTKLNDTQKGSKVHVEAKSGMLTSALRHTW--------------------GFS-------------------
Jnet 218563121          --H----HHHHHHH-------------------------HHHHH-HHH----HHHHHHHHHHHHHHHHH---------------------EEEEEEE---HHHHHHH-------------------------------------------
H. mustelae             ----NFSPSKKRKLTQKN-------------------FKERNEEDFLARNLVDTGYIGRVTKEYIKHSLSFLPLPDGKKE-----------HIRIISGSMTSTMRSFW--------------------GVQ-------------------
A. ebreus               ---KAYRMAKRNRLLRKN-------------------YGVDESKGFIDRNLNDTRYICKFFKNYVEEHLQLAARADGDTAR----------RCVVVNGQLTAFLRARW--------------------GLT-------------------
B. smithii              ---KQFSKKKKQNLLRLR-------------------YEETEEKEFKERNLNDTRYISKFFANFIKEHLKFADGDGGQ-------------KVYTINGKITAHLRSRW--------------------DFN-------------------
C. cellulolyticum       ----HLPQSKEKRLLNRN-------------------FITKDLDSFISRNLNDTRYISRFLKNYIESYLQFSNDSPKS-------------CVVCVNGQCTAQLRSRW--------------------GLN-------------------
HF0070_07E19            ----NWPEAKRNKLLHKS--------------------FGKVAEEWKSRALNDTRYLTSALADHLRHHLPDS-------------------KIQTVNGRITGYLRKQW--------------------GLE-------------------
P. lavamentivorans      KGGTGMSPGKVKRFLAKT-----------------------MPEDFAARQLNDTRYAAKQILAQLKRLWPDMGPEAPV-------------KVEAVTGQVTAQLRKLW--------------------TLN-------------------
N. meningitidis         ----RFPRSKKQRILLQK-------------------FDED---GFKERNLNDTRYVNRFLCQFVADRMRLTGKGKK--------------RVFASNGQITNLLRGFW--------------------GLR-------------------
P. multocida            ----QCSAAKKQRLLTQV-------------------IDDN---KFIDRNLNDTRYIARFLSNYIQENLLLVGKNKK--------------NVFTPNGQITALLRSRW--------------------GLI-------------------
A. muciniphila          EDDRRRKKKRKALLMVRG----------LSHKHQSQNHEAMKEIGMTEGMMTQSSHLMKLACKSIKTSLPDA-------------------HIDMIPGAVTAEVRKAW--------------------DVF-------------------
Actinomyces F0310       KAMGKRKSEVISRLKTEM-----------------------PYEEFDGRSMESVAWMAIELKKRIEGYFNSDRPEGCAAV-----------QVNAYSGRLTACARRAA--------------------HVD-------------------
A. cellulolyticus       SGNMYWTRDEFSRYKKSV----------------VARLKRRTSDPEVIQSIESTGYAAVALRDRLLSYGEKNGVA----------------QVAVFRGGVTAEARRWL--------------------DISIE-----------------
B. longum               YAPREVKAFKQAVIARLQ--------------------QTEDDAAIDNRSIESVAWMADELHRRIDWYFNAKQYVNSASIDDAEAETMKT-TVSVFQGRVTASARRAA--------------------GIE-------------------
A. naeslundii           TSSEDLTRLKKEVIARLR--------------------RTQEDPEIDERSMESVAWMANELHHRI------------------AAAYPET-TVMVYRGSITAAARKAA--------------------GID-------------------

                        --HHHHHHHHHHHHHHH--------------------------------------HHHHHHHHHH------------------HHHHHHH-EEEEEE--HHHHHHHHHH-----------------------------------------
domains                 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN|rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
catalytic residues      NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN|rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr

A. hesperidum           KKK-----NHLLMDKRAI----------------DLEQ-----EGFRERNLSDTRYITRALMNHIQAWLLFDETASTRS-----------KRVVCVNGAVTAYMRARW--------------------GLT-------------------
Alcanivorax W11-5       SSRK----TRLLLLKDFE----------------GEALTDESIDEFADRQLHESSWLAKVTTQWLSSLG---------------------SDVYVSRGSLTAELRRHW--------------------GLD-------------------
B. zoohelcum            Y-------FSGVKRKNLL----------------ATSIP----EDPVQRQIKDTQYIAIRVKEELNKIVGN-------------------ENVKTTTGSITDYLRNHW--------------------GLT-------------------
Porphyromonas F0450     KIS-----YSKHQRLRWL----------------KEDIP----SDFLERQLRLTQYISRQAMAILQQGI---------------------RRVSASEGGVTARLRSLW--------------------GYG-------------------
B. intestinihominis     QISK----TKWQNLLTPV----------------DKIS-----IDFIDRQLRESQYIARKAKEILTSIC---------------------YNVTATSGSVTSFLRHVW--------------------GWD-------------------
O. laneus               K-------AQRFIRRKPQ----------------ES-------NEFISRQLNDTRYISKKAVEYLSAIC---------------------SDVKAFPGQLTAELRHLW--------------------GLN-------------------
Treponema JC4           KKNK----FSPNAMDSFE----------------KD-------SSFIARQLSDNQYIAKAALRYLKCLVENP------------------SDVWTTNGSMTKLLRDKW--------------------EMD-------------------
Rhodovulum PH10         RWR-----FDPAALERFE----------------RE-------GGFLGRQLNETKYLSRLAKIYLGKICDP-------------------DRVYVTPGTLTGLLRARW--------------------GLN-------------------
consensus/95%           ..................          ...........................hh.......h................................G..T..hc..h                    .h....................
consensus/85%           .............h....          .................h..R.h..o.ah.c.h..hh..h.........................h...G.hT..hR..a                    .h....................

#map                    -------------------------------------------**************************----------------------*****************------------------------------------------
```

```
Jnet 331001027          --------HHHHHHHHHHHHH--HEE----------------------------------------------------------------------------------------------------------------------------
P. excrementihominis    --------ADSSEMRKRFAEYRPEFRK---------------------------------------------------------------------------------------------------------------------------
W. succinogenes         --------TEVSELRRQYARQNPLLAK---------------------------------------------------------------------------------------------------------------------------
L. pneumophila          --------EEVHDHRELLSKQEPKLVK---------------------------------------------------------------------------------------------------------------------------
S. wadsworthensis       --------SDAKNLRLKLAQNQPDFEK---------------------------------------------------------------------------------------------------------------------------
HTCC5015                --------WSVSKQRKMLASAEPIWAK---------------------------------------------------------------------------------------------------------------------------
F. novicida             --------RGIAEIRQLYEKVDSDIQAYAK------------------------------------------------------------------------------------------------------------------------
consensus/95%                   ......R..h......h.....
consensus/85%                   ......R..h......h.....

A. coleocanis           ----------SR------------------------------------------------------------------------------------------------------------------------------------------
C. glomerans            ----------KC------------------------------------------------------------------------------------------------------------------------------------------
Acidaminococcus D21     ----------KS------------------------------------------------------------------------------------------------------------------------------------------
V. atypica              ----------KV------------------------------------------------------------------------------------------------------------------------------------------
F. nucleatum            ----------KV------------------------------------------------------------------------------------------------------------------------------------------
F. alocis               ----------KS------------------------------------------------------------------------------------------------------------------------------------------
S. moorei               ----------KF------------------------------------------------------------------------------------------------------------------------------------------
C. catus                ----------KV------------------------------------------------------------------------------------------------------------------------------------------
T. denticola            ----------KC------------------------------------------------------------------------------------------------------------------------------------------
P. duerdenii            ----------KC------------------------------------------------------------------------------------------------------------------------------------------
C. mitsuokai            ----------KN------------------------------------------------------------------------------------------------------------------------------------------
S. thermophilus         ----------KV------------------------------------------------------------------------------------------------------------------------------------------
S. mutans               ----------KV------------------------------------------------------------------------------------------------------------------------------------------
S. pyogenes             ----------KV------------------------------------------------------------------------------------------------------------------------------------------

                        ------------------------------------------------------------------------------------------------------------------------------------------------------
domains                 rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
catalytic residues      rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
domains                 rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
catalytic residues      rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
.
B. bifidum              ----------HK------------------------------------------------------------------------------------------------------------------------------------------
O. kitaharae            ----------KN------------------------------------------------------------------------------------------------------------------------------------------
S. sanguinis            -----------K------------------------------------------------------------------------------------------------------------------------------------------
F. fructosus            -----------K------------------------------------------------------------------------------------------------------------------------------------------
E. yurii                -----------K------------------------------------------------------------------------------------------------------------------------------------------
L. farciminis           -----------K------------------------------------------------------------------------------------------------------------------------------------------
S. pseudintermedius     -----------K------------------------------------------------------------------------------------------------------------------------------------------
L. rhamnosus            ----------KN------------------------------------------------------------------------------------------------------------------------------------------
F. magna                ----------KS------------------------------------------------------------------------------------------------------------------------------------------
P. antarcticus          -----------K------------------------------------------------------------------------------------------------------------------------------------------
consensus/95%                     ..
consensus/85%                     ..
.
Jnet 227501312          ------------------------------------------------------------------------------------------------------------------------------------------------------
M. mobile               ----------SI------------------------------------------------------------------------------------------------------------------------------------------
M. gallisepticum        ----------KK------------------------------------------------------------------------------------------------------------------------------------------
M. synoviae             ----------EE------------------------------------------------------------------------------------------------------------------------------------------
M. ovipneumoniae        ----------PE------------------------------------------------------------------------------------------------------------------------------------------
M. canis                ----------KL------------------------------------------------------------------------------------------------------------------------------------------
consensus/95%                     ..
consensus/85%                     ..

E. rectale              ----------KN------------------------------------------------------------------------------------------------------------------------------------------
E. faecalis             ----------LD------------------------------------------------------------------------------------------------------------------------------------------
S. thermophilus         ----------KT------------------------------------------------------------------------------------------------------------------------------------------
S. lugdunensis          ----------KE------------------------------------------------------------------------------------------------------------------------------------------
E. dolichum             ----------KD------------------------------------------------------------------------------------------------------------------------------------------
R. albus                ----------EK------------------------------------------------------------------------------------------------------------------------------------------
R. inulinivorans        ----------QK------------------------------------------------------------------------------------------------------------------------------------------
L. coryniformis         ----------KV------------------------------------------------------------------------------------------------------------------------------------------
I. polytropus           ----------KV------------------------------------------------------------------------------------------------------------------------------------------
consensus/95%                     ..
consensus/85%                     ..

Bacteroides sp.20       ----------DVWNKIVLPRFIRL------------------------------------------------------------------------------------------------------------------------------
I. album                ----------GVWKDILRPRFKRL------------------------------------------------------------------------------------------------------------------------------
B. fragilis             ----------DV-----------M------------------------------------------------------------------------------------------------------------------------------
N. salsuginis           ----------SK------------------------------------------------------------------------------------------------------------------------------------------
E. minutum              ----------GL------------------------------------------------------------------------------------------------------------------------------------------
S. globus               ----------GI------------------------------------------------------------------------------------------------------------------------------------------
M. trichosporium        ----------WM------------------------------------------------------------------------------------------------------------------------------------------
F. branchiophilum       ----------ES------------------------------------------------------------------------------------------------------------------------------------------
Prevotella C561         ----------SV------------------------------------------------------------------------------------------------------------------------------------------
P. timonensis           ----------SE------------------------------------------------------------------------------------------------------------------------------------------
P. timonensis           ----------SE------------------------------------------------------------------------------------------------------------------------------------------
A. paucivorans          ----------KN------------------------------------------------------------------------------------------------------------------------------------------
P. marinum              ----------SI------------------------------------------------------------------------------------------------------------------------------------------
A. denitrificans        ----------DV------------------------------------------------------------------------------------------------------------------------------------------
R. syzygii              ----------RV------------------------------------------------------------------------------------------------------------------------------------------
D. shibae               ----------SL------------------------------------------------------------------------------------------------------------------------------------------
R. rubrum               TAARRFLDGLTPPPLAKAVEGAVQ------------------------------------------------------------------------------------------------------------------------------
Azospirillum B510       ----------AI------------------------------------------------------------------------------------------------------------------------------------------
C. salinarum            HHAEDACLLLLAGPA---------------------------------------------------------------------------------------------------------------------------------------
N. hamburgensis         ----------GL------------------------------------------------------------------------------------------------------------------------------------------
Bradyrhizobium BTAi1    ----------DL------------------------------------------------------------------------------------------------------------------------------------------
W. succinogenes         ----------SK------------------------------------------------------------------------------------------------------------------------------------------
C. jejuni               ----------AK------------------------------------------------------------------------------------------------------------------------------------------
Jnet 218563121          ------------------------------------------------------------------------------------------------------------------------------------------------------
H. mustelae             ----------EK------------------------------------------------------------------------------------------------------------------------------------------
A. ebreus               ----------KV------------------------------------------------------------------------------------------------------------------------------------------
B. smithii              ----------KN------------------------------------------------------------------------------------------------------------------------------------------
C. cellulolyticum       ----------KN------------------------------------------------------------------------------------------------------------------------------------------
HF0070_07E19            ------------------------------------------------------------------------------------------------------------------------------------------------------
P. lavamentivorans      ----------NI------------------------------------------------------------------------------------------------------------------------------------------
N. meningitidis         ----------KV------------------------------------------------------------------------------------------------------------------------------------------
P. multocida            ----------KA------------------------------------------------------------------------------------------------------------------------------------------
A. muciniphila          ----------GV------------------------------------------------------------------------------------------------------------------------------------------
Actinomyces F0310       ----------KR------------------------------------------------------------------------------------------------------------------------------------------
A. cellulolyticus       ----------RL------------------------------------------------------------------------------------------------------------------------------------------
B. longum               ----------GK------------------------------------------------------------------------------------------------------------------------------------------
A. naeslundii           ----------SR------------------------------------------------------------------------------------------------------------------------------------------

                        ------------------------------------------------------------------------------------------------------------------------------------------------------
domains                 rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
catalytic residues      rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr

A. hesperidum           ----------KD------------------------------------------------------------------------------------------------------------------------------------------
Alcanivorax W11-5       ----------TVIPQVRFESG-MPVVDEEGAEI------------------------------TPEEFEKFRLQWEGHRVTREMRTDRR-------------------------------------------------------------
B. zoohelcum            ----------DKFKLLLKERY-EALLESEKFLEAEYDNYKKDFDSRKKEYEEKEVLFEEQELTREEFIKEYKENYIRYKKNKLIIKGWS-------------------------------------------------------------
Porphyromonas F0450     ----------KILHTLNLDRY-DSMGETERV------------------------------------------SREGEATEELHITNWS-------------------------------------------------------------
B. intestinihominis     ----------TVLHDLNFDRY-KKVGLTEVIEV----------------------------------------NHRGSVIRREQIKDWS-------------------------------------------------------------
O. laneus               ----------NILQSAPDITFPLPVSATENHREYYVITNEQNEVIRLFPKQGETPRTEKGELLLTGEVERKVFRCKGMQEFQTDVSDGKYWRRIKLSSSVTWSPLFAPKPISADGQIVLKGRIEKGVFVCNQLKQKLKTGLPDGSYWISL
Treponema JC4           ----------SIL-------------------------------------------------------------CRKFTEKEVALLGLK-------------------------------------------------------------
Rhodovulum PH10         ----------SIL------------------------------------------------------------------------SDSN-------------------------------------------------------------
consensus/95%           ......................................................................................................................................................
consensus/85%           ......................................................................................................................................................

#map                    ------------------------------------------------------------------------------------------------------------------------------------------------------
```

```
Jnet 331001027          ------------------------------------------------------------------------------------------------------------------------------------------------------
P. excrementihominis    ------------------------------------------------------------------------------------------------------------------------------------------------------
W. succinogenes         ------------------------------------------------------------------------------------------------------------------------------------------------------
L. pneumophila          ------------------------------------------------------------------------------------------------------------------------------------------------------
S. wadsworthensis       ------------------------------------------------------------------------------------------------------------------------------------------------------
HTCC5015                ------------------------------------------------------------------------------------------------------------------------------------------------------
F. novicida             ------------------------------------------------------------------------------------------------------------------------------------------------------
consensus/95%
consensus/85%

A. coleocanis           -----------------------------------------------------------------------------------------------------------------------------------------------------V
C. glomerans            ------------------------------------------------------------------------------------------------------------------------------------------------------
Acidaminococcus D21     ------------------------------------------------------------------------------------------------------------------------------------------------------
V. atypica              ------------------------------------------------------------------------------------------------------------------------------------------------------
F. nucleatum            ------------------------------------------------------------------------------------------------------------------------------------------------------
F. alocis               ------------------------------------------------------------------------------------------------------------------------------------------------------
S. moorei               ------------------------------------------------------------------------------------------------------------------------------------------------------
C. catus                ------------------------------------------------------------------------------------------------------------------------------------------------------
T. denticola            ------------------------------------------------------------------------------------------------------------------------------------------------------
P. duerdenii            ------------------------------------------------------------------------------------------------------------------------------------------------------
C. mitsuokai            ------------------------------------------------------------------------------------------------------------------------------------------------------
S. thermophilus         ------------------------------------------------------------------------------------------------------------------------------------------------------
S. mutans               ------------------------------------------------------------------------------------------------------------------------------------------------------
S. pyogenes             ------------------------------------------------------------------------------------------------------------------------------------------------------

                        ------------------------------------------------------------------------------------------------------------------------------------------------------
domains                 rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
catalytic residues      rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
domains                 rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
catalytic residues      rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
.
B. bifidum              ------------------------------------------------------------------------------------------------------------------------------------------------------
O. kitaharae            ------------------------------------------------------------------------------------------------------------------------------------------------------
S. sanguinis            ------------------------------------------------------------------------------------------------------------------------------------------------------
F. fructosus            ------------------------------------------------------------------------------------------------------------------------------------------------------
E. yurii                ------------------------------------------------------------------------------------------------------------------------------------------------------
L. farciminis           ------------------------------------------------------------------------------------------------------------------------------------------------------
S. pseudintermedius     ------------------------------------------------------------------------------------------------------------------------------------------------------
L. rhamnosus            ------------------------------------------------------------------------------------------------------------------------------------------------------
F. magna                ------------------------------------------------------------------------------------------------------------------------------------------------------
P. antarcticus          ------------------------------------------------------------------------------------------------------------------------------------------------------
consensus/95%                                                                                                                                                                        .
consensus/85%                                                                                                                                                                        .
.
Jnet 227501312          ------------------------------------------------------------------------------------------------------------------------------------------------------
M. mobile               ----------------------------------------------------------------------------------------------------------------------------------------NSNDFLIREEGKND
M. gallisepticum        ------------------------------------------------------------------------------------------------------------------------------------------------------
M. synoviae             ----------------------------------------------------------------------------------------------------------------------------------------------------QV
M. ovipneumoniae        ----------------------------------------------------------------------------------------------------------------------------------------------------KI
M. canis                ----------------------------------------------------------------------------------------------------------------------------------------------------RL
consensus/95%                                                                                                                                                ..............
consensus/85%                                                                                                                                                ..............

E. rectale              ------------------------------------------------------------------------------------------------------------------------------------------------------
E. faecalis             ------------------------------------------------------------------------------------------------------------------------------------------------------
S. thermophilus         ------------------------------------------------------------------------------------------------------------------------------------------------------
S. lugdunensis          ------------------------------------------------------------------------------------------------------------------------------------------------------
E. dolichum             ------------------------------------------------------------------------------------------------------------------------------------------------------
R. albus                ---------------------------------------------------------------------------------------------------------------------------------------------------TWG
R. inulinivorans        ------------------------------------------------------------------------------------------------------------------------------------------------------
L. coryniformis         ------------------------------------------------------------------------------------------------------------------------------------------------------
I. polytropus           ------------------------------------------------------------------------------------------------------------------------------------------------------
consensus/95%                                                                                                                                                                      ...
consensus/85%                                                                                                                                                                      ...

Bacteroides sp.20       --------------------------------------------------------------------------------------------------------------------------NELTESTRFTSINTNNTMIPSMPLELQK
I. album                --------------------------------------------------------------------------------------------------------------------------ESII-NKKLIFQDEDDPNKYHFDLSINP
B. fragilis             ------------------------------------------------------------------------------------------------------------------------------KELNWEKYKALGLVEYFEDRDGRQ
N. salsuginis           ------------------------------------------------------------------------------------------------------------------------------------------------------
E. minutum              ------------------------------------------------------------------------------------------------------------------------------------MIPFAKQLITEKESESFN
S. globus               -----------------------------------------------------------------------------------------------------------------------------------------DDLLGSRHWSKDA
M. trichosporium        ------------------------------------------------------------------------------------------------------------------------------------------------------
F. branchiophilum       ------------------------------------------------------------------------------------------------------------------------------------------------FIDENG
Prevotella C561         ------------------------------------------------------------------------------------------------------------------------------------------------------
P. timonensis           ------------------------------------------------------------------------------------------------------------------------------------------------------
P. timonensis           ------------------------------------------------------------------------------------------------------------------------------------------------------
A. paucivorans          ------------------------------------------------------------------------------------------------------------------------------------------------------
P. marinum              ------------------------------------------------------------------------------------------------------------------------------------------------LRGHNT
A. denitrificans        ------------------------------------------------------------------------------------------------------------------------------------------------------
R. syzygii              ------------------------------------------------------------------------------------------------------------------------------------------------LDGVME
D. shibae               --------------------------------------------------------------------------------------------------------------------------------------------------LSDA
R. rubrum               ------------------------------------------------------------------------------------------------------------------------------ARLAALGRSRVADAGLADALGLTL
Azospirillum B510       ---------------------------------------------------------------------------------------------------------------------------------------LAEADGPPPEVPAET
C. salinarum            -------------------------------------------------------------------------------------------------------------------------------------VVQAFNTEAAQHGADAP
N. hamburgensis         ---------------------------------------------------------------------------------------------------------------------------------LPSDNYAGVQDKAEEFLASTD
Bradyrhizobium BTAi1    ---------------------------------------------------------------------------------------------------------------------------------------------------LPG
W. succinogenes         ------------------------------------------------------------------------------------------------------------------------------------------------------
C. jejuni               ------------------------------------------------------------------------------------------------------------------------------------------------------
Jnet 218563121          ------------------------------------------------------------------------------------------------------------------------------------------------------
H. mustelae             ------------------------------------------------------------------------------------------------------------------------------------------------------
A. ebreus               ------------------------------------------------------------------------------------------------------------------------------------------------------
B. smithii              ------------------------------------------------------------------------------------------------------------------------------------------------------
C. cellulolyticum       ------------------------------------------------------------------------------------------------------------------------------------------------------
HF0070_07E19            ------------------------------------------------------------------------------------------------------------------------------------------------------
P. lavamentivorans      ------------------------------------------------------------------------------------------------------------------------------------------------------
N. meningitidis         ------------------------------------------------------------------------------------------------------------------------------------------------------
P. multocida            ------------------------------------------------------------------------------------------------------------------------------------------------------
A. muciniphila          -----------------------------------------------------------------------------------------------------------------------------------------FKELCPEAADPDS
Actinomyces F0310       ------------------------------------------------------------------------------------------------------------------------------------------------VRLIRL
A. cellulolyticus       -------------------------------------------------------------------------------------------------------------------------------------------------FSRVA
B. longum               -----------------------------------------------------------------------------------------------------------------------------------------------------I
A. naeslundii           ----------------------------------------------------------------------------------------------------------------------------------------------------IN

                        ------------------------------------------------------------------------------------------------------------------------------------------------------
domains                 rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
catalytic residues      rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr

A. hesperidum           ------------------------------------------------------------------------------------------------------------------------------------------------------
Alcanivorax W11-5       ------------------------------------------------------------------------------------------------------------------------------------------------------
B. zoohelcum            ------------------------------------------------------------------------------------------------------------------------------------------------------
Porphyromonas F0450     ------------------------------------------------------------------------------------------------------------------------------------------------------
B. intestinihominis     ------------------------------------------------------------------------------------------------------------------------------------------------------
O. laneus               PVISQTFKEGESVNNSKLTSQQVQLFGRVREGIFRCHNYQCPASGADGNFWCTLDTDTAQPAFTPIKNAPPGVGGGQIILTGDVDDKGIFHADDDLHYELPASLPKGKYYGIFTVESCDPTLIPIELSAPKTSKGENLIEGNIWVDEHTG
Treponema JC4           -----------------------------------------------------------------------------------------------------------------------------------------------------P
Rhodovulum PH10         ------------------------------------------------------------------------------------------------------------------------------------------------------
consensus/95%           ......................................................................................................................................................
consensus/85%           ......................................................................................................................................................

#map                    ------------------------------------------------------------------------------------------------------------------------------------------------------
```

```
Jnet 331001027          ---------------HHHHHHHHHHHHH----------------------------------------------EEEEE--------------------------EEE------------------------------------------
P. excrementihominis    -------PKVQPVASHSIDAMCIYLAACSDPFK-----------------------------------TKRMGSQLAIYEPINFDN--------------LFTGSCQVIQNTPRNFSDKTNIANS-------------------------
W. succinogenes         -------AEKQAPSSHAIDAVMAFVARYQKVFKDGTPP------------------------------NADEVAKLAMLDSWNPASNEPLT--------KGLSTNQKIEKMIKSGDYGQKNMREVFGK----------------------
L. pneumophila          -------SRQQSFPSHAIDATLTMSIGL----------------------------------------KEFPQFSQELDNSWFINH--------------LMPDEVHLNPVRSKEKYNKPNISST-------------------------
S. wadsworthensis       -------PDIQPIASHSIDALCSFAVGSAD--------------------------------------AERDQNGFDYLDGKTVLG--------------LYPQSCEVIHLQAKPQEEKSHFDSV-------------------------
HTCC5015                -------KDPQPAASHVVDAVCTFLEALE---------------------------------------QPHTASRLKTISSTSFEKTGWRS--------ALIPDLIKVDALDRRPKYRRYNIGST-------------------------
F. novicida             ------GDKPQASYSHLIDAMLAFCIAADEHRNDGSIGLEID--------------------------KNYSLYPLDKNTGEVFTKDIFSQ--------IKITDNEFSDKKLVRKKAIEGFNTHR-------------------------
consensus/95%                 ..c.Q...SH.hDA...h..................                          .......................        ..h..........................
consensus/85%                 ..c.Q...SH.hDA...h..................                          .......................        ..h..........................

A. coleocanis           NFIGGNGKTRLDRRHHAMDAATVAMLRNSVA----KTLVLRG----------------NIRASERAIG----------------------AAETWKSFRGENVADRQI-----------------FESW---SENMRVLVEKFNL-----
C. glomerans            --------REANDFHHAHDAFLACRVGLFIQ----KR--HPC---------------VYENPIGLSQVVRNYVRQ---------------QADIFKRCRTIPGSSGFIVNSFMTSGFDKETGEIFKDDW---DAEAEVEGIRRSL-----
Acidaminococcus D21     --------RNINDLHHAKDAFLAIVTGNVYH----ER--FNR-----------------RWFMVNQPY----------------------SVKTKTLFTHSIKNGN-------------------FVAW---NGEEDLGRIV-KM-----
V. atypica              --------RSLNHHHHAKDAYLNIVVGNVYH----EK--FTR---------------NFRLFFKKNGANR--------------------TYNLAKMFNYDVICTNAQD----------------GKAW---DVKTSMNTVK-KM-----
F. nucleatum            --------RELNDTHHAKDAYLNIVAGNVYN----TK--FTE---------------KPYRYLQEIKE----------------------NYDVKKIYNYDI-----------------------KNAW---DKENSLEIVKKNM-----
F. alocis               --------RRVNDYHHAKDAYLNIVVGNVYN----KK--FTS---------------NPIQWMKKNRDT---------------------NYSLNKVFEHDVVING-------------------EVIW---EKCTYHEDTN-TY-----
S. moorei               --------REINDLHHAKDAYLNIVVGNVYD----TK--FTE---------------KFFNNIRNE------------------------NYSLKRVFDFSV-----------------------PGAW---DAKGSTFNTIKKY-----
C. catus                --------REMNDLHHAKDAYLNIVVGNAYF----VK--FTK---------------NAAWFIRNNPGR---------------------SYNLKRMFEFDIERSG-------------------EIAW---KAGNKGSIVTVKK-----
T. denticola            --------REINDFHHAHDAYLNIVVGNVYN----TK--FTN---------------NPWNFIKEKRDNPKIAD----------------TYNYYKVFDYDVKRNN-------------------ITAW---EKGKTIITVK-DM-----
P. duerdenii            --------RTVNDLHHMHDAYLNIVVGNVYN----TK--FTK---------------NPLNFIKDKDNVR--------------------SYNLENMFKYDVVRGS-------------------YTAWIADDSEGNVKAATIKK-----
C. mitsuokai            --------RDINDYHHAHDAYIVALIGGFMR----DR--YPN-------------------MHDSKAV----------------------YSEYMKMFRKNKNDQKRWKDGFVINSMNYPYEVDGKLIWN--PDLINEIKKC--------
S. thermophilus         --------REINDFHHAHDAYLNAVVASALL----KK--YPK-------------LEPEFVYGDYPKYN-------SFRERKSATEKVYFYSNIMNIFKKSISLADGRVIERPLIEVNEETG---ESVW---NKESDLATVR-RV-----
S. mutans               --------REINDYHHAHDAYLNAVIGKALL----GV--YPQ-------------LEPEFVYGDYPHFH-------GHKE-NKATAKKFFYSNIMNFFKKD------DVR-------TDKNG---EIIW---KKDEHISNIK-KV-----
S. pyogenes             --------REINNYHHAHDAYLNAVVGTALI----KK--YPK-------------LESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETG---EIVW---DKGRDFATVR-KV-----

                        -------------HHHHHHHHHHHHHHHHHH----HH-------------------HHHH---------HHHH---------------------------EEE-----EEE---EEE---------EEE---E--HHHHHHH-HH-----
domains                 rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
catalytic residues      rrrrrrrrrrrrr*H...D*rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
domains                 rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
catalytic residues      rrrrrrrrrrrrr*H...D*rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
.
B. bifidum              -------NRDINDYHHAQDALCVGIAGQFAA----NRGFFAD------GEVSDGAQNSYNQYLRDYLRGYREKLSAE-------------DRKQGRAFGFIVGSMRSQDEQKRVNPRTG------EVVW---SEEDKDYLRK-VM-----
O. kitaharae            --------RDINDYHHAFDALLFSTVGQYTE----NSGLMKK------GQLSDSAGNQYNRYIKEWIHAARLNA----------------QSQRVNPFGFVVGSMRNAAPGK-LNPETG------EITP----EENADWSIA-DL-----
S. sanguinis            -------IRELNDYHHAHDAYLAVVVGQTLL----KVYPKLA------PELIYGHHAHFNRHEENKATLRKHLYSN------------------------IMRFFNNPDSK------VS------KDIWD--CNRDLPIIKD-VI-----
F. fructosus            -------HREINSFHHAFDALLITAAGQYMQ----NRYPDRD---------STNVYNEFDRYTNDYLKNLRQLSSRD-------------EVRRLKSFGFVVGTMRKGNED-----------------W---SEENTSYLRK-VM-----
E. yurii                -------SRLINDFHHANDAYLNIVVGNSY-------------------------FVKFTRNPANFIKDARKNPDN-------------------PVYKYHMDRFFERDVK-----SKSEVAWIGQSEG---NSGTIVIVKK-TM-----
L. farciminis           -------NRNVNDFHHAQDAYLASFLGTYRL----RRFPTNE------MLLMNGEYNKFYGQVKELYSKKKKLPDS-------------------RKNGFIISPLVNGTTQ--YDRNTG------EIIW---NVGFRDKILK-IF-----
S. pseudintermedius     -------IRQMNDAHHAIDAYLNGVVYHGAQ----LAYPNVD-------------LFDFNFKWEKVREKWKALGEFN-------------TKQKSREL-FFFKKLEKMEVS-----------------Q---GERLISKIKL-DM-----
L. rhamnosus            --------RDVNHYHHAFDAFLAARIGTYLL----KR--YPK-------------LAPFFTYGEFAKV----------------------DVKKFREFNFIGALTHAKKNIIAKDTG--------EIVW---DKERDIRELD-RI-----
F. magna                --------REVNDLHHAHDAFLNIVAGDVWN----RE--FTS---------------NPINYVKENREGDKV------------------KYSLSKDFTRPRKSKG-------------------KVIWTPEKGRKLIVDTL--------
P. antarcticus          -------IRDYNNKHHAMDALFAAALIQSIL----GKYGKN--------------FLAFDLSKKDRQKQWRSVKGSN------------------KEF-FLFKNFGNLRLQ---SPVTG------EEV------SGVEYMKH-VY-----
consensus/95%           ........R.hN..HHA.DAh....h.....    .......      .................................................................................................
consensus/85%           ........R.hNchHHA.DAhh..hhu..h.    .......      .................................................h...............................................
.
Jnet 227501312          -------------HHHHHHHHHHHHH-HHHH----HH--H-------------------HHHH---------------------------------------------------------------EE---------HHHHH--------
M. mobile               VKDDVLIKDRSFNGHHAEDAYFITIISQYFR----SFKRIER---------------LNVNYRKE-------------------------TRELDDLEKNNIKFKE-------------------KASF---DNFLLINALD-EL-----
M. gallisepticum        --------DRDKNIHHAVDASIISIFSNETK----TL--FNQ----------------LTQFADYKLFKN--------------------TDGSWKKIDPKTGVVTEVT----------------DENW---KQIRVRNQVS-EI-----
M. synoviae             YRFNFWKKDRDQFFHHAVDASIIAIFSLLTK----TLYNKLR---------------VYESYDVQRREDG--------------------VYLINKETGEVKKAD--------------------KDYWKDQHNFLKIRENAIEI-----
M. ovipneumoniae        E------KNRENNEHHAVDAAIVAIIGNKNP----QIANLLT---------------LADNKTDKKFLLHDENYKE------------------NIETGELVKIP--------------------KFE---VDKLAKVEDLKKII-----
M. canis                -------KDRSDFSHHAYDAAIIALFSNKTK----TLYNLID---------------PSLNG--IISKRSEGYWVIE-----------------DRYTGEIKELK--------------------KEDWTSIKNNVQARKIAKEI-----
consensus/95%           .........R....HHA.DA.hh.hhu....    .h.....               ....................             .............h.....                c......c...........h
consensus/85%           .........R....HHA.DA.hh.hhu....    .h.....               ....................             .............h.....                c......c...........h

E. rectale              --------RDESYSHHAVDAMLIGYSELGYE----AYHKLQG---------------EFIDFETGEIL----------------------RKDMWDENMSDEVYADYLY----------------GKKW---ANIRNEVVKA--------
E. faecalis             -------KTRETHHHHAVDATLCAVTSFVKV----SRYHYAV------KEETGEKVMREIDFETGEIVNEMS------------------YWEFKKSKKYERKTY--------------------QVKW---PNF-----RE-QL-----
S. thermophilus         ---------RDTYHHHAVDALIIAASSQLNL--------WKK-----------QKNTLVSYSEDQ-------------------------LLDIETGELISDDEYKESVF---------------KAPY---QHFVDTLKSK-EF-----
S. lugdunensis          --------RNHGYKHHAEDALIIANADFLFK----ENKKLKA---------------VNSVLEKPEIE----------------------SKQLDIQVDSEDNYS---------------------------EMFIIPKQVQ-DI-----
E. dolichum             --------RDEDYGHHAIDALIIASMPKMRL--------LST---------------IFSRYKIEDIY----------------------DESTGEVFSSGDDSMYY------------------DDRY---FAFIASLKAI--------
R. albus                RYDKA-ELKKLTYLDHAADAIIIANCRPEYVVLAGEKLKLNK-----------------MYHQAGKRITPEYEQSKKACIDNLYKLFRMDRRTAEKLLSGHGRLTPIIPNLSEEVDKRLWDKNIYEQFW---KDDKDKKSCE-EL-----
R. inulinivorans        --------DRSTDRHHAMDAVVIACCTDGMI----HK--ISR---------------YMQGRELAYSRNFKFPDEETGEI-------LNRDNFTREQWDEKFGVKV-------------------PLPW---NSFRDELDIRLLNEDPKN
L. coryniformis         --------REDGDLHHAIDATVIAAITPKFI----QQ--VTK---------------YSQHQEVKN------------------------NQALWHDAEIKDAEYAAEAQRMDADLFNKIFNGF-PLPW---PEFLDELLARISD-----
I. polytropus           --------RENGDLHHALDAAVVAVTDQKAI----NN--ISN----------------ISRYKELKNCKDVIPSIEYHA-----------DEETGEVYFEEVKDTRF------------------PMPW---SGFDLELQKRLESENPRE
consensus/95%           ..... ........cHA.DA.h.u...............h..      .................................................................................   ..................
consensus/85%           ..... ...c....HHAhDA.hhA...............h..      ................................................................................a   ..h...............

Bacteroides sp.20       GFNK----KRIDHRHHAMDAIIIACANRNIV----NY--LNN----------------VSASKNTKIT----------------------RRDLQTLLCHKDKTDNNGNYKWVI-----------DKPW---ETFTQDTLTALQK-----
I. album                QLDKEGL-KRLDHRHHALDATIIAATTREHV----RY--LNS----------------LNAADNDE-E----------------------KREYFLSLCNHKIRD--------F-----------KLPW---ENFTSEVKSKLLS-----
B. fragilis             IGRIKDWTKRNDHRHHAMDALTVAFTKDVFI----QY--FNN----------------KNASLDPNAN----------------------EHAIKNKYFQNGRAIA-------------------PMPL---REFRAEAKKHLEN-----
N. salsuginis           --------SRETNFHHAEDALILSTLTRGWQ----NR--LHR--------------MLRDNYGKSEAE----------------------LKELWKKYMPHIEGLTLADYI--------------DEAF---RRFMSKGEES-LF-----
E. minutum              KDVNSNKKIRLDNRHHALDAIVIAYASRGYG----NL--LNK------------------------------------------------MAGKDYKINYSERNWLSKILLPPN-----------NIVW---ENIDADLESF-ES-----
S. globus               DTSPTMRKNRDDNRHHGLDAIVALYCSRSLV----QM--INT------------------------------------------------MSEQGKRAVEIEA----------------------MIPI---PGYASEPNLSFEA-----
M. trichosporium        -KKSTKGDRIPDDRHHALDAIVIAATTESLL----QR-----------------------------------------------------ATREVQEIEDKGLHYDLVKNV--------------TPPW---PGFREQAVEA--------
F. branchiophilum       MKHYK-VKDRSKHTHHTIDAITIACMTKEKY----DV--LAH------------------------------------------------AWTLEDQQNK---KEARSIIEA-------------SKPW---KTFKEDLLK---------
Prevotella C561         ---DE-KKDRSLHSHHAIDATTLTIIPVSAKRD--RM--LEL------------------------------------------------FAKIEEINKMLSFSGSEDRTGL-------------IQEL---EGLKNKLQMEVKVCRIGH
P. timonensis           ---YE-KKCRDNHSHHCMDAITIACIGKREY----DL--MAE------------------------------------------------YYRMEETFK----QGRGSKPKF-------------SKPW---ATFTEDVLN---------
P. timonensis           ----YEKKCRDNHSHHCMDAITIACIGKREY----DL--MAE------------------------------------------------YYRMEETFKQGRGSKPKF-----------------SKPW---ATFTEDVLNI--------
A. paucivorans          --------RKNGDLHHALDAAVLAVASRSFV----YR--LSS------------------------------------------------HNAAWGELPRGREAENGF-----------------SLPY---PAFRSEVLARLCP-----
P. marinum              DDGTPAKKSRDDHRHHAIDAIVVGMTSRGLL----QK--VSK------------------------------------------------AARRSEDLDLTRLFEGR------------------IDPW---DGFRDEVKKH--------
A. denitrificans        -LGLDGEKNRNDHRHHAVDACVIGVTDQGLM----QR--FAT------------------------------------------------ASAQARGDGLTRLVDGM------------------PMPW---PTYRDHVERA--------
R. syzygii              EQGRPAVKNRDDHRHHAIDAVVIGATDRAML----QQ--VAT------------------------------------------------LAARAREQDAERLIGDM------------------PTPW---PNFLEDVRAA--------
D. shibae               GRGAVKAKNRTDHRHHAIDAAVIAATDPGLL----NR--ISR------------------------------------------------AAGQGEAAGQSAELIARDT----------------PPPW---EGFRDDLRVR--------
R. rubrum               ASLGGGGKNRADHRHHFIDAAMIAVTTRGLI----NQ--INQ------------------------------------------------ASGAGRILDLRKWPRTNF-----------------EPPY---PTFRAEVMKQ--------
Azospirillum B510       LDPSPAEKNRADHRHHALDAVVIGCIDRSMV----QRVQLAA------------------------------------------------ASAEREAAAREDNIRRVLEGFK-------------EEPW---DGFRAELERR--------
C. salinarum            DDRPVDLMPTSDAYHQQRRARALGRVPLATV----DAAL---------------------------------------------------ADIVMPESDRQDPETG-------------------RVHW---RLTRAGRGLKRRIDDLTR
N. hamburgensis         DMEFSGVKNRADHRHHAIDGLVTALTDRSLL------------------------------------------------------------WKMANAYDEEHEKFVI------------------EPPW---PTMRDDLKAA--------
Bradyrhizobium BTAi1    SDDRA-AKSRKDHRHHAIDALVAALTDQALL----RR-----------------------------------------------------MANAHDDTRRKIEV---------------------LLPW---PTFRIDLETR--------
W. succinogenes         --------DRESHTHHAVDAIIIAFSTQGMV----QK--LSE------------------------------------------------YYRFKETHREKERPKL-------------------AVPL---ANFRDAVEEATRI-----
C. jejuni               --------DRNNHLHHAIDAVIIAYANNSIV----KA--FSD----------------FKKEQESNSAEL--------------------YAKKISELDYKNKRKF-------------------FEPF---SGFRQKVLDK--------
Jnet 218563121          ---------------HHHHHHHHHHH-HHHH----HH--HHH----------------HHHHH-----------------------------------------------------------------------HHHHHHHH--------
H. mustelae             --------NRDHHLHHAQDAIIIACIEPSMI----QK--YTT----------------YLKDKETHRLK---------------------SHQKAQILREGDHKL--------------------SLRW-PMSNFKDKIQES--------
A. ebreus               --------RGDSDRHHALDAAVVAACTHGMV----KA--LAD----------------YSRRKEISFLQEGFP-----------------DPETGEILNPAAFDRARQHF---------------PEPW---THFAHELKARLFT-----
B. smithii              --------REESDLHHAVDAVIVACATQGMI----KK--ITE---------------------------------------------FYKAREQNKESAKKKEPIF-------------------PQPW---PHFADELKARLSK-----
C. cellulolyticum       --------REESDLHHALDAAVIACADRKII----KE--ITN------------------------------------------------YYNERENHNYKVKY---------------------PLPW---HSFRQDLMET--------
HF0070_07E19            -------KDRDKHTHHAVDAIVVACTTPAIV----QQ--VTL------------------------------------------------YHQDIRRYKKLGEKR--------------------PTPW--PETFRQDVLDV--------
P. lavamentivorans      -LADDGEKTRADHRHHAIDALTVACTHPGMT----NK--LSR------------------------------------------------YWQLRDDPRAEKPAL--------------------TPPW---DTIRADAEKA--------
N. meningitidis         --------RAENDRHHALDAVVVACSTVAMQ----QK--ITR------------------------------------------------FVRYKEMNAFDGKT----IDKETGEVLHQKTHF--PQPW---EFFAQEVMIRVFGKPDGK
P. multocida            --------RENNNRHHALDAIVVACATPSMQ----QK--ITR------------------------------------------------FIRFKEVHPYKIENRYEMVDQESGEIISP--HF--PEPW---AYFRQEVNIRVFD---NH
A. muciniphila          GKILKENLRSLTHLHHALDACVLGLIPYIIP----AH-----------------------------------------------------HNGLLRRVLAMRRI---------------------PEKL---------------------
Actinomyces F0310       KGDDGHHKNRFDRRNHAMDALVIALMTPAIA----RTIAVRE----------------DRREAQQLTR----------------------AFESWKNFLGSEERMQDR-----------------WESW---IGDVEYACDRLNE-----
A. cellulolyticus       IFAQSTSTKRLDRRHHAVDAVVLTTLTPGVA----KT--LAD--------------------ARSRRV----------------------SAEFWRRPSDVNRHSTEEPQSPA------------YRQWKESCSGLGDLLIS--------
B. longum               HFIGQQSKTRLDRRHHAVDASVIAMMNTAAA----QT--LME---------------RESLRESQRLIGLMP------------------GERSWKEYPYEGTSRYES-----------------FHLWL--DNMDVLLELLNDA-----
A. naeslundii           LIGEKGRKDRIDRRHHAVDASVVALMEASVA----KT--LAE---------------RSSLRGEQRLTG---------------------KEQTWKQYTGSTVGAREH-----------------FEMWR--GHMLHLTELFNER-----

                        --------------HHHHHHHHHHH--HHHH----HH--HHH----------------HHHHHHHHHHH-------------------------HHH-----HHHHHH-----------------HHHHH--HHHHHHHHHHHHH-----
domains                 rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
catalytic residues      rrrrrrrrrrrrr*H...D*rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr

A. hesperidum           --------RDAGDKHHAADAVVVACIGDSLI----QR--VTK---------------YDKFKRNALADRNRYVQQVSKSEG--ITQYVDKETGEVFTWESFDERKFL--------------PNEPLEPWP---FFRDELLARLSDDPSKN
Alcanivorax W11-5       ------PDKRIDHRHHLVDAIVTALTSRSLY----QQ--YAK---------------AWKVADEKQRHGR-------------------------------VDVKVEL-----------------PMPIL---TIRDIA-----------
B. zoohelcum            --------KRIDHRHHAIDALIVACTEPAHI----KR--LND---------------LNKVLQDWLVEHKSEFMPNFEGSN--SELLEEILSLPENERTEIFTQIEK--------------FRAIEMPWK---GFPEQV-----------
Porphyromonas F0450     --------KRMDHRHHAIDALVVACTRQSYI----QR--LNR---------------LSSEFGREDKKKE---------------DQEAQEQQATETGRLSNLERWLT-----------------QRPHF---SVRTV------------
B. intestinihominis     --------KRFDHRHHAIDALTIACTKQAYI----QR--LNN---------------LRAEEGPDFNKMS--------------------------------LERYIQ-----------------SQPHF---SVAQV------------
O. laneus               EVRFDPKKNREDQRHHAIDAIVIALSSQSLF----QR--LST---------------YNARRENKKRGLD---------------------------------STEHF-----------------PSPWP---GFAQDV-----------
Treponema JC4           EQIGNYKKNRFDHRHHAIDAVVIGLTDRSMV----QK--LAT---------------KNSHKGNRIEI-----------------------------------------------------------PEF--PILRSDL-----------
Rhodovulum PH10         ------FKNRSDHRHHAVDAVVIGVLTRGMI----QR--IAH---------------DAARAEDQDLDRV----------------------------------FRDV-----------------PVPFE---DFRDHV-----------
consensus/95%           ..............HH..DA..h..........  .......              ..........................  ............................................h.....................
consensus/85%           .........c..c.HHAhDAhhhu.....h...  ....h..              ..........................  ............................................h.....h...............

#map                    --------********************--------------------------------------------------------------------------------------------------------------------------
```

```
Jnet 331001027          ------------------------------------------------EEEE-----EEEE-----------------------------------------------EE-----HHHHHHHHHHHHH--------EEEEEE--------
P. excrementihominis    -------------------------------PIFKETIY---AERFLDIIVSRG--EIFIG---------------YP--SNMPFEEK-----------------PNRISIGGKDPFSILSVLGAYLDKAPSSEKEKLTIYRVV------
W. succinogenes         -------------------------------SIFGENAI---GERYKPIVVQEG--GYYIG---------------YPATVKKGYELK-----------------NCKVVTSKNDIAKLEKII-------KNQDLISLKENQYIKIFSIN
L. pneumophila          -------------------------------PLFKDSLY---AERFIPVWVKGE--TFAIG---------------FS--EKDLFEIK---------------------PSNKEKLFTLLKTYSTKNPGESLQELQAKSKAKWL-YFPIN
S. wadsworthensis       -------------------------------AIFKEGIY---AEQFLPIFTLNE--KIWIG---------------YE--TLNAKGER-----------------CGAIEVSGKQPKELLEMLAPFFNK-PVGDLSAHATYRIL------
HTCC5015                -------------------------------SLFKDGIY---AERFLPILIDEN--GLMAG---------------YD--IDNSLKAK-----------------GADVVFESLSPFLLFK--GEEVGAQSLSDWQERIDGRYL-YMSID
F. novicida             -------------------------------QMTRDGIY---AENYLPILIHKELNEVRKG---------------YT--WKNSEEIKIFKGKKYDIQQLNNLVYCLKFVDKPISIDIQISTLEELRNILTTNNIAATAEYYYI---NLK
consensus/95%                                          .h..-.hh   uE.a..hh.......h..G               a..........c......................................................hh......
consensus/85%                                          .h..-.hh   uE.a..hh.......h..G               a..........c......................................................hh......

A. coleocanis           ------------------------------ALYNDEVSI---FSSLRLQL----------------------------GNGKAHDDTI----------------------------TKLQMHKVGDAWSLTEIDRAST-----PALWCAL
C. glomerans            ---------------------------------NFRQCF---ISRMPFED-----------------------------HGVFWDATIYS--------------------------------PRAKKTAALPLKQGLN-----PSRYGSF
Acidaminococcus D21     -------------------------------LKQNKNTIH--FTRFSFDR----------------------------KEGLFDIQPL------------------------------------KASTGLVPRKAGLD-----VVKYGGY
V. atypica              --------------------------------MASNDVR---VTRRLLEQ-----------------------------SGALADATIYK------------------------ASVAAKAKDGAYIGMKTKYSVFAD-----VTKYGGM
F. nucleatum            ----------------------------------EKNTVN--ITRFIKEE-----------------------------KGELFNLNP-------------------------IKKGETSNEIISIKPKLYDGKDNKL-----NEKYGYY
F. alocis               ---------------------DGGTLDRIRKIVERDNIL---YTEYAYCE-----------------------------KGELFNATI----------------------------------QNKNGNSTVSLKKGLD-----VKKYGGY
S. moorei               --------------------------------MAKNNPI---IAFAPYEV-----------------------------KGELFDQQI-----------------------------------VPKGKGQFPIKQGKD-----IEKYGGY
C. catus                -------------------------------VMQKNNIL---VTRKAYEV-----------------------------KGGLFDQQI-------------------------------MKKGKGQVPIKGNDERLAD-----IEKYGGY
T. denticola            --------------------------------LKRNTPI---YTRQAACK-----------------------------KGELFNQTI--------------------------------MKKGLGQHPLKKEGPFSN-----ISKYGGY
P. duerdenii            -----------------------------VKRELEGKNYR--FTRMSYIG-----------------------------TGGLYDQNL--------------------------------MRKGKGQIPQKENTNKSN-----IEKYGGY
C. mitsuokai            --------------------------------FYYKDCY---CTTKLDQK-----------------------------SGQLFNLTV--------------------------LSNDAHADKGVTKAVVPVNKNRSD-----VHKYGGF
S. thermophilus         --------------------------------LSYPQVN---VVKKVEEQ-----NHGLDR---------------GKPKGLFNANLSS--------------------------------KPKPNSNENLVGAKEYLD----PKKYGGY
S. mutans               --------------------------------LSYPQVN---IVKKVEEQ-----------------------------TGGFSKESI---------------------------------LPKGNSDKLIPRKTKKFYWD--TKKYGGF
S. pyogenes             --------------------------------LSMPQVN---IVKKTEVQ-----------------------------TGGFSKESI---------------------------------LPKRNSDKLIARKKDWD-----PKKYGGF

                        --------------------------------HH----E---EEE---------------------------------------------------------------------------------------------------------
domains                 rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr|ppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppp*W*ppppppppppp|
catalytic residues      rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr|ppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppp*W*ppppppppppp|
domains                 rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr|PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
catalytic residues      rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr|PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
.
B. bifidum              ---------------------------------NYRKML---VTQKVGDD-----------------------------FGALYDETRYA------------------------------ATDPKGIKGIPFDGAKQD-----TSLYGGF
O. kitaharae            ---------------------------------DYLHKV---MNFRKITVTRRLKDQ----------------------KGQLYDESRYP----------------------------SVLHDAKSKASINFDKHKP------VDLYGGF
S. sanguinis            ---------------------------------YNSQIN---FVKRTMIK-----------------------------KGAFYNQNPVG----------------------------KFNKQLAANNRYPLKTKALCLD---TSIYGGY
F. fructosus            ---------------------------------MFKNIL---TTKKTEKD-----------------------------RGPLNKETIFS------------------------------PKS--GKKLIPLNSKRSD-----TALYGGY
E. yurii                ---------------------------------AKNSPL---ITKKVEEG-----------------------------HGSITKETIVGVKEIKFGRNKVEK-------------ADKTPKKPNLQAYRPIKTS-DERLCN-ILRYGGR
L. farciminis           ---------------------------------NYHQCN---VTRKTEIK-----------------------------TGQFYDQTIYS------------------------------PKNPKYKKLIAQKKD-MD-----PNIYGGF
S. pseudintermedius     ---------------------------------NHFKIN---YSRKLANI-----------------------------PQQFYNQTAVS------------------------------PKTAELKYESNKSN---------EVVYKGL
L. rhamnosus            --------------------------------YNFKRML---ITHEVYFE-----------------------------TADLFKQTIYA---------------------------AKDSKERGGSKQLIPKKQGYP-----TQVYGGY
F. magna                ---------------------------------NKPSVL---ISNESHVK-----------------------------KGELFNATIAG---------------------------KKDYKKGKIYLPLKKDDRLQD-----VSKYGGY
P. antarcticus          ---------------------------------FELPWQ---TTKMTQTG-----------------------------DGMFYKESIFS------------------------------PKVKQAKYVSPKTE---------KFVHDEV
consensus/95%                                ..................   ...................               ....u.h....................          ............................. ...a..h
consensus/85%                                ..................   h..................               ....G.h....................          ............................. ...YGuh
.
Jnet 227501312          --------------------------------------E---EEEEEEE------------------------------------------------------------------------------------------------EEEE-
M. mobile               -------------------------------NEKLNQMR---FSRMVITK----------------------------KNTQLFNETLYS----------------------GKYDKGKNTIKKVEKLNLLDNRTDKI-----KKIEEFF
M. gallisepticum        ---------------------AKVIEKYIQDSNIERKAR---YSRKIENK----------------------------TNISLFNDTVYSAKKVGYEDQIKRKNLKTLDIHESAKENKNSKVKRQFVYRKLVNVSLLN-----NDKLADL
M. synoviae             -------------------------KNVLNNVDFQNQVR---YSRKANTK----------------------------LNTQLFNETLYG---------------------VKEFENNFYKLEKVNLFSRKDLRKFIL-----EDLNEES
M. ovipneumoniae         -------------------------QEKYEEAKKHTAIK---FSRKTRTI----------------------------LNGGLSDETLYG----------------FKYDEKEDK---YFKIIKKKLVTSK---------------NEEL
M. canis                -------------------------EEYLIDLD--DEVF---FSRKTKRK----------------------------TNRQLYNETIYG----------------IATKTDEDGITNYYKKEKFSILDDKDIYLRLL-----RE-REKF
consensus/95%                                ................h.   aSR.....                            .N..L..-ThYu................................c...h...........     .......
consensus/85%                                ................h.   aSR.....                            .N..L..-ThYu................................c...h...........     .......

E. rectale              ----------------------------------EKNVK---YWHYVMRK----------------------------SNRGLCNQTI----------------------RGTREYDGKQYKINKLDIRTKEGIKVFA-----KLAFSKK
E. faecalis             -----------------------------KPVNLHPRIK---FSHQVDRK----------------------------ANRKLSDATIYS-----------VREKTEVKTLKSGKQKITTDEYTIGKIKDIYTLDGWE------AFKKKQ
S. thermophilus         ----------------------------------EDSIL---FSYQVDSK----------------------------FNRKISDATI----------------YATRQAKVGKDKADETYVLGKIKDIYTQDGYD--------AFMKIY
S. lugdunensis          --------------------------------KDFRNFK---YSHRVDKK----------------------------PNRQLINDTL--------------------------YSTRKKDNSTYIVQTIKDIYAKDN-----TTLKKQF
E. dolichum             -----------------------------------KVRK---FSHKIDTK----------------------------PNRSVADETI------------------------------YSTRVIDGKEKVVKKYKDIY-----DPKFTAL
R. albus                ----------------YRENVASLYKGDPKFASSLSMPV---ISLKPDHK----------------------------YRGTITGEEA---------------------------IRVKEIDGKLIKLKRKSISEITA-----ESINSIY
R. inulinivorans        FLLTHADVQRELDYPGWMYGEEESPIEEGRYINYIRPLF---VSRMPNHK----------------------------VTGSAHDATI-------------------------------------RSARDYETRGVVI----TKVPLTDL
L. coryniformis         -------------NPVEMMKSRSWNTYTPIEIAKLKPVF---VVRLANHK----------------------------ISGPAHLDTI---------------------RSAKLFDEKGIVLSRVSITKLKINKKGQV-----ATGDGIY
I. polytropus           EFYN---L--LSDKRYLGWFNYEE-----GFIEKLRPVF---VSRMPNRG----------------------------VKGQAHQETI-------------------------------------RSSKKISNQIAVS-----KKPLNSI
consensus/95%           .......................................   h.......                            ....h....h..           .....................................    ........
consensus/85%           .......................................   h......K                            ....h...Th..           ...........................c.........    ........

Bacteroides sp.20       ----------------------ITVSFKQNLRVINKTTNH--YQHYENGKKIVSNQSKGDS---------------WAIRKSMHKETV----------------HGEVNLRMIKTVSFNEALKKPQAIVEMDLKKKIL-----AMLELGY
I. album                -----CVVSYKESKPILSDPFNKYLKWEYKNGKWQKVFA---IQIKNDRW--------------------------KAVRRSMFKEPI-GTVWIKKIKEVSLKEAIKIQAIWEEVKNDP-VRKKKEKYIYDDYAQKVI---AKIVQELGL
B. fragilis             ----------------TLISIKAKNKVITGNINKTRKKGG--VNKNMQQT----------------------------PRGQLHLETI---------YGSGKQYLTKEEKVNASFDMRKIGTVSKSAYRDALLKRLYENDNDPKKAFAGK
N. salsuginis           ------------------------------YRDMFDTIRS--ISYWVDKK---------------------------PLSASSHKETV------------------YSSRHEVPTLRKNILEAFDSLNVIKDRHKLTT-----EEFMKRY
E. minutum              -----------------------------SVKTALKNAF---ISVKHDHS----------------------------DNGELVKGTM----------------------YKIFYSERGYTLTTYKKLSALKLTDPQK-----KKTPKDF
S. globus               ---------------------QRELFRKKILEFMDLHAF---VSMKTDND----------------------------ANGALLKDTV--------------------------YSILGADTQGEDLVFVVKKKIKDI-----GVKIGDY
M. trichosporium        ----------------------------------VEKVF---VARAERRR----------------------------ARGKAHDATI-------------------------------RHIAVREGEQRVYERRKVA-----ELKLADL
F. branchiophilum       --------------------------------IEEEILV---SHYTPDNVK-------KQA---------------KKIVRVRGKKQFVAEVERDVNGKA---VPKKASGKTIYKLDGEGKKLPRLQQGDTIRGSLHQ-----DSIYGAI
Prevotella C561         NVSEIGTF------------------------INDNIIV---NHHIKNQA--------------------------LTPVRRRLRKKGY------------------VGGVDN----------PRWQTGDALRGEIHK-----ASYYGAI
P. timonensis           --------------------------------IYKNLLV---VHDTPNNM--------------------------PKHTKKYVQT---------------------SIGK-------------VLAQGDTARGSLHL-----DTYYGAI
P. timonensis           ---------------------------------YKNLLV---VHDTPNNM--------------------------PKHTKKYVQTSI----------------------------------GKVLAQGDTARGSLHL-----DTYYGAI
A. paucivorans          -----------TREEILLRLDQGGVGYDEAFRNGLRPVF---VSRAPSRR----------------------------LRGKAHMETL----------------------------------RSPKWKDHPEGPRTAS-----RIPLKDL
P. marinum              ----------------------------------IDAII---VSHRPRKK----------------------------SQGALHNDTA-------------------------------YGIVEHAENGASTVVHRVP-----ITSLGKQ
A. denitrificans        ----------------------------------VRHIW---VSHRPDHG----------------------------FEGAMMEETS-------------------------------------------YGIRKDG-----SIKQRRK
R. syzygii              ----------------------------------VARCV---VSHKPDHG----------------------------PEGGLHNDTA-------------------YGIVAGPFEDGRYRVRHRVSLFDLKPGDLSN-----VRCDAPL
D. shibae               ----------------------------------LDRII---VSHRADHGRIDHAARKQGR---------------DSTAGQLHQETA---------------------------------YSIVDDIHVASRTDLLS-----LKPAQLL
R. rubrum               ----------------------------------WDHIH---PSIRPAHR----------------------------DGGSLHAATV---------------------------FGVRNRPDARVLVQRKPVEKLFL-----DANAKPL
Azospirillum B510       ----------------------------------ARTIV---VSHRPEHG----------------------------IGGALHKETA-----------------------------------YGPVDPPEEGFNLVV-----RKPIDGL
C. salinarum            NC--------------------------------VILSR---PRRPSETG----------------------------TPGALHNATH-------------YGRREITVDGRTDTVVTQRMNARDLVALLDNAKIVPAARLDAAAPGDTI
N. hamburgensis         ----------------------------------LEKMV---VSHKPDHG----------------------------IEGKLHEDSA----------------------------YGFVKPLDATGLKEEEAGNLVY-----RKAIESL
Bradyrhizobium BTAi1    ----------------------------------LKAML---VSHKPDHG----------------------------LQARLHEDTA---------------------------------YGTVEHPETEDGANLVY-----RKTFVDI
W. succinogenes         -----------------------ENTETVKEGVEVKRLL---ISRPPRAR----------------------------VTGQAHEQTA--------------------------------------------------------KPYPRI
C. jejuni               ----------------------------------IDEIF---VSKPERKK----------------------------PSGALHEETF--------------------------------------------RKEEEF-----YQSYGG-
Jnet 218563121          ----------------------------------H--EE---EEE-----------------------------------HHHHHHHH--------------------------------------------------------------
H. mustelae             ----------------------------------IQNII---PSHHVSHK----------------------------VTGELHQETV--------------------------------------------RTKEFY-----YQAFGG-
A. ebreus               -----------DDLAALREDMQRLGSYTTEDLGRLRTLF---VSRAPQRR----------------------------SGGAVHKETI--------------------------YAQPESLKQQGGVIEKILLTSLKL-----QDFDKLL
B. smithii              -------------FPQESIEAFALGNYDRKKLESLRPVF---VSRMPKRS----------------------------VTGAAHQETL------------------------RRCVGIDEQSGKIQTAVKTKLSDIK------LDKDGHF
C. cellulolyticum       ----------------------------------LAGVF---ISRAPRRK----------------------------ITGPAHDETI--------------------------------RSPKHFNKGLTSVKIPLT-----TVTLEKL
HF0070_07E19            ----------------------------------EEEIF---ITRQPKKV----------------------------SGGIQTKDTL----------------------------------RKHRSKP-DRQRVALT-----KVKLADL
P. lavamentivorans      ----------------------------------VSEIV---VSHRVRKK----------------------------VSGPLHKETT------------------------------YGDTGTDIKTKSGTYRQFVT-----RKKIESL
N. meningitidis         PEFEEADTPEKL----RTLLAEKLSSRPEAVHEYVTPLF---VSRAPNRK----------------------------MSGQGHMETV-------------------------------------KSAKRLDEGVSVL-----RVPLTQL
P. multocida            PD---------------TVLKEMLPDRPQANHQFVQPLF---VSRAPTRK----------------------------MSGQGHMETI-------------------------------------KSAKRLAEGISVL-----RIPLTQL
A. muciniphila          -------------------------------IPQVRPVAN--QRHYVLND-----------------------------DGRMMLRDL----------------------------SASLKENIREQLMEQRVIQHVP-----ADMGGAL
Actinomyces F0310       ------------------------------LIDADKIPV---TENLRLR-----------------------------NSGKLHADQP----------------ESLKKARRGSKRPRPQRYVLGDALPADVINRVTD-----PGLWTAL
A. cellulolyticus       --------------------------------TAARDSIA--VAAPLRLR----------------------------PTGALHEETL----------------------------RAFSEHTVGAAWKGAELRRIVE-----PEVYAAF
B. longum               --------------------------------LDNDRIAV--MQSQRYVL----------------------------GNSIAHDATIHPL-----------------------EKVPLGSAMSADLIRRASTPALWCA----LTRLPDY
A. naeslundii           --------------------------------LAEDKVYV--TQNIRLRL----------------------------SDGNAHTVNPSKL-----------------------VSHRLGDGLTVQQIDRACTPALWCA----LTREKDF

                        --------------------------------HH----EE--EE----------------------------------------------------------------------EEEEHHH---HHHHH----HHHHHH----HH-----
domains                 rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr|ppppppppppppppppppppppppppppppiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii*W*iiiiiiiiiiii
catalytic residues      rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr|ppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppppp*W*ppppppppppp|

A. hesperidum           IRAIGLLTYSE---TEQ---------------IDPIFVSR--MPTRKV-------------------------------TGAAHKETIRSP----------------------RIVKVDDNKGTEIQ--VVVSKVALTE----LKLTKDG
Alcanivorax W11-5       --------------LEA---------------VRSVRISH--KPDRYP-------------------------------DGRFFEATAYGI----------------------AQRLDERSGEKVD---WLVSRKSLTD----LAPEKKS
B. zoohelcum            --------------EQK---------------LKEIIISH--KPKDKLLLQYNKAGDRQIKL-----------------RGQLHEGTLYGI----------------------SQGKEAY-------------RIPLTK----FGGSKFA
Porphyromonas F0450     --------------SDK---------------VAEILISY--RPGQRVVTRGRNIYRKKMADGREVSCVQRGVL---VPRGELMEASFYGK----------------------ILSQGRV---------RIVKRYPLHD----LK-----
B. intestinihominis     --------------REA---------------VDRILVSF--RAGKRAVTPGKRYIRKNRKRISVQSVLIP--------RGALSEESVYGV----------------------IHVWEKDEQGHVIQKQRAVMKYPITS----INRE---
O. laneus               --------------RQS---------------VVPLLVSY--KQNPKTLCKISKTLYKDGKKIHSCGNAV---------RGQLHKETVYGQ----------------------RTAPGATEK-------SYHIRKDIRE----LKT----
Treponema JC4           --------------IEK---------------VKNIVVSF--KPDHGA-------------------------------EGKLSKETLLGK----------------------IKLHGKE---------TFVCRENIVS----LS-----
Rhodovulum PH10         --------------RER---------------VSTITVAV--KPEHGK-------------------------------GGALHEDTSYGL----------------------VPDTDPNAALG-----NLVVRKPIRS----LT-----
consensus/95%           ........................................  ................................  ..........................................................................
consensus/85%           .....................................h..  ................................  ....u.hh..oh..............................................................

#map                    ------------------------------------------------------------------------------------------------------------------------------------------------------
```

```
Jnet 331001027          -HHHHHHHHHHH------------HHHHHH----HHHHH-EEEEEEEEEEEEHHH-------HHHHH------------------------------------------------HHH-HHHHHHHHHH---EEE---EEEEEE---HHH
P. excrementihominis    KNKAFELFSKVAG----SKFTAEEDKAAKI---LEALHFVTVKQDVAATVSDLIKSKKELSKDSIEN------------------------------------------------LAKQKGCLKKVEYSSKEFKFKGSLIIPAAVEWGKV
W. succinogenes         KQTISELSNRYFNMNYKNLVERDKEIVGLLEFIVENCRYYTKKVDVKFA----PKYIHETKYPFYDD------------------------------------------------WRRFDEAWRYLQENQNKTSSKDRFVIDKSS--LNE
L. pneumophila          KTLALEFLHHYFH---KEIVTPDDTTVCHF---INSLRYYTKKESITVKI--------------LKE------------------------------------------------PMPVLSVKFESSKKNVLGSFKHTIALPATKDWERL
S. wadsworthensis       KKPAYEFLAKAAL----QPLSAEEKRLAAL---LDALRYCTSRKSLMSLFMA-ANGKSLKKREDVLK------------------------------------------------PKLFQLKVELKGEKS--FKLNGSLTLPVKQDWLRI
HTCC5015                KVKAFDYLQEKVG---------EKDIAAEL---LNSIHFTQRKTELRAKFSD-DSGKKMKTLDAIRK------------------------------------------------SLKLTVTVNEIGKRKEKCGFSGTIGIPAKSAWENL
F. novicida             TQKLHEYYIENYN---TALGYKKYSKEMEF---LRSLAYRSERVKIKS----------------------------------------------------------------------IDDVKQVLDKDSNF--IIGKITLPFKKEWQRL
consensus/95%           ...h.-h..c.h..........c......h...h....a...c..h.....................                                                .......................h.h.........
consensus/85%           ...h.-h..c.h..........c......h...h....a...c..h.....................                                                .......................h.h.........

A. coleocanis           TR-QPDFTWKDGL--------------------------PANEDRTIIVNGTHYGPLDKVGIF------------------------------------------------------------------------GKAAASLLVRGG---
C. glomerans            SREQFAYFFIYKARNPRK---------------------EQTLFEFAQVPVRLSAQIRQDENA-----------------------------------------------------------------------LERYARELAKD-----
Acidaminococcus D21     DKSTAAYYLLVRFTLED----------------------KKTQHKLMMI---PVEGLYKARID--------------------------------------------------------------------------HDKEFLTDYA---
V. atypica              TKIKNAYSIIVQYTGKK----------------------GEEIKEIVPLPIYLINRNATDIEL-----------------------------------------------------------------IDYVKSVIPKAKDISIKYR---
F. nucleatum            TSLKAAYFIYVEHEKKNK---------------------KVKTFERITRIDSTLIKNEKNLIK-------------------------------------------------------------------YLVSQKKLLNPKIIKKI---
F. alocis               FSANTSYFSLIEF--------------------------EDKKGDRARH---IIGVPIYIANM---------------------------------------------------------------------------LEHSPSAFL---
S. moorei               NKLSSAFLFAVEYKGKKAR--------------------ERSLETVYIK---DVELYLQDPIK-------------------------------------------------------------------------------YCESV---
C. catus                NKAAGTYFMLVKSLDKKGK--------------------EIRTIEFVPL---YLKNQIEINHE----------------------------------------------------------------------SAIQYLAQERGLNS---
T. denticola            NKVSAAYYTLIEYEEK-----------------------GNKIRSLETIPLYLVKDIQKDQDV-----------------------------------------------------------------------LKSYLTDLLGKKE---
P. duerdenii            NKASSAYFALIESDGKAG---------------------RERTLETIPIMVYNQEKYGNTEAV------------------------------------------------------------------------DKYLKDNLELQD---
C. mitsuokai            SGLQYTIVAIEGQKKKGK---------------------KTELVKKISGVPLHLKAASINEKI----------------------------------------------------------------------NYIEEKEGLSDVRI---
S. thermophilus         AGISNSFTVLVKGTIEKG---------------------AKKKITNVLEFQGISILDRINYRK------------------------------------------------------------------------DKLNFLLEKGYK---
S. mutans               DSPIVAYSILVIADIEKG---------------------KSKKLKTVKALVGVTIMEKMTFER------------------------------------------------------------------------DPVAFLERKGYR---
S. pyogenes             DSPTVAYSVLVVAKVEKG---------------------KSKKLKSVKELLGITIMERSSFEK------------------------------------------------------------------------NPIDFLEAKGYK---

                        ----EEEEEEE------------------------------------EEEEEEHHHHHHHHHH-------------------------------------------------------------------------HHHHHHHH------
domains                 tttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
catalytic residues      tttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
domains                 PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
catalytic residues      PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
.
B. bifidum              SSAKPAYAVLIESKG------------------------KTRLVNVTMQ---EYSLLGDRPSD-----------------------------------------------------------------------DELRKVLAKKKSE---
O. kitaharae            SSAKPAYAALIKFKN------------------------KFRLVNVLRQ--WTYSDKNSEDYI-----------------------------------------------------------------------------LEQIRGK---
S. sanguinis            GPMNSALSIIIIAERFNE---------------------KKGKIETVKEFH-DIFIIDYEKFN-----------------------------------------------------------------------NNPFQFLNDTSENGFL
F. fructosus            SNVYSAYMTLVRANG------------------------KNLLIKIPISIA-NQIEVGNLKIN-----------------------------------------------------------------------D----YIVNNPAIKKF
E. yurii                TSISISGYCLVEYVK------------------------KRKTIRSLEAIP-VYLGRKDSLSE-----------------------------------------------------------------------EKLLNYFRYNLNDGGK
L. farciminis           SGDNKSSITIVKIDN------------------------NKIK---PVAIP-IRLI-NDLKDK-----------------------------------------------------------------------KTLQNWLEENVKHKKS
S. pseudintermedius     TPYQTYVVAIKSVNK------------------------KGKEKMEYQMID-HYVFDFYKFQN-----------------------------------------------------------------------GNEKELALYLAQRENK
L. rhamnosus            TQESGSYNALVRVAEA-----------------------DTTAYQVIKISAQNASKIASANLK----------------------------------------------------------------SREKGKQLLNEIVVKQLAKR---
F. magna                KAINGAFFFLVEHTKS-----------------------KKRIRSIELFPLHLLSKFYEDKNT-----------------------------------------------------------------------VLDYAINVLQ------
P. antarcticus          KNHSICLVEFTFMK-------------------------KEKEVQETKFID-LKVIEHHQFLK-----------------------------------------------------------------------EPESQLAKFLAEKETN
consensus/95%           .........h.........                    ........................                                                                .......................
consensus/85%           ......h..h.........                    ........................                                                                .......................
.
Jnet 227501312          ----EEEEEEEE-EE----------------------------EEEEEEE----HHH-----------------------------------------------------------------------------HHHHHHHHH-------
M. mobile               DEDKLKENELTKLHIFNH---------------------DKNLYETLKIIWNEVKIEIKNKNL--------------------------------------------------------------NEKNYFKYFVNKKLQEGKISFN---
M. gallisepticum        FAEKEDILMYRANPWVINLAEQIFNEYTE----------NKKIKSQNVFEKYMLDLTKEFPEK----------------------------------------------------------------FSEFLVKSMLRNKTAIIYDD---
M. synoviae             EKNKKNENGSRKRILT-----------------------EKYIVDEILQILENEEFKDSKSDI---------------------------------------------------------------------NALNKYMDSLPSKFS---
M. ovipneumoniae        KKYFENPFGKKADGKSEYTVLMAQSHLSEFNKLKEIFE----KYNGFSNKTG--------------------------------------------------------------------------------NAFVEYMNDLALKEP---
M. canis                VINQSNPE-------VIDQIIEIIESYGKENNIPSRDEAINIKYTKNKINYN--------------------------------------------------------------------------------LYLKQYMRSLTKSLD---
consensus/95%           ...............................................................                                                              ......................
consensus/85%           ...............................................................                                                              ......................

E. rectale              DSDRERLLVYLN---------------------------DRRTFDDLCKIYEDYSDAANPFVQ-----------------------------------------------------------------------YEKETGDIIRKYS---
E. faecalis             DKLLMKDL-------------------------------DEKTYERLLSIAETTPDFQEVEEK--------------------------------------------------------NGKVKRVKRSPFAVYCEENDIPAIQKYA---
S. thermophilus         KKDKSKFLMYRH---------------------------DPQTFEKVIEPILENYPNKQINEK-------------------------------------------------------------GKEVPCNPFLKYKEEHGYIRKYS---
S. lugdunensis          DKSPEKFLMYQH---------------------------DPRTFEKLEV---IMKQYANEKNP--------------------------------------------------------------------LAKYHEETGEYLTKYS---
E. dolichum             AEDILNNAYQEKYLMALH---------------------DPQTFDQIVKVVNYYFEEMSKSEK----------------------------------------------------YFTKDKKGRIKISGMNPLSLYRDEHGMLKKYS---
R. albus                TD-------------------------------------DKILIDSLKT---IFEQADYKDVG-------------------------------------------------------------------------DYLKKTNQHFF---
R. inulinivorans        KLNKDNEIEGYYDKDS-----------------------DRLLYQALVR---QLLLHGNDGKK-------------------------------------------------------------------------AFAEDFHKPKA---
L. coryniformis         DPENSNNG-------------------------------DKVVYSAIRQALEAHNGSGELAFP-----------------------------------------------------------------------------DGYLEYV---
I. polytropus           KLKDLEKMQGRDT--------------------------DRKLYEALKN---RLEEYDDKPEK-------------------------------------------------------------------------AFAEPFYKPTN---
consensus/95%           ..................                     D...h..h................                                                    ................................
consensus/85%           ..................                     D...a..h................                                                    ...........................h....

Bacteroides sp.20       DTKRIKNYFEENKDTWQDINPS-----------------KIKVYYFTKETKDRYFAVRKPIDTSFDKKKIKESIT------------------------------------------------DTGIQQIMLRHLETKDNDPTLAFS---
I. album                SSSMRKQDDEKLNKFINEAKVSAGVNKNLNTTNKTI---YNLEGRFYEKIKVAEYVLYKAKRMPLNKKEYIEKLSL-------QKMFNDLPNFILEKSILDNYPEILKELESDNKYIIEPHKKNNPVNRLLLEHILEYHNNPKEAFS---
B. fragilis             NSLDKQPIWLDKEQMRKV---------------------PEKVKIVTLEAIYTIRKEISPDLK-------------------------------------------------------VDKVIDVGVRKILIDRLNEYGNDAKKAFSNLD
N. salsuginis           DKEIRQKLWLHRIGNT-----------------------NDESYRAVEERATQIAQILTRYQL------------------------------------------------------------MDAQNDKEIDEKFQQALKELITSP---
E. minutum              LETALLKFKGRESEMKNEKIKSAIEN-------------NKRLFDVIQDNLEKAKKLLEEENE-----------------------------------------------------------------KSKAEGKKEKNINDASIYQ---
S. globus               EEVASAIRGRIT---------------------------DKQPKWYPMEMKDKIEQLQSKNEA----------------------------------------------------------------------ALQKYKESLVQAAAVLE
M. trichosporium        DRVKDAER-------------------------------NARLIEKLRNWIEAGSPKDDPPLS---------------------------------------------------------------------------------------
F. branchiophilum       -KNPLNTDE----------------IKYVIRKDL--------ESI-KGSDVES-IVDEVVKEKIKEAIANKVLL-----------------------------------------------------LSSNAQQKNKLVGTVWM--NEE-
Prevotella C561         --TQFAKDDEGKVLMKEGRPQVNPTIKFVIRRELKYKKSAADSGFASWDDLGKAIVDKELFALMKGQFPAETSF------------------------------------------------------------KDACEQGIYMIKKGKN
P. timonensis           -------ERDGE-------------IRYVVRRPL--------SSFTKPEELEN-IVDETVKRTIKEAIADKNF-------------------------------------------------------------KQAIAEPIYM--NEE-
P. timonensis           ERDGE--------------------IRYVVRRPLSSFTKPEELENIVDETVKRTIKEAIADKN----------------------------------------------------------------------FKQAIAEPIYM------
A. paucivorans          NLEKLERMVGKDR--------------------------DRKLYEALRE---RLAAFGGNGKK-------------------------------------------------------------------------AFVAPFRKPC----
P. marinum              SDIEKVR--------------------------------DPLIKSALLNETAGLSGKSFENAV-----------------------------------------------------------------------------------Q---
A. denitrificans        ADGSAGR---------------------------------EISNLIRIH---EATQPLRHGVS---------------------------------------------------------------------------------------
R. syzygii              QAELEPIFEQD----------------------------DARAREVALTALAERYRQRKVWLE--------------------------------------------------------------------------ELMSVLPIRP---
D. shibae               DEPGRSGQVR-----------------------------DPQLRKALRV---ATGGKTGKDFE----------------------------------------------------------------------NALRYFASKPGPYQ---
R. rubrum               PADKIAEIIDGFA--------------------------SPRMAKRFKALLARYQAAHPEVPP------------------------------------------------------------------------ALAALAVARDPA---
Azospirillum B510       SKDEINSVR------------------------------DPRLRRALIDRLAIRRRDANDPAT--------------------------------------------------------------------------ALAKAAEDLA---
C. salinarum            LKEICTEIADRHDRVVDPEGTHARRWISARLAALVPAHAEAVARDIAELADLDALADADRTPEQEAR------------------------------------------------RSALRQSPYLGRAISAKKADGRARAREQEILTRAL
N. hamburgensis         NENEVDRIR------------------------------DIQLRTIVRD---HVNVEKTKGVA--------------------------------------------------------------------LADALRQLQAPSDDYP---
Bradyrhizobium BTAi1    SEKEIDRIR------------------------------DRRLRDLVRA---HVAGERQQGKT-----------------------------------------------------------------------LKAAVLSFAQRRD---
W. succinogenes         KQVKNKKKWRLAPI-------------------------DEEKFESFKA-----------------------------------------------------------------------------------------------------
C. jejuni               ---------------------------------------KEGVLKAL-----ELGKIRKVNGK---------------------------------------------------------------------------------------
Jnet 218563121          -------------------------------------------------------EEEE-------------------------------------------------------------------------------------------
H. mustelae             ---------------------------------------EEGVKKAL-----KFGKIREINQG---------------------------------------------------------------------------------------
A. ebreus               NPESNDHFVEPHR--------------------------NERLYAAIRQ---RLEQFGGRADK----------------------------------------------------------------------AFGPDNLFHKPDKN---
B. smithii              PMYQKES--------------------------------DPRTYEAIRQ---RLLEHNNDPKK-------------------------------------------------------------------------AFQEPLYKPKK---
C. cellulolyticum       ETMVKNTKGGIS---------------------------DKAVYNVLKN---RLIEHNNKPLK-------------------------------------------------------------------------AFAEKIYKPLK---
HF0070_07E19            ERLVEKDA---S---------------------------NRNLYEHLKQ---CLEESGDQPTK-------------------------------------------------------------------------AFKAPFYMP-----
P. lavamentivorans      SKGELDEIR------------------------------DPRIKEIVAA---HVAGRGGDPKK------------------------------------------------------------------------------AFPPYPC--
N. meningitidis         KLKDLEKMVN-RER-------------------------EPKLYEALKA---RLEAHKDDPAK-------------------------------------------------------------------------AFAEPFYKYDK---
P. multocida            KPNLLENMVN-KER-------------------------EPALYAGLKA---RLAEFNQDPAK-------------------------------------------------------------------------AFATPFYK--Q---
A. muciniphila          LKETMQRVLSV----------------------------DGSGEDAMVSLSKKKDGKKEKNQV--------------------------------------------------------------------------KASKLVGVFP---
Actinomyces F0310       VR-APGFDSQLGL--------------------------PADLNRGLKLRGKRISADFPIDYF------------------------------------------------------------------------PTDSPALAVQGGYV-
A. cellulolyticus       LALTDPGGRFLKVSPS-----------------------EDVLPADENRHIVLSDRVLGPRDR---------------------------------------------------------------------VKLFPDDRGSIRVRG---
B. longum               DEKEGL---------------------------------PEDSHREIRVHDTRYSADDEMGFF------------------------------------------------------------------------ASQAAQIAVQEG---
A. naeslundii           DEKNGL---------------------------------PAREDRAIRVHGHEIKSSDYIQVF-------------------------------------------------------------SKRKKTDSDRDETPFGAIAVRGG---

                        ---------------------------------------------EEEE--EEEE---EEEE------------------------------------------------------------------------------EEEE--E---
domains                 iittttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt
catalytic residues      tttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt

A. hesperidum           EI----------------KDYFRP------------EDDPRLYNTLRERL-VQFGGDAKAAF--------------------------------------------------------------------------KEPVYKISKD----
Alcanivorax W11-5       ID----------------VDKVRA------------NISRIVGEAIRLHISNIFEKRVSKGMT------------------------------------------------------------------------PQQALREPIEF----
B. zoohelcum            T-----------------EKNIQK------------IVSPFLSGFIANHL-KEYNNKKEEAFS------------------------------------------------------------------------AEGIMDLNNKLAQYR
Porphyromonas F0450     ----------------------GE------------VVDPHLRELITTYN-QELKSREKGAP--------------------------------------------------------------------------IPPLCLDKDK----
B. intestinihominis     ------------------MLDKEK------------VVDKRIHRILSGRL-AQYNDNPKEAF--------------------------------------------------------------------------AKPVYIDKEC----
O. laneus               ------------------SKHIGK------------VVDITIRQMLLKHLQENYHIDITQEFN------------------------------------------------------------------------IPSNAFFKEGV-YRI
Treponema JC4           ------------------EKNLDD------------IVDEKIKSKVKDYV-AKHKGQKIEAV--------------------------------------------------------------------------------LSDF----
Rhodovulum PH10         ------------------AGEVDR------------VRDRALRARLGALA-APFRDESGRVRD------------------------------------------------------------------------AKGLAQALEAF----
consensus/95%           ............................................................................        ..................................................................
consensus/85%           ............................................................................        ..................................................................

#map                    ------------------------------------------------------------------------------------------------------------------------------------------------------
```

```
Jnet 331001027          HHHHHHHH--------HHHHHHHHHH-------------------------------------------HHHHHHH-----EEE----E----------------------EEEEEE------EEEEEE---------------------
P. excrementihominis    LWNVFKENTAEELKDENALRKALEAAWPSSFGTR----------------------------------NLHSKAKRVFSLPVVATQSGA----------------------VRIRRKTAFGDFVYQSQDTNN-LYSSFPVKNGKLDWSSP
W. succinogenes         YYQPDKNEYKLD-VDTQPIWDDFCRWYFLDRYKTAND-----------------------------KKSIRIKARKTFSLLAESGVQGK---------------------VFRAKRKIPTG-YAYQALPMDNNVIAG---DYANILLEAN
L. pneumophila          FNHPNFLALKANPAPNPKEFNEFIRKYFLSDNNPNSDIPN---------------------NGHNIKPQKHKAVRKVFSLPVIPGNAGT---------------------MMRIRRKDNKGQPLYQLQTIDDTPSMGIQINEDRL-VKQE
S. wadsworthensis       CDSPELADAFGKPCSADELTSKLARIWKRPVMRD----------------------------------LAHAPVRREFSLPAIDNPSGG----------------------FRIRRTNLFGNELYQVHAINAKKYRGFASAGSNVDWSKG
HTCC5015                LDEPLLETYWGTKMPPQEIWEKVYRKHFPRNIPN----------------------------------QAHRKVRKDFSLPVVDSVSGG----------------------FRVKRKTPNG-YNYQLLAIDGYSAVGFKKEGDNVDFKSP
F. novicida             YREWQNTTIKDD--------YEFLKSFFNVKSIT----------------------------------KLHKKVRKDFSLP-ISTNEGK----------------------FLVKRKTWDNNFIYQILNDSDSRADGTKPFIPAFDISKN
consensus/95%           ......................h...a.............                     .........c..hcc.FSL.......G.                     .h.hcR........YQ..........u.......h.h...
consensus/85%           ......................h...a.............                     .........c..hcc.FSL.......G.                     .h.hcR........YQ..........u.......h.h...

A. coleocanis           --------SVDIGSAIHH------ARIY-------------------RIAGKKP------------TYGMVRVFAPDLLR----------------YRNEDLFNVELPP------------------------QSVSMRYAEPKVREAIR
C. glomerans            --------QGLEFIRIER------SKILKNQLIEIDGDRLC------ITGKEEVRNACEL------AFAQDEMRVIRMLVSEK-----------------PVSRECVIS--------------------------LFNRILLHGDQASRR
Acidaminococcus D21     --------QTTISEILQK----DKQKVINI-----------------MFPMGTR------------HIKLNSMISIDGFYLSI---GGKSSKGKSVLCHAMVPLIVPHK-----------------IECYIKAMESFARKFKENNKLRIV
V. atypica              --------KLCINQLVKV---NGFYYYLGG-----------------KTNDKIYIDN---------AIELVVPHDIATYIKLL---DK--------YDLLRKENKTLKA-----------------SSITTSIYNINTSTVVSLSNKVGI
F. nucleatum            --------YKEQTLIIDS------YPYTFTGVDSNKKVELKNKKQ--LYLEKKYEQ----------ILKNALKFVEDNQGETE-------------ENYKFIYLKKRNN-----------------NEKNETIDAVKERYNIEFN-----
F. alocis               --------EYCEQKGYQN------VRILVE-----------------KIKKNSL------------LIINGYPLRIRGENEVD---TS--------FKRAIQLKLDQKN-----------------YELVRNIEKFLEKYVEKKGNYPID
S. moorei               --------LGLKEPQIIK------PKILMGSLFSINNKKLVVT-----------------------GRSGKQYVCHHIYQLSI-------------NDEDSQYLKNIAK-----------------YLQEEPDGNIERQNILNITSV---
C. catus                --------PEILLSKIKI------DTLFKV----DGFKM--------WLSGRTGNQL---------IFKGANQLILSHQEAAI-------------LKGVVKYVNRKNE---------------------NKDAKLSERDGMTEEKL---
T. denticola            --------FKILVPKIKI------NSLLKI----NGFPC--------HITGKTND-----------SFLLRPAVQFCCSNNEV-------------LYFKKIIRFSEIR-----------------SQREKIGKTISPYEDLSFRSYIKE
P. duerdenii            --------PKILKDKIKI------NSLIKL----DGFLY--------NIKGKTGD-----------SLSIAGSVQLIVNKEEQ-------------KLIKKMDKFLVKK------------------------KDNKDIKVTSFDN----
C. mitsuokai            --------IKDNIPVNQM------IEMDGGEYLLTSPTE--------YVNARQLV-----------LNEKQCALIADIYN--------------------AIYKQDYDN--------------------------LDDILMIQLYI----
S. thermophilus         --------DI--ELIIEL------PKYSLF-----------------ELSDGSRRMLASIL---STNNKRGEIHKGNQIFLSQ-------------KFVKLLYHAKRIS-----------------NT-----INENHRKYVENHK----
S. mutans               --------NVQEENIIKL------PKYSLF-----------------KLENGRKRLLASA----------RELQKGNEIVLPN-------------HLGTLLYHAKNIH-----------------KV-----DEPKHLDYVDKHK----
S. pyogenes             --------EVKKDLIIKL------PKYSLF-----------------ELENGRKRMLASA----------GELQKGNELALPS-------------KYVNFLYLASHYE-----------------KLKGSPEDNEQKQLFVEQHK----

                        ----------------EE---------EEE-----------------EE---EEEEEEE------------EEEE-------H-------------HHHHHHHHHH--------------------------------------------
domains                 tttttttttttttttttttttttttt|cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
catalytic residues      ttttttttttttttttttttttttt|cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
domains                 PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
catalytic residues      PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
.
B. bifidum              --------YAKANILLRHVPKMQLIRYGGGLMVIKSAGELNN------------------------AQQLWLPYEEYCYFDDL---S---------QGKGSLEKDDLKK--------------------------LLDSILGSVQC---L
O. kitaharae            --------YPKAEMVLSHIPYGQLVKKDGALVTISSATELHN------------------------FEQLWLPLADYKLINTL---L---------KTKEDNLVDILHNRLDLPEMTIESAFY-----------KAFDSILSFAFN---R
S. sanguinis            KKNNIN---RVLGFYRIP-KYSLMQKIDGTRMLFESKSN------LHK------------------ATQFKLTKTQNELFFHM---K---------RLLTKSNLMDL-KSKSAIKESQNFIL-------------KHKEEFDNISNQLSA
F. fructosus            EK------ILISKL-----PLGQLVNEDGNLIYLASNEY------RHN------------------AKQLWLSTTDADKIASI---S---------ENSSDEELLEA-----------------------------YDILTSENVKNRFP
E. yurii                DS------VSDIRLCLPFISTNSLVKIDGYLYYLGGKNDD--RIQLYN------------------AYQLKMKKEEVEYIRKI---E---------KAVSMSKFDEIDREKNPVLTEEKNIE-------------LYNKIQDKFENTVFS
L. farciminis           IQ------IIKNNV----PIGQIIYSKKVGLLSLNSDRE------VAN------------------RQQLILPPEHSALLRLL---Q-----------IPDEDLDQI----LAFYDKNILVE-------------ILQELITK-MKKFYP
S. pseudintermedius     DE------VLDAQI-VYSLNKGDLLYINNHPCYFVSRKE------VIN------------------AKQFELTVEQQLSLYNV---M---------NNK-ETNVEKL-------------LI-------------EYDFIAEK-VINEYH
L. rhamnosus            --------RKNWKPSANS------FKIVIP-----------------------RFGMGTLFQN---AKYGLFMVNSDTYYRN--------------YQELWLSRENQKL--------------------------LKKLFSIKYEKTQMN
F. magna                --------LQDPKIIIDK------INYRTEIIIDNFSYLIS-------------------------TKSNDGSITVKPNEQMY-------------WRVDEISNLKKIE------------------------NKYKKDAILTEEDRKIM
P. antarcticus          SP------IIHARI-IRTIPKYQKIWIEHFPYYFISTRE------LHN------------------ARQFEISYELMEKVKQL---S---------ERSSVEELKIV-----------------------------FGLLIDQ-MNDNYP
consensus/95%           ......  ....................................................... ...................   .....................................   ........................
consensus/85%           ......  ..................h.................................... ...................   .....................................   ........................
.
Jnet 227501312          -----------------------------EEEE---EEEEE-------------------------E------EEEE---EEE-------------E--HHHHHHHHHH------------------------HHHHH-----HHHHHH-
M. mobile               ---EWVPILDNDFKIIRK------IRYIKFSSEEKETDEIIFSQSNFL-----KIDQRQNFS----FHNTLYWVQIWVYKNQK---DQ--------YCFISIDARNSKF--------------------------EKDEIKINYEKLKTQ
M. gallisepticum        --------KKNIVHRIKR------LKMLSSELKENKLSNVII-----------RSKNQSGTKLS--YQDTINSLALMIMRSID---PTAKKQ----YIRVPLNTLNLHL----------------------GDHDFDLHNMDAYLKKPKF
M. synoviae             --------EFFSQDFINKCKKENSLILTFDAIKHNDPKKVIKIKNLKFFREDATLKNKQ-------AVHKDSKNQIKSFYESY-------------KCVGFIWLKNKND-----------------LEESIFVPINSRVIHFGDKDKDIF
M. ovipneumoniae        --------T-LKAEIESAKSVEKLLYYNFKPSDQFTYHDNINNKSFKRFYKNIRIIEYKSIPIK--FKILSKHDGGKSFKDTL-------------FSLYSLVYK-VYENG---------------KESYKSIPVTSQMRNFGIDEFDFL
M. canis                --------Q-FSEEFINQMIANKTFVL-YNPTKNTTR----KIK-FLRLVNDVKINDIRKNQVI--NKFNGKNNEPKAFYENI-------------NSLGAIVFK----NS---------------ANNFKTLSINTQIAIFGDKNWDIE
consensus/95%              ..........c..........h.h.....c...............................  ............h....   ......    .....h.........               ........................
consensus/85%              ..........c..........h.h.....c...............................  ............h....   ......    .....h.........               ........................

E. rectale              --------KKHNGPRIDK------LKYKDG----EVGACIDISHKY-------GFEKGSKKV----ILESLVPYRMDVYYKEE---NHS-------YYLVGVKQSDIKF-----------------EKGRNVIDEEAYARILVNEKMIQP
E. faecalis             --------KKNNGPLIRS------LKYYDG----KLNKHINITKDSQGRPVEKTKNGRKV------TLQSLKPYRYDIYQDLE---TKA-------YYTVQLYYSDLRF--------------------------VEGKYGITEKEYMKK
S. thermophilus         --------KKGNGPEIKS------LKYYDS---K-LGNHIDI-----------TPKDSNNKV----VLQSVSPWRADVYFNKT---TGK-------YEILGLKYADLQF------------------------EKGTGTYKISQEKYNDI
S. lugdunensis          --------KKNNGPIVKS------LKYIGN----KLGSHLDVTHQF-------KSSTKKL------VKLSIKPYRFDVYLTDK---G---------YKFITISYLDVLK--------------------------KDNYYYIPEQKYDKL
E. dolichum             --------KKGDGPAITQ------MKYFDG----VLGNHIDISAHY-------QVRDKKV------VLQQISPYRTDFYYSKE---NG--------YKFVTIRYKDVRW------------------------SEKKKKYVIDQQDYAMK
R. albus                --------TTSSGKRVNK------VTVIEK----VPSRWLRK-----------EIDDNNFS-----LLNDSSYYCIELYKDSK---GD--------NNLQGIAMSDIVH------------------------DRKTKKLYLKPDFNYP-
R. inulinivorans        --------DGTEGPVVRK------VKIEKK----QTSGV--------------MVRGG------TGIAANGEMVRIDVFRENG---K---------YYFVPVYTADVVR------------------------KVLPNRAATHTKP---Y
L. coryniformis         --------DHGTKKLVRK------VRVAKKVSLPVRLKNKA-------------------------AADNGSMVRIDVFNTGK---K---------FVFVPIYIKDTVE------------------------QVLPNKAIARGKS----
I. polytropus           --------SGKRGPLVRG------IKVEEK----QNVGV--------------YVNGG--------QASNGSMVRIDVFRKNG---K---------FYTVPIYVHQTLL------------------------KELPNRAINGKPY----
consensus/95%                   .......h..      h.h..................................  .........h..-ha....   ...       .....h.......                 ........................
consensus/85%                   ....G..h..      hch.c................................  ...h.....hRhDha....   ...       a..h.h...D...                 ...........h............

Bacteroides sp.20       --------PDGIDEMNRN------ILILNKGKKHQPIYKVRVYEKAEKFTVGQKGNKRTKFV----EAAKGTNLFFAIYETEEIDKDTKKVIRKRSYSTIPLNVVIERQ-----------------KQGLSSAPEDENGNLPKYILSPND
I. album                --------TEGLEKLNKK------AI----NKIGKPIKYITRLDG----DINEEEIFRGAVF----ETDKGSNVYFVMYE----NNQTKDREFLKPNPSISVLKAIEHK-----------------NKIDFFAP--NRLGFSRIILSPGD
B. fragilis             KNPIWL--NKEKGISIKR------VTISGISNAQSLHVKKDKDGK--------PILDENGRNIPVDFVNTGNNHHVAVYYRPV---IDKRGQLVVDEAGNPKYELEEVV-----------------VSFFEAVTRANLGLPIIDKDYKTT
N. salsuginis           --------IEVTGKLLRK------MRFVYD----KLNAM--------------QIDRGL-------VETDKNMLGIHISKGPN---EK--------LIFRRMDVNNAHE------------------------LQKERSGILCYLN----
E. minutum              --------KAISLSGDKY------VQLSKK----EPGKFFAISK---------PTPTTTGY-----GYDTGDSLCVDLYYDNK---GK--------LCGEIIRKIDAQQ-----------------------------KNPLKYK-----
S. globus               ESNRKLIESGKKPIQLSE------KTISKKALELVGGYYYLISNNKRTKTFVVKEPSNEVKGF---AFDTGSNLCLDFYHDAQ---GK--------LCGEIIRKIQAMN------------------------PSYKPAYM---------
M. trichosporium        ----------PKGDPIFK------VRLVTKSKVNIALDT--------------GNPKRPG------TVDRGEMARVDVFRKAS---KKGKYE----YYLVPIYPHDIAT--------------------------MKTPPIRAVQAYKPE
F. branchiophilum       -----------KRIAINK------VRIYANSVKNPLHIKEHSLL---------SKSKHVHKQKV--YGQNDENYAMAIYE---LDGK---------------RDFELIN------------------------IFNLAKLIKQGQGFYPL
Prevotella C561         GMPD------IKLHHIRH------VRC--EAPQSGLKIKEQT-----------YKSEKEYKRYF--YAAVGDLYAMCCYT----NGK--------------IREFRIYS------------------------LYDVSCHRKSDIEDIP-
P. timonensis           -----------KGILIKK------VRCFAKSVKQPINIRQHR-----------DLSKKEYKQQY--HVMNENNYLLAIYEGL-VKNKV-------------VREFEIVS------------------------YIEAAKYYKRSQDRNIF
P. timonensis           --------NEEKGILIKK------VRCFAKSVKQPINIRQHR-----------DLSKKEYKQQY--HVMNENNYLLAIYEGLV--------------KNKVVREFEIVS------------------------YIEAAKYYKRSQDRNIF
A. paucivorans          --------RSGEGPLVRS------LRIFDSGYSGVE-----------------LRDGGEVYA----VADHESMVRVDVYAKKN---R---------FYLVPVYVADVAR------------------------GIVKNRAIVAHKS---E
P. marinum              --------KWCADNSIKS------LRIVETVSII-------------------PITDKEGVAYK--GYKGDGNAYMDIYQDPT---SSK-------WKGEIVSRFDANQ----------------------------KGFIPSWQSQ---
A. denitrificans        --------ADGQPLAYK-------------------------------------------------GYVGGSNYCIEITVNDK---GK--------WEGEVISTFRAYG--------------------------VVRAGGMGRLRNPHE
R. syzygii              --------RGEDGKTLPD----------------SAPYK---------------------------AYKGDSNYCYELFINER---GR--------WDGELISTFRANQ----------------------AAYRRFRNDPARFRRYT--
D. shibae               --------------AIRR------VRIIKPLQAQARVPV--------------PAQDPIK------AYQGGSNHLFEIWRLPD---GE--------IEAQVITSFEAHT------------------------LEGEKRPHPAAKRL---
R. rubrum               --------FGPRGMTANT------VIAGRSDGDGEDAGLIT------------PFRANPKAAVRTMGNAVYEVWEIQVKGRPR-------------WTHRVLTRFDRTQ-------------------------------PAPPPP----
Azospirillum B510       --------AQPASRGIRR------VRVLKKESNPIRVEHGG------------NPSGPRSGGPFHKLLLAGEVHHVDVALRAD---GRR-------WVGHWVTLFEAHG------------------------GRGADGAAAPPRL----
C. salinarum            LDPHWGPRGLRHL-IMREARAPSLVRIRANKTDA-FGRPVPDAAVWVKTDGNAVSQLWRLTSVVTDDGRRIPLPKPIEKRIEISNLEYARLNGLDEGAGVTGNNAPPRPLRQDIDRLTPLWRDHGTAPGGYLGTAVGELEDKARSALRGK
N. hamburgensis         ----------QFKHGLRH------VRILKK----EKGDYLV------------PIANRASGVAYK-AYSAGENFCVEVFETAG---GK--------WDGEAVRRFDANK------------------------KNAGPKIAHAPQWRDA-
Bradyrhizobium BTAi1    --------IAGHPNGIRH------VRLTKS----IKPDYLV------------PIRDKAGRIYK--SYNAGENAFVDILQAES---GR--------WIARATTVFQANQ------------------------ANESHDAPAAQPI----
W. succinogenes         --------DRVASANQKN------------------------------------------------FYETSTIPRVDVYHKKG---K---------FHLVPIYLHEMVL------------------------NELPNLSLGTNPE----
C. jejuni               ------------------------------------------------------------------IVKNGDMFRVDIFKHKK---TNK-------FYAVPIYTMDFAL------------------------KVLPNKAVARSKK-GEI
Jnet 218563121          ----------------------------------------------------------------------EEEEEEEEE---------E-------EEEEE----HHH-------------------------------EEEE-------
H. mustelae             ------------------------------------------------------------------IVDNGAMVRVDIFKSKD---KGK-------FYAVPIYTYDFAI------------------------GKLPNKAIVQGKKNGII
A. ebreus               --------NQPTGPVVRS------IKLVRG----KQTGI--------------PIRGG--------LAKNDSMLRVDIFTKAG---K---------FHLVPVYVHHRVT-------------------------GLPNRAIVAFKD---E
B. smithii              --------NGEPGPVIRT------VKIIDT----KNKVV--------------HLDGSKT------VAYNSNIVRTDVFEKDG---K---------YYCVPVYTMDIMK------------------------GTLPNKAIEANKP---Y
C. cellulolyticum       --------NGTNGAIIRS------IRVETP----SYTGV--------------FRNEGKG------ISDNSLMVRVDVFKKKD---K---------YYLVPIYVAHMIK------------------------KELPSKAIVPLKP---E
HF0070_07E19            -----------SGPEAKQ-------RPILSKVTLLREKPEPP---------KQLTELSGG-----RRYDSMAQGRLDIYRYKP---GGKRKD---EYRVVLQRMIDLMR------------------------GEENVHVFQKGVP---Y
P. lavamentivorans      --------VSPGGPEIRK------VRLTSK----QQLNL--------------MAQTGNG------YADLGSNHHIAIYRLPD---GK--------ADFEIVSLFDASR------------------------RLAQRNPIVQRTR----
N. meningitidis         --------AGNRTQQVKA------VRVEQV----QKTGV--------------WVRNHNG------IADNATMVRVDVFEKGD---K---------YYLVPIYSWQVAK------------------------GILPDRAVVQGKD---E
P. multocida            --------GG---QQVKA------IRVEQV----QKSGV--------------LVRENNG------VADNASIVRTDVFIKNN---K---------FFLVPIYTWQVAK------------------------GILPNKAIVAHKN---E
A. muciniphila          --------EGPSK-----------LKALKA------------------------------------AIEIDGNYGVALDPKPV-----------------VIRHIKVFK-----------------------------------------
Actinomyces F0310       ------------GLEFHH------ARLYRIIGPK-----------------------EKV------KYALLRVCAIDLCG----------------IDCDDLFEVELKP------------------------SSISMRTADAKLKEAMG
A. cellulolyticus       --------GAAYIASFHH------ARVFRWGSSHSPSFALL------------RVSLADL------AVAGLLRDGVDVFTAELPPWTPA-------WRYASIALVKAVE----------------------SGDAKQVGWLVPGDELDFG
B. longum               --------SADIGSAIHH------ARVYRC-----------------------WKTNAKGVRKY--FYGMIRVFQTDLLRACH----DDLFTVPLPPQSISMRYGEPRV-----------------VQALQSGNAQYLGSLVVGDEIEMD
A. naeslundii           --------FVEIGPSIHH------ARIYRV----------------------------EG-KKP--VYAMLRVFTHDLLSQRH----GDLFSAVIPPQSISMRCAEPKL-----------------RKAITTGNATYLGWVVVGDELEIN

                        --------EEE----EEE------EEEEEE------------------------------------EEEEEEEEHHHHHHH------------------HHHH---HHH-----------------HHHHH---EEEEEEE----EEE--
domains                 tttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt|ccccc
catalytic residues      tttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttttt|ccccc

A. hesperidum           ---------GSVRTPVRK------VKIQEKLTL-----------------------------GV--PVHGGRGIAENG-GMVR----IDVF---AKGGK------YYFVPIY--------------VADVLKRELPNRLATAHKP-----
Alcanivorax W11-5       -----------QGNILRK------VRCFYSKADDCV--------------------------RI--EHSSRRGHHYKMLLNDG----FAYMEVPCKEGI------LYGVPNL--------------V-------RPSEAVGIKRA-----
B. zoohelcum            NEK----GELKPHTPIST------VKIYYKDPSKNKKKKDEEDLSLQKL------------DRE--KAFNEKLYVKTG-DNYL----FAVLEGEIKTKKTSQIKRLYDIISF--------------FDATNFLKEEFRNAPDKKT-----
Porphyromonas F0450     ------------KQEVRS------VRCYAKTLSLDKAIPMC---------------------FD--EKGEPTAFVKSA-SNHH----LALY--RTPKGK------LVESIVT--------------FWDAVDRARYGIPLVITHPREVME
B. intestinihominis     ------------RIPIRT------VRCFAKPAINTLVPLK----------------------KD--DKGNPVAWVNPG-NNHH----VAIY--RDEDGK------YKERTVT--------------FWEAVDRCRVGIPAIVTQPDTIWD
O. laneus               FLP----NKHGEPVPIKK------IRMKEELGNA----------------------------ER--LKDNINQYVNPR-NNHH----VMIY--QDADGN------LKEEIVS--------------FWSVIERQNQGQPI----------
Treponema JC4           ----------SKENGIKK------VRCVNRVQTPI---------------------------EI--TSGKISRYLSPE-DYFA----AVIWEIPGEKKT---FKAQYIRRNE--------------VEKNSKGLNVVKPAVLENG-----
Rhodovulum PH10         ----------GAENGIRR------VRILKPDASVVTIA------------------------DR--RTGVPYRAVAPG-ENHH----VDIV--QMRDGS------WRGFAAS--------------VFEVNRP-----GW----------
consensus/95%           ......................................................................................................................................................
consensus/85%           .........................................................................h............................................................................

#map                    ------------------------------------------------------------------------------------------------------------------------------------------------------
```

```
Jnet 331001027           ---------------------HHHHH--------------HHE-----------EEEEE--------EEEEEE--HHHHHHHH------------EE------------------EEHHHHHH---------------------------
P. excrementihominis     IIHPALQNRNLTA--------YGYRFVDHDRSIS-MS--EFREVYNKDD-----LMRIELAQGTSSRRYLRVEMPGEKFLAWFGE--------NSISLGSSFKFSVSEVFD--NKIYTENAEFTKFLPKPREDNKHN-------------
W. succinogenes          SKTLSLVPKSGISIEKQLDKKLDVIKKTDVRGLA-IDNNSFFNADFDTHGIRLIVENTSVKVGNFPISAIDKSAKRMIFRALFEKEKGKRKKKTTISFKESGPVQDYLKVF-LKKI----------------------------------
L. pneumophila           VLMDAYKTRNLST--------IDGINNSEGQAYATFD--NWLTLPVSTF--KPEIIKLEMKPHSKTRRYIRITQSLADFIKTIDE---------ALMIKPSDSIDDPLNMP--NEIVCKNKLFGNELKPRD-------------------
S. wadsworthensis        ILFNELQHENLTE--------CGGRFITSADVTP-MS--EWRKVVAEDN------LSIWIAPGTEGRRYVRVETTFIQASHWFEQSVENWAITSPLSLPASFKVDKPAEFQ--KAV---GTELSELLGQPR-------------------
HTCC5015                 ALVPQIAESKSVT-----PISSELVHLDKNEIVY-FD--EWRKIDISDSDLKQFVSSLELAPGSQNRFYIRFTVDEDQFERHF---------------KSALRVNGIQDLDTVNKTFDWNREIPSLLIPPR-------------------
F. novicida              EIVEAIIDSFTSKNIFWLPKNIELQKVDNKNIFA-IDTSKWFEVETPSDLRDIGIATIQYKIDNNSRPKVRVKLDYVIDDDS--------------------KINYFMNHSLLKSRY------------PD-------------------
consensus/95%            .....h.............................h....a..h...............h.........hc................................h.....h...........................
consensus/85%            .....h.............................h....a..h...............h.........hc................................h.....h...........................

A. coleocanis            ----------------------EGKAEYLGWLVVGDELLLDLSSETSGQIA----ELQQDFPGTTHWTVAGFFSPSR--------------LRLRPVYLAQEGLGEDVSEGSKSIIAG-------------------QGWRPAVNKVFGS
C. glomerans             L---------------------SKQLKLALLSEAFSEASDNVQRN----------VVLGLIAIFNG-------STNM--------------VNLSDIGGSKFAGNVRIKYKKELASPK-------------------------------V
Acidaminococcus D21      EKFDKITV--------------EDNLNL---YELFLQKLQHNPYNKFFS------TQFDVLTNGRSTFTKLSPEEQVQTL-----------LNILSIFKTCRSSGCDLKSINGSAQAA----------------RIMISADLTGLSKKYS
V. atypica               DVFDYF----------------MSKLRTPLYMKMKGNKVDELSST----------GRSKFIKMTL--------EEQS--------------IYLLEVLNLLTNSKTTFDVKPLGITGS--------------------RSTIGVKIHNLD
F. nucleatum             ----------------------EMYDKFL--EKLSSKDYKNYINN----------KLYTNFLNSKEKFKKLKLWEKS--------------LILREFLKIFNKNTYGKYEIKDSQTKE------------KLFSFPEDTGRIRLGQSSLG
F. alocis                ENRDHITH--------------EKMNQL---YEVLLSKMKKFNKKGMADPS----DRIEKSKPKFIKLEDLIDKINV--------------INKMLNLLRCDNDTKADLSLIELPKNA-------------------GSFVVKKNTIGKS
S. moorei                ----------------------NNIKLFDVLCTKFNSNTYEIILN----------SLKNDVNEGREKFSELDILEQC--------------NILLQLLKAFKCNRESSNLEKLNNKKQ-------------------AGVIVIPHLFTKC
C. catus                 ----------------------LQLYDTF--LDKLSNTVYSIRLS----------AQIKTLTEKRAKFIGLSNEDQC--------------IVLNEILHMFQCQSGSANLKLIGGPGS-------------------AGILVMNNNITAC
T. denticola             NLWKKTKNDEIGE---------KEFYDLLQKKNLEIYDMLLTKHKDTIYKKRPNSATIDILVKGKEKFKSLIIENQF--------------EVILEILKLFSATRNVSDLQHIGGSKY-------------------SGVAKIGNKISSL
P. duerdenii             ----------------------IKEEELIKLYKTLSDKLNNGIYSNKRN------NQAKNISEALDKFKEISIEEKI--------------DVLNQIILLFQSYNNGCNLKSIGLSAK-------------------TGVVFIP-KKLNY
C. mitsuokai             ----------------------ELTNKMKVLYPAYRGIAEKFE------------SMNENYVVISKEEKANIIKQML--------------IVMHRGPQNGNIVYDDFKISDRIGRLK--------------------------TKNHNL
S. thermophilus          ----------------------KEFEELFYYILEFNENYVGAKKNG---------KLLNSAFQSWQNHSIDELCSSFIGPTGSERK------GLFELTSRGSAADFEFLGVKIPRYRDY-------------------------TPSSLL
S. mutans                ----------------------DEFKELLDVVSNFSKKYTLAEGNL---------EKIKELYAQNNGEDLKELASSF--------------INLLTFTAIGAPATFKFFDKNIDRKRY--------------------------TSTTEI
S. pyogenes              ----------------------HYLDEIIEQISEFSKRVILADANL---------DKVLSAYNKHRDKPIREQAENI--------------IHLFTLTNLGAPAAFKYFDTTIDRKRY--------------------------TSTKEV

                         ----------------------HHHHHHHHHHHHHHHHH---HHHH---------HHHHHHHHH-----HHHHHHHH--------------HH-HHHH-------EEE--EEE-------------------------------------
domains                  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
catalytic residues       cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
domains                  PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
catalytic residues       PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP
.
B. bifidum               YPWHRFTE--------------EELADLHVAFDKLPEDEKKNVIT----------GIVSALHADAKTANLSIVGMTG-----------------------------------------------------------SWRRMNNKSGYTFS
O. kitaharae             YALHQNAL--------------VKLQAHRDDFNALNYEDKQQTLE----------RILDALHASPASSDLKKINLSS-----------------------------------------------------------GFGRLFSPSHFTLA
S. sanguinis             -------FSQKMLGN------TTSLKNLIKGYNERKIKEIDIRDETIKYFY----DNFIKMFSFVKSGAPKDINDFFDNKCTV-------------------------------------------------------ARMRPKPDKKL-
F. fructosus             -------FFKKD---------IDKLSQVRDEFLDS--------DKRIAVIQ----TILRGLQIDAAYQAPVKIISK-KVSDW--------------------------------------------------------HKLQQSGGIKLS
E. yurii                 KRMSLVKYNKKDLSF------GDFLKNKKSKFEEIDL------EKQCKVLY----NIIFNLSNLKE----VDLSDIGGSKST--------------------------------------------------------GKCRCKKNITNY
L. farciminis            -------FYKGE---------REFLIANIENFNQATT------SEKVNSLE----ELITLLHANST-SAHLIFNNI-EKKAFG-------------------------------------------------------RK---THGLTL-
S. pseudintermedius      -------HYLNSKLK------EKRVRTFFSESNQTH-------EDFIKALD----ELFKVVTASATRSD--------KIGSRK-------------------------------------------------------NSMTHRAFLGKG
L. rhamnosus             HDALQVYK--------------AIIDQVEKFFKLYDINQFRAKLS----------DAIERFEKLPINTDGNKIGKTE--------------TLRQILIGLQANGTRSNV-KNLGIKTD-------------------LGLLQVGSGIKLD
F. magna                 E---------------------SYIDKIYQQFKAGKYKNRRTTDT----------IIEKYEIIDLDTLDNKQLYQLL--------------VAFISLSYKTSNNAVDFTVIGLGTECG------------------------KPRITNLP
P. antarcticus           -------IYTKS-----------SIQDRVQKFVDTQLYDF---KSFEIGFE----ELKKAVAANAQ-----------RSDTFG-------------------------------------------------------SRISKKP---KP
consensus/95%            ...............      .................................................................     ............................           ....................
consensus/85%            ...............      .......................................h.........................     ............................           ....................
.
Jnet 227501312           -----------------------H-HHHHHHHH--------HHHH----------HHHHHHHHHH-------HHHHH--------------HHHHHH-H--------EEEE------E------------------------EEEE----
M. mobile                KEKLQI----------------INEEPI---LKINKGDLFENE------------EKELFYIVGRDE------KPQK--------------LEIKYILGKKIKDQKQIQKPVKKYFPN--------------------------------
M. gallisepticum         VKYLKAN---------------EIGDEYKPWRVLTSGTLLIHKKD----------KKLMYISSFQN-------LNDV--------------IEIKNLIETEYKENDDSDSKKKKKANR-------------------FLMTLSTILNDYI
M. synoviae              DFDS------------------YNKEKLLNEINLKRPENKKFNSI----------NEIEFVKFVKPGALLLNFENQQ--------------IYYISTLESSSLRAKIKLLNKMDKGKA----------------------VSMKKITNPD
M. ovipneumoniae          DEN------------------LYNKEKL---------DIYKSDFAKPIP------VNCKPVFVLKKGSILKKKSLDIDDFKETKETEEGN-YYFISTISKRFNRDTAYGLKPLKLSVVKPV-------------------AEPSTNPIFK
M. canis                 DFK------------------TYNMEKI-----------EKYKEIYGID------KTYNFHSFIFPGTILLDKQNKE--------------FYYISSIQTVRDIIEIKFLNKIEFKDENKN----------QDTSKTPKRLMFGIKSIMN
consensus/95%            .......              ....-.h.....................      .....h............................. h.h...h...............c.......          ...................
consensus/85%            .......              ....-.h.....................      .....h............................. h.h...h...............c.......          ...................

E. rectale               GQSRAD----------------LENLGFKFKLSFYKNDIIEY-------------EKDGKIYTERLVSRTMPKQRNY--------------IETKPIDKAKFEKQNLV------------------------------------GLGKTK
E. faecalis              VAEQTK----------------GQVVRFC--FSLQKNDGLEIEWK----------DSQRYDVRFYNFQ-----SANS--------------INFKGLEQEMMPAENQFKQKPYNNGAI--------------------NLNIAKYGKEGK
S. thermophilus          KKKEGV----------------DSDSEFK--FTLYKNDLLLVKDTET--------KEQQLFRFLSRTMPK---QKHY--------------VELKPYDKQKFEGGEALIKVLGNVANS--------------------GQCKKGLGKSNI
S. lugdunensis           KLGKAI----------------DKNAKFI--ASFYKNDLIKLD------------GEIYKIIGVNSD------TRNM--------------IELDLPDIRYKEYCELNNIKGEPRIKK-------------------------TIGKKVN
E. dolichum              KAEKKI----------------DDTYEFQ--FSMHRDELIGITKAEG--------EALIYPDETWHNFNFFFHAGET--------------PEILKFTATNNDKSNKIEVKPIHCYCK--------------------MRLMPTISKKIV
R. albus                 ----------------------DDYYTHV--MYIFPGDYLRIKSTSKKSGEQL--KFEGYFISVKNV------NENS--------------FRFISDNKPCAKDKRVSITKKDIVIKL--------------------AVDLMGKVQGEN
R. inulinivorans         SEWRVM-----------------DDANFV--FSLYSRDLIHV--------------KSKKDIKTNLVNGGLLLQ-----------------KEIFAYYTGADIATASIAGFANDSNFKF----------------------RGLGIQSLE
L. coryniformis          -LWYQI----------------TESDQFC--FSLYPGDMVHIESKT---------GIKPKYSNKENNTSVVPIKNFY--------------GYFDGADIATASILVRAHDSSYTARSI--------------------------GIAGLL
I. polytropus            KDWDLI----------------DGSFEFL--YSFYPNDLIEIEF-----------GKSKSIKNDNKLTKTEIPEVNL--------------SEVLGYYRGMDTSTGAATIDTQDGKIQ-----------------------MRIGIKTVK
consensus/95%            ......                .....a...h.h...-.h.h...........  ......................              ............................                   ............
consensus/85%            ......                .....F...hoha..Dhh.h...........  ......................              ..h.........................                   ............

Bacteroides sp.20        LVYVPTQ---------------EEINKGEVVMPIDRDRIYKMVDSS---------GITANFIPASTANLIFALPKAT--------------AEIYCNGENCIQNEYGIGSPQSKNQKA------------------------ITGEMVKE
I. album                 LVYVPTNDQYVLIKDNSSNETIINWDDNEFI----SNRIYQVKKFT---------GNSCYFLKNDIASLILSYSASN--------------GVGEFGSQNI--SEYSVDDPPIR---------------------------------IKD
B. fragilis              EGWQFL----------------FSMKQNE--YFVFPNEKTGFNPKEIDLL-----DVENYGLISPNLFRVQKFSLKN--------------YVFRHHLETTIKDTSSILRGITWIDFR--------------------------SSKGLD
N. salsuginis            --------------------------EML--FIFNKKGLIHY-------------GCLRSYLEKGQGSKYIALFNPR--------------FPANPKAQPSKFTSDSKIKQVGIGSAT-------------------GIIKAHLDLDGHV
E. minutum               ----------------------EQGFTLF--ERIYGGDILEVDF-----------DIHSDKNSFRNNTGSAPENRVF--------------IKVGTFTEITNNNIQIWFGNIIKSTGG-------------------QDDSFTINSMQQY
S. globus                ----------------------KQGYSLY--VRLYQGDVCELRASDLTE------AESNLAKTTHVRLPNAKPGRTF--------------VIIITFTEMGSGYQIYFSNLAKSKKGQ--------------------DTSFTLTTIKNY
M. trichosporium         DEWPEM----------------DSSYEFC--WSLVPMTYLQVISSKGE-------IFEGYYRGMNR-------SVGA--------------IQLSAHSNSSDVVQGI-------------------------------------GARTLT
F. branchiophilum        HKKKEIK-------GKIVFVPIEKRNKRDVV--LKRGQQVVFYDKEVENPKDI--SEI--VDFKGRIYIIEGLSIQRIV-RPSGKVDEYGVIMLRYFKEARKAD-DIKQDNFKPDGVFKLG-----------------------ENKPTR
Prevotella C561          EFITDKK---------------GNRLMLDYK--LRTGDMILLYK---DNPAEL--YDLDNVNLSRRLYKINRFESQ-------------SNLVLMTHHLSTSKE---RGRSLGKTVDYQNL-----------------------PESIRS
P. timonensis            SSIVPTH---------------STKYGLPLKTKLLMGQLVLMFE---ENPDEI--QVDNTKDLVKRLYKVVGIEKD-------------GRIKFKYHQEARKEGLPIFSTPYKNNDDY--------------------------APIFRQ
P. timonensis            SSIVPTH---------------STKYGLPLKTKLLMGQLVLMFEENPDEI-----QVDNTKDLVKRLYKVVGIEKDGR-------------IKFKYHQEARKEGLPIFSTPYKNNDDY--------------------------APIFRQ
A. paucivorans           EEWDLV----------------DGSFDFR--FSLFPGDLVEIEKKDG--------AYLGYYKSCHR-------GDGR--------------LLLDRHDRMPRESDCGTFYVS--------------------------------TRKDVL
P. marinum               ----------------------FPTARLI--MRLRINDLLKLQDG----------EIEEIYRVQRL-------SGSK--------------ILMAPHTEANVDARDRDKNDTFKLTSK------------------------SPGKLQSA
A. denitrificans         G---------------------QNGRKLI--MRLVIGDSVRLEVD----------GAERTMRIVKIS-----GSNGQ--------------IFMAPIHEANVDARNTDKQDAFTYTSKYA------------------------GSLQKA
R. syzygii               ----------------------AGGRPLL--MRLCINDYIAVGTA----------AERTIFRVVKM-------SENK--------------ITLAEHFEGGTLKQRDADKDDPFKYLT-----------------------KSPGALRDL
D. shibae                -------------------------------LRVHKGDMVALERDGR--------RVVGHVQKMDIANG----------------------LFIVPHNEANADTRNNDKSDPFKWIQI------------------------GARPAIAS
R. rubrum                ----------------------PENARLV--MRLRRGDLVYWPLESG--------DRLFLVKKMAVDGRLALWPARL--------------ATGKATALYAQLSCPNINLNGDQGYCV-----------------------QSAEGIRKE
Azospirillum B510        ----------------------GDGERFL--MRLHKGDCLKLEHK----------GRVRVMQVVKLEP-----SSNS--------------VVVVEPHQVKTDRSKHVKISC--------------------------------DQLRAR
C. salinarum             AMRQTLTDAGITAEAGWRLDSEGA-------------------------------------------------------------------------------------------------------------------------VCDLE
N. hamburgensis          ----------------------NEGAKLV--MRIHKGDLIRLDHE----------GRARIMVVHRLDA-----AAGR--------------FKLADHNETGNLDKRHATNNDIDPFRW---------------------LMASYNTLKKL
Bradyrhizobium BTAi1     -------------------------------MRVFKGDMLRIDHA----------GAEKFVKIVRLSP-----SNNL--------------LYLVEHHQAGVFQTRHDDPEDSF-----------------------RWLFASFDKLREW
W. succinogenes          ----------------------AMDENFFK-FSIFKDDLISIQTQGTPKKPAK--IIMGYFKNMHGANMVLSSINNSPCEGFTCT-----PVSMDKKHKDKCKLCPEENRIAGRCLQGFLDYWSQ-----EGLRPPRKEFECDQGVKFAL
C. jejuni                KDWILM----------------DENYEFC--FSLYKDSLILIQTKDMQE------PEFVYYNAFTSSTVSLIVSKHD--------------NKFETLSKNQKILFKNANEKEVIAKSI--------------------------GIQNLK
Jnet 218563121           ---------------------------EE--EEE----EEEEEE-----------EEEEEEE------EEEEEEE-----------------------HHHHH---------EEEEE--------------------------------E
H. mustelae              KDWLEM----------------DENYEFC--FSLFKNDCIKIQTK----------EMQEAVLAIYKSTNSAKATIEL--------------EHLSKYALKNEDEEKMFTDTDKEKNKT--------------------MTRESCGIQGLK
A. ebreus                DEWTLI----------------DESFAFL--FSVYPNDYVKVTLKKE--------QQSGYYSGADR-------STGA--------------MNLWAHDRAASVGKDGLIRGI--------------------------------GVKTAL
B. smithii               SEWKEM----------------TEEYTFQ--FSLFPNDLVRIV------------LPREKTIKT--S------TNEE--------------IIIKDIFAYYKTIDSATGGLELISHDR-------------------NFSLRGVGSKTLK
C. cellulolyticum        SQWELI----------------DSTHEFL--FSLYQNDYLVIKTKKG--------ITEGYYRSCHR-------GTGS--------------LSLMPHFANNKNVKIDI------------------------------------GVRTAI
HF0070_07E19             DQGPEI----------------EQNYTFL--FSLYFDDLVEFQRSADSE------VIRGYYRTFNI-------ANGQ--------------LKISTYLEGRQDFDFF-------------------------------------GANRLA
P. lavamentivorans       ----------------------ADGASFV--MSLAAGEAIMIPEG----------SKKGIWIVQGVW------ASGQ--------------VVLERDTDADHSTTTRPMP----------------------------------NPILKD
N. meningitidis          EDWQLI----------------DDSFNFK--FSLHPNDLVEVITKKA--------RMFGYFASCHR-------GTGN--------------INIRIHDLDHKIGKNGILEGI--------------------------------GVKTAL
P. multocida             DEWEEM----------------DEGAKFK--FSLFPNDLVELKTKKE--------YFFGYYIGLDR-------ATGN--------------ISLKEHDGEISKGKDGVYR-V--------------------------------GVKLAL
A. muciniphila           --------------------------RIMALKEQNGGKPVRILKK----------GMLIHLTSSKDP------KHAG--------------VWRIESIQDSKGGVKLDLQRAHCAVPK--------------------------NKTHEC
Actinomyces F0310        ----------------------NGSAKQIGWLVLGDEIQIDPTKFPKQSIG----KFLKECGPVSSWRVSALDTPSK--------------ITLKPRLLSNEPLLKTSRVGGHESDLVVAECVEKIMKK--------TGWVVEINALCQS
A. cellulolyticus        PEGVTTAA--------------GDLSMFLKYFPERHWVVTGFEDD----------KRINLKPAFLSAEQAEVLRTER--------------SDRPDTLTEAGEILAQFFPRCWRATVA--------------------------KVLCHP
B. longum                FSSLDVD---------------GQIGEYLQFFSQFSGGNLAWKHW----------VVDGFFNQTQLRIRPRYLAAEG--------------LAKAFSDDVVPDGVQKIVTKQGWLPPV--------------------------NTASKT
A. naeslundii            VDSF-TK---------------YAIGRFLEDFPNTT-------RW----------RICGYDTNSKLTLKPIVLAAEG--------------LENPSS------AVNEIVELKGWRVAI--------------------------NVLTKV

                         ----------------------HHHHHHHHH------------EE----------EEEEEEE--EEEEEE----HHH-------------------H------HHHHHH----EEEEH--------------------------HHHHH-
domains                  cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc
catalytic residues       cccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccccc

A. hesperidum            ----------YSEWRVVDDSYQF---KFSLYPNDAV-------MIKPSREVD-ITYKDRKEPVGCRIM-----------------YFVSANIASASIS----------LRTHDNSGELEGL-----------------------GIQGLE
Alcanivorax W11-5        -----------------PESGDF---I-RFYKGDTV-------KNIKTGRVY-TIK-----------------------------QILGDGGGKLILTP-----------------------------------------------VTET
B. zoohelcum             ----FDKDLLFRQYFEERNKAKL---LFTLKQGDFVYLPNENEEVILDKESP-LYNQYWGDLKERGKNIYVVQKFSKKQ-----IYFIKHTIADIIKK----------DVEFGSQNCYETV--------------------EGRSIK--E
Porphyromonas F0450      QVLQRGDIPEQVLSLLPPSDWVF---VDSLQQDEMV-------VIGLSDEEL-QRALEAQNYRKISEHLYRVQKMSSSY------YVFRYHLETSVAD---DKNTSGRIPKFHRVQSLKAY-----------------------EER---
B. intestinihominis      NILQRNDISENVLESLPDVKWQF---VLSLQQNEMF-------ILGMNEEDY-RYAMDQQDYALLNKYLYRVQKLSKSD------YSFRYHTETSVED---KYDGKPNLKLSMQMGKLKRV-------------------SIKSLLG--L
O. laneus                -------------YQLPREGRNI---VSILQINDTF-------LIGLKEEEP-EV--YRNDLSTLSKHLYRVQKLSGMY------YTFRHHLASTLNN----------EREEFRIQSLEAW-----------------------KRA---
Treponema JC4            ---------------KPHPAAKQ---VCLLHKDDYL-------EFSDKGKMY-FCRIAGYAATNNKLDIRPV-------------YAVSYCADWINSTN--ETMLTGYWKPTPTQNWVSVN-----------------------VLFDKQ
Rhodovulum PH10          ----------RPEWEVKKLGGKL---VMRLHKGDMV-------ELSDKDGQR-RVKVVQQIEISANRV-----------------RLSPHNDGGKLQD----RHADADDPFRWDLATIPLL-----------------------KDR---
consensus/95%            ................................................................................................................................. ....................
consensus/85%            ................................................................................................................................. ....................

#map                     ------------------------------------------------------------------------------------------------------------------------------------------------------
```

```
Jnet 331001027          ---EE-EEEEE----EEEEEEEE--HHHHHHHHHHHH----
P. excrementihominis    --GTI-FFELVG--PRVIFNYIVGGAASSLKEIFSEAGKER
W. succinogenes         --VKI-QLRTDGSISNIVVRKNAADFTLSFRSEHIQKLLK-
L. pneumophila          --GKM-KIVSTG--KIVTYEFESDSTPQWIQTLYVTQLKKQ
S. wadsworthensis       --SEI-FIENVGNAKHIRFWYIVVSSNKKMNESYNNVSKS-
HTCC5015                --SNL-FLLETG--QKITFEYIANGANAEVKKAYSLRRA--
F. novicida             --KVL-EILK----QSTIIEFESSGFNKTIKEMLGMKLAGI
consensus/95%             ..h .h..........h..........h...h.......
consensus/85%             ..h .h..........h..........h...h.......

A. coleocanis           AMPEVIRRDGLGR-KRRFSYSGLPVSWQG------------
C. glomerans            NVHLI-DQSVTGM-FERRTKIGL------------------
Acidaminococcus D21     DIRLV-EQSASGL-FVSKSQNLLEYL---------------
V. atypica              EFKII-NESITGL-YSNEVTIV-------------------
F. nucleatum            NNKELLEESVTGL-FVKKIKL--------------------
F. alocis               KIILV-NQSVTGL-YENRREL--------------------
S. moorei               SVFKVIHQSITGL-FEKEMDLLK------------------
C. catus                KQISVINQSPTGI-YEKEIDLIKL-----------------
T. denticola            DNCILIYQSITGI-FEKRIDLLKV-----------------
P. duerdenii            KECKLINQSITGL-FENEVDLLNL-----------------
C. mitsuokai            NNIVFISQSPTGI-YTKKYKL--------------------
S. thermophilus         KDATLIHQSVTGL-YETRIDLAKLGEG--------------
S. mutans               LNATLIHQSITGL-YETRIDLNKLGGD--------------
S. pyogenes             LDATLIHQSITGL-YETRIDLSQLGGD--------------

                        ---EEEEE-------EEEE----------------------
domains                 cccccccccccccccccccccccccccccccccccccccc|
catalytic residues      cccccccccccccccccccccccccccccccccccccccc|
domains                 PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP|
catalytic residues      PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP|
.
B. bifidum              DEDEFIFQSPSGL-FEKRVTVGELKRKAKKEVNSKYRTNEK
O. kitaharae            DTDEFIFQSVTGL-FSTQKTVAQLYQETK------------
S. sanguinis            LNATLIHQSITGL-YETRIDLSKLGED--------------
F. fructosus            DNSEMIYQSATGI-FETRVKISDLL----------------
E. yurii                KEFKLIQQSITGL-YSCEKDLMTI-----------------
L. farciminis           NNTDFIYQSVTGL-YETRIHIE-------------------
S. pseudintermedius     KDVKIAYTSISGLKTTKPKSLFKLAESRNEL----------
L. rhamnosus            KDTQIVYQSPSGL-FKRRIPLADL-----------------
F. magna                DNTYLVYKSITGI-YEKRIRIK-------------------
P. antarcticus          EEVAIGYESITGLKYRKPRSVVGTKR---------------
consensus/95%           ....h...S.oGh............................
consensus/85%           ....h...S.oGh.a.....h....................
.
Jnet 227501312          --EEEEE--------EEEEEE--------------------
M. mobile               -WKKV-NLTYMGE-IFKK-----------------------
M. gallisepticum        LLDAKDNFDILGL-SKNRIDEILNSKLGLDKIVK-------
M. synoviae             EYKIIEHVNPLGI-NLNWTKKLENNN---------------
M. ovipneumoniae        EYIPI-HLDELGN-EYPVKIKEHTDDEKLMCTIK-------
M. canis                NYEQV-DISPFGI-NKKIFE---------------------
consensus/95%           .h.....h..hG. ....................
consensus/85%           .h.....h..hG. ....................

E. rectale              FIKKY-RYDILGN-KYSCSEEKFTSFC--------------
E. faecalis             KLRKF-NTDILGK-KHYLFYEKEPKNIIK------------
S. thermophilus         SIYKV-RTDVLGN-QHIIKNEGDKPKLDF------------
S. lugdunensis          SIEKL-TTDVLGN-VFTNTQYTKPQLLFKRGN---------
E. dolichum             RIDKY-ATDVVGN-LYKVKKNTLKFEFD-------------
R. albus                NGKGISCGEPLSL-LKEKN----------------------
R. inulinivorans        IFEKC-QVDILGN-ISVVRHENRQEFH--------------
L. coryniformis         KFEKY-QVDYFGR-YHKVHEKKRQLFVKRDE----------
I. polytropus           NIKKY-QVDVLGN-VYKVKREKRQTF---------------
consensus/95%           ........-.hu. ..................
consensus/85%           .hcKh...DhhG. ..................

Bacteroides sp.20       ICFPI-KVDRLGN-IIQVGSCILTN----------------
I. album                VCIKI-RVDRLGN-VRPL-----------------------
B. fragilis             TIVKV-RVNHIGQ-IVSVGEY--------------------
N. salsuginis           RSYEVFGTLPEGS-IEWFKEESGYGRVEDDPHH--------
E. minutum              NPRKL-ILSSCGF-IKYRSPILKNKEG--------------
S. globus               DVRKV-QLSSAGL-VRYVSPLLVDKIEKDEVALCGE-----
M. trichosporium        EFKKF-NVDRFGR-KHEVERELRTWRGETWRGKAYI-----
F. branchiophilum       KMNHN-QFTAFVEGIDFKVLPSGK--FEKI-----------
Prevotella C561         SVKSL-NFLIMGENRDFVIKN-GKIIFNHR-----------
P. timonensis           SINNI-NILVDGI--DFTIDILGKVTLKE------------
P. timonensis           SINNI-NILVDGI-DFTIDILGKVTLKE-------------
A. paucivorans          SMSKY-QVDPLGE-IRLVGSEKPPFVL--------------
P. marinum              SARKV-HISPTGL-IREG-----------------------
A. denitrificans        KTRRV-TISPIGE-VRDPGFKG-------------------
R. syzygii              GARRI-FVDLIGR-VLDPGIKGD------------------
D. shibae               GIRRV-SVDEIGR-LRDGGTRPI------------------
R. rubrum               KIRTT-SCTALGR-LRLSKKAT-------------------
Azospirillum B510       GARRV-TVDPLGR-VRVHAPGARVGIGGDAGRTAMEPAEDI
C. salinarum            VAKGD-TVKKDGKTYKVGV--ITQGIFGMPVDAAGSAPRTP
N. hamburgensis         AAVPV-RVDELGR-VWRVMPN--------------------
Bradyrhizobium BTAi1    NAELV-RIDTLGQ-PWRRKRGLETGSEDATRIGWTRPKKWP
W. succinogenes         DVKKY-QIDPLGY-YYEVKQEKRLGTIPQMRSAKKLVKK--
C. jejuni               VFEKY-IVSALGE-VTKAEFRQREDFKK-------------
Jnet 218563121          EEEEE-EEE-----EEE------------------------
H. mustelae             VFQKV-KLSVLGE-VLEHKPRNRQNIALKTTPKHV------
A. ebreus               SVEKF-NVDVLGR-IYLAPPETRSGLA--------------
B. smithii              RFEKY-QVDVLGN-IHKVKGEKRVGLAAPTNQKKGKTVDSL
C. cellulolyticum       SIEKY-NVDILGN-KSIVKGEPRRGMEKYNSFKSN------
HF0070_07E19            HFAKV-QVNLLGK-VIK------------------------
P. lavamentivorans      DAKKV-SIDPIGR-VRPSND---------------------
N. meningitidis         SFQKY-QIDELGK-EIRPCRLKKRPPVR-------------
P. multocida            SFEKY-QVDELGK-NRQICRPQQRQPVR-------------
A. muciniphila          NWREVDLISLLKK-YQMKRYPTSYTGTPR------------
Actinomyces F0310       GLIRVIRRNALGE-VRTSPKSGLPISLNLR-----------
A. cellulolyticus       GLTVI-RRTALGQ-PRWRRGHLPYSWRPWSADPWSGGTP--
B. longum               AVRIV-RRNAFGE-PRLSSAHHMPCSWQWRHE---------
A. naeslundii           HPTVV-RRDALGR-PRYSSRSNLPTSWTIE-----------

                        --EEE---------------------EE-------------
domains                 cccccccccccccccccccccccccccccccccccccccc|
catalytic residues      cccccccccccccccccccccccccccccccccccccccc|

A. hesperidum           VFEKY-VVGPLGD-THPVYKERRMPFRVERKMN--------
Alcanivorax W11-5       KPADL-LSAKWGR-LKVGGRNIHLLRLCAE-----------
B. zoohelcum            NCFKL-EIDRLGN-IVKVIKR--------------------
Porphyromonas F0450     NIRKV-RVDLLGR-ISLL-----------------------
B. intestinihominis     NPHKV-HISVLGE-IKEIS----------------------
O. laneus               NPVKV-QIDEIGR-ITFLNGPLC------------------
Treponema JC4           KARLV-TVSPIGR-VFRK-----------------------
Rhodovulum PH10         GCVAV-RVDPIGV-VTLRRSNV-------------------
consensus/95%           ...........G.............................
consensus/85%           ....h..h..hG.............................

#map                    --****************-----------------------
```

Supplementary Figure S5. Multiple alignment of Cas1 sequences associated with type II CRISPR-Cas systems (Please see Supplementary Materials and Methods for details); Sequences are denoted by Genbank identifiers and complete organism name.

```
16130662_Escherichia_coli_str-_K-12_substr-_MG1655                     ---------------------------------------MTWLPLNPIPLKDRVSMIFLQYGQIDVIDGAFVLIDKT-------------------GIRT
11500011_Archaeoglobus_fulgidus_DSM_4304                               ---------------------------------------MRK----------KNYYLVSDG-KLRRHENTIYFENE--------------------DGKR
18977490_Pyrococcus_furiosus_DSM_3638                                  ---------------------------------------MRK----------KSLTIFSDG-TLLRRENTLYFENV--------------------NGRK
46447796_Desulfovibrio_vulgaris_str-_Hildenborough                     -------------------------------------MKKLL----------NTLYVTTQGTYLAKEGECIVVRVGD-------------------EVRL
54296139_Legionella_pneumophila_str-_Paris                             -----------------------------------------M----------KHLTLSEQGSFCKLQGERILVFNKD-------------------QMVF
331001028_Parasutterella_excrementihominis_YIT_11859                   -----------------------------------------M----------RHLTVSGFGDFLGVHGNRLIVRNGD-------------------NVIL
319941582_Sutterella_wadsworthensis_3_1_45B                            -----------------------------------------M----------RYLTVSSFGTALGLSGERLVVHEAD-------------------GRTW
118497353_Francisella_novicida_U112                                    -----------------------------------------M----------KHLIISEYGIYLGLESGRLVVKNK--------------------DDKK
34557933_Wolinella_succinogenes_DSM_1740                               ----------------------------------------MT----------KHLYIDEYGSFVGKTSERLTVSQGG-------------------ELKA
187736488_Akkermansia_muciniphila_ATCC_BAA-835                         ---------------------------------------MSY----------HILSIDAYTCHLSCDKGQLRCADGE-------------------NSPR
407803668_Alcanivorax_sp-_W11-5                                        --------------------------------------MSEQ----------HILLVESPC-ALNIHLECLKVSRKG-------------------MEPL
325972002_Sphaerochaeta_globus_str-_Buddy                              ----------------------------------------MW----------RVIDLGGTGYYLHTVHNKLIVEKD--------------------GEES
187250661_Elusimicrobium_minutum_Pei191                                ----------------------------------------MW----------RVLDIPGDGYHLCVKNNNFSAVKDR-------------------EEKL
47458867_Mycoplasma_mobile_163K                                        ---------------------------------------MGW----------KIVEINTDE-YVHLYLNNLYIKRK--------------------NEKI
363542551_Mycoplasma_ovipneumoniae_SC01                                ----------------------------------------MK----------KIIEISESE-YVYLFLNNIVIKKD--------------------SEKF
71894593_Mycoplasma_synoviae_53                                        --------------------------------------MMAK----------KVVEIKETE-YLFFHLNSLIVKKE--------------------NNKI
384393287_Mycoplasma_canis_PG_14                                       ---------------------------------------MSK----------KVLEISQSE-YISYFLNSLIVKSN--------------------EKRI
284931711_Mycoplasma_gallisepticum_str-_F                              ----------------------MQLPHPPEHLCSGFLFIMSK----------KVLEVSETE-FISLHLDTLIVRKH--------------------GARV
117929157_Acidothermus_cellulolyticus_11B                              -------------------------------------MTGPW----------RVVDLSELSGEVHAAQGALLVGDE------------------------
317482065_Bifidobacterium_sp-_12_1_47BFAA                              -------------------------------------MQNQW----------RVIDCTDLDGSISYERGRMRVVKNDT------------------GECV
315605739_Actinomyces_sp-_oral_taxon_180_str-_F0310                    -------------------------------------MNSQW----------RVIDLCGFEGTLRSTRGGVEVCAQE-------------------GLPT
294086112_Candidatus_Puniceispirillum_marinum_IMCC1322                 -----------------------------------------------------MIEISGQPRHISVRRGAIIVSDNE-------------------TELG
282880053_Prevotella_timonensis_CRIS_5C-B1                             ---------------------------------------MLK----------RSVVFSHPA-SLSLRYNQLVITMKDLP-----------------DERR
374384762_Odoribacter_laneus_YIT_12061                                 ---------------------------------------MLK----------RTLFFGKPS-RLSVKDKQLVIEIKET------------------GELH
423317188_Bergeyella_zoohelcum_ATCC_43767                              ---------------------------------------MLY----------RSIYIGNPA-YLKLKDQQMKIVCPETK-----------------AEKG
385811610_Ignavibacterium_album_JCM_16511                              ---------------------------------------MIK----------RTLYFGNPA-YLSLKNMQMLIKLPISGNDNEATDEDKNEKYKDRNNQN
402847305_Porphyromonas_sp-_oral_taxon_279_str-_F0450                  ---------------------------------------MMK----------RTIYIGNPA-YLSLRLKQMEVRQPLDDRE---------------TTVR
345885719_Prevotella_sp-_C561                                          ---------------------------------------MIK----------KTLYFGNPA-YLSLRNGQLVIRLPEVEKAE-LPDIIK------QETIR
404487227_Barnesiella_intestinihominis_YIT_11860                       ---------------------------------------MIK----------RTLYFGNPS-YLSLRNAQLVLRLPEVSNNDTLPDTFK------KEAER
301311870_Bacteroides_sp-_20_3                                         ---------------------------------------MIK----------KTFYFGNPA-YLSLRNQQLVIKLPEIVKNDTVPESFK------RQSEI
60683388_Bacteroides_fragilis_NCTC_9343                                ---------------------------------------MIK----------KTLYFGNPV-YLSLRNAQLVIKLPDVEKATVLPETLK------KQAEV
402849996_Rhodovulum_sp-_PH10                                          ---------------------------------------MLG----------RIIEVEGEGRRLCLERGFLRITGPA-------------------GLLG
154250554_Parvibaculum_lavamentivorans_DS-1                            ---------------------------------------MIR----------KTVEFATPGTRLSVANKQLVIERPD-------------------LPKA
159042957_Dinoroseobacter_shibae_DFL_12                                ----------------------------------------MD----------QIVDIATDGRHLSRDRGFLKVSEGA-------------------REIG
288957740_Azospirillum_sp-_B510                                        ---------------------------------------MTG----------RIVEIAEDGRHLSLSRGFLVVTAEG-------------------RELG
330822846_Alicycliphilus_denitrificans_K601                            ---------------------------------------MIG----------RIVEIADDRRHLFVNRGFLVVRDTEGER----------------KELG
344171926_Ralstonia_syzygii_R24                                        ---------------------------------------MIG----------RIVEIATNNRYLGLDRGFMVVKSAD-------------------EEVG
319957207_Nitratifractor_salsuginis_DSM_16511                          --------------------------------------MAAW----------KTIRITKPC-RLSIENGNLVVEDE--------------------KVRF
315149831_Enterococcus_faecalis_TX0012                                 -------------------------------------MLLGF----------RTVFVKNGE-KLKVKLDNLEVIKE--------------------GNTY
116627543_Streptococcus_thermophilus_LMD-9                             ---------------------------------------MTW----------RVVHVSQSE-KMRLKLDNLLVQKM--------------------GQEF
238924076_Eubacterium_rectale_ATCC_33656                               ---------------------------------------MAF----------RVILIENEV-TIKVKLNNLIITKC--------------------GEDI
218563120_Campylobacter_jejuni_subsp-_jejuni_NCTC_11168_-_ATCC_700819  -----------------------------------MSYDEAF----------KTLLISSNA-KLNLELNHLVIKQD--------------------ENIA
291276264_Helicobacter_mustelae_12198                                  ------------------------------------MFDEAF----------RTLFMSKPA-KLSHKNNNLIIHGK--------------------EDIG
325677757_Ruminococcus_albus_8                                         ----------------------------------------MW----------RTVIVTKGE-KLTLRDNWLVVYSD--------------------NNEQ
296446028_Methylosinus_trichosporium_OB3b                              ---------------------------------------MAW----------KGLHLSRPA-RLNFSDNQIVVAQD--------------------DGEA
34557789_Wolinella_succinogenes_DSM_1740                               --------------------------------------MAAW----------RTVVIATSC-HLRVRDRQLVIESRE-------------------GERT
222109284_Acidovorax_ebreus_TPSY                                       ---------------------------------------MVW----------RSIVINRPA-KLKREHFALVVEQE--------------------Q-SA
336393380_Lactobacillus_coryniformis_subsp-_torquens_KCTC_3535         ---------------------------------------MAF----------RNLYIQNPA-RLSISRSQLVIQQE---------------------ERY
218767587_Neisseria_meningitidis_Z2491                                 ---------------------------------------MTW----------RSLLIQNGG-KLSLQRRQLLIQQN--------------------GESH
15602991_Pasteurella_multocida_subsp-_multocida_str-_Pm70              MMSSVKIDKFADLSKDNLCVKLYHPDFLYIGNRDFLGVIMSW----------RSILISQGG-KLSLQRKQMLIQQS--------------------NNSF
310780383_Ilyobacter_polytropus_DSM_2926                               ---------------------------------------MGW----------RNIFITSST-KISTRDSGLLITKE--------------------GGEV
312879014_Aminomonas_paucivorans_DSM_12260                             ---------------------------------------MGW----------RSVFIANPA-RLSTGGGRLQVEQE--------------------GEAA
403744859_Alicyclobacillus_hesperidum_URH17-3-68                       ---------------------------------------MSH----------RHVMVTRSA-KLSCRNNQLVVRQEDESSGE--------------TCEY
421874296_Brevibacillus_laterosporus_GI-9                              --------------------------------------------------------------MSIRLNQLVVKQ----------------------EEEV
220930481_Clostridium_cellulolyticum_H10                               ---------------------------------------MGW----------RNIIVSNPT-KLKLKQNNLWVEQ---------------------SDGF
225377803_Roseburia_inulinivorans_DSM_16841                            ---------------------------------------MGF----------RNIMISSNM-KLRIQDQQLVVTA---------------------GDEV
315659847_Staphylococcus_lugdunensis_M23590                            ---------------------------------------MSW----------RIVYVSDVN-HMSLNLNSLKIKKG--------------------DLET
160915783_Eubacterium_dolichum_DSM_3991                                ---------------------------------------MGF----------RVLYIENQY-HLQLYLDNLKVETS--------------------QGDI
323463802_Staphylococcus_pseudintermedius_ED99                         ---------------------------------------MSF----------RTVIITKES-KLSLRMNQLIVKSE---------------------NLS
389815358_Planococcus_antarcticus_DSM_14505                            ---------------------------------------MSW----------RTVILAKEA-KLTLRMGHLIVSGE---------------------TVN
328956316_Coriobacterium_glomerans_PW2                                 ---------------------------------------MAF----------RNVLIESAC-KCTYQGGYLVVRKE--------------------DDAA
422884107_Streptococcus_sanguinis_SK49                                 ---------------------------------------MGW----------RTVVVNTHS-KLSYKNNHLIFKDA--------------------SRTE
116628212_Streptococcus_thermophilus_LMD-9                             --------------------------------------MAGW----------RTVVVNIHS-KLSYKNNHLIFRNS--------------------YKTE
24379808_Streptococcus_mutans_UA159                                    ---------------------------------------MGW----------RTVVVNTHS-KLSYKNNHLIFKDA--------------------YQTE
13622194_Streptococcus_pyogenes_M1_GAS                                 --------------------------------------MAGW----------RTVVVNTHS-KLSYKNNHLIFKDA--------------------YKTE
227824982_Acidaminococcus_sp-_D21                                      ---------------------------------------MSY----------RTIVITQRC-KLDFCMNYVEVQTA--------------------VSKK
169823756_Finegoldia_magna_ATCC_29328                                  ---------------------------------------MGW----------RTVLISGRA-KLDLKLNNLVIRK---------------------EEIV
320528779_Solobacterium_moorei_F0204                                   ---------------------------------------MGW----------RTVIISNTA-KLDYKMDYLVVRSS--------------------NQVS
303229394_Veillonella_atypica_ACS-134-V-Col7a                          ---------------------------------------MSW----------RTVVISSRS-KLDFKMNYMIIRKD--------------------FETT
304438953_Peptoniphilus_duerdenii_ATCC_BAA-1640                        ---------------------------------------MTW----------RVVVIDNRA-KLELKLSHLVVRQDG-------------------NKVQ
374307737_Filifactor_alocis_ATCC_35896                                 ---------------------------------------MSW----------REVIITGHA-KLDYKMDYLVVRKQ--------------------DGTD
256845020_Fusobacterium_sp-_3_1_36A2                                   --------------------------------------MSGW----------RVIVITGRS-KLDLRYNSISIRRD--------------------NGTD
306821690_Eubacterium_yurii_subsp-_margaretiae_ATCC_43715              ---------------------------------------MSW----------RTIEISNSA-KLDYQMGYMVVRQ---------------------ESTT
42525844_Treponema_denticola_ATCC_35405                                ---------------------------------------MSW----------RTVVISNRA-KLDLHLNHLVVRG---------------------EKTQ
291520706_Coprococcus_catus_GD-7                                       ---------------------------------------MSW----------RIIVISKRA-KLDLQLGKMVVRSD--------------------EITK
258509198_Lactobacillus_rhamnosus_GG                                   ---------------------------------------MGW----------RSFIITQHC-KVTTKQHSLVVQTN--------------------GEVY
336394883_Lactobacillus_farciminis_KCTC_3681                           ---------------------------------------MGW----------RTIYITQKA-KLSYKANHLLIQTN--------------------MDIK
339625080_Fructobacillus_fructosus_KCTC_3544                           ---------------------------------------MAW----------RTVFVNEHA-KINYKMNDINIQTK--------------------EGGT
310286727_Bifidobacterium_bifidum_S17                                  ---------------------------------------MAW----------RNVVVTRHC-KISTKMRLLVIQTD--------------------DDVC
366983954_Oenococcus_kitaharae_DSM_17330                               ---------------------------------------MAW----------RNILINQHC-KISYKMNMCVVQTE--------------------TDVH
#map of conserved positions used for tree reconstruction               --------------------------------------------------------------*************---------------------****
```

Continued

```
16130662_Escherichia_coli_str-_K-12_substr-_MG1655                     HIPVGSVACIMLE-PGTRVSHAAVRLAAQVGTLLVWVGEAGVRVYASGQPGG-ARSDKLLYQAKLALDEDLRLKV-VRKMFELRFGEPAP----------
11500011_Archaeoglobus_fulgidus_DSM_4304                               PIPINSIYAIYALG-SLSITSKAISLLAKEGVCIHFFNRYGYYIGSFYPRESLVSGEVVLRQAEHHLDSEKRLHL-ARAFVEGAILNMARVLKK--AEQD
18977490_Pyrococcus_furiosus_DSM_3638                                  PLAIEGIYDIYIYG-HVNITSQALHYIAQKGILIHFFNHYGYYDGTFYPRETLLSGDLIIKQAEHYLDKNKRLFL-AKSFVVGGAKNMEKNLKN--WGIS
46447796_Desulfovibrio_vulgaris_str-_Hildenborough                     RVPVHSLGGVVCFG-QVSCSPFLMGFAAERGLGFSFLTEHGRFLARVQGPVS-GNVLLRREQYRRADSPEASAEV-ARSIVSAKVVNARGVLQR--AMRD
54296139_Legionella_pneumophila_str-_Paris                             EGTLNRLKTISIVKQGVTLTSNLLVACAMRGIQIFILDYSGKAVCSLQG--G-LHRTAKVRESQFKYIYTKEASNMAARFILAKIKKQSAVLKY--FAKA
331001028_Parasutterella_excrementihominis_YIT_11859                   ETALSRLRTIRID-KSVSLSSNLILACAARGIRIYFLDWKGVGVVEVSGMH--RHAVVGVRKSQFECLKSLQAQLLAKEIIFSKIRNQRAVLLY--FSKY
319941582_Sutterella_wadsworthensis_3_1_45B                            ETALSRLRTIRIEKKGISVSTNLMLACAARGIRLFLVDWRGVGVVALSGLHQ-HAVVSIREAQFLAIHSSQGRTI-AAEMIIAKIRNQRTVLLY--FWKY
118497353_Francisella_novicida_U112                                    YFPLNRLATLSIAKKGVSFSSDLVEQFSLRGIKLFFLDFRGVAHSMLVGANQ-HAVVQARI-NQYRYIDRNALTL-SIKLIAAKIKNQRATLNY--FNKH
34557933_Wolinella_succinogenes_DSM_1740                               EIPLSRLKSVNILKGGISLSSDIIEACSLRGIKIFFLDFRNVAHTAITSTAN-HAVANVRKAQFLTIESNHGREI-ARTLIVGKLKNQRATLQYFYKYLK
187736488_Akkermansia_muciniphila_ATCC_BAA-835                         TIPLEDVGAVVLSSFKATLTSNLLIELARKRIGFVLCESYR-PAVLLLPAD--RSTDTGLLRHLADMPARLRNRL-WQKTLDAKCGNQTALAQA--WNPH
407803668_Alcanivorax_sp-_W11-5                                        RLLPSDIAVLCLHHHTVQLSVHVLRKLTEAGAAILVTDDRHQPAGLMLQDYG-RPNAVVRLRQQIALEQSGQDRLVWQQLVSWKLRTQSRLLGK--LERR
325972002_Sphaerochaeta_globus_str-_Buddy                              LIPFDDIHSIICHGIGFRYSDKFFKSCLHHKVPVTFCDEKHIPLGMLLPMLQ-HKDTSQRMDIQLKASLPRRKQA-WQQIITEKLKNQRDFLQH--LQIP
187250661_Elusimicrobium_minutum_Pei191                                HCLFDDINSIILYGNNITISNTCIQKCLEHKVPVIFCDKTYNPAGMLLSSFT-TNIYGRRLQLQINASKPQIKQA-WQQIITSKLNNQAEVLKR--FDTL
47458867_Mycoplasma_mobile_163K                                        LLNIRDIDTILFNNQYSTISIRLLTFLAKNNVNIIFMNEKNEPNSYLIPIEG-NHNSLKVVAEQVKWTKEYKAIL-WKDIIKNKIHNQKNLLIK--NKLF
363542551_Mycoplasma_ovipneumoniae_SC01                                VFPIDTVDVLIFENDRATISIPVINELVEKKVNIIICKNHL-PQSLIIPYSG-YYNN-KIFQEQIKWDIPYKTKT-WQEIIKLKIQRSISVLKS--IQKI
71894593_Mycoplasma_synoviae_53                                        VLPLSDIDTILISNPYCTISVPLINAIVSNNINLIICNKDFEPNVQLLSISG-YYSN-KNFLSQINWTQEFKDKT-WEKIIKLKTTNYVNLIYF--FGLL
384393287_Mycoplasma_canis_PG_14                                       IVPISQIDTVLVSNPRCNISVPLLNELVAQNVNIIICDSKFKPTVQISPIQG-YYSN-KNFLNQINWSKDFKGFV-WREIIKTKIVNYTKLIYY--LDIT
284931711_Mycoplasma_gallisepticum_str-_F                              TLPLKTLDTILITNPYCNISVPLLNKIVENNINLIVCNTKFQPTMQILPISS-YYSN-KNYLWQIRWTDEQKNQI-WKQLVKLKIINYARLINH--LNIT
117929157_Acidothermus_cellulolyticus_11B                              RVPLVDVA-MMLTGPYVSLHGSVIDRAAAFGVGVVHCDWRGVPVAATLPWST-HNRVAARHRAQAELSLPRQKNA-WMNIVKTKIRNQAAVLRA--LRRD
317482065_Bifidobacterium_sp-_12_1_47BFAA                              DVPLAQIAVVLL-GLHVRCSAAVLHEMARYGVSVMLCDWRGVPDAALHAWTDMPTTVTTRQLAQSNMSLPRKKSA-WARIVQMKIRGQAACLDS--SKLE
315605739_Actinomyces_sp-_oral_taxon_180_str-_F0310                    TIPVAELAVVLV-GMKVNLSAAVLHRLAEADVAVLFCDWRGIPEGGCYSWSD-HGRVAARHRAQAEVTLPRKKNA-WARLIRAKIEGQASVLEN--IKIR
294086112_Candidatus_Puniceispirillum_marinum_IMCC1322                 QIDLDGVLSVIVTSRGATVTTPMITEAAARNIPIIICNDRFQPVSLSLPLVQ-HSDQTRRFETQALAKKGIKNRI-WQKIIKGKVKNQATLLVL--YQSA
282880053_Prevotella_timonensis_CRIS_5C-B1                             TIPIEDIGFLMVDNPMVSISVPLINALTENNVSVVFCNQKGMPTSMLFNLDN-NNTQGETLRNQLSASQSFNKQL-WKQIIECKIKNQAQLLEK--LGKE
374384762_Odoribacter_laneus_YIT_12061                                 KAPIEDIGFVVIENQQISVSLPLLQEFMSNNVAVIFCDASHMPNSMLLNLDG-HHLQSEMFRFQINASEPLKKNL-WKQTVEAKIRNQAALLDK--SGKN
423317188_Bergeyella_zoohelcum_ATCC_43767                              SVPVEDLGLLMLDHFQITISHQLIQWLMGNNVVIISCDAHHLPHGQMLPLHG-NAIYSQRIKDQIEASEPLKKQL-WKQTIECKIANQAAVLER--LGTF
385811610_Ignavibacterium_album_JCM_16511                              TFPIEDIGIVILDNQQITITQGLLSSLLDNNSAIVTCDSRHHPAGLFLPLES-NLEQSQRFQEQIEASVPLKKQL-WQQTISSKIINQALLLNK--IGVK
402847305_Porphyromonas_sp-_oral_taxon_279_str-_F0450                  TIPIEDIGVVLLDHPQITITHALLGALLENNCAIVTCSATHMPSGLLLPLDG-HTLQTERFRAQLDASLPLRKQL-WQQTVQAKILNQAHALHT--VLGE
345885719_Prevotella_sp-_C561                                          TVPIEDIGVVVLDNKQITITQGALAELVANSAAVITCDSKCMPTGMLLPLSG-NTLHQERFRNQIEAAIPLRKQL-WQQTVKAKIENQAYCLQK--NTSK
404487227_Barnesiella_intestinihominis_YIT_11860                       TIPIEDIGVVVLDHSQITITQGLLEALLENNCAIITCDKSRLPVGLILPLCG-NTTQNERFRDQLNASLPLKKQL-WQQTVQQKIANQAAILES--QGII
301311870_Bacteroides_sp-_20_3                                         TKPIEDIAVIVLDNKQITITQGVLEALLENNCAVITCDGRSMPVGLMLPLYG-NTTQNERFRDQLDASLPLKKQL-WQQTIQAKINNQASVLCS--CKNE
60683388_Bacteroides_fragilis_NCTC_9343                                TKPIEDIGIVVLDNKQITITSGVLEALLENNCSVITCDSRSMPVGLMLPLYG-NTTQNERFRKQLDTSLPLKKQL-WQQTIQAKINNQASVLKD--CMDE
402849996_Rhodovulum_sp-_PH10                                          KIPLDDIEAVILTHPAASLTTQAVAALAARGTPLVVSGPDFRPVAFLLPIDG-HHGQGHRLEAQAEASRPTRKRV-WAEIVRAKVTAQADMVEH--VGGA
154250554_Parvibaculum_lavamentivorans_DS-1                            TLPIEDLGVVIVDDLRATYTQAVFIELLEAGATVMVTGRDHLPAGMMLPLDA-HHIQTERHRAQVEASEPTKKRA-WQALIRSKIAQQGIVLAH--FTGE
159042957_Dinoroseobacter_shibae_DFL_12                                RIPLDQIAGVIVHAHGTTWTTSLLTELADRGAPVVLCGANHAPRSVLMPLDG-HHAQGARLRAQWQARAPLVKQA-WKQTVIAKIAMQAAALEA--MGEP
288957740_Azospirillum_sp-_B510                                        RVPLDDIAVVMANAHGLTYSNALLMALMERGAVVVMCGPNHLPAAFVWPVDA-HHVQSRRMRAQVDAPAPLAKRL-WQALVRAKIRQQAAALEA--RGLP
330822846_Alicycliphilus_denitrificans_K601                            QVPLDDIAAVIANAHGLTYTNNLLVALAERGAPFVLCGPNHNAMGMLLPIDG-HHVQAKRIEAQIAASLPVHKRL-WAAVVKSKLEQQAAALEA--AAAP
344171926_Ralstonia_syzygii_R24                                        RVPLDDLAAVIIAAHGITHSSNLLVALAERGVPFVLTAANMQPVAMLWSVDG-NYQQAGRMEAQLGASLPQNKRL-WQQVVRSKLQMQAVVLES--VGVP
319957207_Nitratifractor_salsuginis_DSM_16511                          KLSLSDTDSIIFEGDRFTLSAKVLAALSRHKVATLFCDEYYMPTAILHPYHQ-SSLATETLKAQLSLPDDLRDRL-WQEIIRVKILNQSAVLSH--FAKK
315149831_Enterococcus_faecalis_TX0012                                 VIPLTDIESVILEGDQTVITTRLLAKFAQHHIDTVICDNTFMPVGVFLGIGQ-YHRSAKRAIWQSNLKEEQKQVA-WCEIVTQKIQNQIAVAKY--LGTD
116627543_Streptococcus_thermophilus_LMD-9                             TVPLSDISIIVAEGGDTVVTLRLLSALSKYNIALVVCDNEHLPTGIYHSQNG-HFRAYKRLKEQLDWSQKQKDKA-WQIVTYYKINNQEDVLAM--FEKS
238924076_Eubacterium_rectale_ATCC_33656                               WIPLDDISMIVLDNLASNLSTRLMCQLAEQGIGLMLYNQQHLPTGFYSAYDN-HSRASKVIGFQIDKEQDYYGKL-WQQIVKIKIENQAKAYQI--MTRD
218563120_Campylobacter_jejuni_subsp-_jejuni_NCTC_11168_-_ATCC_700819  KLFLKDINIIVLESLQISISSALFNAFAKYKIILLTCDETHSINGVFTPFLG-HFQSAKIAKEQMNVSAQKKAIL-WQKIIKNKILNQAFILKK--HNKI
291276264_Helicobacter_mustelae_12198                                  -IPLRDIHTIVIDTPQITLTSSLMKELAQYKILVFFCDDSHLPSGVFSPFLA-HYRSSAILDLQIHFSKQNKSIL-WQKIIQAKIQNQTNLLMQ--ENKT
325677757_Ruminococcus_albus_8                                         CVPLGDIYSVVIDNRAALVSVSLMAALAEANIHVIFCDSKHTPTALMLPMNT-HYKPFGVIKRQLGLSDEFKAVL-WQRIVKQKLTNQIICLKL--AGIN
296446028_Methylosinus_trichosporium_OB3b                              RLPLEDVAYIILDAPHATLTSTLISACMEAGVVIVSVDARHTPNGLALPFHS-HHRQAGVAAKQVLLSEPFKKRC-WQGVVVAKIDNQAAHLES--LGRN
34557789_Wolinella_succinogenes_DSM_1740                               QLPLADIGVVIVENPQVTLSAALLSALAQEGTALMSCDSSHLPDGVFAPFGT-HSRHTRAARVQAVWSEPFKKRC-WQKIVRAKVSSQAEVLRR--VGAE
222109284_Acidovorax_ebreus_TPSY                                       RVPFEDIAVIVLNHREITLTHPVLSACGEYGIGLYSTGDNHQPNGVFMPFLQ-HSRATRMQRLQLDLDKPSAKRA-WAHIVQVKIGNQARCMEL--LGTV
336393380_Lactobacillus_coryniformis_subsp-_torquens_KCTC_3535         TVPVEDISSIMIEHKQCTITAPTLSFLAEHQVAVFTCDQKHLPCAVVLPLGQ-YYRKLEALEAQLHVRTKTKARL-WQKLMKQKIENQAKVLEL--NHCK
218767587_Neisseria_meningitidis_Z2491                                 TVPLEDIAVIIIENRETLITAPLLSALAEHGATLLTCDEQFLPCGQWLPYAQ-YHRQLKILKLQLNISEPLKKQL-WQHIVRQKILNQAFVADE--TGND
15602991_Pasteurella_multocida_subsp-_multocida_str-_Pm70              TVPLEDIAVIVIESKETLITAPLLSALAQYGVTLLTCDEQFLPCGQWLPFGQ-YYRQLKTLKLQLETTLPQKKQL-WQKIVQQKIRNQGFVLEQ--VGYF
310780383_Ilyobacter_polytropus_DSM_2926                               NIPLEDISSIILDNNSVTITSRVLEEIASRDIVLLTSDEKHLPNGIFQGYGR-HSRQLAVAKAQINMSVPRKKRL-WQEIVTKKLENQGRVLEL--MRCE
312879014_Aminomonas_paucivorans_DSM_12260                             SVPLEDLGALVLESPQVSLTVPLLRELAENRVPLFVCDRRHLPCGVLLPFHA-HCRQEKVLRAQMGLSRPFAKNC-WQRVVRRKILNQAACLRF--LDLP
403744859_Alicyclobacillus_hesperidum_URH17-3-68                       TIPLEDIRTLMLEDRQSIITSALLARLAEFRIAVFTCDEKHLPNGVLLPYCQ-HSRGPKIMQSQLEMKRPLVKRL-WQAVVRRKIENQAACLTL--CGCS
421874296_Brevibacillus_laterosporus_GI-9                              TIPLEDIASIVVEENQVVLTAALLSKCSEYGIGLYTCDEKHLPNGVLLPFQS-HSRQLRVLETQYNLSKPFKKRV-WQKIITTKIMNQALCLSY--LDKT
220930481_Clostridium_cellulolyticum_H10                               SIPIDDINTIVLDSADVTITSALLSKLAEEDIALYSCDGKHTPNGVLLPFSC-HSRQYKIVKTQINLSAPFKKRC-WQRVVQQKIENQAFCLNI--LELK
225377803_Roseburia_inulinivorans_DSM_16841                            HIPLEDINCILIENQSVTLSAYLLQKVADSGIALYVCDEKHLPNAVLLPMVR-HSRHFKILKYQMETGKPLQKRL-WQQVVIQKIRNQALCLQL--LELD
315659847_Staphylococcus_lugdunensis_M23590                            KIPLADIFALVIEDLTVTITSRLLVELSNYNILVIFCNQKHLPECSLQPISG-HFNQYIQMKEQLKWDEARKNEL-WSAIVKRKIHNQIECMKY--LDID
160915783_Eubacterium_dolichum_DSM_3991                                QFPISDIQILVIDHYRSTLTVPLVNKLTENNVCVIICGIDHLPKSYILPMNG-HFAQSGNIMKQITWSNEIKQIL-HQQIVKAKIFNQIEILKT--NQCK
323463802_Staphylococcus_pseudintermedius_ED99                         RIPLNEILCLVIEHPNVSMTGHLMNALSDQKIVTILCNTKHLPNTLLLPLYG-HHRQARLIQNQMNWKNERKGQL-WQLLVKHKILNQKQVIQE--CFKE
389815358_Planococcus_antarcticus_DSM_14505                            RVPLEEISFLLIENPNIQLTGHLINALSHHHIVTLICDANMKPSSVVNSVYG-HHRQSRIIAQQATWTEERKGML-WQEITKQKIENQRQLLIH--LKKE
328956316_Coriobacterium_glomerans_PW2                                 KIHLSEISSVTLQSTQVFVSAYLLSELAKAKITFVISDEKRNPVGQYLPIYG-AHNTSKRIGEQISWEEPIKKRV-WQRVVRDKITNQARLLEI--RGRS
422884107_Streptococcus_sanguinis_SK49                                 LIHLSEVDILLLETTDIVLSTMLIKRLVDENILVIFCDDKRLPTAYLMPYYG-RHDSSLQLSRQITWNEETKAQV-WTTIIAQKILNQAFYLGS--CGFL
116628212_Streptococcus_thermophilus_LMD-9                             MIHLSEIDILLLETTDIVLTTMLVKRLVDENILVIFCDDKRLPTAFLTPYYA-RHDSSLQIARQIAWKENVKCEV-WTAIIAQKILNQSYYLGE--CSFF
24379808_Streptococcus_mutans_UA159                                    MIHLSEIDILLLETTDIVLSTMLIKRLVDENILVIFCDDKRLPTAMLMPYYA-RHDSSLQLSHQISWTEEVKCDV-WTTIIAQKILNQSCYLGE--CFYF
13622194_Streptococcus_pyogenes_M1_GAS                                 LIHLSEIDILLLETTDIVLSTMLVKRLVDENVLVIFCDDKRLPTAMLMPFYG-RHDSSLQLGKQMSWSETVKSQV-WTTIIAQKILNQSCYLGA--CSYF
227824982_Acidaminococcus_sp-_D21                                      RIFIDEIKTLIIENTGVAITAYLLSELMNRKVNLIFCDKAHNPQSSLLPLHA-RFDSIRKIKQQMIWSQEIKDEV-WDCIVKAKIRQQALFLDE--LEKK
169823756_Finegoldia_magna_ATCC_29328                                  KIHIPEIAVLIIDSTMVSFTTALIEKLIAEKVKIIFCGKDHNTIGEIVNYYN-KHNSPLMIRKQISWQNNSKQKT-WREITKEKINRQALLLKF--IGNE
320528779_Solobacterium_moorei_F0204                                   RVHLSEISLIIVETPSVSLTAYLLCELAANKIGIIFCDEKRYPNGQYIPFYG-SHDTSEKIRKQYNWNLNSKMII-WKEVVIRKIAGQMGVLNRYNHVKS
303229394_Veillonella_atypica_ACS-134-V-Col7a                          KIYIDEIYMLVIESTAVSLTAVLLNELIKKKVAIVFCDEKRNPASEVLPLHG-SHNSSKRCREQADWSKDVQAYI-WTEIVRRKIMNQAELLKY--KELK
304438953_Peptoniphilus_duerdenii_ATCC_BAA-1640                        KIYLKDISHIIVNTTEASITFALLNEIVKNKIKLIFCDETRNPTAEVSPLYS-SYNTSEKIRLQIKWDEYTKGEV-WQQIVKQKIKNQRNLLEA--AGKP
374307737_Filifactor_alocis_ATCC_35896                                 RVYIPDIGFLMIESTRVSLTTTLISELVKNKVKIIFCDANHNPQSELLPYHG-SHDTSLKVREQVSWSDEIKQLV-WTEIVKEKIKNQMELLEE--YQNF
256845020_Fusobacterium_sp-_3_1_36A2                                   FIHIGEVNTLILETTTISITAALMCELIKQKVKVIFCDEKSNPHFELLPFYG-SHDCSAKIKEQIAWTDFLKESL-WTIIVTEKIENQMKLLKK--LNKE
306821690_Eubacterium_yurii_subsp-_margaretiae_ATCC_43715              KIHLSEIHTLILATTAISITCALLSELTKQKIKVVFCDEKRNPSSELIPYYG-SHDSSAKVRTQIQWDNEMKLEV-WTQIVKEKIRQQALHLKS--RSKS
42525844_Treponema_denticola_ATCC_35405                                KVFIEEISVLIIETTAVSITAALLNELIKQKVKVIFCDEKRNPASELIGYYG-SHDTSEKIRLQIKWDKNIKQLV-WTEVVTEKIRQQKYLLEK--LNLP
291520706_Coprococcus_catus_GD-7                                       -IVLSEISTIIIESTAVSITASLISELAKRKIKVIFCDEKRNPSCELVNYYG-SHDTSNKIRKQIAWKQQTKEAI-WTEIVTDKIRKQKDLLEL--LRRK
258509198_Lactobacillus_rhamnosus_GG                                   TMPIDDLNQVIFTTTRALISADAIAALADANAKVIFSGRDGQPVTETKNLYS-DRRRAEQVRLQVNWPLDLIETV-WTKIVAAKISNQAQVAKL--CGHD
336394883_Lactobacillus_farciminis_KCTC_3681                           QIPFHQINCIIIATTQSVITGYLISKLVTENIKVIFCAEDHNPCAELNGYYN-NIQRNQNIENQVSWSLPTKQDL-WTRVVKNKLQNQRSVLVR--ADLS
339625080_Fructobacillus_fructosus_KCTC_3544                           RIPIDDIRVLILGNTRTVITSYAVNELQKKNVKIINCDEKGMPVGEFLSYCD-NTKQNINIWKQFDWAENRKNLL-WQKIVQTKIFNQHQALGE--KVIA
310286727_Bifidobacterium_bifidum_S17                                  QLPMDDIGMLMIETTRAVITTCAIADLLKRDVKIVFCDDKHLPIGETSPYET-ESSRRAYIIQQIRWDEARKDLL-WQHIIREKIAHQVSVLEK-NCDSL
366983954_Oenococcus_kitaharae_DSM_17330                               QIPVEDIRLLVIATTRAVITTYAVMELLKRQVGILFTDDKAFPIGEINNYLG-GGNRNKNIQQQVDWPIERKQNL-WAEIVKSKILNQSKCLDF--FSEN
#map of conserved positions used for tree reconstruction               ****************************************************-**********************-******************------
```

Continued

```
16130662_Escherichia_coli_str-_K-12_substr-_MG1655                     ------------------ARRSVEQLRGIE--------------GS-RVRA----------TYALLAKQYGVTWNG-----RRYDP---KDWEKG-----
11500011_Archaeoglobus_fulgidus_DSM_4304                               D------------SD--VIASLQHLSSAKS--------------IV-ELMG----------AEAAARNAYYTK-FDEILKNFEFGKRSRMPPE-------
18977490_Pyrococcus_furiosus_DSM_3638                                  S------------DF---SNHLKELQGAKK--------------VT-EIMN----------VEGRIRQEYYARWDESLPGEFRIGKRTRRPPK-------
46447796_Desulfovibrio_vulgaris_str-_Hildenborough                     H-GDKVDGVALEAEVLHLRACLMRLQQPAG--------------LD-AVRG----------IEGEAAKGYFSVFDN-----LILTREAAFRFEGRSRRPP
54296139_Legionella_pneumophila_str-_Paris                             K----NRPENEKIILHAASQKMAGIVANIT--------------ET-NWIKLQDWRFTLLGFEGSAARIYWETVKEVNLVCSDFEG---RRPRVNK----
```

```
331001028_Parasutterella_excrementihominis_YIT_11859                   LMKRETAYGLLLRNAGTRLEEISNLLNTRA--------------SGPDWKN------TVMGLEGSASLIYWNALKEARLLPETFQE---REGRGST----
319941582_Sutterella_wadsworthensis_3_1_45B                            LNKSAPEKAEKLREA---ADTIAKIVTKTKAEIENLQTWNEGIWST-ILMG----------YEGAAATLYWQALRTAELVPETFLT---REGRGSV----
118497353_Francisella_novicida_U112                                    H---------KSINLLNAIEELKRVAQLIKNAKT----------LN-DVLG----------YEGYAANIYFSSLAK-----DKFLS---ASFANREGRGS
34557933_Wolinella_succinogenes_DSM_1740                               N------TDKQSKKLESAIEKIKETIESLK--------------NI-ELSNNKHWSQIIMGKEGGMAAIYFDAIGAELLEECDFTT---RIGRGAE----
187736488_Akkermansia_muciniphila_ATCC_BAA-835                         H-------------------PAIAELKRMA--------------VT-EKTA----------REAECARLFW-SVFADTWANSDFRR---GRHE-------
407803668_Alcanivorax_sp-_W11-5                                        S-----------------TARLARLAEAVE--------------PG-DAGH----------LEGQGARLYWRDLFD-----VPFKR---QKRGAE-----
325972002_Sphaerochaeta_globus_str-_Buddy                              G-----------------VKRLDVLSQTVL--------------SG-DSSN----------NEAQGARVYFESIYG-----TSFFR---SDED-------
187250661_Elusimicrobium_minutum_Pei191                                K------------AA----ETIFNMAREVR--------------SG-DATF----------KEGVGAKVYFENLF------NDFHRNTDDK---------
47458867_Mycoplasma_mobile_163K                                        N------------NSEFGIEYFDDLINNVN--------------LF-DISN----------REGHAAKVYWNMTYS-----KEFIR---NNSATKLKFDK
363542551_Mycoplasma_ovipneumoniae_SC01                                N------------SED--EAKMWDYFHDVQ--------------PY-DTNN----------REGHAAKLYFKLLFG-----KGFIR---DQNGD------
71894593_Mycoplasma_synoviae_53                                        N------------KED--VEKFNFYYKKII--------------PG-DKNS----------MEGHIAKLTFKNLYG-----STFNR---TDKE-------
384393287_Mycoplasma_canis_PG_14                                       S------------EEE--IDKLIKYYESVL--------------IN-DRSN----------CEGHAAKLSFKILYG-----ANFSR---DD-E-------
284931711_Mycoplasma_gallisepticum_str-_F                              T------------DDN--IEKLTNYYTNVV--------------DG-DLNN----------CEGHAAKLSFKLLYG-----EDFSR---SQKD-------
117929157_Acidothermus_cellulolyticus_11B                              G-----------------VAQLERLAAQVR--------------SG-DASN----------AEGAAARVYWARLFQD----KHFRRVPRAR---------
317482065_Bifidobacterium_sp-_12_1_47BFAA                              G-----------------GGLLRGIAASVR--------------SG-DPSN----------CEGHAAREYWRRMFPQD---EKFRR---IP-GQG----
315605739_Actinomyces_sp-_oral_taxon_180_str-_F0310                    G-----------------SGELLALADQVR--------------SG-DPGN----------IEAQAARLYWSRTLG-----KGVGRQPAAGQL-------
294086112_Candidatus_Puniceispirillum_marinum_IMCC1322                 S-----------------VERLNRLHTLIK--------------SG-DPQN----------IEAQAAQVYWPALFG-----RDFRR---DRHK-------
282880053_Prevotella_timonensis_CRIS_5C-B1                             G------------------KCLFPYYNNVK--------------SG-DSTN----------REGAAAKLYWTQLFG-----KDFIR---DRTQ-------
374384762_Odoribacter_laneus_YIT_12061                                 G------------------NILKPFYCNVK--------------SG-DTDN----------REGAAARVYWTEMFG-----TSFIR---DRYG-------
423317188_Bergeyella_zoohelcum_ATCC_43767                              S------------------SPMYEYLNQVK--------------SG-DTSN----------MEGIAAQHYWKYLIH-----PDFLR---ERYG-------
385811610_Ignavibacterium_album_JCM_16511                              S------------------DNMKKWARDVK--------------SG-DPDN----------LEGRAAAYYWKNLFPEE---LQFQR---DRYG-------
402847305_Porphyromonas_sp-_oral_taxon_279_str-_F0450                  P-----------------AKNMLAWSRQVK--------------SG-DADN----------LEARAVAFYWRNLFISL---PGFVW---EREE-------
345885719_Prevotella_sp-_C561                                          S-----------------FVPLHVLARKVR--------------SD-DADN----------HEAQAAAYYWKNLFS-----DGFTR---DKDG-------
404487227_Barnesiella_intestinihominis_YIT_11860                       A------------------KNMRAWVGEVR--------------SG-DIDN----------MEARAAAYYWSSIFPHL---DSFRR---DRDG-------
301311870_Bacteroides_sp-_20_3                                         E-----------------IKCMRIWANDVR--------------SG-DPDN----------LEGRAAAYYWKYLFGHI---EGFTR---DREG-------
60683388_Bacteroides_fragilis_NCTC_9343                                E-----------------VKCMRIWATNVR--------------SG-DPDN----------LEARAAAYYWKSLFSDV---DSFTR---EREG-------
402849996_Rhodovulum_sp-_PH10                                          S------------------GAIRALLPRVR--------------SG-DPDN----------VEALASQRYFPAFFG-----TGFRR---DRNA-------
154250554_Parvibaculum_lavamentivorans_DS-1                            H------------------GGLLPMARRVR--------------SG-DPDN----------LEAQAAQRYWPRLFG-----KDFRR---DRDL-------
159042957_Dinoroseobacter_shibae_DFL_12                                H------------------APVGMLARKVT--------------SG-DATN----------VEAQAARLYWPRMMG-----TEFRR---DRTA-------
288957740_Azospirillum_sp-_B510                                        A------------------GGLVEMARRVR--------------SG-DPDN----------LEAQAARRYWPALFG-----REFRR---DQDG-------
330822846_Alicycliphilus_denitrificans_K601                            T------------------APLLALVSKVK--------------SG-DPEN----------IEGQGARRYWGLLFG-----NTFRR---DQNG-------
344171926_Ralstonia_syzygii_R24                                        S------------------APLSALVSKVR--------------SG-DPEN----------YEAQGARRYWVLLFG-----DTFRR---DRQQ-------
319957207_Nitratifractor_salsuginis_DSM_16511                          N------------------PKLERYAEQIR--------------PG-DPYR----------TEAKSARLYWNKLF------DGLKR---EA-DSL----
315149831_Enterococcus_faecalis_TX0012                                 S------------ER---VEVLEKLSEGIL--------------PG-DTTN----------REGHVAKVYFHSLYG-----VGFTR---EE-E-------
116627543_Streptococcus_thermophilus_LMD-9                             L------------DN---IRLLSDYKEQIE--------------PG-DRTN----------REGHAAKVYFNELFG-----KQFVRVT-QKEA-------
238924076_Eubacterium_rectale_ATCC_33656                               S------------DG---PEKIIEFSKNIL--------------IG-DKSN----------REAHAAKVYFNLLMG-----TSFSR---GNED-------
218563120_Campylobacter_jejuni_subsp-_jejuni_NCTC_11168_-_ATCC_700819  E----------------QSNELINLAKKVS--------------LN-DSKN----------IEAVAAALYFKTLFG-----TSFSR---DE-L-------
291276264_Helicobacter_mustelae_12198                                  S------------TA----TKLQALQKAVK--------------LA-DGTN----------NEAKASALYFPSLFG-----RGFSR---KD-F-------
325677757_Ruminococcus_albus_8                                         T------------DE---IKPIEEMISEVL--------------PN-DKTN----------REAAAAKRYFIALFG-----NSFRR---IK-D-------
296446028_Methylosinus_trichosporium_OB3b                              A------------------EALRAMAKLVG--------------SG-DPDN----------VEARAARDYWRALY------KDFTR---DDAC-------
34557789_Wolinella_succinogenes_DSM_1740                               D----------------AARRLENLVGKVT--------------SG-DTTG----------VEAQAAQVYWRSLF------VNFKR---WDGGL------
222109284_Acidovorax_ebreus_TPSY                                       G-----------------TDRLASYARRVR--------------SG-DGGN----------LEAQASAYYFPQVFG-----RSFHR---SQ-T-------
336393380_Lactobacillus_coryniformis_subsp-_torquens_KCTC_3535         N-----------------VAELNFLATQVT--------------EG-DKTF----------KESRAARLYFMGLFG-----DQFNR---RN-E-------
218767587_Neisseria_meningitidis_Z2491                                 L----------------AAKRLRTLASEVR--------------SG-DTGN----------REAQAAALYFQALFG-----EKFTR---ND-N-------
15602991_Pasteurella_multocida_subsp-_multocida_str-_Pm70              N----------------ETKRLNRMANQVK--------------SG-DKEN----------VEAQAALIYFQTLFG-----KTFKR---SE-E-------
310780383_Ilyobacter_polytropus_DSM_2926                               G-----------------SKYLYDLSKRVK--------------SG-DDGN----------LEALGAKYYFQKLFG-----DDFTR---RD-H-------
312879014_Aminomonas_paucivorans_DSM_12260                             G-----------------AEELHALASQVR--------------SG-DPDQ----------REGVAARLYFSRLM------PGRTR---DD-P-------
403744859_Alicyclobacillus_hesperidum_URH17-3-68                       G-----------------DEALLALAGRVA--------------SG-DKGS----------IESQAARMYFKHLFG-----KAFKR---SS-D-------
421874296_Brevibacillus_laterosporus_GI-9                              G-----------------ESELLAIRNTVD--------------SG-DSTN----------REAYAAKVYFQHLFG-----KRFIR---RE-D-------
220930481_Clostridium_cellulolyticum_H10                               G-----------------RDELINLSKSVL--------------SG-DSTN----------VEAHAAKYYFSVLF------TNFKR---GM-Q-------
225377803_Roseburia_inulinivorans_DSM_16841                            G-----------------HEELLRMCKEVQ--------------SG-DRTH----------VEAKAAAYYFRCMYG-----LGFSR---GN-D-------
315659847_Staphylococcus_lugdunensis_M23590                            Q------------LR---INKMTDLYHSVA--------------LF-DRNN----------IEGQAAKYYFNSFF------KDFIR---DNDE-------
160915783_Eubacterium_dolichum_DSM_3991                                Y------------EV---IKKLYEFYDTVD--------------LG-DATN----------REGLAAKMYFREMFG-----NDFIR---FE-D-------
323463802_Staphylococcus_pseudintermedius_ED99                         R----------------DSELFDAFANNVE--------------LD-DSTN----------REGHAAKVYFNIVLD-----PGYAR---GD-D-------
389815358_Planococcus_antarcticus_DSM_14505                            G-----------------IDELSMFRDQVE--------------VF-DTTN----------REGHAAKVYFNRLYG-----QDFTR---SI-D-------
328956316_Coriobacterium_glomerans_PW2                                 E----------------EAGRIGAIVSDVR--------------SG-DTTN----------REAYAARCYFQSLFG-----DSFSR---DD-D-------
422884107_Streptococcus_sanguinis_SK49                                 E----------------KSQSVIDLYHGLD--------------LF-DPSN----------REGHAARIYFNALFG-----NDFSR---EQ-D-------
116628212_Streptococcus_thermophilus_LMD-9                             E----------------KSQSIMELYHGLE--------------RF-DPSN----------REGHSARIYFNTLFG-----NDFTR---ES-D-------
24379808_Streptococcus_mutans_UA159                                    E----------------KSQSIMDLYHDLE--------------PF-DPSN----------REGHSARIYFNTLFG-----NVFSR---EQ-D-------
13622194_Streptococcus_pyogenes_M1_GAS                                 E----------------KSQSIMDLYHGLE--------------NF-DPSN----------REGHAARIYFNTLFG-----NDFSR---DL-E-------
227824982_Acidaminococcus_sp-_D21                                      E----------------QSKMLMSYLDDVV--------------SA-DAHN----------REGHAAKVYFNALFG-----NSFTR---DL-D-------
169823756_Finegoldia_magna_ATCC_29328                                  N-----------------ANLLLEYKSQVQ--------------DG-DVSN----------REGHAAKVYFNSLFG-----KDFSR---SQ-D-------
320528779_Solobacterium_moorei_F0204                                   G------------------EKLKQYSYDVE--------------IG-DTTN----------REGHAAKVYFNSLFG-----KDFSR---DDKD-------
303229394_Veillonella_atypica_ACS-134-V-Col7a                          E-----------------ADLLYQYLDELT--------------LN-DETQ----------REGHAAKVYFNALFG-----KSFSR---EQ-D-------
304438953_Peptoniphilus_duerdenii_ATCC_BAA-1640                        E-----------------YALLEKYIEEVE--------------LK-DATN----------REGHAAKVYFNALFG-----KSFSR---SD-D-------
374307737_Filifactor_alocis_ATCC_35896                                 D----------------RAELLKTYIEEME--------------PG-DKNN----------REGHAAKVYFNTVFG-----IDFTR---ND-E-------
256845020_Fusobacterium_sp-_3_1_36A2                                   E-----------------YKILQEYASQIE--------------HN-DNTN----------REGHSAKIYFSALFG-----NNFSR---NK-E-------
306821690_Eubacterium_yurii_subsp-_margaretiae_ATCC_43715              Q-----------------YEMLEDYVRQVQ--------------FA-DSTN----------REGHAAKVYFNALFG-----LDFSR---RA-D-------
42525844_Treponema_denticola_ATCC_35405                                Q-----------------ASLLAEYITDID--------------IN-DKTN----------KEAHAAKAYFAALFG-----AGFSR---SL-D-------
291520706_Coprococcus_catus_GD-7                                       E-----------------SALLDTYLKKIT--------------WN-DGTN----------REGHAAKVYFNSLFG-----LDFTR---TE-D-------
258509198_Lactobacillus_rhamnosus_GG                                   A------------------QGLLDDLDALE--------------IN-DVSN----------REASAARRYFKLMFG-----DDFSR---SDIC-------
336394883_Lactobacillus_farciminis_KCTC_3681                           P------------------SFLDNEYRFIQ--------------EN-DSSN----------REAVIARKYFLALFN-----ENFSR---GK-D-------
339625080_Fructobacillus_fructosus_KCTC_3544                           N-----------------YGKLKEIGLEVK--------------EA-DEDN----------REGQAARLYFPRIFD-----KDFVR---SNDS-------
310286727_Bifidobacterium_bifidum_S17                                  D-----------------LEKMKSLVVDVK--------------PG-DSDN----------REAVAAHMYFPRLFT-----YEFVR---SDEE-------
366983954_Oenococcus_kitaharae_DSM_17330                               D------------------RSLEQLIIQVN--------------QN-DNSN----------REAVAARMYFTRLFG-----PSFVR---HNDD-------
#map of conserved positions used for tree reconstruction               -------------------***********--------------**-****----------***************------------------------
```

Continued

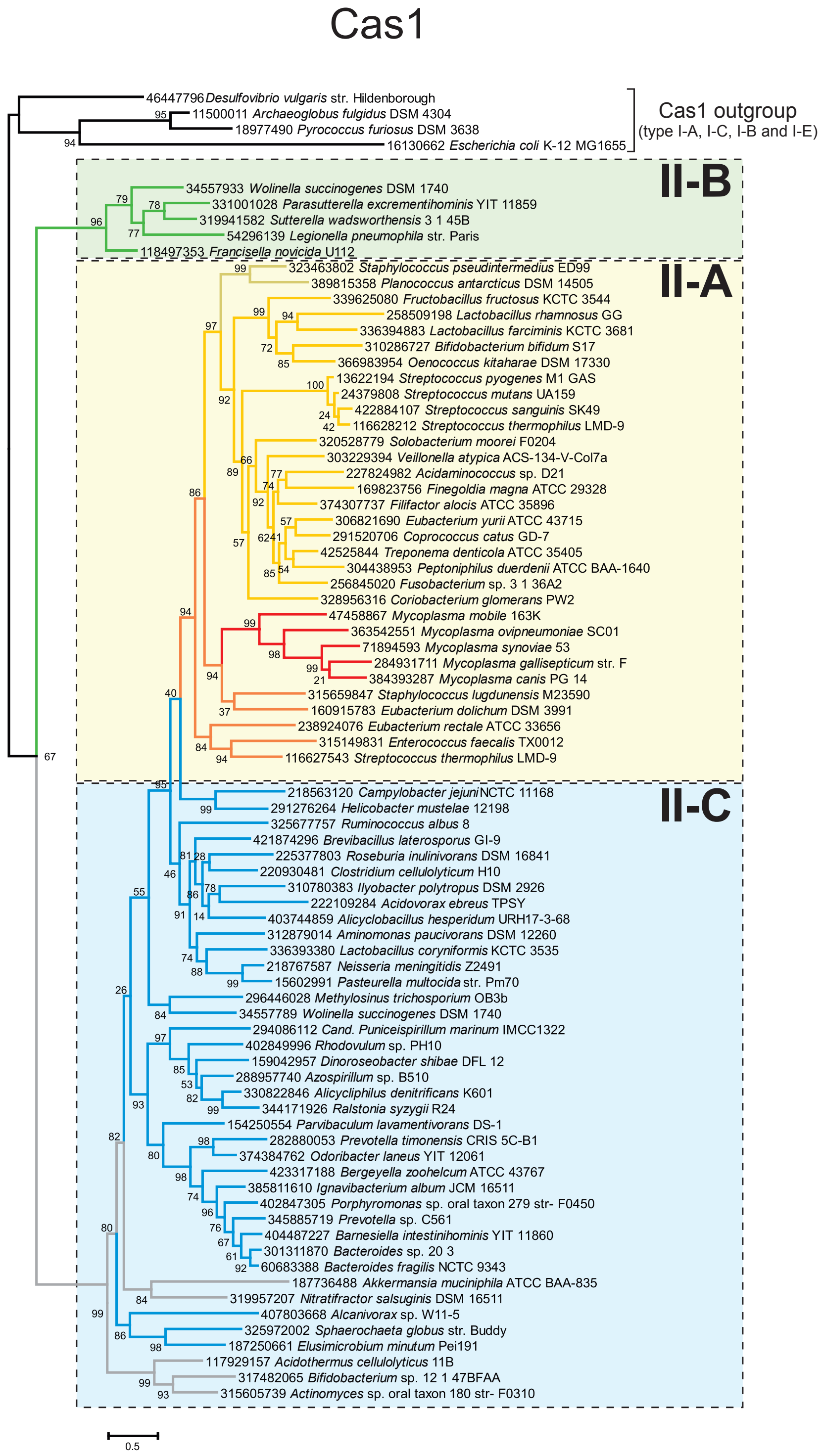
```
16130662_Escherichia_coli_str-_K-12_substr-_MG1655                     -DTINQCISAATSCLYGVTEAAILAAGYAPAIGFVHTGK--PLSFVYDIADIIKF---DTVVPKAFEIARRNPG------E----PDREVRLACRDIFRS
11500011_Archaeoglobus_fulgidus_DSM_4304                               -NEVNAMISFGNSLLYSAVLSEIYHTQLNPAISYLHEPSERRFSLALDIAELFKPVIVDRLIFYLVNNGIVTESDFDSRLGGILLSE-EGKKKFVRHFNE
18977490_Pyrococcus_furiosus_DSM_3638                                  -NEMNALISFLNSRLYATMISEIYNTQLAPTISYLHEPSERRFSLALDLSEIFKPIIADRIANRLVKKGIIKKDHFREDLNGVLLTD-EGMKIVTKAYNQ
46447796_Desulfovibrio_vulgaris_str-_Hildenborough                     LDRVNCLLSFIYTLLGHDVRSALEGVGLDSAVGFLHRDRPGRHGLALDVMEEFRAVVADRLALSLINLGKLKKSDFEIQETGAVRMTDDARKALLVAYQK
54296139_Legionella_pneumophila_str-_Paris                             -DITNSMLNYAYSILSSWVWQAITNAGLELYAGILHTSKPGKPSLVLDLMEEFRPWCADRIIFKLHARAQKQNE-L----TPALKK--TLIGEMANLYAK
331001028_Parasutterella_excrementihominis_YIT_11859                   -EITNAALNYGYAILHSYIQNALGNAGLEIYAGLLHADRAGKPSLVLDCIEEYRAWVVDRNVISLRCKLGSEQNFL----D----T--ALKKQLVELIDK
319941582_Sutterella_wadsworthensis_3_1_45B                            -EIVNAALNYGYAILQSYIWSALDNAGFELYAGFLHRRRPGKPSLVLDVMEEYRAWVVDRNIIKLRSMLKNSKRGL----T----Q--EIKSAIVNAIDE
118497353_Francisella_novicida_U112                                    QEIANSMLNFGYAILSSYILNAITNAGLEPYLGFLHQKRPGKMSLVLDLMEEYRAWVVDRVVIKLREQYKNKQY-I----D----T--KLKSILISEIQA
34557933_Wolinella_succinogenes_DSM_1740                               -DIMNSALNYGYAILTSFVWQSIINAGLEAYAGVLHTNRPGKPSLVLDVMEEYRAWVVDRNIIKLRSSLAGKKE-L----D----A--ELKKKIITEINA
187736488_Akkermansia_muciniphila_ATCC_BAA-835                         -EGFNNLFNYAYAILLSCILQYLFALGLDPCFGIFHQSREHAAPLAYDLMEPFRPAF-DANVARWIHLCLREGKTE----ERAGEITREFRQHITATLQA
407803668_Alcanivorax_sp-_W11-5                                        -DPINTRLNFGYALLRSLVARELAFAALRPEIGVGHRSTENPFNLADDFLEPYRFLV-DEIVVSYPPDENSEFDAQ------------SRKRLISDLFEQ
325972002_Sphaerochaeta_globus_str-_Buddy                              -NAINGKLDYGYSILRSAVARAVVDTGLLPSCGVFHSSRINPFCLVDDLMEPFRPMV-DSLVLAMCSKYGVIED-----------LESREKKELMRITAR
187250661_Elusimicrobium_minutum_Pei191                                -DIINSALNYGYAIVRSSIARAVVSAGLNPAIGIFHSKNHNPFCLIDDLIEPLRPLI-DFMVKNKLDVLTQEESLS---------P--SAKKYMASVIES
```

```
47458867_Mycoplasma_mobile_163K                                        FEIVNAILNYGYSILRSSAIQSIIKKGLDPRFSIFHKSFSNFFALASDLIEIFRPLV-DEIAYLHKDETFFDV---------------KIKDELINVLNK
363542551_Mycoplasma_ovipneumoniae_SC01                                -DNINILLNYGYIVLLSYVARIICGKGLDNRLGIYHKSFNNNFPLACDLMEPYRYWV-DQIVYNHINAEFLNFQ--------------DFKEALFKSFSN
71894593_Mycoplasma_synoviae_53                                        -NEINKFLNYGYTILMTYVSRNLVKKGYDNRIGVFHKSFNNHFALATDLMEPFRFLI-DKLVYELLIIEKNYDF-I------------NFKKKVFLIFEE
384393287_Mycoplasma_canis_PG_14                                       -NIINTFLDYGYTVLMTYVSRSIIKNGLDNRIGIFHKSFNNHFALSCDLMEPLRALI-DKLVYTHIYIYGKKDF-I------------LFKKDLFCMFEE
284931711_Mycoplasma_gallisepticum_str-_F                              -DIINKYLDYGYIVLMTYVSRTLVKNGLDNRIGIFHKSFNNHFALSCDIMEPFRPII-DYLTFSYLIKNQNDDFK-------------SYKKDLFISFEN
117929157_Acidothermus_cellulolyticus_11B                              -DVVNGLLDYGYAILRGCCLRAVVGAGLAPSLGLWHRRHDNPFTLVDDLIEPFRPAV-DKTVIEIVTAGASG---L----D----R--PTKRLLVAVLDH
317482065_Bifidobacterium_sp-_12_1_47BFAA                              -YGRNAQLDYAYMVLRGFAVKAVIAAGLIPSLGVNHHGRGNYFCLADDLLEVYRPAV-DYRVSLLNDEDSLQEK--------------AVKKHLVDAVNQ
315605739_Actinomyces_sp-_oral_taxon_180_str-_F0310                    -IGANACLDYGYSVLRGHLMRAVLAAGLAPALGVFHRGRGNAFALADDLIEPFRPAI-DEVALQLPPTASPSDR--------------SVKQLLVAAASQ
294086112_Candidatus_Puniceispirillum_marinum_IMCC1322                 -GDVNALLNYGYAIIRSAMVSAVLSAGLHPTFGMFHKNKNNPLCLVDDLMEPYRPLV-DQLVKRLVERGHSN---V----T----T--EVKRCLASIVTS
282880053_Prevotella_timonensis_CRIS_5C-B1                             -SGTNALLNYGYTILRAAVTRAIMSAGLFPALGIFHHNRSNAFPLADDMMEPYRPFV-DEIVYHLTQNGQTT---L----T----T--EVKGILIKLLYC
374384762_Odoribacter_laneus_YIT_12061                                 -ECPNNLLNYAYIVLRAAVARAITGSGLLPTLGLFHHNRYNAFCLADDLMEPYRPYA-DEVVSGLFQKGEKE---L----T----T--AVKAEILKILAN
423317188_Bergeyella_zoohelcum_ATCC_43767                              -DSPNQFFNFGYSVLRSIVARAIVETGLLPVLGIFHKNKYNAYCLADDLMEPYRPFV-DLLVIRWLEKNPEAEE-L----T----K--EAKAYFLQIATK
385811610_Ignavibacterium_album_JCM_16511                              -KPPNNLLNYGYAILRAVVARSLVASGLLPTLGIHHHNKYNAFCLADDIMEPYRPFV-DDIIVTLVRDGEDFTE-L----T----P--SIKKQLLTIPVI
402847305_Porphyromonas_sp-_oral_taxon_279_str-_F0450                  -AAPNSLLNYGYAILRAIIARALVGCGLLPTLGIHHHNKYNAYCLADDIMEPYRPYV-DLLVVEICEEEGFPTE-L----T----K--EIKARLLALPTL
345885719_Prevotella_sp-_C561                                          -IPPNNLLNYGYAILRAVIARALVGSGLLPVYGIHHHNRYNAYCLADDIMEPYRPFV-DDLVLSTMKKMEISDD-L----T----V--DLKREMLSIPVL
404487227_Barnesiella_intestinihominis_YIT_11860                       -DPPNNLLNYGYAILRAVVARSLVISGLLPTLGIHHHNRYNAYCLADDVMEPYRPYV-DGLVCEIMESGKDYSM-L----T----T--ELKAKLLSLPVV
301311870_Bacteroides_sp-_20_3                                         -IPPNNLLNYGYAILRAIVARGLVTSGMLPTLGIHHHNRYNAYCLADDIMEPYRPYV-DELVFGLIRTKGISPEIL----T----K--EWKASLLSIPTL
60683388_Bacteroides_fragilis_NCTC_9343                                -IPPNNLLNYGYAILRAVVARGLVISGLLPTLGIHHHNRYNAYCLADDIMEPYRPYV-DELVFSLIQEYGKNAE-L----T----K--EMKTRLLTVPTL
402849996_Rhodovulum_sp-_PH10                                          -AGANALLNYGYTVLRAATARAIVAAGLHPSLALHHRSRGDALRLADDVMEPFRPTV-DLAVKMLLDEGRED---L----D----Q--GAKRRLAGVLHA
154250554_Parvibaculum_lavamentivorans_DS-1                            -EGVNALLNYGYAVVRAATARATVAAGLIPSLGVFHRNRANPFCLADDLLEPYRPYV-DWRVRLLANQMGEEAPSL---------DDRDTRAALLSIFNE
159042957_Dinoroseobacter_shibae_DFL_12                                -PDLNALLNYGYTVLRAATARAVVAAGLHPTIGLHHSNRGNAFALADDLMEPFRPLV-DCCVRGLAARNGPQVD-----------P--AAKQSLARLIAL
288957740_Azospirillum_sp-_B510                                        -DGANALLNYGYAILRSATARAVTAAGLHPSIGLHHANRGNPLCLVDDLMEPFRPFV-DLAVLRLAEAGHRE---V----T----R--ETKRYLALVPSL
330822846_Alicycliphilus_denitrificans_K601                            -DGVNALLNYGYTILRSATARAVVAAGLHPSIGLHHSNDSNAMRLVDDLMEPFRPVV-DLKVWQLLAQGEAH---V----T----P--ETKRALVRVLYD
344171926_Ralstonia_syzygii_R24                                        -EGINTLLNYGYTVLRAATARAVIAAGLHPSVGLHHSNTQNAMRLVDDLMEPFRPLI-DLQVHACWRAGDIALDR-------------ETKKRLADVMYL
319957207_Nitratifractor_salsuginis_DSM_16511                          -DVRNQALNYAYAILRSLITRDLSAAGFLPALGLWHDNRYNAFNLSDDLMEPFRPIV-DAAVYLLLPLCDEEYLTP------------EFKKAIISIFDS
315149831_Enterococcus_faecalis_TX0012                                 -CLPNACMNYGYAVIRAQMARCVVALGLLPMLGIFHKNEYNSFNLVDDLMEPFRPLM-DWYIHQTILKKNEKYLTY------------HSRLTLVEFLHQ
116627543_Streptococcus_thermophilus_LMD-9                             -DVINAGLNYGYAIMRAQMARIVAGYGLNGLLGIFHKNEYNQFNLVDDLMEPFRQIV-DVWVYDNLRDQEFLKY--------------EYRLGLTDLLNA
238924076_Eubacterium_rectale_ATCC_33656                               -ILLNSGLDYGYAVIRGYIARACVGYGLNTQIGIHHKSEYNRFNLVDDLMEPLRPIV-DIVAYNSMKNDKYFTP--------------EHRRKLVSILNM
218563120_Campylobacter_jejuni_subsp-_jejuni_NCTC_11168_-_ATCC_700819  -CFENSALNYGYAIIRACIIRAVCISGLLPWLGIKHDNIYNSFALCDDLIEVFRASV-DDCVLKLKGESEFLSK--------------DDKRALIGNLQS
291276264_Helicobacter_mustelae_12198                                  -SPINAALNYGYAIIRGCIARNLVGSGLIPSLGIFHSNQFNPFNLADDLIEPYRPFV-DSRVLALRIENELTL---------------SHRVELVNLLNT
325677757_Ruminococcus_albus_8                                         -DVTNAALNYGYAIIRSAVAKTLVAYGYNCVLGIHHINEQNAYNLADDMMEPLRPLV-DLWVDNSRDDLFEELT-----------K--TNRRDLVDLVNV
296446028_Methylosinus_trichosporium_OB3b                              -DLRNKMMNYGYAVVRACVARALVAYGLLPALGLHHVSVTNAFNLADDILEPFRPFV-DAIAVARAEKRKKTDDLT---------I--DDRRAMAGALHV
34557789_Wolinella_succinogenes_DSM_1740                               -DARNGGLNYGYAVLRGALGRAIAGAGLIPAFGLHHAGELNSFNLADDLIEPFRPFV-DLLVWGLFLERTGDDP-L----T----R--EERMRIAQILGE
222109284_Acidovorax_ebreus_TPSY                                       -GWSNAALDYGYAVMRGACARALVAHGMLPSLGLFHRSEQNAFNLADDLIEPYRPVV-DLHVAQHRPADEDAELQP------------SDKVALVGLLNV
336393380_Lactobacillus_coryniformis_subsp-_torquens_KCTC_3535         -NPMNGALNYGYAIIRGIIARDICAYGFEPALGVFHKNELNPFNLADDLIEPFRAYV-DLWAYEHQTDLKAG---L----T----P--TNKQALFQLLFT
218767587_Neisseria_meningitidis_Z2491                                 -NAVNAALNYTYAVLRAAVARALTLYGWLPALGLFHRSELNPFNLADDFIEPLRPLA-DLTVIHLYEQGRLKTE-L----T----L--GIKQHLIKILYY
15602991_Pasteurella_multocida_subsp-_multocida_str-_Pm70              -NNINAHLNYAYTVLRSAIARALVLYGWLPQLGIFHRSEVNPFNLADDFVEPFRPLV-DLMVWQLWHTNQLTSG-L----S----P--NTKQKLISLLHH
310780383_Ilyobacter_polytropus_DSM_2926                               -DIKNFALNYGYAVIRGRIARSLTSYGYICSLGIQHKNEYNAFNLADDIIEPYRSVV-DLWVANNIDKLSDELTT-------------DNKVSLVEILNA
312879014_Aminomonas_paucivorans_DSM_12260                             -TSRNAALDYGYSVLRGALARSLAAHGFLPALGIQHHNELNPFNLADDLLELFRPQV-DLWVFRNVSDDDPFSP--------------EHRASLVGLLHR
403744859_Alicyclobacillus_hesperidum_URH17-3-68                       -DFINSAMNYGYALIRGQVARSLCAFGFQPSIGLQHHSELNNFNLADDLMEPYRPIV-DYVVASELDVDSPILSKL-------------HKRRLMQILTQ
421874296_Brevibacillus_laterosporus_GI-9                              -HIINHSLNYGYSILRGNVARVLVQYGFFPCLGIDHHNENNAFNLADDFMEVFRPLV-DLFVVKHFHEEMEFTP--------------KWRAGLYNVVNH
220930481_Clostridium_cellulolyticum_H10                               -DNTNYALNYGYSILRGAVARTIASYGFIPSIGIHHRSELNNFNLADDFIEPFRPIV-DMWVKQNINEDTLLTP--------------KHKLNLISLLGY
225377803_Roseburia_inulinivorans_DSM_16841                            -HIINSALNYGYAIIRGMIARSIICYGLEPSIGIFHHSELNNYNLADDMIEPFRPLV-DLYVSSHFDISEIDST-L----T----P--EMKKGLFGIINF
315659847_Staphylococcus_lugdunensis_M23590                            -LIENAVLNYGYTIFNAAIARTIVAKGLIPAIGIHHIGGRNHFNLASDLIEPFRPIV-DLFLLKYPPEDFLTK---------------DYRVKLINLLHA
160915783_Eubacterium_dolichum_DSM_3991                                -DVINAGLNYGYSIFRSLISSIIVGKGYLPNLGIFHKGKTNMFNLSDDIIEVFRPIV-DKYVYDNLKEELLFKS--------------KHREELIQLTVK
323463802_Staphylococcus_pseudintermedius_ED99                         -TAQNAAMDYGYQVLLAIVARTIVSKGYLTELGIKHHNEFNIYNLASDFMEIFRPLI-DYLVLKNVEQQFET----------------EQKRAILDILNR
389815358_Planococcus_antarcticus_DSM_14505                            -EPRNWALNYGYVVIHSLIARQIVSRGYLTELGIFHANEFNQMNLASDFVEVFRPLV-DYIVHEHITDTFGK----------------DEKRKVLKMLEY
328956316_Coriobacterium_glomerans_PW2                                 -LAINAALNYGYAIILSAVNREIVSRGYLTQLGVCHRNEYNQFNLSCDLMEPFRPIV-DRIVFDNVQDDFTR----------------NDRRLLVDLMNG
422884107_Streptococcus_sanguinis_SK49                                 -NDVNAALDYGYTLILSMFAREIVLSGCMTQFGLKHANQFNQFNLASDIMEPFRPII-DKIVYENQNSSFV-----------------KIKRELFTIFSD
116628212_Streptococcus_thermophilus_LMD-9                             -NDINAALDYGYTLLLSMFAREVVVCGCMTQIGLKHANQFNQFNLASDIMEPFRPII-DRIVYQNRHNNFV-----------------KIKKELFSIFSE
24379808_Streptococcus_mutans_UA159                                    -NDINAGLDYGYTLLLSMFAREVVVSGCMTQFGLKHANQFNQFNFASDIMEPFRPIV-DRIVYENRNNSFI-----------------KIKRELFSMFSD
13622194_Streptococcus_pyogenes_M1_GAS                                 -HPINAGLDYGYTLLLSMFAREVVVSGCMTQFGLKHANQFNQFNFASDIMEPFRPLV-DKIVYENRNQPFP-----------------KIKRELFTLFSD
227824982_Acidaminococcus_sp-_D21                                      -SPINAGLNYGYSLLLSLFNREIVSAGYLTQLGIFHENTYNPYNFSCDLMEPFRILV-DRYVYNMNPTTFAK----------------DEKREIINLFQE
169823756_Finegoldia_magna_ATCC_29328                                  -NPTNAMLNFGYSILLSIFTREITSSGYLTQLGIFHENQFNYYNLSSDLMEPLRPMI-DNIVYFYEPEKFDK----------------EEKIATLKIMDE
320528779_Solobacterium_moorei_F0204                                   -NIINAQLNYGYSILLSAVCREIVKNGYITQIGIHHDNIFNSYNLACDLMEPLRPFI-DLLILKLEHIVLSA----------------DIKQSILTFLNR
303229394_Veillonella_atypica_ACS-134-V-Col7a                          -NPVNAALNYGYAILLSTVNREILSLGYITQLGLNHCNQFNPYNLGSDLMEPLRGFI-DEIVINLMPTEFNR----------------DIKLSLIELLSK
304438953_Peptoniphilus_duerdenii_ATCC_BAA-1640                        -SNINAALNYGYAILLASINRQVVSNGYLTQIGIFHDNMFNQFNFSSDLIEPWRVLV-DNTVYIDDIKIFGK----------------EQKIELINLINE
374307737_Filifactor_alocis_ATCC_35896                                 -NTINAALNYGYNILLSCFNREIVSNGYLTQIGLWHSSRFNSFNLASDLMEPFRIIV-DRKVLTLNLENFDS----------------HNKYELVSILNG
256845020_Fusobacterium_sp-_3_1_36A2                                   -NSLNAFLNYGYQLLLSTFNKEIVANGYLTQIGIFHKNMFNYYNLSSDLMEPFRVIV-DELAYKENPQKFEK----------------DEKRKLQNILNL
306821690_Eubacterium_yurii_subsp-_margaretiae_ATCC_43715              -IVVNSALNYGYAIIASMFSREITSCGYMTQLGLFHDNMFNHFNLASDLMEPFRVLV-DRKVCSMEMEEFTS----------------EEKLKLVNIMQD
42525844_Treponema_denticola_ATCC_35405                                -IPINAALNYGYSILLSAFNREIIANGYITQLGIFHDNMFNPFNLGSDLMEPFRPLV-DAEVFKLNPQKFEH----------------EEKLKIVSVINK
291520706_Coprococcus_catus_GD-7                                       -SLINAALNYGYSVLMSAFAREIVASGYLTQLGIFHDNMFNQFNLASDLMEPFRILI-DREVLMMTLEVFQH----------------SEKMQLVNVLNH
258509198_Lactobacillus_rhamnosus_GG                                   --ATNAALNYGYSVLLSTTNRAIVSTGHLTEVGLHHHSASNPYNLGSDLMEPFRPAV-DYWVAQHRIKELTP----------------DVKYGLIDLLNL
336394883_Lactobacillus_farciminis_KCTC_3681                           -TKINAMLNYGYTILLSATNQEIIAQGYLTQLGIHHHSLKNCFNLSSDLMEPFRPFV-DDKVFQKSDYNFDE----------------YMKLELVDILNQ
339625080_Fructobacillus_fructosus_KCTC_3544                           -LKANAHLNYGYSILLSAVSREIVAAGFLTQIGVHHQSQTNPFNLASDLMEPFRPYV-DRIVKWQRDSPLTL----------------EHKLNLVSVLNE
310286727_Bifidobacterium_bifidum_S17                                  -NGINGILDYGYAIILSEVARKITEIGYLTEFGIHHDNEKNPFNLACDLMEPFRPFM-DERVFQLGAEHLDS----------------KVKVELMRFMRS
366983954_Oenococcus_kitaharae_DSM_17330                               -EPINALLNYGYSILLSTVAREISSNGFLTELGIHHNSQENVFNLACDLMEPFRPII-DRAVAQMTDFDLSL----------------QHKIELIELFDD
#map of conserved positions used for tree reconstruction               -********************************************************_*************_________________************
```

Continued

```
16130662_Escherichia_coli_str-_K-12_substr-_MG1655                     S-------KTLAKLIPLIEDVLAAGEIQPPAPPEDAQ----PVAIPLPVSLGDAHGRSS---------------
11500011_Archaeoglobus_fulgidus_DSM_4304                               RLEKTVKHRKLNRKVSYQRLIRLECYKLVKHFTAVEK----YSPFVMWW-------------------------
18977490_Pyrococcus_furiosus_DSM_3638                                  ELEKTVKHPKLGKGVTRRRLIRLEAYKIIKHLVGVEE----YSPLVAWF-------------------------
46447796_Desulfovibrio_vulgaris_str-_Hildenborough                     RKQDEIVHPFLNERIPLGLVFHVQAMLMARWLRGDLD----GYP-PFVWK------------------------
54296139_Legionella_pneumophila_str-_Paris                             PL------PYHKRRISLESIIQRQVYHLSAVFSGEGK----YKGYLFKW-------------------------
331001028_Parasutterella_excrementihominis_YIT_11859                   TMST--KINYSQKQCKLENALQRQVYKLAGAIVEGKR----YRGIRFKW-------------------------
319941582_Sutterella_wadsworthensis_3_1_45B                            TMAN--AVSWQGKSVRLENALQRQAYRLAGTVVDGKR----YKSIRFRW-------------------------
118497353_Francisella_novicida_U112                                    TIAKKY--IYNGKKLKLEHIIQRQVYRLSGEFAGEHN----YKPYIFKW-------------------------
34557933_Wolinella_succinogenes_DSM_1740                               TMNK--KHLYNGKNLSLQGVMQRQVYRLVGAFMQQNR----YKSYIFKW-------------------------
187736488_Akkermansia_muciniphila_ATCC_BAA-835                         SV------MYRDKQLPLKAAVEAVCRSFRKAVLAGQS----EPYEPWLMTTIKWAG------------------
407803668_Alcanivorax_sp-_W11-5                                        TV------CLNGQDFRLPAAIRKTVASYLRILEAGGG----GRRRKLELPECP---------------------
325972002_Sphaerochaeta_globus_str-_Buddy                              EV------VFEEQDVEFSYALRLYVLSYLSFLSRNRD----AIQFPKYRYAWAV--------------------
187250661_Elusimicrobium_minutum_Pei191                                NL------YFEDGAFNLTAGIQKYIQSYIAFLEEREN----RIIFPAILK------------------------
47458867_Mycoplasma_mobile_163K                                        KV------IYKGKYFYVNDTFDQIADNLVNKRSWEWV----ELWD-----------------------------
363542551_Mycoplasma_ovipneumoniae_SC01                                HI------EYKGKHIKFSKYMEFEIMDILNLKENYKE----ITTESDQS-------------------------
71894593_Mycoplasma_synoviae_53                                        KI------LLNKSPISVNEYICKLIENFINKDFNFES----LEIDWNKK-------------------------
384393287_Mycoplasma_canis_PG_14                                       KI------AIKNTKLSVNEYIDKIVKSVINNNLICED----FIIEWS---------------------------
284931711_Mycoplasma_gallisepticum_str-_F                              FI------FPNGKN--LNQVIDMLIKAIINNKINERD----FNIDWS---------------------------
117929157_Acidothermus_cellulolyticus_11B                              QF------DASGATVG--TAVERFAQQVGRYVEGEIR----SLRPPAMELSHA---------------------
317482065_Bifidobacterium_sp-_12_1_47BFAA                              Q--------FNGSGLTIPSSLNDFAQQFGLYCEGKID----RLQVPEYVGES----------------------
315605739_Actinomyces_sp-_oral_taxon_180_str-_F0310                    --------RFDGDGHGIPAVAEALAQSFGRYVEGDID----RLCVLAWQGPSSVAVED----------------
```

```
294086112_Candidatus_Puniceispirillum_marinum_IMCC1322                DQ------AVVGATSPLFQHMSQLSYSLWENLDGQKT----TMPMPQLIPELEIEAMVSEC--------------
282880053_Prevotella_timonensis_CRIS_5C-B1                            DT------YFINVTRPLSVGLTMTMASLVKYYAKEVK----KISLPLLK--------------------------
374384762_Odoribacter_laneus_YIT_12061                                DT------YFENTTRPLMVGLSQTTASLARCFSGEQK----KMVYPEFK--------------------------
423317188_Bergeyella_zoohelcum_ATCC_43767                             DV------KIDEQVRPLLVAVKSTVSSLYKCYTGEKR----QISYPKML--------------------------
385811610_Ignavibacterium_album_JCM_16511                             DV------FIEEQKRPLMIAVQRTTSSLAKCFSGNQR----KIIYPVFRL-------------------------
402847305_Porphyromonas_sp-_oral_taxon_279_str-_F0450                 DV------EIEGQRRPLMLAATQTATSLLRCFTGEAR----RLAYPLLSCEPSA---------------------
345885719_Prevotella_sp-_C561                                         DV------VIGGKRSPLMIAAGLTTASLAKCYNGEAR----EISYPSFL--------------------------
404487227_Barnesiella_intestinihominis_YIT_11860                      EV------GIQGRRSPLMVAVAQTTASLYKCFNGELR----KIIYPEL---------------------------
301311870_Bacteroides_sp-_20_3                                        EV------KIGGKRSPLMIAVAQTTASLYKCFSGEQR----RIVYPER---------------------------
60683388_Bacteroides_fragilis_NCTC_9343                               EV------IIGGKRSPLMVAVGQTTASLYKCFNGELR----RISYPER---------------------------
402849996_Rhodovulum_sp-_PH10                                         DF------ATSDGRTPLSNVLARLAVSLAQVFTRERR----GLAFPETLVPLPTAEDDGSANETAEHAVGVSDLV
154250554_Parvibaculum_lavamentivorans_DS-1                           TV------LVGGRRMPLLLALHASAASLCRALTGGEA----ALALPEGMPLAPDLLDNDGEG-------------
159042957_Dinoroseobacter_shibae_DFL_12                               DL------PLGDSLTPVSVALGKLAISLGQSFESGTL----DLALPAPPDALTLAGLGA----------------
288957740_Azospirillum_sp-_B510                                       DM------LTAQGTTPIGTVLIRLAGSLAQVVEEGKG----DLDLPALDRPWAPMPLEWPAAPPAGGETGC----
330822846_Alicycliphilus_denitrificans_K601                           DM------QTSVGVTPVMVCMQKLAISLAQVYLGERT----KLDLPLQGLPLALAGSSESD--------------
344171926_Ralstonia_syzygii_R24                                       DM------PSPQGITPVMGAVQRLCTSLAQLYLGERK----ELELPLCPTPIDLAAMFVPLA-------------
319957207_Nitratifractor_salsuginis_DSM_16511                         EYL-----IYDAGRSTLRTVATRYIQGFKAVVRSRKI----ENLHFPTIDFGALRECF-----------------
315149831_Enterococcus_faecalis_TX0012                                KI------KVKNKKIYMNQAMSEYVAAFVKAMETGDF----ENVQKIHLVNMLECEGK-----------------
116627543_Streptococcus_thermophilus_LMD-9                            KI------KYGKETCSVTVAMDKYVKGFIKYISEKDS----SKFHCPVVSSLEWRK-------------------
238924076_Eubacterium_rectale_ATCC_33656                              KI------MYREKKMYMCNMIENYIEQYASLIMKRTD----NIVFPDVYNLIGEEYG------------------
218563120_Campylobacter_jejuni_subsp-_jejuni_NCTC_11168_-_ATCC_700819 KI------NFDGQNYPLNRAINHYVANFKNALLYEDE----LKIVKFDD--------------------------
291276264_Helicobacter_mustelae_12198                                 KV------LLQGKFYPFHRSITRMVWSISNFICDRGK----MIELPILDKESNGREVYESFSDV-----------
325677757_Ruminococcus_albus_8                                        PI------KMGKKKTRVRYAIDKYISSLTSAIEKNDP----GLLQLPTLIQIDEMFEDDEDG-------------
296446028_Methylosinus_trichosporium_OB3b                             NA------RVGKEIVSLLVASEKAAEGLVQAMESGSA----AILKLPFFVTPKGSSP------------------
34557789_Wolinella_succinogenes_DSM_1740                              SC------IMTERYETLLSATQIIARSLVSAGEQKES----RRLI-LPWMGEDDVQGEVYEADGDV---------
222109284_Acidovorax_ebreus_TPSY                                      DV------AMPRGQMSVLASIEQAAESLARLYDGGSE----QVLELPRLIGLRQHVFEI----------------
336393380_Lactobacillus_coryniformis_subsp-_torquens_KCTC_3535        EV------RIDGEFHSLSNAVHKLVSSYTTCCRNNSA----AGLKLPELIPLVAHEYE-----------------
218767587_Neisseria_meningitidis_Z2491                                QT------SIERQHFSTLAAIDKMISSFQAGVTDKNA----KQLKLPEILPLKEYQYE-----------------
15602991_Pasteurella_multocida_subsp-_multocida_str-_Pm70             QM------SFQEETFSVLAAIDRTVGSFQAALTTPNA----NLLKLPEILPLKEHQYE-----------------
310780383_Ilyobacter_polytropus_DSM_2926                              DV------KLTGKKYSVNISIDEVIKRYTKCIRGKGG----NLILPSIIPIQEHRYV------------------
312879014_Aminomonas_paucivorans_DSM_12260                            RM------EIQGERHTILRGTELCAEGFAAACREKNP----DLLALPALISPDG---------------------
403744859_Alicyclobacillus_hesperidum_URH17-3-68                      MV------WLENERYDLTASIEAMVSSFSNAISTEDP----SALQLPRLTEVQMGVYE-----------------
421874296_Brevibacillus_laterosporus_GI-9                             QV------ELEGERYSCNRAIEEMIKSYVTACRTSMY----ERLRTPTLLPLKIHSYE-----------------
220930481_Clostridium_cellulolyticum_H10                              EC------VFEGKIISIRSAIEKVISSFSSSCAKNDY----SLLKLPEIIPLEVHANE-----------------
225377803_Roseburia_inulinivorans_DSM_16841                           DM------SVKGDMRIVSNCIDMLIASYSSALQGNRT----ELDLPELIQLQVHSYE------------------
315659847_Staphylococcus_lugdunensis_M23590                           RI------EIDGKMQTIIRAIEIMIQSIIEYFKNGHL----DIVKFPNLRKYQFYEL------------------
160915783_Eubacterium_dolichum_DSM_3991                               KI------EIDGKLQTVPNAVNEYVESILKVIETGKI----TDFKFPSPHVVDYDL-------------------
323463802_Staphylococcus_pseudintermedius_ED99                        KL------KIDNKKYYFLNAVEIYVDSLFKYLSNGNR----EVIKIPTWTY------------------------
389815358_Planococcus_antarcticus_DSM_14505                           KV------QIRNSKQFLQAAVQVYLDSCLQYLNDGHK----EKLLFPILDFPKAR--------------------
328956316_Coriobacterium_glomerans_PW2                                CI------AYRDGTYKTSSVISLYVQDCLNALNKRLA---IDEIDPFDIV-------------------------
422884107_Streptococcus_sanguinis_SK49                                TF------HYNGKDMYLSNIVSDYTKKVIQALNQPEK----GVPEFRI---------------------------
116628212_Streptococcus_thermophilus_LMD-9                            TY------LYNGKEMYLSNIVSDYTKKVIKALNQLGE----EIPEFRI---------------------------
24379808_Streptococcus_mutans_UA159                                   TY------LYNNKEMYLTNIVSDYTKKVIKALNNDGK----GVPEFRI---------------------------
13622194_Streptococcus_pyogenes_M1_GAS                                TF------SYNGKEMYLTNIISDYTKKVVKALNNEGK----GVPEFRI---------------------------
227824982_Acidaminococcus_sp-_D21                                     IL------YIDDSRQFLANAINIYANSIFQALEKCDT----SKIKIYEI--------------------------
169823756_Finegoldia_magna_ATCC_29328                                 EV------FIGKTRYTLLNAVNEYCKSVFLALEKGEE----QLIKFIDYEL------------------------
320528779_Solobacterium_moorei_F0204                                  KV------SIMGREQYMLNAIEIYVKSILDAISSENP----SLIKFVEYEFTCYESDSVL---------------
303229394_Veillonella_atypica_ACS-134-V-Col7a                         DV------KINDTIQTFSAAVRIYCKSVFDALRVQAA----NEIRFIEYEL------------------------
304438953_Peptoniphilus_duerdenii_ATCC_BAA-1640                       FV------TIDNKYCRINQALEIYIRSVFDALNNRDI----SLLRFPSYEL------------------------
374307737_Filifactor_alocis_ATCC_35896                                EV------RIDSKVSYVNNAIKLYCKSIFEALHYEDT----SYIKFYTNEL------------------------
256845020_Fusobacterium_sp-_3_1_36A2                                  KF------RINNVNHYLSDIIKIYTKSIFDALSANDL----SLVRFFEDEL------------------------
306821690_Eubacterium_yurii_subsp-_margaretiae_ATCC_43715             SV------IINGRTEVLTNAVKIYCKSIFDALEQKDI----SLIRFYENEF------------------------
42525844_Treponema_denticola_ATCC_35405                               KV------LINNKEHYLNKAIEIFVHSIFDALNEKDI----SQINFYRNEL------------------------
291520706_Coprococcus_catus_GD-7                                      EV------RIDGKVQYVNNAIKIYCKSVFDALSEDDS----ALIRFYKIEL------------------------
258509198_Lactobacillus_rhamnosus_GG                                  EI------TFNGKKALLQNAVADHVENCLAFLDGKAKVCRIEVKFPDEVSSHAINDHV-----------------
336394883_Lactobacillus_farciminis_KCTC_3681                          EI------SYGSKTFILKNALSQYIRDCISYLDSNQE----VEIKKVGFKDEE----------------------
339625080_Fructobacillus_fructosus_KCTC_3544                          EI------TMNSKRYELAQAVQKHVQECLKYLCSDDS----ELPEMDFEL-------------------------
310286727_Bifidobacterium_bifidum_S17                                 D--------VPELGTTITQLISVFVRDALRYLSGEEC----LPELGFVRCGTELCE-------------------
366983954_Oenococcus_kitaharae_DSM_17330                              EV------GYLQHQATVNSAITEYVRNALDYLCGKRP----EVAVSVAI--------------------------
#map of conserved positions used for tree reconstruction              *-------*****************************--------------------------------------
```


Cas1

Supplementary Figure S6. Phylogenetic tree of representative Cas1 sequences. The tree was reconstructed from selected, informative positions of the Cas1 multiple sequence alignment (see Supplementary Figure S5 and Supplementary Materials and Methods), and was rooted to selected Cas1 orthologs of the type I CRISPR-Cas outgroup. Shaded boxes group Cas1 sequences associated with three major type II CRISPR-Cas subtypes. Colored branches indicate the CRISPR-Cas loci with similar architecture. Compare with Figure 4. Note the similar topology of Cas1 and Cas9 trees (Figure 4).

Supplementary Figure S7. Analysis of Csn2 orthologs. (**A**) Complete multiple sequence alignment of selected Csn2 orthologs and selected ATPases. See Figure 5. Underlined, regions of homology with short Csn2 ortholog according to HHPRED. (**B**) HHPRED hits for selected Csn2 orthologs representative of five families. See Figure 5. (**C**) Selected proteins with structures similar to long and short Csn2 structures (3zth and 3toc) according to PDBe and their structural alignment with long and short Csn2 representatives (3toc and 3zth). Walker A, Walker B and signature motifs are highlighted in boxes. See Figure 5 and Supplementary Materials and Methods.

A

```
3ZTH_SS     ------------------EEEE-------EEE-E------E--EEEE---HHHHHHHHHHHHHH---HH-----HHHHHHH------EE------------------EE-EEE--HHH----HHHH--------HHHHHHHH--HH-------HHHHHHH-HHHHHHHHHHHHH-HH-------EE--EE--------------HHH--HHHH-----EE
3ZTH        -----------------MKFFVQHPYKERIEL-NI--G-AI--TQIVGQNNELKYYTWQILSWY---FGGKKYSSEDLSIFDYEEPTILDEA-REI---VKRSS-Y-HY-IDISSFKD----LLEQ-MEYKKGTLAQGYLRK--IVNQ-----VDIVGHL-EKINEQVELIEEA-MN--RHINLNCGQVEYH-LE-NL--P-LTLDQ--LLTKNFSPFFA
365905381   -----------------MRLKIEIDDQSFFEL-DF--G-DI--VYISGYNHQNMWKIYRSLYYY---FNKSPM--LSTNIYGDDNIEISLDD-VSP----STKN-T-QT-YFISNRDS----IYQQ-MIYKKGNLLFENLNT--LNDS-----FEITRIL-EELNDTNTKLSLL-VQ--EHINNYANNLIID-FE-DV-----TYFD--VLKYFLKVSYE
406671115   ---------------MMQEFRIEYRKGTFLEF-EI--D-DY--VIITGNTYQLKREVNYALERL---FYGKTLSDMEMNYFGTNGI-LTIDG-KNI----SDKE-A-CY-LSITQFSS----LEKV-YSIKKGSLLYDRLIR--LSQE-----KIITQQL-ERVNDELLRLEDQ-IN--LLAGMYLKNITIN-IN-QF-----EFEE--LIKSYSSYIYR
422729713   ----------------MKQIRLEYQYQQFIEF-SL--E-DY--VFFYGGENQWRRKILRTLKRF---AKQKALSDLEESVYGDNGIEIFLEE-KKL----KAKS-I-DF-YFLEDNTS----IFQQ-MNFIKGTLLHQELLY--LQDD-----FDITNQM-SNLNDQLLKLENL-MN--DKLNNYLTNIEPH-FL-EM-----NFQD--ILKYSLSLNYL
406838713   -----------------MNLKIEYENNQYLNF-DA--G-LI--NFFLGPNSEIKWKLFRGLQRF---NMGKALNELEENVYGEDGISIECDG-KKL----NSKN-N-LI-MILDSRDS----FLKH-YTLSKGSLLLHYFEN--LKND-----FDLNKSI-YELNDKILNLETI-FQ--DKIENELHNFVPS-LH-AI-----EFDE--ILKKLLIINYL
238924078   -----------------MNVEIT-EFEKRFCF-EL--C-PV--TQLCGQNVRKKNYIFESLRRY---FGTFKYSESKNKWRDN----VFIDN-NQV----GRKF-F-SV-LSISNKTD----IIQM-IKMTKQSLLMEYVKN--VIQD-----FDWQLHL-RNLSEEIEIMFQI-IN--DQVNK-VGDIELS-YS-MA-----DVWD--MVQKSEVSGID
381184142   ----------------MSTLTFSHPKLKHIEM-GF--S-QL--TQVVGDNTELKQQLWQNISWY---LSKHKYSESELSILNYNEPQIWEDG-LDL----SRSK-F-RT-IIIESISD----IHEL-LGANKGTPTFDVLNQ--VMKN-----LEISKNI-EKINQEIIQIENQ-LNASEEIQQLNGEKKFK-WK-VA--V-EEFNSKILLQKNVSLMPT
116100822   -----------------MKFFVRHPYKDRIEL-NI--G-AI--TQIVGQNNELKYYILQILSWY---FGGKKYSREDLSIFDYEEPTILDEA-REI---VKRSS-Y-HY-IDISTFKD----LLEQ-MEYKKGTLAHGYLRK--IVNQ-----VDIVGHL-EKINEQVELIEEA-MN--RHINLNCGQVEYH-LE-NL--P-LTLDQ--LLTKNFSPFFT

4I99_SS     ----------------------------EEEE--------E--EEEE-----HHHHHHHHHHHH---H----H----------------------------------------------------------------------------------------------------------HHHHHHHH---EE-EE-EE-------HH--HHH-EEEEEE-
4I99        ----------------------------VVIP-FS--K-GF--TAIVGANGSGKSNIGDAILFV---LGGLSA----------------------------------------------------------------------------------------------------------SEIFAKLSPGGSA-RL-IL--E-NPEDP--FSGGLEIEAKP
2ONK        ----------------------------NVDF-EM--GRDY--CVLLGPTGAGKSVFLELIAGI---VKP--------------------------------------------------------------------------------------------------------------------------------------------------KLGI
2O5V        ----------------------------PGTL-NF--PEGV--TGIYGENGAGKTNLLEAAYLA---LTGQTD----------------------------------------------------------------------------------------------------------ETYAADLRGRRAE-EL-AR--G-STVTG--PHRDDLLLTLG

3TOC_SS     ------------------EEE----------E-E------E--EEEEE--HHHHHHHHHHHH--------------------------E----E---------H-H-HE-EEE----------------H---HHHHHHHHH--HHHHH---HHHHHHHH-HHHHHHHHHHHHH-HH--H------------------------HHH--HHHHH------
3TOC        -----------------MNLNFSL---LDEPI-PL--R-GG--TILVLEDVCVFSKIVQYCYQY--------------------ESELK----FFM---KTIKE-S-EI-MLVTDILG----FD-V-NSS---TILKLIHAD--LESQFNEKPEVKSMID-KLVATITELIVFE-CL--ENE--------LD-LE-YD--E-ITILE--LIKSLGVKVE-
13622196    -----------------MNLNFSL---LDEPI-PL--R-GG--TILVLEDVCVFSKIVQYCYQ------YEEDSE------------LK----FFDHKMKTIKE-S-EI-MLVTDILG----FD-V-NSS---TILKLIHAD--LESQFNEKPEVKSMID-KLVATITELIVFE-CL--ENE--------LD-LE-YD--E-ITILE--LIKSLGVKVE-
339625078   ---------------MKNEQISIE---TVGVL-KM--N-SG-LNLITVANSELYLRLIAKLS-------NDIEDD------------II----YSYDNKIKPLC-K-SG-LLIGDPVS----GR-D-IRG---FYSKFVPKI--LAQSIS--EEGIHELF-QLNQKIQSVIASE-II--DNN--------LP-FE-IS--QGWELER--LLHSQNIVVE-
333398276   ---------------MNYSIVCFP---NN-PI-KI--D-SG-LTMISIANHELYWQLSTGL--------KETSED------------VT----LGQNGHIVDVN-K-TV-MFIGDIGA----QL-D-VNK---LYQKNIVNT--MKNYAD--DQALEVFY-KINSELNHVLENI-IL--ENN--------LP-FY-FQ--TEFNIVE--LINEKKIRIE-
372325148   ---------------MNYQITRYP---FD-PV-AL--D-TG-LNILSVEDTQLFFDFCQDF--------QSEEGK------------TI----ITSGGKQLVMQ-K-NM-TFFGDVTR----VF-D-FNQ---LYLKAAIKT--LSKNFD--EMDLQQLD-DLNGQINRLFTKV-IF--SSN--------LP-ID-LN--LEFDLME--TIKSRKIVID-
423349697   ---------------MISRITIPP---LE-PF-DL--D-KG-LNIISTNSLSVFFSLSQGIKG------KNEMMH------------FY----DDDNNRIEKLE-S-EI-LLVGDIAG----ML-N-YDE---WFKKQLLSA--YIREFD--DEQLAALH-NIHAQLFSVLDKA-AL--DSM--------LP-IA-LN--QEFDTKN--LVSIFKPTVD-
310286726   ---------------MIDSLVAYP---LD-PV-QL--E-MG-LNVIGCASPKIFYELSAGL--------QEDADS------------VV----LGVGEKTVLAS-K-SI-IYIGDCAD----EV-D-YNA---LFQKLAVKK--IIEGFT--PEQISRLL-SLQAELKTIIQDE-IW--GQN--------LP-LT-IP--CDLDLKA--AIAMAKPRID-
336394885   ---------------MTNTITVFP---YK-PF-EI--N-DG-ITLLNFQNIDEYSNLIFRINKL---KNGETFSDDTNRIIH-----FY----TNDDEPIEEHL-K-DI-SFIGNLSS----FN-L-NAS---SQMRIILKK--IIEDSQ---IMLNEIE-KVNADLNSVVFDS-IT--RYS--------LP-LT-LD--MYANYEK--LLKSKDIKID-
408790125   -----------------MNLTVYP---QA-PF-AV--D-SGKLTFLQTNNPKFYSTIVLG---------CKNLDK------------II----ISENDEILELS-S-HL-AYLGNGID----NV-D-LIT---AFKTKLEIE--VIENLT--ESERQRLF-QLDREVKSIFLGA-AY--LEE--------LP-LA-ID--DEWNLKK--MAKYCGINIL-
258509196   -----------------MNLAYYP---FQ-LI-TT--K-PSVITVIDTGLPKVLLDLVQAFR-------NDDDGI------------VL----SNEQMQIQELR-K-AS-LWIGDPVL----EL-D-LDK---LFQRLIIKK--VSQLIE--DRRLIELI-EQSQRLAMSLLRE-PL--LSE--------LP-VT-VE--PGGKLEQ--IMKYCNVHFD-
260662217   -----------------MKLQYHS---HQ-PI-EM--T-PGKVVILQIENPSVNTELIQGLQ-------GYNDCL------------RM----IDDEYQELNLS-K-VM-KWEGDLVL----NP-T-KLD---SYQTALDKM--IVKNLS--IQKRDRMN-DLARGVFSEMQDS-LF--EFD--------LP-LE-VR--YDESLLR--LYKYAKIKYL-
354808133   -----------------MNITFYP---YH-PF-SV--Q-NKKITVCSTSSPEVYTKLIIGMS-------GAQETV------------KF----SNDQYELMDYS-K-AV-IWGGDVAA----IN-L-N-H---LFQNRLIKK--FAADLT--DQQRQKLT-QLDSEIRSTILDA-AF--MYE--------LT-LD-VN--QEWDLQR--MIKFYNLQFS-
392947439   -----------------MNLTYYP---FP-PF-KI--N-TGQVNILDIGNSQMYADICQGFR-------DRTDDL------------RV----SNDNLELQTMS-S-RC-SWYGDLML----SV-D-LNH---LFMKKIQAQ--IIKLMA--DEQQVAVL-DQSRAMIAKIMEA-SF--LLD--------LP-LD-VE--EAPGLEQ--ALKFAGIHFP-
333394443   -----------------MKLIYYP---FK-SF-TI--D-AGVPTILETDNHEVYRNLLINFT-------DMADNL------------NF----SDDTLELISTT-K-AL-KWFGDASI----CT-D-LNK---LFLPQLYKQ--IKLNLT--AEQNRIIT-DQSRTLKNVVLEA-TY--SLD--------LP-LN-IG--DQIATEK--VLKFCDLHFD-
323340065   -----------------MKIAYVT---HD-SL-DI--E-SGKVTAICTASKSVYLEIIRSFN-------NADDKL------------FL----CSDDYKEMELK-D-GI-DWEGDVME----ID-V--GK---KYLNELLKK--SVEFMP--DELRKQVY-ELERQTATALQDS-IY--MFD--------LP-LI-VS--MDNDLKK--TLKNYHFSVD-
380344195   -----------------MKLYYTS---HK-NI-DI--N-TGKITVLGTNNVNVYKEFIDTFLN------GYGSNI------------QL----SDDKYNRKDIS-T-SI-DWDGDVML----TD-R-IGK---KYMNILIKK--IIENLT--DDERQAIL-KSVNSLYDRIREV-LY--KID--------IP-LQ-VD--YDNDLTR--LFKYCQVHTE-
328956318   -----------------MRIVFSG---LERPI-DI--E-PGLPTVLQVENSSLFTRLCLSLRSG---EEREAIEP------------YV----IWRGEKEIATR-D-AL-MFIGDPML----LP-W-DDR---ALMGSISKR--FEKELIEDEELRRTLE-DAAVALSSQVASL-GL--SMY--------GD-YS-FA--IEWDLKR--FIKAFGFGAH-
374307735   -----------------MNLVYYD---LNMDI-EL--E-EERLNFLVIENPVKLEDFVLSIHSA---K-QKITED------------VS----IFEKFEKINFA-K-QV-DILFSPMS----IS-Y-HTK---EIQKKLLEY--VLEKVQ-DSYLAEDFI-DISGSLVSLCSKI-YL--ELE--------YD-IE-YL--DYVEVED--ILKLMDIRLK-
224543315   -----------------MKLRIMG---FE-EI-DF--N-DQDNTILIIKNHKFYAHILSLLSSI---LYDKNKST------------EL----LVIDGEK-NIT-T-NV-LLITDLYK----ID-F-NDK---NLLKEIYSE--ISKSITSDEEKCMRYQ-ELVDDLVAILENE-LE--DYN--------LE-FD-YK--NDLTIEN--YLKAVSLKIS-
227824980   -----------------MKLLIPD---YQINL-TW--K-DYSVPTLIIENPHQFFEIINDLVLQ---IKEGNNGK------------AI----YSDKDKPVPLE-K-AF-EILIDYLS----ID-S-VPK---MLLTKIQKR--MEFQAT-NGPLFENTQ-ILLAQMEHLAEEI-AM--SLD--------LE-ID-CE--K-LTVGN--LLKALGITIN-
169823758   -----------------MNIKYDL---LNEPF-FI--D-YGDICNITIENKLEYRKIIENIY---------NQET------------IE----LYDDDELITKH-K-----VIENSAI----LD-F-DDK---KITTKIIKD--LMEIIS-DEKHYEQKI-NLNSSIVNFINDI-TL--DYP--------YM-LE-IG--DEINFES--LLKVVPIKPS-
303229343   -----------------MKLMYQD---SILEF-EV--L-RDKVTVVSFENRKVFRRMVTELDIQ---C-DSGVGP------------WI----LNDNDKAFNLD-K-YS-HIILNPLY----VD-V-NSK---SLLTKLQNQ--LTKEAL---LMTEEVA-DIVNRLHAFYYSL-EF--GYP--------LS-IQ-HK--MEMGAAE--IIKLGSFNFE-
315926106   -----------------MILVNEK---YDLKI-EM--R-ENRSCCLVVEKPEYLADMMMDLIAQ---S-QGDEGG------------FV----LSDHGSLLPIK-H-YL-DLQTDIFH----LD-L-NQR---KVMKGIYAA--LEEVAQ---EEVEAKG-EITTAIISVLERV-TS--RAA--------YDNLT-FD--FDFEWEN--LFKFFHLRIE-
306821688   -----------------MRLVYQD---YNFAA-NI--S-QDEMTVIIAENPKTFCDFVSNIREM---I-DGKGDK------------VL----LTDNSGAMIST-D-KC-DIILSPMD----FQ-L-NSK---KILNSIYKE--ISQIVK--EELWMQEQ-ELNNAIMDFLDRA-VL--KLP--------YE-LI-YE--KNIDTQD--FLKLYKLNLL-
320528781   -----------------MKLANSK---FFETL-DF--E-KG-SQTLVVENVELLRTIILQLKFQ---I-NNKIGD------------FI----LSDKDAILDIS-K-NI-LLITDVFE----IS-S-LSK---QLKNKLQQY--VESNYD-NDDLYQDLY-Q---KLIDFGNDL-IN--SSP--------YP-LC-FN--QSITKID--MIKLLDIQLE-
342218249   -----------------MKVLYAL---SDMEI-EL--S-ENYVATLYIENMLYFRHILQELAAQ---I-AGAEGH------------WV----FTQAQQQLDCK-K-DV-ILVTDLFH----ID-G-NQK---SVLSKLHAY--LTSIAE---EESELLY-EINDKLRSLYTEI-TL--QAP--------VE-VG-YK--AALSAQD--LIKAGDFKFE-
227501315   -----------------MILTSEY---LEKSF-EI--K-EQRLNTIVLENKIYYRKIIQAVRDT---I-NLDLKE------------FV----LYEDEEEVKLN-K-IA-ELITNVFN----IN-P-NDN---KILAKLYEK--LGQEIT-YEDIYCYNR-NFKSALSKLLEEL-SM--LSE--------YD-FK-YE--D-VDYKY--IFKAVGLEID-
304438951   -----------------MNLVTKY---LENSL-EI--K-SKTINTLVVEDIRYFSIFLKGLIES---T-EIESDE------------FE----LIEDYKTVDMT-K-YV-EIIFDIFN----LE-A-NSA---NILKKMYSE--LEEDLN-TQEVYTKKV-ELESIIANITDEL-IY--RSR--------FS-LK-AG--G-INYQN--LFKAIEIEFD-
34762595    -----------------MTFQYKG---FNFKI-DF--E-EKNIFSLIVENKRAYRKIIEDLVNN---S-NIEDGD------------II----LSKNNKLVIPE-K-EI-FIFSDYFN----FD-I-N-K---FVLNKYYKE--LKNLSE--NEFLNETL-EIKEILKDYINKL-I---ENN--------YS-LK-FE--DDLDVSQ--ILKAFSIKFE-
241889886   -----------------MKIINKN---WQRKI-EI--E-DNIIYTLVFENKKYYRENIIELVNQ---H-KSNEGS------------YI----LSNDNKEISFD-K-NS-YIITDIFN----ID-I-NNK---KVLSKIYSS--LLKEIV---EDISSYN-ELSTNIRVYFEKL-IF--NSS--------LE-IE-QG--EEIDMSS--LLKLGDFKIH-
291520707   -----------------MKMVHSE---LSGEI-IS--E-SAAFTEWIIESPKDFSEYLHELYCQ---Y-EKDEGR------------FV----LSQRDKILELS-K-YL-EIIVNPFT----VE-I-NSR---KILNKLYIK--LEKVSK-EEQMYEETL-KLTAYIHEYLMQL-EQ--QTN--------YI-LS-FD--DNLEIPA--LLKAVGIKHE-
42525846    -----------------MKLVHPD---IKNQI-IF--S-ENKVNVLTIENKTFFTEFVSELFSQ---Y-SGNEGK------------FI----LSEDDKELSIP-K-TC-ELILNPIN----LD-I-NNK---QNLNKIYSI--LKQSIS-DEYNYLKTC-TIKSEILKHIDDV-LL--SLD--------IP-LK-RS--EELDLIS--IFKAVNISFE-
299144349   ---------------------------------------------MYIEDKKYFRKIITELNEQ---I-NGNIGK------------FI----MSEDDREILIS-K-NV-DIIIDFLN----TN-P-NNK---KVINRLLSE--LTENAK---IEYEKLN-IIINQLNIFFDDL-LF--SGD--------FE-LE-RE--DDIKVEE--LIKLGGFYFN-
116101487   -----------------MNINFSL---LDEPL-EI--N-QG--TVLVIEDVSVFAQLVKEFYQ------YDEQSN------------LT----IFDSKIRSIRS-S-EL-LLITDILG----YD-I-NTS---QVLKLLHTD--IVSQLNDKPEVRSEID-SLVSLITDIIMAE-CI--ENE--------LD-IE-YD--E-ITLLE--LIKALGVRIE-

MYCO        ----------------HHHHHHH-----------------E--EEEEE--HHHHHHHHH-----------HHHH-------------EE-----------H----H--E-EEE---HHHHHHH-----HHH--HHHHH-HHH----HH-HHHHHHHHHHH-HHHHHHHHHH--H-HE--E-----------H------------HHH--HHHHH-H----
350546883   --------------MNFLNILYNN---KNYEI-SF--N-GI--KIFQVRNPNKFSIDFI-----------NQA------KG------SR----FNDVNIFK----N-NI-ILISDFTKINDLIS-L-NKSS--QIFKE-IIN--Y-IE-TNKILGNDKIH-NIKEYINSKIGFD-VI--EEK--------NE------------MQK--IISNL-FEIN-
193216855   -----------------MMIIRTN---PECENPNF--K-GI--VTIATSDSDAFLRELY-----------LWEINNNDNAA------FS----IEHKDIYL----K-DC-FIFSQLTKSYELFN-L-SAKN--FLTKR-IYDDEL-WD-DSKVINFSVLE-KICENINENIEED-FL--RIE--------SD------------FNK--MLKTF-FKVN-
47458866    -----------------------------MMKINF--A-EN--KIIYTNN---YFEFLL-----------NLE--------------------------NI----N-DF-LIINPFSKVSEILN-F-KSKS--FFNNF-FMDEYS-IN-FLNLISHEKIN-EMSNDINQKLGFE-FV--STN--------LD------------LEK--NTKSF-LEID-
385326557   -------------------MKIKI---HNQEIDNF--N-GL--LLIDVKNTSEYLKRLF-----------MYE-KQHETSV------FE----INNVNVDI----S-DC-LIITPFSKYSDLIS-Y-TAKN--VF-TK-LLGNIN-FE-HDKILNEEYLDKEVVAKLNETLGRD-II--SLD--------TS------------YSK--ILKSI-IKIS-
490549417   -----------------MWNFFND---INFNLEAK--Y-NF--KIVEEDNTDYLIRNLI-----------DYE-NNTSKKS------IR----INKNTFSI----R-DC-LIISNLSYVSDILN-S-ISKS--WIQEK-INNNEN-WN-TNIILDQEKVE-TIIEQINLKIGFE-YL--FKD--------ID------------NNK--LIKSI-FTIN-
363542553   -----------------MLKFFNN---LNNTVQLK--T-NV--KIIETENTDLFLSHLF-----------NYE-YESESKA------FE----LMSRKISL----K-DS-VLITNLTKFSEFLS-L-SAKN--WLGDS-IINDEY-WS-ESIFFRNEKID-GIVDKINQKLGFE-FL--NYE--------ID------------NVK--LIKAI-FNIE-
71894595    -------------------MQFLN---LKGLVKNY--N-GI--VCIETNNSDLFIRKLF-----------EFE-HAENQSS------IN----INNNKYSI----K-DF-IIIDNLTKYHDLYN-F-NSKG--LL-NQ-WINDLD-FE-NQKIANEKLVL-EIKNLLNNKIGFE-FV--SIE--------EN------------NSK--YLKYL-FNLE-
419703977   ------------------MLKIQI---HNKLIENF--S-GV--ACINVDNTTAYLRKLY-----------EYE--NDNRKV------FN----IGGNEISI----E-DF-MIISPLTTLDSLYS-F-TNKN--IL-NK-IIGEHN-LD-TDKLINKNYLQ-NIANKVNKKIGYD-LV--NIN--------ES------------TSK--LFKSM-FDIK-

STAPH_S     --------------------H-H-------EE-EEE---EE--EEEEE--HHHHHHHHHHHHHHHH---------------------EE----EEE---EEE--------EEEE--------HH-HHHHHH--HHHHH-HHH--HHHHHH---HHHHHHH-HHHHHHHHHHHH--EE--E----------EE-EE-----H-HHHHH--HHHH----HH-
389815356   -----------------MKWH-SN---LFEVI-PLVEG-EV--NVIAFADLEKQFAYRQLVKSWSAHQRMDY-PF------------CL----LTIEGRE-VAA-K-ELNILSLQGGG----LE-FFKDKH--HAKQL-IEL--MKFELENSPELLEVFT-VLQGELEKAFEQC-VV--EFEG-------ME-LQ-PSTEL-ISFEQ--LIKIAPIEVR-
386318633   -----------------MKIA-SH---VTEVL-EIKEQ-SL--NIITFEDTDLCGMFIQNFMDYMTNKPVVENQF------------AH----FISDDYETINS-K-NFHPIILNGGD----RD-VLGQKL--IEEKV-YQI--FEEQIIKCETTNYEFL-KFEQSLTHLKSSL-GI--YDSD-------YQ-IE-LQSST-FKLRK--LLKLFGLDFH-
414160480   -----------------MKLV-GE---FNKPI-EIVEQ-KL--NVISFQNTSKLNALTDGLIKYSQSKSKNYDEN------------VR----LVADNYQTILP-N-NVNLIIVPCGD----IN-LFDSKI--FKEAL-YHR--FEQQIFNEETTNLLYN-KVESTIESFLNSL-EI--KNES-------YQ-VD-FQSSN-IVLKK--LLGMFKPDFH-

STAPH_L     -HHHHHHHHHHHHHHHHHHEE---------EE-EE--E----EEEEE----HHHHHHHHHHHHHHH---------------------EEEHHHHHHHH--H-----EEE-EEE----H----HH-H-HHHH--HHHHH-HHH--HHHHHHH---HHHHHH-HHHHHHHHHHHHH------------------------------HHH--HHHHH-EEEH-
315659845   ------------MTNLVWQIKYES---L-PNI-EL--N-IGKYTAIYNEYETIEEDLLTIVDGYFKRR--NSNSK------------EVSIIDTLNQE--LVSHHLFES-VIIDNNLI----EN-E-HLLG--ASSIL-NKK--IQRDFVNNLESNGYLQ-SINSLLEDLLELLNYT--DLP--------LR-TKQFD------IKQ--FVKMLKFEYD-
403411239   ------------MYESILEINIDM---V-PKV-IL--N-TSVYTSIYNEDSTHEEYILDMLKNYFQSR--NSNKD------------IVTITDISDHE--RVSHNRFRG-FLLDHSTI----EN-E-HQLS--SSSIL-TKK--IVREMQSDLEIDGYIN-SINVLLDDLRDNLKGT---LP--------LS-IKSFD------LKQ--FIKLLSFKYD-
425737239   MIIEHCYSKAGVAMKQIWQIRSSQ---L-NHV-TL--N-LGEYTSIYNENETNEELILSLIYQYFQPR--NRNKD------------VL-IKELIEDV--EVSSKSHVA-FMISHNDI----EQ-E-HQLA--ASSIL-NKK--LQYDMNQNIEVSGYIN-SINVLMEDLLEEVKRK---LP--------LK-VKPFD------IKQ--FIKQLNFEYE-
```

```
3ZTH_SS     E--EEE-HH--H---HH--HHHHHHHHH----H-HH-----HH------EEEEEE--HHH--HH-HHHHHHHHHHHHHHH---EEEEEE----------------------HHHEEEE-----EEE---HHHHHHHHHHH--------HHHHHHHHHH--HHHH------------HHHHHHHHHHHHH----------------HHHHHHH---------
3ZTH        IENKNLSFE--WVS-NI--DKLSLFLEM----L-DH-----LLSQTTEKYLIVLKNIDGFISEE-SYTIFYRQICHLVKKYPNLTFILFP-------SDQGYL--KIDEENSRFVNILSDQ--VEHLYDVEFMYERVMKYYPSNDFPTREGFRMSLET--VTPYLLTKMLRQPSLSLVDSVILNILNQLFHFSY--RIRCSQTPDKELLQKFLE-------
365905381   DSNRTFPLE--FME-TE--SLVDELLNL----I-ES-----SLKKANNDTWIVLYNLDSFVSRS-AKKTILNKLKEFTDRF-ELKVIYLG-------NSLTDC--SITSAETEKIVVAADE--FHQLLPIENLSKSIKMRYPNDFPHSTEQLVQSLIQ--VIPYVGNK--ENVSLNPRDLVLLKIVNDILGYET--SLDYSDQILNSAETKFLLD------
406671115   YDGQEVPLD--TFT-LN--LAYQDMLEL----I-KF-----RVEEEGRLVILLL-DLEPFRTINFEVAEFITQLIEIANNTHLIKLIIFD-------DGDLNQ--KI-PLSISKTIIVTDD--IQQMPEDEIFYQSIKRNYPDKITCSDSNLRDRFYC--ICKYLGHQIYKNISLSSKDVVLLYVVKDLLADNT--YIETSVDSLSEAEKFYLASKNTRY-
422729713   ENGQSLPLE--MMD-TG--ELLAEFIYL----L-QR-----SISRREQPVWLVITNPESFLETE-SIHYLFQQLKKLAQETKQLKFFIFS-------NRSLPL--PYTEEDVEKTILLYDK--YQQLPVFDVFRQSIERHYPDQLSWTNQQLIDAFYR--VCHFVGDQ-YTKNYLLPKDMILLKLLKELLDDNS-ECVETSMEKLTVLEEQYFRERLLRKE
406838713   ENSQVFPLQ--MLD-VN--KTIEDFCRL----L-RN-----FLISESKPIWLWIKNPNSFLAKG-SLKKFIAVLRKITNETKLLRVFIVS-------DDFLDI--AYEPEDMENTVLIYNE--IQQLPPYDIFLNSIKNNYPDNEFDDEEDLLGNLFK--IFPYIDQKNYKNVFLDKKNMLLLKLVDELLGCNEIQLIKGNRNELTDCEENFLYGK-----
238924078   DTE----LS--DKS-NA--ELILILLNI----V-DN-----VLKNNPKKTLILFENLDHLVSLK-EYVEIIHAADTISKKY-DLYFIFSS-------SLNKYV--CCDIDFLEGISVFGDV--YFQMPEYNRLY-----LFLKENYPYEKEFSNEHVKKIICNIIQNIGQSNFLNTVEENVVCKLINKTL------LLDDKIPDEKMPEMAFLKA------
381184142   YEERTYATE--YVD-GL--TKYMSLLDL----L-EI-----NLESSSEMILLMIKNIDDTLSYQ-EYEEVMNSMKRLTTKHPNFYCMVFP-------SQIGYV--YISEEHIESILVVGEE--AHSLADGRNLYTRVCAYYPDQTIPSYNEFLGRLKM--IAPYLFTSKKQSIQMNVKELVLIKIINELFQYFD--LIPVAVLQHSQLEYDFLYES-----
116100822   IENKNLSFE--WVS-NI--DKLSLFLEM----L-DH-----LLLQTTEKYLIVLKNIDGFISEE-SYTIFYRQICHLVKKYPNLTFILFP-------SDQGYL--KIDEENSRFVNILSDQ--VEHLYDVEFMYERVMKYYPSNDFPTREGFRMSLET--VTPYLLTKMLRQPSLSLVDSVILNILNQLFHFSY--RIRYLQKSDEKLLQKFLESKD----

4I99_SS     ------HHH--HHH-HH--HHHHHHHHH----H-HH-----HHH-HH---EEEEE------HHH-HHHHHHHHHHHHHH-----EEEEE------------------------------------------------------------------------------------------------------------------------------------
4I99        AGKDVKRIE--AXS-GG--EKALTALAF----V-FA-----IQK-FKPAPFYLFDEIDAHLDDA-NVKRVADLIKESSK---ESQFIVIT-----------------------------------------------------------------------------------------------------------------------------------
2ONK        AHLLDRKPA--RLS-GG--ERQRVALAR----A-LV-----I-----QPRLLLLDEPLSAVDLK-TKGVLMEELRFVQREFD-VPILHVT-----------------------------------------------------------------------------------------------------------------------------------
2O5V        DFPASDYAS--RGE-GR--TVALALRRA----ELEL-----LREKFGEDPVLLLDDFTAELDPH-RRQYLLDLAASV------PQAIVTG-----------------------------------------------------------------------------------------------------------------------------------

3TOC_SS     ---------------HH--HHHHHHHHH----H-----------------EEEEE--HHH--HH-HHHHHHHHHHH---------EEEEE------------------------EEEE-----EEE-----------------------------------------------------------------------------------------------
3TOC        -------TQ--SDT-IF--EKCLEILQI----F-KY-----LT----KKKLLIFVNSGAFLTKD-EVASLQEYISLTN-----LTVLFLE-------PRELYD--FP-------QYILD---LITKNM---------------------------------------------------------------------------------------------
13622196    -------TQ--SDT-IF--EKCLEILQI----F-KY-----LT----KKKLLIFVNSGAFLTKD-EVASLQEYISLTN-----LTVLFLE-------PRELYD--FP-------QYILDEDYFLITKNMV-------------------------------------------------------------------------------------------
339625078   -------KG--NNSTIF--GKIEEVIHV----M-GM-----LD----ESRYLFLTNASLYCKLS-DLNQLHECLLAEG-----VNLISID---LSLEKPNFSDEHYC-------HFHIDQDFVLYS-----------------------------------------------------------------------------------------------
333398276   -------TL--TNSSGF--GKIEDVVSV----A-GE-----FS----EQRLIVFTNLYLLLSVD-QIDYLNNLAKTMN-----LCLISLN---LTQNPVMTKQGLPP-------EIFIDEDFVQFGAD---------------------------------------------------------------------------------------------
372325148   -------MP--ILNTPY--DKIRSTIDL----A-AE-----LS----DPHLLVFTNAFQYLTMG-QTNQLAEYSRSME-----RNVLFLE----RFDSKIDLTEAAN-------SYFFDKDFVQF------------------------------------------------------------------------------------------------
423349697   -------VP--EYENFY--DEIQAVIDI----A-GA-----LS----EKRMLVLFHISEYCTND-QLEYLVSDIKRQE-----LQVLYLE---RTDHLLRFGD--ER-------SYYVDEDFVQFS-----------------------------------------------------------------------------------------------
310286726   -------TS--SLGSLY--DKISTAVDT----A-GA-----LA----ESRMLVTLHTTQYCDND-QLLYLHRTLLRYQ-----LQLLDLE---CCSKKVMLTE--GM-------SHYVDEDYVQFS-----------------------------------------------------------------------------------------------
336394885   -------FS--SWANYC--DKIMDVISF----Y-SD-----FT----NKKMLIFNNIGRLLNVN-QLNEIHAYLKSVD-----LKLVSLESYPMIFKEEKLN---AK-------VYSIDNDHVRFDY----------------------------------------------------------------------------------------------
408790125   -------TS--SMDNPY--DIIKVIIKL----Y-HK-----FY----PKKLMVLNGLFNYLNSS-QINDLIQMIKALD-----LYVLILD-FSADFRKELADK--CR-------YYRVDSDFVRFNY----------------------------------------------------------------------------------------------
258509196   -------EE--VLTEPF--SKVEAAIKT----L-AK-----LE----EKRLVIFTNISHYLDVD-SLNKLLDQVKETS-----LELLLIE-FSDVERRKFFER--CQ-------YIYIDKDFMDSRELID-------------------------------------------------------------------------------------------
260662217   -------TQ--VIRQPY--GIIETDLKL----H-LE-----LK----DCQVLGYCNVANYLLPE-QIQEFATLVEKTE-----LAVLLVE---FSEKRETLQQLKQD-------IYRIDRDFVDWHE----------------------------------------------------------------------------------------------
354808133   -------TV--VQHDPY--GIIESIVQT----A-VE-----LN----ESKIITLINVSHYLSIN-QFNELVRLVATLN-----VKLFLIE-FSEEVKSDQYQK--CY-------YYHIDNDYVEWRYE---------------------------------------------------------------------------------------------
392947439   -------TD--LTHNPS--AVLEALVQT----H-VE-----LG----LKQCPVMTNVSHYLNPE-QWAVFEQQVADLG-----TTVIIIE-YSEMNRMEKFKD--CC-------YYYVDEDLVDWRDMN--------------------------------------------------------------------------------------------
333394443   -------TS--LSTNCY--GIIETILKT----A-KE-----LN----EQRILGFMNISDYLTLQ-ELQDLFELIRTLD-----LTVLIIK-FSENQRSVEFNN--CR-------YYYIDQDYVDWRYE---------------------------------------------------------------------------------------------
323340065   -------SS--IAYDPC--AIIETDIKI----H-AE-----CA----LKKAIGVSGLATYLTPE-EFSYIKDVVKEMD-----VPLLMID-YNVSGWKSYFDN--CN-------VMYIDEDFIDWH-----------------------------------------------------------------------------------------------
380344195   -------AL--LWKNAY--DRISSDVKL----H-VE-----LN----RKRIIGLTNVAHYLTEG-EFQELVSLVKMTN-----SSMFIIE-FTEKNGQRFFEN--CD-------NYYIDEDYVDWY-----------------------------------------------------------------------------------------------
328956318   -------PN--PQETFL--DNLIDFLSF----V-LD-----IQ----CSKVMVFVNLKTFLTEI-ELEKLYEHVFFSK-----TRVLLLE---NKADPNRHEN--EI-------KTTIDLDFIEK------------------------------------------------------------------------------------------------
374307735   -------EE--EGR-FV--ERLIDYSKT----M-FE-----LL----HKNIFILVGCSGFLTKI-DFEYLRQHQFSQK-----ILFLFVE----SHQIDLESP--KK-------QTIIDIDLCEI------------------------------------------------------------------------------------------------
224543315   -------RK--AEEKLY--DRLMNYVEL----V-TE-----LV----KKPVLVLYNCFDYLDDE-ELMELIKYKNYKH-----LHLLFIE-----NKNREVDSIAFR-------KYIIDEDLCEDYDKIEEVQSD--------------------------------------------------------------------------------------
227824980   -------LD--YNG-LG--EKIFVLMDL----I-SR-----FE----GDKIFILVNLRSYLSSE-EIQLFVDTALAHD-----LKFLLID----NRRYPLFHQ--EK-------RVLIDNDLCVL------------------------------------------------------------------------------------------------
169823758   -------VD--YTD-FL--DKFVSFLDI----Y-QN-----VL----KIDVFYTINLLQYFTEE-ELKELSDYIKIKN-----ICIINYD------NILVDSNYISK-------KLLFDNDLCRVL-----------------------------------------------------------------------------------------------
303229343   -------FN--RKG-EV--MDLMSYIEV----V-DT-----LL----SPMIYIMINLDLVLNDD-EINAFYQNMLSRQ-----LRLVCLT--TGSFDRESLDKNLIN-------GYILDNDFCII------------------------------------------------------------------------------------------------
315926106   -------EQ--SLS-LE--EKLAEYFKI----L-SL-----VL----RKPILCLLHAHIYFSEK-QIRELIQLSKYQK-----LQLLFIE----AREPKKIPDEDED-------IYILDQDRCVIIK----------------------------------------------------------------------------------------------
306821688   -------EN--YDC-LL--EKLINYVKI----I-SQ-----VS----FTKCLFLVNIHDFLSFK-DLDMLYEIALYND-----VSLIVIE----SHQREKSSY--EH-------TYILDDNNCLIEF----------------------------------------------------------------------------------------------
320528781   -------HN--YSS-LL--EEIVDYIDI----F-SQ-----II----KTKLFVFISLRSYLTGE-EFNNFIKIMDYKG-----IRILLIE-----RYLSLDNSIVDN-------VQIIDKDLCVI------------------------------------------------------------------------------------------------
342218249   -------WG--EME-LL--ERLVTYLDI----I-CD-----LL----KSKLIVLVNVRKYMTEE-ELKLVYEHCIRKK-----IPIFCIE---SGYTCHSVDELLEK-------VYTIDKDLCFIY-----------------------------------------------------------------------------------------------
227501315   -------TK--FTG-LM--ANIIEYLDI----S-HK-----LL----NKKIFFIVGIDDYFEER-ELYDLNEYLCYND-----IIVICLQ---NRLDRDLLPN--EN-------LRIIDEDLGEI------------------------------------------------------------------------------------------------
304438951   -------YE--KNS-VL--ERLIEYIKV----T-SE-----LL----KTKVYIIVNLDSFLSEE-ELVELEKFLLYND-----IKVLALQ---NAIRREVIPS--EN-------LRIVDKDLCEI------------------------------------------------------------------------------------------------
34762595    -------RS--ED--LL--LNLFEWLKI----L-NE-----IL----GYEIFFFINIENFLSED-ELLEFSKFIVYNK-----YKVVFLE---NFYRNKLSDN--DN-------LIVIDNDLCEIF-----------------------------------------------------------------------------------------------
241889886   -------VE--NDD-IL--ENFVKFLKV----L-TE-----LC----CFKIIFVVGLHTVFTQE-EIIEIYKEVCLNK-----INIINVE--YQQFNNLSAKNYKEI-------VYIFDKDNCEI------------------------------------------------------------------------------------------------
291520707   -------EM--EDD-FF--ERIIRYVKI----A-VE-----VL----STKVFVFVNIRSYLTDE-QIQELIKEIRYQE-----AKILFME----SQERACLEG--GM-------RYIIDRDGCEIY-----------------------------------------------------------------------------------------------
42525846    -------TE--YTN-LL--EKIIDYISA----S-VE-----LN----QIKCFIFINLKQFLHFK-DLSELYKFACYIK-----TNLFLIE---GSYTSTKEDC--EK-------SYIIDKDLCEIY-----------------------------------------------------------------------------------------------
299144349   -------FD--YGN-TI--DRLISYMVV----L-RK-----YI----GIKLFILINIEKILTVC-EMKEFFEYCIMNE-----FKVLFIESGKADYENLDLKY--IK-------RYIIDEDLCDLY-----------------------------------------------------------------------------------------------
116101487   -------TK--SCT-VF--EKIFDILQI----F-KY-----LV----KKRILVFVNSLSYFSKD-EIYQILEYTKLSQ-----ADVLFLE-------PRQIEG--IQ-------QFILDKDYILMPYNN--------------------------------------------------------------------------------------------

MYCO        ----------------H--HHHHHHHHH----H-H-------------EEEEEEE-----------HHHHHHHH---------EEEEEEE--HHHHHHHHHHHHHHHHH------EEEEE--HHHHHHHHHHHH-----HHHHHHHHHHHHHH--HHHHHHHHHHH-------------------------------------------------------
350546883   -------DL--INI-NK--NIFISLIKN----N--------LFD---ERMTFVFCDIDWLN-----INDISEYI--NS-----HNFIFVTND-LRKNIVNITQIESCVICNDF---LVEIYDYNKLISYLELKMNTLINN-KVIENYLKNEES--VSNFYINYLIKYIR----------------------------------------------------
193216855   -------ET--SLL-TR--DLFFKWLE-----N--Y-----DEV---GQKVVILKNLPWVK-----IEDLVQYL--DK-----FVFIILTDN-FFDNCIRFEALETCAFWEKGR--LVLVHSLNVIINWIEKKINKNIDLQEEIKNIFNSTFD--GEKLFFNLKSEVFS----------------------------------------------------
47458866    -------KD--LLI-NE--INILNILKI----I-DK-----YSN---ETISILIYGIKLKN-----FSEFI--F--NK-----LNIIYLIDK--LAYNLSFENIEQLVFANKKE--IVEIFDKEKLINYLESGTKTLITV-EKINNYLKGEES--FDSFLINKFLEKI-----------------------------------------------------
385326557   -------ED---YI-DH--EFIYSYLEL----L--H-----WSK---EVKTVILKDFDNID-----LKRLSNLT--QT-----TNLIIMKNDIIYDLEKNFEFFETYCLIDHDYENMIKIDNMPSFEYLLEDIFKVMVDE-EFKANMSFHQLEIIKKELKKQIN---------------------------------------------------------
490549417   -------EK--LLI-SK--NNLLNILDI----L--S-----SSN---FKPLIIVKDLTYIN-----IEELIKY---SN-----LNFLILTSD-FTKYIQSYNQLELVSFYNGEI--LFDIKSCTPIINFFENIKNKEIK--ENENFFLNKEEEKQFKFNFFDIKKSFFE----------------------------------------------------
363542553   -------ND--ILI-NE--ENFENLAEI----I--F-----STM---KQTIVILKDLDYIK-----FEKLFKY---NN-----VTFLILTND-FTKYISKFSELELVAFYSQKT--VIDVIDNLPIIRYFENLKNRPVN--EDELIYTSEEEKMAAKMHFYKIKTSFLN----------------------------------------------------
71894595    -------ND--KFI-DN--KSLIKWMEN----Q--K-----YNN---QKINLIIKNFDFVL-----INELIKFS--NN-----FNIIVLTND-FFKHINNFDYIESVAFTNEDLNRIVSIEEKDSFEYFLEEIFQDSIRNIEKFDFFSQNNKDLIKKSLKINFF---------------------------------------------------------
419703977   -------TK--EFI-TS--SLLYSFLRN----I--------SQA---EKQNIIIKDMDDIS-----LSLLTEFS--ND-----FNIIVMTNN-SFNIIKSSSEVLLTNFVDENLYTLKNIDNESVFEYFIEEYFQKPFDNVNPELIYTKETMENMKKALFLK-----------------------------------------------------------

STAPH_S     -------H-------HHHHHHHHHHHHHHHHHH-----------------EEEEE--HHH--HH-HHHHHHHHHHHH------EEEEEEE---------HHHHHH----EEEEEE-------HHHHHHHHHH-------HHHHHHHHHHHHHHHHH-HHHHHHHHHHHHHH--------------------------------------------------
389815356   -------TV--SDE-PMGAYQYQSFLLKSWLHL-VH-----NK----ATNVCFYDFPENELNKS-ELKELFQFATSQA-----CTMICLT-------TSPQVINQVGLSNVHLIKRTGERYSIETLHEELKLFFEDTPES-LKSLAKNLAYSDLHEEYIMLDQKWWEFLTSRMC-----------------------------------------------
386318633   -------NQ--LKD-LNELNKRKAFIDL----L-IH-----PT----KENILIILFPESHLGIH-DIKKFIHLIKRYK-----FTTIIFT-------NHPSII--INEENIFLCKSNKKCHDIEKIYEELMLFVPEEVVS--HQLVNMIAYHEFVGENLSEYERYFKFINHYN------------------------------------------------
414160480   -------ND--LDD-LNAIEQRKLFIDT----L-TS-----TE----KENVLVLLFPEAHLGLS-DMKAFNKYLQTLN-----MFSIIVT-------SHPYFI--IESDTLALIKTNGESVNTKQIREELLLFKEKETITDLDKTSKMIAYHEFCQHPYLDDEFFIQFMNKY-------------------------------------------------

STAPH_L     -------HHHHHHH-HH--HHHHHHHHH----H-HHHHHHHH------EEEEEEE-------HH-HHHHHHHHHH----------EEEE----------HHHHHHHH----EEE--HHHHHHHHHHHHHH-------HHHHHHHHHHHHHHHH-EEEE--EE---EEEEEEEEE---EEEEHHHHHH--HHEEE--------HHHHHHHHHH---------
315659845   -------LKDDYTR-LI--VRIEQVLPL----V-IEELKKQHN----NQLLLIYLYPEANLSPK-EQIRLKQLLISLD-----VNIIVLT-------GSMHFLAQQWKNNNYIRYDKQKLTSNFIDRLVWDAPLNFEQLEIENSLNQFTLTYQDKIEVHPTISNYQIAPIMLFNAIDLYVGINYMQHCQHQFNFDIDETLLPKTMFKYFSSYLK-------
403411239   -------YSEDYAR-LI--TRLEQVVPL----V-VDEMNYQVK----EKAFVIYLYPEAGLSIK-EQIRFRKILNNLP-----VPVIVLT-------ESPRFLSESIKGLNYFINNVQMITESFFEDSYWNAPINVEKEQIEERFEYLFVKYCSVLEIKPIISNYLLADIIVFNDIDLYILVSFLKHCNLEYKLDIDKTKISLALSSYLFS----------
425737239   -------LEADYSR-LI--VRLKQLLPL----L-VEQMNNLSN----NKGLLIYLYPEANLSIK-ERLIFREVLNDLP-----IKIIVLT-------QSRYFLADSIDGLNYLRNSKQMITVDFIENCYWEAPLNFKREDIVVSLKNVFDQYSHCFEIHPTITNYHLGDIMVFKDIDIYVIVRFLRFCKLDYILDIDDTKISEPLALYLGLSTS-------
```

B

HHPRED hits for Csn2 long

```
 No Hit                             Prob E-value P-value  Score    SS Cols Query HMM  Template HMM
  1 4i99_A Chromosome partition pr  95.0     2.5 1.7E-05   43.4  14.0  230    2-245     1-305 (354)
  2 g1xew.1 c.37.1.12 (X:,Y:) Smc   94.6     3.1 2.1E-05   42.3  13.3  219    4-236     6-278 (329)
  3 3auy_A DNA double-strand break  81.5     8.8 5.9E-05   37.8   4.9   66    1-67      4-72  (371)
  4 4ad8_A DNA repair protein RECN  76.4      18 0.00012   40.1   5.9   47    1-48     39-88  (517)
  5 4aby_A DNA repair protein RECN  75.0      21 0.00014   38.5   5.7   46    1-47     39-87  (415)
  6 3kta_B Chromosome segregation   73.2      71 0.00048   32.3   8.4  115  113-242     9-123 (173)
  7 g1f2t.1 c.37.1.12 (A:,B:) Rad5  72.6 1.1E+02 0.00074   30.8   9.4  216   13-244    15-263 (292)
  8 d2za4b_ c.9.1.1 (B:) automated  72.3      70 0.00047   29.0   7.3   62  175-243    26-89  (90)
  9 2fnaA01                         71.5      60  0.0004   31.6   7.3   87  161-249    95-181 (208)
 10 d1e69a_ c.37.1.12 (A:) Smc hea  70.9      13 8.7E-05   37.8   3.1   47    4-51      5-55  (308)
 11 PF02463 SMC_N:  RecF/RecN/SMC   69.4      23 0.00016   35.2   4.4   46    1-47      1-51  (220)
 12 d2hxxa_ c.9.1.1 (A:) automated  69.3 1.1E+02 0.00073   27.8   7.8   52  191-243    34-88  (89)
 13 PF13304 AAA_21:  AAA domain; P  68.1 1.6E+02  0.0011   27.9   9.0  215   22-244     1-295 (303)
 14 1w1w_A Structural maintenance   67.3 1.3E+02 0.00086   31.5   8.9  124  113-244   269-393 (430)
 15 3zgx_A SMC, chromosome partiti  66.5      20 0.00014   36.5   3.4   44    4-48      5-52  (426)
```

```
No 1
>4i99_A Chromosome partition protein SMC; winged-helix domain and SMC head domain, chromosome condensa SCPB, DNA binding protein; 2.30A {Pyrococcus
furiosus}
Probab=95.02  E-value=2.5  Score=43.40  Aligned_cols=230  Identities=11%  Similarity=0.085  Sum_probs=0.0

Q ss_pred              ce---EEEEEeccccce----EEEeccc-eEEEecCcHHHHHHHHHHHHHHhCCCCccHHHhh------------------
Q gi|365905381|r    1 -M---RLKIEIDDQSF----FELDFGD-IVYISGYNHQNMWKIYRSLYYYFNKSPM--LSTN------------------   51 (340)
Q Consensus         2 mM---~l~i~h~~~~~~~~~---ie~~~g~~~tqi~G~N~qlK~yi~q~L~~yF~gkkyse~e~~~~~~~~~~~~~~~~~~~~~~~~~   55 (365)
                      ||   +++|. .|+.+    ++++|++ +|.|+|.|.-=|-=|+++|.|.++|.......-.
T Consensus         1 m~yI~~i~I~-NFksf~~~~~~i~f~~~inviiG~NgsGKSniLeAi~~~l~~~~~~~~~r~~~~~d~i~~~~~~~~~~~~~~~~~~   79 (354)
T 4i99_A            1 MPYIEKLELK-GFKSYGNKKVVIPFSKGFTAIVGANGSGKSNIGDAILFVLGGLSAKAMRASRISDLIFAGSKNEPPAKY   79 (354)
T ss_dssp              -CEEEEEEEE-SBGGGCSSCEEEECCSSEEEEEECCTTSSHHHHHHHHHHHTTCCCTGGGTCSSGGGGBCCCCSSCCCCSC
T ss_pred              CCccCEEEEE-CccccCCCEEEEcCCCCeEEEECCCCCCHHHHHHHHHHHhcCCchhhccccchhHHhhccccCCccce


Q ss_pred              -------------------------hhcCCCCeEEecCccccCcc--------------ceEEEEEccHHHHHH-----
Q gi|365905381|r   52 -------------------------IYGDDNIEISLDDVSP-STK--------------NTQTYFISNRDSIYQ-----   85 (340)
Q Consensus        56 -------------------------if~~~ep~I~~d~~~l~~rk----------------~~~~i~Is~~~di~e-----   90 (365)
                                                +..+.+-+.+++|+.+ .++              .+.++.-.....+..
T Consensus        80 ~~i~~~~~~~~~~~~~~~~~~~~~~~~i~~~~~~~~~g~~~~~~in~~~~-~~~~i~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  158 (354)
T 4i99_A           80 AEVAIYFNNEDRGFPIDEDEVVIRRRVYPDGRSSYWLNGRRA-TRSEILDILTAAMISPDGYNIVLQGDITKFIKMSPLE  158 (354)
T ss_dssp              EEEEEEEECTTCCSSSSSSEEEEEEEECTTSCEEEEETTEEE-CHHHHHHHHHHHTTCCTTSTTEECTTSHHHHHHCCHHH
T ss_pred              EEEEEeeccccccCcccccceEEEEEEEEecCCCceEEEECCcee-hHHHHHHHHHHhCCCccceEEEEccchhhhhhcChHH


Q ss_pred              ---HHHhccccchHHHHHHHHhhhHHHHHHHHHHHHHHHHHHHHHHhChHHHHHhhcCCeEEEhhccccchhHHHHHHhhccc
Q gi|365905381|r   86 ---QMIYKKGNLLFENLNTLNDSFEITRILEELNDTNTKLSLLVQ--EHINNYANNLIIDFEDV--TYFD--VLKYFLKV  156 (340)
Q Consensus        91 ---qm~~kKGTll~~yL~~~~~~~dI~~~le~iNd~l~~i~~~iN~~~~in~~~~~i~~~~e~~~~~~~~~~ll~k~~~~  167 (365)
                         ......+......-....... .....+.+.+....+...++  +-......+-..+..-.  +..+  .+......
T Consensus       159 ~~~~~~~~~~~~~~~~~~~~~~~~~-~~~~~~~~~~~~~~~~~~~~~~--~~~~~~~~~~~~~~l~l~--~~~~--~~~~~i~~  231 (354)
T 4i99_A          159 RRLLIDDISGIAEYDSKKEKALEE-EKEKKNVFMRTFEAISRNFS--EIFAKLSPGGSARLILE--NPED--PFSGGLEI  231 (354)
T ss_dssp              HHHHHHHHHTC---------------CCHHHHHHHHHHHHHHHHHH--HHHHHHSTTCEEEEEES--STTS--GGGSCEEE
T ss_pred              hhhhhhhhccccchhHHHHHHHHHH-HHHhhhhhhhhhhHhHHHH--HHhhhhcCCCcceeeee--cccC--ccccceeeE


Q ss_pred              ccccCCcee-ceeecchHHHHHHHHHHHHHHHHhcCCCCcEEEEEeChHHHHhccHHHHHHHHHHHHHHHhcCCCEEEEEEEc
Q gi|365905381|r  157 SYEDSNRTF-PLEFMETESLVDELLNLIESSLKKANNDTWIVLYNLDSFVSRS-AKKTILNKLKEFTDRF-ELKVIYLG  232 (340)
Q Consensus       168 ~~~~~~~~~~-~le~v~~~~~~~~~~L~ll~~~l~~~~~~~lLvlkNiD~fls~~~s~~~i~~~l~~lt~k~~~l~~iif~  245 (365)
                    .....++.. +++.++....-...|-++-...+....+ +++|.-||..|... .-.++.+.+++++   ++.-+|+..
T Consensus       232 ~~~~~~~~~~~~~~~~~s~g~~~~~~l~l~~~~~~~~~~~~~-i~liDEpe~~Lhp~-~~~~~l~~~L~~~~~---~~~QviitT  305 (354)
T 4i99_A          232 EAKPAGKDVKRIEAMSGGEKALTALAFVFAIQKFKPAP-FYLFDEIDAHLDDA-NVKRVADLIKESS---KESQFIVIT  305 (354)
T ss_dssp              EEESSSSSCCCGGGSCHHHHHHHHHHHHHHHHHHSCCS-EEEEESTTTTCCHH-HHHHHHHHHHHHH---TTSEEEEEC
T ss_pred              EeccccccccccccccchhHHHHHHHHHHHHHHHhhcCcc-hhhhccHHhhcCHH-HHHHHHHHHHHhC---cCCeEEEEE

No 2
>g1xew.1 c.37.1.12 (X:,Y:) Smc head domain {Pyrococcus furiosus [TaxId: 2261]}
Probab=94.58  E-value=3.1  Score=42.31  Aligned_cols=219  Identities=12%  Similarity=0.084  Sum_probs=0.0

Q ss_pred              EEEEEeccccce----EEEeccc-eEEEecCcHHHHHHHHHHHHHHhCCCCcc----------------------------
Q gi|365905381|r    2 RLKIEIDDQSF----FELDFGD-IVYISGYNHQNMWKIYRSLYYYFNKSPM----------------------------   47 (340)
Q Consensus         4 ~l~i~h~~~~~~~~~---ie~~~g~~~tqi~G~N~qlK~yi~q~L~~yF~gkkys----------------------------   50 (365)
                      +++|+ .|+.+    ++++|++ +|.|+|.|--=|-=|+++++|.++|+...
T Consensus         6 ~i~i~-NFks~~~~~~~~~l~f~~~lnvi~G~NGsGKS~il~AI~~~L~~~~~~~~~~~r~~~~~~~~~~~~~~~~~~~~~~~~~~~~   84 (329)
T g1xew.1          6 KLELK-GFKSYGNKKVVIPFSKGFTAIVGANGSGKSNIGDAILFVLGGLSAKAMRASRISDLIFAGSKNEPPAKYAEVAI   84 (329)
T ss_dssp              EEEEE-SBTTBCSSCEEEECCSSEEEEEEECTTSSSHHHHHHHHHHTTSSCCCC-------------------CSEEEEE
T ss_pred              EEEEE-CeEcccCCEEEEeCCCCeEEEECCCCCHHHHHHHHHHHHCCCcchhhccccchhhhhhcccccccccchhhhhhh


Q ss_pred              -----HHHhhhhcCCCCeEEecCccccCccceEEEEE----ccHHHHHHHHH-----------hccchHHHHHHHHhhh
Q gi|365905381|r   48 ------LSTNIYGDDNIEISLDDVSP-STKNTQTYFI----SNRDSIYQQMI-----------YKKGNLLFENLNTLNDS  105 (340)
Q Consensus        51 -----e~e~~if~~~ep~I~~d~~~l~~rk~~~~i~I----s~~~di~eqm-----------~kKGTll~~yL~~~~~~  110 (365)
                           +.......+    ..+-...+ .+..-..+.+    .+..++.+...           ...|.+.-=.-..-...
T Consensus        85 ~~~~~~~~~~~~~~~~----~~~~~~r~i-~~~g~~~~~~~~~~~~~~~~i~~~~~~~~~~~~~~~~~~i~qg~i~~~~~~~~~~~~  159 (329)
T g1xew.1         85 YFNNEDRGFPIDED----EVVIRRRV-YPDGRSSYWLNGRRATRSEILDILTAAMISPDGYNIVLQGDITKFIKMSPLER  159 (329)
T ss_dssp              EEECTTCCSSSSSS----EEEEEEEE-CTTSCEEEEETTEEECHHHHHHHHHHTTCCTTCTTEECTTHHHHHHHCCHHHH
T ss_pred              hhhhhcccCCccCC----cceEEEEE-ecCCCEEEEeCCeEccHHHHHHHHHHcCCCCCceEEEeehhhhHhhcCCchhh


Q ss_pred              HHHHHHHHHHHHHHHHHHHHHhChHHHHHhhcCCeEEEhhccccchhHHHHHHhhccccccCCcee-ceeecchHHHHHH
Q gi|365905381|r  106 FEITRILEELNDTNTKLSLLVQ--EHINNYANNLIIDFEDV--TYFD--VLKYFLKVSYEDSNRTF-PLEFMETESLVDE  178 (340)
Q Consensus       111 ~dI~~~le~iNd~l~~i~~~iN~~~~in~~~~~i~~~~e~~~~~~~~~~ll~k~~~~~~~~~~~~~~~-~le~v~~~~~~~~~  189 (365)
                    ..+..++....+.+..+...++  +-.....+.-.-..+-.  +-++  .+..-+.......++.+ +++.++...+-..
T Consensus       160 ~~~~~~~~~~~~~~~~~~~~~~~~--~~~~~l~~~~~~~l~l~--~~~~--~~~~~l~~~~~~~~~~~~~~~~~~ls~~~~~l~  233 (329)
T g1xew.1        160 RLLIDDISVFMRTFEAISRNFS--EIFAKLSPGGSARLILE--NPED--PFSGGLEIEAKPAGKDVKRIEAMSGGEKALT  233 (329)
T ss_dssp              HHHHHHHCCHHHHHHHHHHHHH--HHHHHHSTTCEEEEEES--CSSS--GGGSCEEEEEECTTSCCEEGGGSCHHHHHHH
T ss_pred              hhhhhhccccchhhhHHhhhhh--hhhheecCCCeEEEecc--CCcc--ccccCcEEEEECCCCceeehhhhhhHHHHHH


Q ss_pred              HHHHHHHHHhcCCCCcEEEEEeChHHHHhccHHHHHHHHHHHHHHHhcC
Q gi|365905381|r  179 LLNLIESSLKKANNDTWIVLYNLDSFVSRS-AKKTILNKLKEFTDRF  224 (340)
Q Consensus       190 ~L~ll~~~l~~~~~~~lLvlkNiD~fls~~~s~~~i~~~l~~lt~k~  236 (365)
                    .+......+... .+.++++..+|.-|... .-.++.+.++++++++
T Consensus       234 ~l~~~~~~~~~~~~~~~~-~~~~~~~ide~e~~L~~~-~~~~~l~~~l~~~~~~~~  278 (329)
T g1xew.1        234 ALAFVFAIQKFK-PAPFYLFDEIDAHLDDA-NVKRVADLIKESSKES  278 (329)
T ss_dssp              HHHHHHHHHHS-CCSEEEEESTTTTSCHH-HHHHHHHHHHHHHTTTS
T ss_pred              HHHHHHHHhhc-ccchhhhhHHhhcCHH-HHHHHHHHHHHhccCC
```

```
No 3
>3auy_A DNA double-strand break repair RAD50 ATPase; DNA repair, ABC transporter ATPase domain-like; HET: DNA ADP; 2.70A {Methanocaldococcus
jannaschii} PDB: 3aux_A* 3av0_B*
Probab=81.48  E-value=8.8  Score=37.79  Aligned_cols=66  Identities=14%  Similarity=0.233  Sum_probs=0.0

Q ss_pred              CceEEEEEeccce--EEEeccc-eEEEecCcHHHHHHHHHHHHHHhCCCCccHHHhhhhcCCCCeEEec
Q gi|365905381|r    1 --MRLKIEIDDQSF--FELDFGD-IVYISGYNHQNMWKIYRSLYYYFNKSPM--LSTNIYGDDNIEISLD   63 (340)
Q Consensus         1 ~mM~l~i~h~~~~~--ie~~~g~-~tqi~G~N~qlK~yi~q~L~~yF~gkkyse~e~~if~~~ep~I~~d   67 (365)
                      |.-+++|+ .|+.+  .+++|++ +|.|+|.|.-=|-=|++++.|.++|..-...-..++...+....++
T Consensus         4 ~i~~i~i~-nFr~~~~~~i~f~~~~~~i~G~NgsGKStil~Ai~~~L~~~~~~~~~~~~~~~~~~~~~~v~   72 (371)
T 3auy_A            4 ILKEIRMN-NFKSHVNSRIKFEKGIVAIIGENGSGKSSIFEAVFFALFGAGSNFNYDTIITKGKKSVYVE   72 (371)
T ss_dssp              EEEEEEEE-EETTEEEEEEEECCSEEEEEEEECTTSSHHHHHHHHHHHHHHCCC-CCCTTTTBCTTCSEEEEE
T ss_pred              EEEEEEEe-CcccCCCeEEECCCCeEEEECCCCCCHHHHHHHHHHHhcCCCCCCcHHHHhhCCCCcEEEE


No 4
>4ad8_A DNA repair protein RECN; DNA binding protein, ATPase domain; HET: DNA; 4.00A {Deinococcus radiodurans}
Probab=76.35  E-value=18  Score=40.08  Aligned_cols=47  Identities=19%  Similarity=0.240  Sum_probs=0.0

Q ss_pred              CceEEEEEecccc--eEEEeccc-eEEEecCcHHHHHHHHHHHHHHhCCCC
Q gi|365905381|r    1 --MRLKIEIDDQS--FFELDFGD-IVYISGYNHQNMWKIYRSLYYYFNKSP   46 (340)
Q Consensus         1 ~mM~l~i~h~~~~--~ie~~~g~-~tqi~G~N~qlK~yi~q~L~~yF~gkk   48 (365)
                      |.-+++|+ .+.-  .++++||+ +|.|+|.|--=|-=|+.+|+|.+||+.
T Consensus        39 mL~rL~i~-Nf~~i~~~~i~F~~glnvItG~nGsGKS~IldAI~~~LG~ra   88 (517)
T 4ad8_A           39 RLSRLEIR-NLATITQLELELGGGFCAFTGETGAGKSIIVDALGLLLGGRA   88 (517)
T ss_dssp              CCCEEEEE-SBTTBSCEEEECCCSEEEEEESHHHHHHHHTHHHHHHTCSCC
T ss_pred              ccCeEEEE-ccccEEEEEEEcCCCeEEEECCCCCCHHHHHHHHHHHhCCCC


No 5
>4aby_A DNA repair protein RECN; hydrolase, double strand break repair, ATPase, nucleotide binding domain; HET: DNA; 3.00A {Deinococcus
radiodurans}
Probab=75.04  E-value=21  Score=38.48  Aligned_cols=46  Identities=20%  Similarity=0.269  Sum_probs=0.0

Q ss_pred              CceEEEEEeccce--EEEeccc-eEEEecCcHHHHHHHHHHHHHHhCCC
Q gi|365905381|r    1 --MRLKIEIDDQSF--FELDFGD-IVYISGYNHQNMWKIYRSLYYYFNKS   45 (340)
Q Consensus         1 ~mM~l~i~h~~~~~--ie~~~g~-~tqi~G~N~qlK~yi~q~L~~yF~gk   47 (365)
                      |.-+++|+ .|.-+  ++++|++ +|.|+|.|--=|-=|+++|.+.+||+
T Consensus        39 mL~~L~Ik-NF~~i~~l~i~f~~glnVltGeNGsGKS~lLdAL~ll~G~r   87 (415)
T 4aby_A           39 RLSRLEIR-NLATITQLELELGGGFCAFTGETGAGKSIIVDALGLLLGGR   87 (415)
T ss_dssp              CCCEEEEE-EETTEEEEEEEECCSSEEEEEESHHHHHHHHTHHHHHHTTCC
T ss_pred              ccceEEEe-ccccEEEEEEEcCCCcEEEECCCCCCHHHHHHHHHHHhCCC


No 6
>3kta_B Chromosome segregation protein SMC; structural maintenance of chromosomes, ABC ATPase, CFTR adenylate kinase, AP5A, transferase; HET: AP5;
1.63A {Pyrococcus furiosus} PDB: 1xew_Y 1xex_B*
Probab=73.19  E-value=71  Score=32.26  Aligned_cols=115  Identities=13%  Similarity=0.128  Sum_probs=0.0

Q ss_pred              HHHHHHHHHHHHHHHHHHHhChHHHHHhhcCCeEEEhhccccchhHHHHHHhhccccccCCceeceeecchHHHHHHHHH
Q gi|365905381|r  108 ITRILEELNDTNTKLSLLVQ--EHINNYANNLIIDFEDV--TYFD--VLKYFLKVSYEDSNRTFPLEFMETESLVDELLN  181 (340)
Q Consensus       113 I~~~le~iNd~l~~i~~~iN~~~~in~~~~~i~~~~e~~~~~~~~~~~~ll~k~~~~~~~~~~~~~~~~le~v~~~~~~~~~L~  192 (365)
                      ..+-+++||..+..++..+.       -.++..+.++.....+..  =+.-.+.|   ......|++.+|.+++-...|-
T Consensus         9 f~~~f~~I~~~f~~~f~~l~-------~~~~a~L~l~~~~~~~~~~~~--gi~~~v~~---~~~~~~~~~~~~LSGGEks~~aLa   76 (173)
T 3kta_B            9 FMRTFEAISRNFSEIFAKLS-------PGGSARLILENPEDPFSG--GLEIEAKP---AGKDVKRIEAMSGGEKALTALA   76 (173)
T ss_dssp              HHHHHHHHHHHHHHHHHHHS-------TTCEEEEEESCSSSGGGS--CEEEEEET---TSSSCCCGGGCCHHHHHHHHHH
T ss_pred              HHHHHHHHHHHHHHHHHHHc-------CCCEEEEEecCcccccc--ceEEEEEc---CCCCCCchhhcCccHHHHHHHH


Q ss_pred              HHHHHHhcCCCCcEEEEEeChHHHHhccHHHHHHHHHHHHHHHhcCCCEEEE
Q gi|365905381|r  182 LIESSLKKANNDTWIVLYNLDSFVSRS-AKKTILNKLKEFTDRF-ELKVI  229 (340)
Q Consensus       193 ll~~~l~~~~~~~~lLvlkNiD~fls~~~s~~~i~~~l~~lt~k~~~l~~i  242 (365)
                      ++=.+.+..+-| +++|-.+|..|+.. .-..+++.|+++++++ ++=||
T Consensus        77 l~lal~~~~~~p-f~ilDE~da~LD~~-~~~~v~~~~L~~~~~~~~~-Q~ivi  123 (173)
T 3kta_B           77 FVFAIQKFKPAP-FYLFDEIDAHLDDA-NVKRVADLIKESSKES-QFIVI  123 (173)
T ss_dssp              HHHHHHHHSCCS-EEEEESTTTTCCHH-HHHHHHHHHHHHHTTTS-EEEEE
T ss_pred              HHHHHhccCCCC-ceEeccccccCCHH-HHHHHHHHHHHHhCCCC-EEEEE


No 7
>g1f2t.1 c.37.1.12 (A:,B:) Rad50 {Archaeon Pyrococcus furiosus [TaxId: 2261]}
Probab=72.62  E-value=1.1e+02  Score=30.76  Aligned_cols=216  Identities=11%  Similarity=0.093  Sum_probs=0.0

Q ss_pred              ceEEEeccc-eEEEecCcHHHHHHHHHHHHHHHh----CCCCccHHHhhhhcCCCCeE----EecCcc-------------
Q gi|365905381|r   11 SFFELDFGD-IVYISGYNHQNMWKIYRSLYYYF----NKSPM--LSTNIYGDDNIEI----SLDDVS-------------   66 (340)
Q Consensus        13 ~~ie~~~g~-~tqi~G~N~qlK~yi~q~L~~yF----~gkkyse~e~~if~~~ep~I----~~d~~~~-------------   70 (365)
                      +.++++|++ +|-|+|.|--=|-=|++++.|-+    ..+..+..+..-.|..++.|    ..+|+.
T Consensus        15 ~~~~l~f~~~ln~IvG~NGsGKStiL~Ai~~~l~~~~~~~~~~~~~~~~~~~~~~~~g~~~~~~iei~f~~~~~~~~~~~i~~~~~~~~~~~~~~   94 (292)
T g1f2t.1         15 SDTVVEFKEGINLIIGQNGSGKSSLLDAILVGLYWPLRIKDIKKDEFTKVGARDTYIDLIFEKDGTKYRITRRFLKGYSS   94 (292)
T ss_dssp              SSEEEECCSEEEEEEECCTTSSHHHHHHHHHHHHHHCSSCCTTSSCCCSCSTTCCCEEEEEEEEEETTEEEEEEEEEECCC---
T ss_pred              cCeEEEcCCCeEEEECCCCCCHHHHHHHHHHHhcCCccccCccccccccccccccccccccccccccccccccccccccccch


Q ss_pred              --------ccCcccEEEEEccHHHHHHHHHhccccchHHHHHHHHhhhHHHHHHHHHHHHHHHHHHHHHHhChHHHHHhhcC
Q gi|365905381|r   67 --------P-STKNTQTYFISNRDSIYQQMIYKKGNLLFENLNTLNDSFEITRILEELNDTNTKLSLLVQ--EHINNYAN  135 (340)
Q Consensus        71 --------l~~rk~~~~i~Is~~~di~eqm~~kKGTll~~yL~~~~~~~~~dI~~~le~iNd~l~~i~~~iN~~~~in~~~~~  142 (365)
                              + ....+..+.-.....+-+.+.-..+--.+..--..-+.--..-....-++.+..+.....  ........
T Consensus        95 ~~~~~~~~~~~-~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~-~~~~~l~~  171 (292)
T g1f2t.1         95 GEIHAMKRL-VGNEWKHVTEPSSKAISAFMEKLIPYNIFLNAIYIRQGQIDAILESAREAALSKIGELAS--EIFAEFTE  171 (292)
T ss_dssp              -EEEEEEEE-ETTEEEEESSCSSHHHHHHHHHHHHSCHHHHHHHHTEECTTHHHHHTCCCHHHHHHHHHHHHHHH--HHHHHHTT
T ss_pred              hhHHHHhhh-cccceeecccchHHHHHHHHHHhhhhhhhheeecccccchhhhcccccccccccccccc--hhhhhhhh


Q ss_pred              CeEEEhhccccchhHHHHHHhhccccccCCceeceeecchHHHHHHHHHH---HHHHHhcCCCCcEEEEEeChHHHHhcc
Q gi|365905381|r  136 NLIIDFEDV--TYFD--VLKYFLKVSYEDSNRTFPLEFMETESLVDELLNL---IESSLKKANNDTWIVLYNLDSFVSRS  208 (340)
Q Consensus       143 ~i~~~~e~~~~~~~~~~~~ll~k~~~~~~~~~~~~~~~~le~v~~~~~~~~~L~l---l~~~l~~~~~~~lLvlkNiD~fls~~  219 (365)
                      +---.....        .........+...++..|...++.+++-...+-+   +-+.+...+.  +|+|--|+.-|...
T Consensus       172 ~~~~~~~~~~~--------~~~~~~~~~~~~~~~~~~~~~~~~~lSgGe~~~~~la~~la~~~~l~~~~~~--llllDEp~~~Ld~~  241 (292)
T g1f2t.1        172 GKYSEVVVR--------AEENKVRLFVVWEGKERPLTFLSGGERIALGLAFRLAMSLYLAGEIS--LLILDEPTPYLDEE  241 (292)
T ss_dssp              TSCCEEEEE--------ETTSSEEEEEEEETTEEECGGGSCHHHHHHHHHHHHHHHHHHHHHSSCS--EEEEESCSCTTCHH
T ss_pred              hhhhhcccc--------cccccccccchhhhcCchhhcchHHHHHHHHHHHHHHHhhhcCCCC--EEEEeCCccccCHH


Q ss_pred              HHHHHHHHHHHHHhcCCCEEEEEEE
Q gi|365905381|r  209 -AKKTILNKLKEFTDRF-ELKVIYL  231 (340)
Q Consensus       220 ~s~~~i~~~l~~lt~k~~~l~~iif  244 (365)
                       .-.++++.+++++++..  .+|+.
T Consensus       242 -~~~~l~~~l~~~~~~~~--qviv~  263 (292)
T g1f2t.1        242 -RRRKLITIMERYLKKIP--QVILV  263 (292)
T ss_dssp              -HHHHHHHHHHHTGGGSS--EEEEE
T ss_pred              -HHHHHHHHHHHHHHhCCC--EEEEE
```

```
No 8
>d2za4b_ c.9.1.1 (B:) automated matches {Bacillus amyloliquefaciens [TaxId: 1390]}
Probab=72.27  E-value=70  Score=29.00  Aligned_cols=62  Identities=10%  Similarity=0.091  Sum_probs=0.0

Q ss_pred             eeceeecchHHHHHHHHHHHHHHHhcCCCcEEEEEeChHHHHhc--cHHHHHHHHHHHHHHHhcCCCEEEEE
Q gi|365905381|r  164 TFPLEFMETESLVDELLNLIESSLKKANNDTWIVLYNLDSFVSR--S-AKKTILNKLKEFTDRF-ELKVIY  230 (340)
Q Consensus       175 ~~~le~v~~~~~~~~~~L~ll~~~l~~~~~~~~lLvlkNiD~fls~--~~s~~~i~~~l~~lt~k~~~l~~ii  243 (365)
                     .+|=.|-.|-+++.+.|.=+      .+.|.-|+++|.+.+...  + .+..++..+.+-+.+.|+|++++
T Consensus        26 ~~P~~fG~NlDAL~D~L~~~------~~~p~~I~~~~~~~~~~~~~l~~-~~~~l~~vl~da~~~~~~l~i~~   89 (90)
T d2za4b_         26 ALPEYYGENLDALWAALTGW------VEYPLVLEWRQFEQSKQLTEN-GAESVLQVFREAKAEGADITIIL   89 (90)
T ss_dssp             TCCTTCCCSHHHHHHHHHHT------CCSSEEEEEETHHHHHHHC-C-HHHHHHHHHHHHHHHTTCCEEEEE
T ss_pred             CCCcccCCCHHHHHHHhcCC------CCCCCEEEeCcHHHHHHCHH-HHHHHHHHHHHHHHHHCCCceEec


No 9
>2fnaA01
Probab=71.53  E-value=60  Score=31.56  Aligned_cols=87  Identities=24%  Similarity=0.332  Sum_probs=0.0

Q ss_pred             HHhhcccccccCCceeceeecchHHHHHHHHHHHHHHHhcCCCcEEEEEeChHHHHhccHHHHHHHHHHHHHHHhcCCCEE
Q gi|365905381|r  150 LKYFLKVSYEDSNRTFPLEFMETESLVDELLNLIESSLKKANNDTWIVLYNLDSFVSRS-AKKTILNKLKEFTDRF-ELK  227 (340)
Q Consensus       161 l~k~~~~~~~~~~~~~~~~~le~v~~~~~~~~~~L~ll~~~l~~~~~~~~lLvlkNiD~fls~~~s~~~i~~~l~~lt~k~~~l~  240 (365)
                     +.+.+...-........+..-........+++.+...+++.+++++++|-.++......  . ..+...+....++.+++.
T Consensus        95 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~DE~~~~~~~~-~-~~~~~~l~~~~~~~~~~~~  172 (208)
T 2fnaA01         95 LKKALKKIEGIEIRGLSLELKLKKESLEELFELLEKLLEKKGKKVILIIDEIQKLAKLE-G-KDLLSALKSILDNLKNVL  172 (208)
T ss_pred             HHHHcCCcccCCCCCCCccccCCCCCCHHHHHHHHHHHHHHHcCCCEEEEECChhHhhccc-c-hhHHHHHHHHHHHcCcCEE


Q ss_pred             EEEEcCCCC
Q gi|365905381|r  228 VIYLGNSLT  236 (340)
Q Consensus       241 ~iif~s~~~  249 (365)
                     +++.+|..+
T Consensus       173 ~~~~gs~~~  181 (208)
T 2fnaA01        173 FIATGSSVG  181 (208)
T ss_pred             EEEeeCcHh


No 10
>d1e69a_ c.37.1.12 (A:) Smc head domain {Thermotoga maritima [TaxId: 2336]}
Probab=70.93  E-value=13  Score=37.79  Aligned_cols=47  Identities=21%  Similarity=0.330  Sum_probs=0.0

Q ss_pred             EEEEEecccce---EEEeccc-eEEEecCcHHHHHHHHHHHHHHhCCCCccH
Q gi|365905381|r    2 RLKIEIDDQSF---FELDFGD-IVYISGYNHQNMWKIYRSLYYYFNKSPM--   47 (340)
Q Consensus         4 ~l~i~h~~~~~~---ie~~~g~-~tqi~G~N~qlK~yi~q~L~~yF~gkkyse   51 (365)
                     +++|+ .|+.+   .+++|++ +|.|+|.|--=|-=|+++|.|.+++.....
T Consensus         5 ~i~I~-NFksf~~~~~i~f~~~lnvlvG~NgsGKS~iL~Ai~~~lg~~~~~~   55 (308)
T d1e69a_          5 KLYLK-GFKSFGRPSLIGFSDRVTAIVGPNGSGKSNIIDAIKWVFGEQSKKE   55 (308)
T ss_dssp             EEEEE-SBTTBCSCEEEEECCSSEEEEEECCTTTCSTHHHHHHHHHTSCC-----
T ss_pred             EEEEE-CeEEECCCEEEeCCCCeEEEECCCCCcHHHHHHHHHHHhCCCchhh


No 11
>PF02463 SMC_N:  RecF/RecN/SMC N terminal domain;  InterPro: IPR003395 This domain is found at the N terminus of structural maintenance of
chromosomes (SMC) proteins, which function together with other proteins in a range of chromosomal transactions, including chromosome condensation,
sister-chromatid cohesion, recombination, DNA repair and epigenetic silencing of gene expression []. The domain is also found in RecF and RecN
proteins, which are involved in DNA metabolism and recombination.; PDB: 3HTK_A 1W1W_C 2WD5_A 3L51_A 1XEW_Y 3KTA_B 3NWC_B 1XEX_A 1GXL_C 1GXK_A ....
Probab=69.41  E-value=23  Score=35.22  Aligned_cols=46  Identities=24%  Similarity=0.339  Sum_probs=0.0

Q ss_pred             CceEEEEEecccce----EEEeccc-eEEEecCcHHHHHHHHHHHHHHhCCC
Q gi|365905381|r    1 --MRLKIEIDDQSF----FELDFGD-IVYISGYNHQNMWKIYRSLYYYFNKS   45 (340)
Q Consensus         1 ~mM~l~i~h~~~~~----ie~~~g~-~tqi~G~N~qlK~yi~q~L~~yF~gk   47 (365)
                     |.-+++|+ +++.+    .+++|++ +|.|+|.|-.=|-=+..+|.|.++++
T Consensus         1 ~I~~l~i~-nFr~~~~~~~~~~~~~~~~~i~G~NGsGKS~ileAi~~~l~~~   51 (220)
T PF02463_consen    1 MIKSLEIE-NFRNFKGKNAELSFSPGLNVIVGPNGSGKSNILEAIEFVLGGR   51 (220)
T ss_dssp             EEEEEEEE-SBTTC-SCEEEEETTSSEEEEEESTTSSHHHHHHHHHHHTTSS
T ss_pred             CccEEEEc-CceEEecCeEEEecCCCCEEEEcCCCCCHHHHHHHHHHHHHc


No 12
>d2hxxa_ c.9.1.1 (A:) automated matches {Bacillus amyloliquefaciens [TaxId: 1390]}
Probab=69.32  E-value=1.1e+02  Score=27.82  Aligned_cols=52  Identities=10%  Similarity=0.179  Sum_probs=0.0

Q ss_pred             HHHHHHHHhcCCC-cEEEEEeChHHHHhc--cHHHHHHHHHHHHHHHhcCCCEEEEE
Q gi|365905381|r  180 LNLIESSLKKANN-DTWIVLYNLDSFVSR--S-AKKTILNKLKEFTDRF-ELKVIY  230 (340)
Q Consensus       191 L~ll~~~l~~~~~~-~~lLvlkNiD~fls~--~~s~~~i~~~l~~lt~k~~~l~~ii  243 (365)
                     |+-|--.|..... |+-|+++|.+.+...  + .+..++..+.+-+.++|+|++++
T Consensus        34 lDAL~D~L~~~~~~p~~I~~~~~~~~~~~~~l~~-~~~~l~~v~~da~~~~~~l~i~~   88 (89)
T d2hxxa_         34 LDALXDALTGXVEYPLVLEXRQFEQSKQLTEN-GAESVLQVFREAKAEGADITIIL   88 (89)
T ss_dssp             HHHHHHHHHHTCCSSEEEEEETHHHHHHTTTT-HHHHHHHHHHHHHHHTTCCEEEEE
T ss_pred             HHHHHHHHcCcCCCCcEEEECChHHHHHCcH-HHHHHHHHHHHHHHHHCCCceEec


No 13
>PF13304 AAA_21:  AAA domain; PDB: 3QKS_B 1US8_B 1F2U_B 1F2T_B 3QKT_A 1II8_B 3QKR_B 3QKU_A.
Probab=68.12  E-value=1.6e+02  Score=27.89  Aligned_cols=215  Identities=16%  Similarity=0.181  Sum_probs=0.0

Q ss_pred             eEEEecCcHHHHHHHHHHHHHHHh----------------CCCCccHHHhhhhcCCCCeEEecCccccCccceEEEEEccH
Q gi|365905381|r   20 IVYISGYNHQNMWKIYRSLYYYF----------------NKSPM--LSTNIYGDDNIEISLDDVSP-STKNTQTYFISNR   80 (340)
Q Consensus        22 ~tqi~G~N~qlK~yi~q~L~~yF----------------~gkkyse~e~~if~~~ep~I~~d~~~l~~rk~~~~i~Is~~   85 (365)
                     +|.|+|.|.-=|--++++|..++                ....+...-.........+...+.... ....+.+..-...
T Consensus         1 i~viiG~N~sGKS~il~ai~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~-~~~~~~~~~~~~~~~   79 (303)
T PF13304_consen    1 INVIIGPNGSGKSNILEAIYFLFSRKFSNNSKDDIIFNSNIDSISFSFDSESFEIDIEFEIEDFES-DNIIYEYSFSFNK   79 (303)
T ss_dssp             --------------------------------------------------------------------------------
T ss_pred             CCcccccccccccccccccccccccccccccccccccccccccccccchhccccceeEeeecccc-cceeeeeeecccc


Q ss_pred             HHH----------------------------------------------------------HHHHHhcccchHHHHHHH-
Q gi|365905381|r   81 DSI----------------------------------------------------------YQQMIYKKGNLLFENLNT-  101 (340)
Q Consensus        86 ~di----------------------------------------------------------~eqm~~kKGTll~~yL~~-  106 (365)
                     ...                                                          .+...........-.-..
T Consensus        80 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~  159 (303)
T PF13304_consen   80 SDIIEEWLSEYDNKKKIFIFERKNSDEEIDESNEKFDKIKKIRNIFNKIFLSLNKNNSLFFSEILKNISKNFNSISKNND  159 (303)
T ss_dssp             ------------------------------------------------------------HHHHHHHHHHHHHHHHHHHH
T ss_pred             ccchhhhhhcccccccceeeeccccccccccccccccccccccccccccccccchhhhhHHhhhhhhhhhhhhhhhchhh


Q ss_pred             -----HhhhHHHHHHHHHHHHHHHHHHHHHhChHHHHHhhcCCeEEEhhccccchhHHHHHHhhcccccccCCceeceeec
Q gi|365905381|r  102 -----LNDSFEITRILEELNDTNTKLSLLVQ--EHINNYANNLIIDFEDV--TYFD--VLKYFLKVSYEDSNRTFPLEFM  170 (340)
Q Consensus       107 -----~~~~~dI~~~le~iNd~l~~i~~~iN~~~~in~~~~~i~~~~e~~~~~~~~~~~~~ll~k~~~~~~~~~~~~~~~~le~v  181 (365)
                          ....+.............................+.............  .............+..+|+..+
T Consensus       160 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~-~~~~~~~~~~~~~~~~~~~~~  237 (303)
T PF13304_consen  160 KIRNSEIDNFNDFKDEINESEIIDLLNKLIKEINDIFIDISDIEIEIEKSIKIDED--INESSIIIKIKGDGKEIPLSSL  237 (303)
T ss_dssp             HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHTTTSS-EEEEEEETTEE------------EEEEEETTEEEEGGGS--
T ss_pred             hhhhhhhhhhHHHHHHHHHHhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccc--cccceeEEEEEcCCCCeEEeeccC
```

```
Q ss_pred             chHHHHHHHHHHHHHHHHhcCCCcEEEEEeChHHHHhccHHHHHHHHHHHHHHHhcCCCEEEEEE
Q gi|365905381|r  171 ETESLVDELLNLIESSLKKANNDTWIVLYNLDSFVSRS-AKKTILNKLKEFTDRF-ELKVIYL  231 (340)
Q Consensus       182 ~~~~~~~~~L~ll~~~l~~~~~~~~lLvlkNiD~fls~~~s~~~i~~~l~~lt~k~~~l~~iif  244 (365)
                      +.+++-..  .++..+.....+..+++|.-||..|.-. --.++++.+.++++  .+..+|+.
T Consensus       238 S~G~~~~l--~l~~~l~~~~~~~~illiDEpE~~LHp~-~q~~l~~~l~~~~~--~~~Qviit  295 (303)
T PF13304_consen  238 SSGEKRLL--SLLSLLLSAKKNGSILLIDEPENHLHPS-WQRKLIELLKELSK--KNIQVIIT  295 (303)
T ss_dssp             -HHHHHHH--HHHHHHHTTTTT-SEEEEESSSTTSSHH-HHHHHHHHHHHTGG--GSSEEEEE
T ss_pred             CHHHHHHH--HHHHHHhCcCCCCeEEEecCCcCCCCHH-HHHHHHHHHHhhCc--cCCEEEEe


No 14
>1w1w_A Structural maintenance of chromosome 1; cohesin, chromosome segregation, cell adhesion, kleisin, MIT cell cycle; HET: AGS; 2.90A
{Saccharomyces cerevisiae} SCOP: c.37.1.12
Probab=67.30  E-value=1.3e+02  Score=31.47  Aligned_cols=124  Identities=11%  Similarity=0.092  Sum_probs=0.0

Q ss_pred             HHHHHHHHHHHHHHHHHHhChHHHHHhhcCCeEEEhhccccchhHHHHHHhhcccccccCCcee-ceeecchHHHHHHH
Q gi|365905381|r  108 ITRILEELNDTNTKLSLLVQ--EHINNYANNLIIDFEDV--TYFD--VLKYFLKVSYEDSNRTF-PLEFMETESLVDELL  180 (340)
Q Consensus       113 I~~~le~iNd~l~~i~~~iN~~~~in~~~~~i~~~~e~~~~~~~~~~~~ll~k~~~~~~~~~~~~~~~-~le~v~~~~~~~~~~L  191 (365)
                      +.+.+..+|+++..++..+.........+..-..++.-.  +.++  .+...+.......++.. |++.+|.+++-...|
T Consensus       269 ~~~~~~~i~~~~~~if~~L~~~~~~~~~~~~~~~~~~~l~~~--~~~~--~~~~~i~~~~~~~~~~~~~~~~~lS~Ge~~~~~l  344 (430)
T 1w1w_A          269 FEKTFDYVSDHLDAIYRELTKNPNSNVELAGGNASLTIE--DEDE--PFNAGIKYHATPPLKRFKDMEYLSGGEKTVAAL  344 (430)
T ss_dssp             ---CHHHHHHHHHHHHHHTC-----------CEEEEC----------------CEEEECTTCCCCCGGGSCHHHHHHHHH
T ss_pred             HHHHHhHHHHHHHHHHHHHHHhhhhhhhhhcccccccccc--cccc--hhhcCceEEEecCCccceehhhcCccHHHHHHH


Q ss_pred             HHHHHHHhcCCCcEEEEEeChHHHHhccHHHHHHHHHHHHHHHhcCCCEEEEEE
Q gi|365905381|r  181 NLIESSLKKANNDTWIVLYNLDSFVSRS-AKKTILNKLKEFTDRF-ELKVIYL  231 (340)
Q Consensus       192 ~ll~~~l~~~~~~~~lLvlkNiD~fls~~~s~~~i~~~l~~lt~k~~~l~~iif  244 (365)
                      -++-...+..+.| +++|.-||..|.-. --..+++.+.++  ..++.-+|+.
T Consensus       345 al~la~~~~~~~~~~-i~llDEpd~~Ld~~-~~~~l~~~l~~~~--~~~~~Q~Ivt  393 (430)
T 1w1w_A          345 ALLFAINSYQPSP-FFVLDEVDAALDIT-NVQRIAAYIRRH--RNPDLQFIVI  393 (430)
T ss_dssp             HHHHHHHTSSCCS-EEEESSTTTTCCHH-HHHHHHHHHHHH--CBTTBEEEEE
T ss_pred             HHHHHHHhcCCCC-EEEEECCCCCCCHH-HHHHHHHHHHHH--hCCCCEEEEE


No 15
>3zgx_A SMC, chromosome partition protein SMC; cell cycle; 3.40A {Bacillus subtilis}
Probab=66.45  E-value=20  Score=36.46  Aligned_cols=44  Identities=20%  Similarity=0.338  Sum_probs=0.0

Q ss_pred             EEEEEecccce---EEEeccc-eEEEecCcHHHHHHHHHHHHHHHhCCCC
Q gi|365905381|r    2 RLKIEIDDQSF---FELDFGD-IVYISGYNHQNMWKIYRSLYYYFNKSP   46 (340)
Q Consensus         4 ~l~i~h~~~~~~---ie~~~g~-~tqi~G~N~qlK~yi~q~L~~yF~gkk   48 (365)
                      +++|+ .|+.+   .+++|++ +|.|+|.|--=|-=|+++|.|.+|++.
T Consensus         5 ~i~i~-nFks~~~~~~i~f~~~~~vivG~NGsGKS~iLeAi~~~lg~~~   52 (426)
T 3zgx_A           5 RLDVI-GFKSFAERISVDFVKGVTAVVGPNGSGKSNITDAIRWVLGEQS   52 (426)
T ss_dssp             CEEEE-SSSSBCSCCCBCCCSSEEEEECC---CTTHHHHHHHHHHTCCC
T ss_pred             EEEEE-CceeecCCEEEecCCCeEEEECCCCCcHHHHHHHHHHHhCCCc
```

HHPRED hits for Csn2 Mycoplasma

```
 No Hit                             Prob E-value P-value  Score    SS Cols Query HMM  Template HMM
  1 PIRSF037215 Peptidase_T         21.0 2.9E+02   0.002   31.2   3.1   33   90-122   293-325 (419)
  2 PF07693 KAP_NTPase:  KAP famil  17.6 1.8E+03   0.012   22.8   7.1   81  143-237   157-246 (325)



No 9
>PIRSF037215 Peptidase_T
Probab=20.96  E-value=2.9e+02  Score=31.18  Aligned_cols=33  Identities=15%  Similarity=0.278  Sum_probs=0.0

Q ss_pred             cchHhhhhHHHHHHHHHHHHHHhhhcchheecc
Q gi|350546883|r   83 -IETNKILGNDKI-HNIKEYINSKIGFDVIEEK  113 (250)
Q Consensus        90 ~~~~dki~~~~~~i~~~I~~~iN~kiGfdfv~~e  122 (265)
                     +|+.++.--++.+++++|+++|++-|..-|.++
T Consensus       293 DFd~~~F~~rK~~~~~~v~~~n~~~g~~~~~~e  325 (419)
T PIRSF037215     293 DFDRDGFEARKELIKEIVDKLNAKYGRARIKLE  325 (419)
T ss_pred             CCChhHHHHHHHHHHHHHHHHHhhCCCCeEEEE


No 10
>PF07693 KAP_NTPase:  KAP family P-loop domain;  InterPro: IPR011646 The KAP (after Kidins220/ARMS and PifA) family of predicted NTPases are
sporadically distributed across a wide phylogenetic range in bacteria and in animals. Many of the prokaryotic KAP NTPases are encoded in plasmids
and tend to undergo disruption to form pseudogenes. A unique feature of all eukaryotic and certain bacterial KAP NTPases is the presence of two or
four transmembrane helices inserted into the P-loop NTPase domain. These transmembrane helices anchor KAP NTPases in the membrane such that the P-
loop domain is located on the intracellular side [].
Probab=17.57  E-value=1.8e+03  Score=22.80  Aligned_cols=81  Identities=16%  Similarity=0.138  Sum_probs=0.0

Q ss_pred             HHHHHHHHHHhhcccceEEEEEecCCcccHHHHHHHhc---------ceeEEEEeCcHHHHHhhhHHHHHHHhhccCCcc
Q gi|350546883|r  134 KNIFISLIKNNLFD-ERMTFVFCDIDWLNINDISEYIN---------SHNFIFVTND-LRKNIVNITQIESCVICND---  199 (250)
Q Consensus       143 ~~~~~s~len~~~~~e~~~vIlKd~d~I~i~~L~e~~n---------~~n~IiltNd~~~~~i~~f~~iE~~~f~~~~~~~  213 (265)
                     ++.+-.+.+.+.=...+++|++-|+|-..-+...+.+.         ++.+|+..|.  -......+.--...+
T Consensus       157 ~~~~~~~~~~l~~~~~~iViiIDdLDR~~~~~i~~~l~~ik~~~~~~~~i~~Il~~D~--~~l~~ai~~~~~~~~~------  228 (325)
T PF07693_consen  157 EELISKIKKKLKESKKRIVIIIDDLDRCSPEEIVELLEAIKLLLDFPNIIFILAFDP--EILEKAIEKNYGEGF------  228 (325)
T ss_pred             HHHHHHHHHhhhcCCceEEEEEcchhcCCcHHHHHHHHHHHHhcCCCCeEEEEEecH--HHHHHHHHhhcCccc------


Q ss_pred             eEEEecchHHHHHHHHHHhcCccc
Q gi|350546883|r  200 FLVEIYDYNKLISYLELKMNTLIN  223 (250)
Q Consensus       214 ~iv~I~n~~~~~~y~lE~~~~~~~i~  237 (265)
                           +...-..|||+.+|.|+.
T Consensus       229 ------~~~~~~~~yLeKiiq~~~~  246 (325)
T PF07693_consen  229 ------DEIDGREYLEKIIQVPFS  246 (325)
T ss_pred             ------ccccHHHHHHhhcCeEEE
```

HHPRED hits for Csn2 Staphylococcus long

```
 No Hit                             Prob E-value P-value  Score    SS Cols Query HMM  Template HMM
  1 PF13304 AAA_21:  AAA domain; P  91.0     4.4   3E-05   33.6   7.1   59  175-234   241-302 (303)
  2 COG3593 Predicted ATP-dependen  89.8     1.9 1.3E-05   48.4   5.6   91  132-236   255-347 (581)
  3 3zgx_A SMC, chromosome partiti  87.6     9.7 6.5E-05   35.0   7.4   60  176-235   334-395 (426)
  4 COG3910 Predicted ATPase [Gene  87.5     3.5 2.3E-05   42.5   5.2   54  182-235   136-192 (233)
  5 g1ii8.1 c.37.1.12 (A:,B:) Rad5  86.8     2.9 1.9E-05   37.0   3.7   35  201-235   311-348 (369)
  6 PIRSF034888 P-loop_UCP034888    86.4     5.8 3.9E-05   40.2   6.0   61  175-236   308-372 (444)
  7 4i99_A Chromosome partition pr  84.0     4.2 2.8E-05   38.5   3.7   35  201-235   276-312 (354)
  8 g1f2t.1 c.37.1.12 (A:,B:) Rad5  82.0     6.5 4.4E-05   35.4   3.9   35  201-235   234-271 (292)
  9 SUPFAM0046693 c.37.1 P-loop co  80.5     8.3 5.6E-05   30.1   3.5   34  201-234   235-271 (293)
 10 d1e69a_ c.37.1.12 (A:) Smc hea  79.8      33 0.00022   32.3   7.5   60  176-235   225-286 (308)
 11 PF13175 AAA_15:  AAA ATPase do  78.9      21 0.00014   35.2   6.3   57  178-234   351-415 (415)
 12 PF11398 DUF2813:  Protein of u  71.3      17 0.00011   39.3   4.1   36  201-236   293-328 (373)
 13 SUPFAM0050337 c.37.1 P-loop co  69.9      24 0.00016   30.2   3.8   36  201-236   363-401 (427)
 14 SUPFAM0050943 c.37.1 P-loop co  66.1      31 0.00021   28.1   3.6   34  201-234   256-291 (330)
 15 SUPFAM0036790 c.37.1 P-loop co  65.9      34 0.00023   27.5   3.7   35  201-235   250-286 (308)


No 1
>PF13304 AAA_21:  AAA domain; PDB: 3QKS_B 1US8_B 1F2U_B 1F2T_B 3QKT_A 1II8_B 3QKR_B 3QKU_A.
Probab=90.97  E-value=4.4  Score=33.65  Aligned_cols=59  Identities=29%  Similarity=0.388  Sum_probs=0.0

Q ss_pred             HHHHHHHHHHHHHHhhcCeEEEEEEcccCCCCHHHHHHHHHHHhcC---CCcEEEEcCChhHH
Q gi|315659845|r  163 IEQVLPLVIEELKKQHNNQLLLIYLYPEANLSPKEQIRLKQLLISL---DVNIIVLTGSMHFL  222 (333)
Q Consensus       175 leQvlPL~veemn~q~~~k~llIYLYPEAnLS~KEqir~r~~l~~L---~i~iIVLT~S~~FL  234 (345)
                     ..|++ .++-.+-....+..+++.==||++|-|+-|.+|-++|.++   ++.+|+-|.||++|
T Consensus       241 ~~~~l-~l~~~l~~~~~~~~~illiDEpE~~LHp~~q~~l~~~l~~~~~~~~~~QviitTHSp~il  302 (303)
T PF13304_consen  241 EKRLL-SLLSLLLSAKKNGSILLIDEPENHLHPSWQRKLIELLKELSKKNIQVIITTHSPFIL  302 (303)
T ss_dssp             HHHHH-HHHHHHHTTTTT-SEEEEESSSTTSSHHHHHHHHHHHHHTGGGSSEEEEEES-GGG-
T ss_pred             HHHHH-HHHHHHhCcCCCCeEEEecCCcCCCCHHHHHHHHHHHHhhCccCCEEEEeCccchhc


No 2
>COG3593 Predicted ATP-dependent endonuclease of the OLD family [DNA replication, recombination, and repair]
Probab=89.82  E-value=1.9  Score=48.39  Aligned_cols=91  Identities=29%  Similarity=0.317  Sum_probs=0.0

Q ss_pred             HHHHHhccCCCccccccccHHHHHHHhceeehhhhhHHHHHHHHHHHHHHHHHHHHHHhhcCeEEEEEEcccCCCCHHHHH
Q gi|315659845|r  120 DLLELLNYTDLPLRTKQFDIKQFVKMLKFEYDLKDDYTRLIVRIEQVLPLVIEELKKQHNNQLLLIYLYPEANLSPKEQI  199 (333)
Q Consensus       132 DLl~~l~~t~LPL~iK~FDiKQFiK~L~FeY~l~~DYsrLIvRleQvlPL~veemn~q~~~k~llIYLYPEAnLS~KEqi  211 (345)
                     |+...++....  .+|++++++    .++.|-=.-.    +.+++-..++.+++...-.    +|+.==|||.|=|.-|.
T Consensus       255 ~~~k~~~~~~~~--~~~~~~~~~----~~lgl~nl~~----~~~l~~~~~~~~~~~~r~~----ILLiEEPEahLHPq~q~  320 (581)
T COG3593        255 DLSKYLKILLR--NTKQIEIES----NGLGLNNLIA----MQLLQAIGSLELERVARPN----ILLIEEPEAHLHPQLQA  320 (581)
T ss_pred             hHHHHHHHHHH--Hhccchhhh----hchhhhhhHH----HHHHHHHhHhhhhhcccc----eEEEeCchhhcCHHHHH


Q ss_pred             HHHHHHhcCC--CcEEEEcCChhHHHH
Q gi|315659845|r  200 RLKQLLISLD--VNIIVLTGSMHFLAQ  224 (333)
Q Consensus       212 r~r~~l~~L~--i~iIVLT~S~~FLae  236 (345)
                     .+.++|+.+|  +.+||-|.|+||+|.
T Consensus       321 ~l~~ll~~l~~~~~Q~IvTThS~~~~s~  347 (581)
T COG3593        321 VLWDLLNNLPLGLQRIVTTHSPHLLSL  347 (581)
T ss_pred             HHHHHHhcCCcceEEEEEcCCcccccc


No 3
>3zgx_A SMC, chromosome partition protein SMC; cell cycle; 3.40A {Bacillus subtilis}
Probab=87.60  E-value=9.7  Score=35.02  Aligned_cols=60  Identities=17%  Similarity=0.113  Sum_probs=0.0

Q ss_pred             HHHHHHHHHHHHHHhhcCeEEEEEEcccCCCCHHHHHHHHHHHhcC--CCcEEEEcCChhHHH
Q gi|315659845|r  164 EQVLPLVIEELKKQHNNQLLLIYLYPEANLSPKEQIRLKQLLISL--DVNIIVLTGSMHFLA  223 (333)
Q Consensus       176 eQvlPL~veemn~q~~~k~llIYLYPEAnLS~KEqir~r~~l~~L--~i~iIVLT~S~~FLa  235 (345)
                     .+++.|..--.-...++.++++-==||++|.|.-|-++-+.|.++  ...|||-|.||+|++
T Consensus       334 k~~~~lal~~a~~~~~~~~~~~~~ilDEpe~~Lhp~~~~~~l~~~l~~~s~~~QiiitTHsp~~~~  395 (426)
T 3zgx_A         334 RALTAIALLFSILKVRPVPFCVLDEVEAALDEANVFRFAQYLKKYSSDTQFIVITHRKGTME  395 (426)
T ss_dssp             HHHHHHHHHHHHHHHSCCSCEEEESTTTSCTHHHHHHHHHHHHHCSSSSCEEEECSCHHHHH
T ss_pred             HHHHHHHHHHHHHhcCCCCEEEEECCCCCCCHHHHHHHHHHHHHHhCCCCCEEEEeCcHHHHH


No 4
>COG3910 Predicted ATPase [General function prediction only]
Probab=87.49  E-value=3.5  Score=42.49  Aligned_cols=54  Identities=35%  Similarity=0.431  Sum_probs=0.0

Q ss_pred             HHHHHHHhhcCeEEEEEEcccCCCCHHHHHHHHHHHHhcC---CCcEEEEcCChhHHH
Q gi|315659845|r  170 VIEELKKQHNNQLLLIYLYPEANLSPKEQIRLKQLLISL---DVNIIVLTGSMHFLA  223 (333)
Q Consensus       182 ~veemn~q~~~k~llIYLYPEAnLS~KEqir~r~~l~~L---~i~iIVLT~S~~FLa  235 (345)
                     +..=+-+..++|++.|-==|||.|||--|+.+-..|.+|   |--||+-|.||--||
T Consensus       136 f~~i~~~rf~~~GiYiLDEPEa~LSp~RQlella~l~~la~sGaQ~IiATHSPiLlA  192 (233)
T COG3910        136 FLAIFHNRFNGQGIYILDEPEAALSPSRQLELLAILRDLADSGAQIIIATHSPILLA  192 (233)
T ss_pred             HHHHHHHhccCceEEecCccccCCHHHHHHHHHHHHHHhcCCeEEEEecChhhee


No 5
>g1ii8.1 c.37.1.12 (A:,B:) Rad50 {Archaeon Pyrococcus furiosus [TaxId: 2261]}
Probab=86.76  E-value=2.9  Score=36.99  Aligned_cols=35  Identities=11%  Similarity=0.219  Sum_probs=0.0

Q ss_pred             ccCCCCHHHHHHHHHHHhcC---CCcEEEEcCChhHHH
Q gi|315659845|r  189 PEANLSPKEQIRLKQLLISL---DVNIIVLTGSMHFLA  223 (333)
Q Consensus       201 PEAnLS~KEqir~r~~l~~L---~i~iIVLT~S~~FLa  235 (345)
                     ||+.|.|+-|.++-++|..+   +--+||.|.||.++.
T Consensus       311 pe~~Lh~~~~~~~l~~~l~~~~~~~~~~~QviitTHs~~~~~~  348 (369)
T g1ii8.1        311 PTPYLDEERRRKLITIMERYLKKIPQVILVSHDEELKD  348 (369)
T ss_dssp             CSSSSCSHHHHHHHHHHHHTGGGSSEEEEEESCGGGGG
T ss_pred             CCCCCCHHHHHHHHHHHHHHhcCCEEEEEechHHHH


No 6
>PIRSF034888 P-loop_UCP034888
Probab=86.44  E-value=5.8  Score=40.15  Aligned_cols=61  Identities=36%  Similarity=0.480  Sum_probs=0.0

Q ss_pred             HHHHHHHHHHHHHHhhcCeEEEEEEcccCCCCHHHHHHHHHHHhcC---C-CcEEEEcCChhHHHH
Q gi|315659845|r  163 IEQVLPLVIEELKKQHNNQLLLIYLYPEANLSPKEQIRLKQLLISL---D-VNIIVLTGSMHFLAQ  224 (333)
Q Consensus       175 leQvlPL~veemn~q~~~k~llIYLYPEAnLS~KEqir~r~~l~~L---~-i~iIVLT~S~~FLae  236 (345)
                     +-++||+++-== ...+.+.++|===|||.|=|+-|..+.++|..+   + +.+||=|.|-|++..
T Consensus       308 ~~~~LPi~La~~-~~~K~~~~viIEEPEaHLHP~~Q~~~~~~~l~~~~~~~~gr~~v~~~THSd~~~~~~
T PIRSF034888     308 VAYVLPILLALR-LLSKEGSLVIIEEPEAHLHPKAQSVLADLLADLVKSGRVQVVVTTHSDHLLNR  372 (444)
T ss_pred             HHHHHHHHHhc-cccCCCcEEEeCcccccCHhHHHHHHHHHHHHHHcCCceEEEECCHHHHHH
```

```
No 7
>4i99_A Chromosome partition protein SMC; winged-helix domain and SMC head domain, chromosome condensa SCPB, DNA binding protein; 2.30A {Pyrococcus
furiosus}
Probab=84.00  E-value=4.2  Score=38.51  Aligned_cols=35  Identities=23%  Similarity=0.264  Sum_probs=0.0

Q ss_pred             ccCCCCHHHHHHHHHHhcC--CCcEEEEcCChhHHH
Q gi|315659845|r  189 PEANLSPKEQIRLKQLLISL--DVNIIVLTGSMHFLA  223 (333)
Q Consensus       201 PEAnLS~KEqir~r~~l~~L--~i~iIVLT~S~~FLa  235 (345)
                     ||++|.|+-|.++.+.|..+  ..-+|+-|.||.+++
T Consensus       276 pe~~Lhp~~~~~l~~~L~~~~~~~~QviitTHsp~~l~  312 (354)
T 4i99_A          276 IDAHLDDANVKRVADLIKESSKESQFIVITLRDVMMA  312 (354)
T ss_dssp             TTTTCCHHHHHHHHHHHHHHTTSEEEEECCCHHHHT
T ss_pred             HHhhcCHHHHHHHHHHHHhCcCCeEEEEECCHHHHH


No 8
>g1f2t.1 c.37.1.12 (A:,B:) Rad50 {Archaeon Pyrococcus furiosus [TaxId: 2261]}
Probab=82.01  E-value=6.5  Score=35.39  Aligned_cols=35  Identities=11%  Similarity=0.217  Sum_probs=0.0

Q ss_pred             ccCCCCHHHHHHHHHHhcCC---CcEEEEcCChhHHH
Q gi|315659845|r  189 PEANLSPKEQIRLKQLLISLD---VNIIVLTGSMHFLA  223 (333)
Q Consensus       201 PEAnLS~KEqir~r~~l~~L~---i~iIVLT~S~~FLa  235 (345)
                     ||++|.|+.|.++-++|..+-   .-|||.|.|++++.
T Consensus       234 p~~~Ld~~~~~~l~~~l~~~~~~~~~~qviv~TH~~~~~~  271 (292)
T g1f2t.1         234 PTPYLDEERRRKLITIMERYLKKIPQVILVSHDEELKD  271 (292)
T ss_dssp             CSCTTCHHHHHHHHHHHHTGGGSSEEEEEESCGGGGG
T ss_pred             CcccCCHHHHHHHHHHHHHHhCCCEEEEEeecHHHHH


No 9
>SUPFAM0046693 c.37.1 P-loop containing nucleoside triphosphate hydrolases (52540) SCOP seed sequence: d1f2t.1.
Probab=80.50  E-value=8.3  Score=30.15  Aligned_cols=34  Identities=35%  Similarity=0.573  Sum_probs=0.0

Q ss_pred             ccCCCCHHHHHHHHHHhcC---CCcEEEEcCChhHH
Q gi|315659845|r  189 PEANLSPKEQIRLKQLLISL---DVNIIVLTGSMHFL  222 (333)
Q Consensus       201 PEAnLS~KEqir~r~~l~~L---~i~iIVLT~S~~FL  234 (345)
                     |.++|.+..+.++.++|.++   +.-+|+.|.++.++
T Consensus       235 p~~~LD~~~~~~~~~~l~~~~~~~~~~~~~~H~~~~~  271 (293)
T SUPFAM0046693   235 PTAGLDPETREELLELLKELAKEGKTVIVVTHDLELL  271 (293)
T ss_pred             CcccCCHHHHHHHHHHHHHHhcCCEEEEEEccHHHH


No 10
>d1e69a_ c.37.1.12 (A:) Smc head domain {Thermotoga maritima [TaxId: 2336]}
Probab=79.78  E-value=33  Score=32.35  Aligned_cols=60  Identities=17%  Similarity=0.143  Sum_probs=0.0

Q ss_pred             HHHHHHHHHHHHHhhcCeEEEEEEcccCCCCHHHHHHHHHHH--hcCCCcEEEEcCChhHHH
Q gi|315659845|r  164 EQVLPLVIEELKKQHNNQLLLIYLYPEANLSPKEQIRLKQLL--ISLDVNIIVLTGSMHFLA  223 (333)
Q Consensus       176 eQvlPL~veemn~q~~~k~llIYLYPEAnLS~KEqir~r~~l--~~L~i~iIVLT~S~~FLa  235 (345)
                     .+++-++..-..-+..+..+++-==||+.|.|+.|-++.++|  .+-..-|||-|.||.++.
T Consensus       225 ~~l~~~~~~~~~~~~~~~~~~~~~~~~iDEpe~~Lhp~~~~~~l~~~l~~~~~~~~QviitTHsp~~~~  286 (308)
T d1e69a_         225 KALVGLALLFALMEIKPSPFYVLDEVDSPLDDYNAERFKRLLKENSKHTQFIVITHNKIVME  286 (308)
T ss_dssp             HHHHHHHHHHHTTTSCCSEEEEESCCSSCCHHHHHHHHHHHHTTTSEEEEECCCTTGGG
T ss_pred             hhHHHHHhhhhhhhccCchhhhhccccCCHHHHHHHHHHHHHhccCCEEEEEEECCHHHHH


No 11
>PF13175 AAA_15:  AAA ATPase domain
Probab=78.91  E-value=21  Score=35.16  Aligned_cols=57  Identities=26%  Similarity=0.446  Sum_probs=0.0

Q ss_pred             HHHHHHHHHHhhcCeE----EEEEEcccCCCCHHHHHHHHHHHhcC----CCcEEEEcCChhHH
Q gi|315659845|r  166 VLPLVIEELKKQHNNQL----LLIYLYPEANLSPKEQIRLKQLLISL----DVNIIVLTGSMHFL  222 (333)
Q Consensus       178 vlPL~veemn~q~~~k~----llIYLYPEAnLS~KEqir~r~~l~~L----~i~iIVLT~S~~FL  234 (345)
                     .+-+++..+-....++.    +++.==|||.|.|+-|..|++.|..+    ++-||+-|.|||++
T Consensus       351 ~~~~~~~~~~~~~~~~~~~~~~~~illidEPE~~LHp~~q~~~~~~~L~~~~~~~~~~~QiiitTHSp~ii  415 (415)
T PF13175_consen  351 YISLLINFLRENKESNNNNYNILLIDEPELHLHPQAQRKFIDFLKKLSKNNNIQIIITTHSPFII  415 (415)
T ss_pred             HHHHHHHHHHhhhccCcceeEEEEeCccccCCHHHHHHHHHHHHHHhccCCCEEEEECCChhhC


No 12
>PF11398 DUF2813:  Protein of unknown function (DUF2813);  InterPro: IPR022602  This entry contains YbjD from Escherichia coli (strain K12), which
is a conserved protein with a nucleotide triphosphate binding domain.
Probab=71.25  E-value=17  Score=39.29  Aligned_cols=36  Identities=36%  Similarity=0.271  Sum_probs=0.0

Q ss_pred             ccCCCCHHHHHHHHHHhcCCCcEEEEcCChhHHHH
Q gi|315659845|r  189 PEANLSPKEQIRLKQLLISLDVNIIVLTGSMHFLAQ  224 (333)
Q Consensus       201 PEAnLS~KEqir~r~~l~~L~i~iIVLT~S~~FLae  236 (345)
                     |||+|=|--+...-.+|+.||.-.|+-|.|+..||+
T Consensus       293 PEarLHP~ml~~aW~ll~~lP~QkI~TTnSgeLLS~  328 (373)
T PF11398_consen  293 PEARLHPIMLAIAWSLLSLLPGQKITTTNSGELLSQ  328 (373)
T ss_pred             cccccCHHHHHHHHHhhCCcccEEeecCCHHHhhh


No 13
>SUPFAM0050337 c.37.1 P-loop containing nucleoside triphosphate hydrolases (52540) SCOP seed sequence: d1w1wa_.
Probab=69.89  E-value=24  Score=30.21  Aligned_cols=36  Identities=36%  Similarity=0.515  Sum_probs=0.0

Q ss_pred             ccCCCCHHHHHHHHHHhcC---CCcEEEEcCChhHHHH
Q gi|315659845|r  189 PEANLSPKEQIRLKQLLISL---DVNIIVLTGSMHFLAQ  224 (333)
Q Consensus       201 PEAnLS~KEqir~r~~l~~L---~i~iIVLT~S~~FLae  236 (345)
                     |.++|.+.-+.++-++|.++   +.-+|+.|.++.+++.
T Consensus       363 ~~~~Ld~~~~~~~~~~~l~~~~~~~~~~~~~~h~~~~~~~~  401 (427)
T SUPFAM0050337   363 PTAGLDPENRERLLELLRELASEGGQVIVVTHDPELAAR  401 (427)
T ss_pred             CcccCCHHHHHHHHHHHHHHhcCCEEEEEEccHHHHHh


No 14
>SUPFAM0050943 c.37.1 P-loop containing nucleoside triphosphate hydrolases (52540) SCOP seed sequence: d1xew.1.
Probab=66.11  E-value=31  Score=28.06  Aligned_cols=34  Identities=35%  Similarity=0.563  Sum_probs=0.0

Q ss_pred             ccCCCCHHHHHHHHHHhcC--CCcEEEEcCChhHH
Q gi|315659845|r  189 PEANLSPKEQIRLKQLLISL--DVNIIVLTGSMHFL  222 (333)
Q Consensus       201 PEAnLS~KEqir~r~~l~~L--~i~iIVLT~S~~FL  234 (345)
                     |.++|.+.-+.++.++|.++  +.-+|+.|.++.++
T Consensus       256 ~~~~Ld~~~~~~~~~~~l~~~~~~~~~~~~~h~~~~~  291 (330)
T SUPFAM0050943   256 PTAGLDPETRERLLELLKELAEGRQVIVVTHDLELA  291 (330)
T ss_pred             ccccCCHHHHHHHHHHHHHHhCCCEEEEEEccHHHH
```

```
No 15
>SUPFAM0036790 c.37.1 P-loop containing nucleoside triphosphate hydrolases (52540) SCOP seed sequence: d1e69a_.
Probab=65.91  E-value=34  Score=27.51  Aligned_cols=35  Identities=34%  Similarity=0.552  Sum_probs=0.0

Q ss_pred             ccCCCCHHHHHHHHHHHhcC--CCcEEEEcCChhHHH
Q gi|315659845|r  189 PEANLSPKEQIRLKQLLISL--DVNIIVLTGSMHFLA  223 (333)
Q Consensus       201 PEAnLS~KEqir~r~~l~~L--~i~iIVLT~S~~FLa  235 (345)
                     |.++|.+..+.++.++|.++  +.-+|+.|..+.++.
T Consensus       250 p~~~Ld~~~~~~~~~~l~~~~~~~~~~~~~~~h~~~~~~  286 (308)
T SUPFAM0036790   250 PTSGLDPETREALLELLRELAKGRTVIVVTHDLELLA  286 (308)
T ss_pred             ccccCCHHHHHHHHHHHHHHhcCCEEEEEEccHHHHh
```

HHPRED hits for Csn2 Staphylococcus short

```
 No Hit                             Prob E-value P-value  Score    SS Cols Query HMM  Template HMM
  1 g1ii8.1 c.37.1.12 (A:,B:) Rad5  96.9   0.017 1.1E-07   49.4   7.8   79  158-237   282-363 (369)
  2 g1f2t.1 c.37.1.12 (A:,B:) Rad5  96.3   0.045   3E-07   47.6   6.8   59  177-237   224-286 (292)
  3 cd03275 ABC_SMC1_euk Eukaryoti  96.2    0.32 2.2E-06   45.2  11.9  123  110-237   111-237 (247)
  4 3qkt_A DNA double-strand break  95.6    0.34 2.3E-06   43.3   8.8   76  158-235   252-330 (339)
  5 PRK11147 ABC transporter ATPas  95.5    0.31 2.1E-06   51.7   9.8   93  138-236   420-515 (635)
  6 3auy_A DNA double-strand break  95.5     1.2   8E-06   38.9  11.5  153   77-237   209-365 (371)
  7 4i99_A Chromosome partition pr  95.5    0.33 2.2E-06   44.8   8.7   82  155-237   244-327 (354)
  8 SUPFAM0046693 c.37.1 P-loop co  95.5    0.58 3.9E-06   35.8   8.4   73  161-235   207-284 (293)
  9 cd03293 ABC_NrtD_SsuB_transpor  95.3    0.75   5E-06   41.5  10.1   93  138-236   112-211 (220)
 10 d1e69a_ c.37.1.12 (A:) Smc hea  95.2    0.36 2.4E-06   43.8   8.0   76  161-237   224-301 (308)
 11 TIGR02211 LolD_lipo_ex lipopro  95.2    0.82 5.5E-06   41.1  10.0   94  138-239   122-221 (221)
 12 SUPFAM0050337 c.37.1 P-loop co  95.2    0.74   5E-06   38.1   8.7   74  161-235   337-413 (427)
 13 3zgx_A SMC, chromosome partiti  95.0    0.42 2.8E-06   42.7   7.7   77  161-239   333-411 (426)
 14 1w1w_A Structural maintenance   95.0    0.79 5.3E-06   41.5   9.2   77  161-239   338-418 (430)
 15 PRK10247 putative ABC transpor  94.9       1   7E-06   41.2  10.0   94  138-238   117-217 (225)


No 1
>g1ii8.1 c.37.1.12 (A:,B:) Rad50 {Archaeon Pyrococcus furiosus [TaxId: 2261]}
Probab=96.90  E-value=0.017  Score=49.40  Aligned_cols=79  Identities=10%  Similarity=0.151  Sum_probs=0.0

Q ss_pred             HHHHHHHHHhhhhhhhccCCcccEEEEECChhhcCHHHHHHHHHHHHhh---ceeEEEEeCCHHHHHhcCccceeeeeecC
Q gi|389815356|r  156 GAYQYQSFLLKSWLHLVHNKATNVCFYDFPENELNKSELKELFQFATSQ---ACTMICLTTSPQVINQVGLSNVHLIKRT  232 (288)
Q Consensus       158 ~~~qrk~Fid~~~~~L~h~~keNvl~~lFPE~hLg~sd~K~f~~~~~~~~---~~t~Ii~T~hP~~I~~i~~~nv~LiK~n  234 (291)
                        |--++=+....-++-.....+.+++++|=||+||-.+-..+++++++..   ..+.|+.|.+|++++... .-+.+-|.+
T Consensus       282 Ge~~~~~la~~l~~~~~~~~~~~~illiDEpe~~Lh~~~~~~~l~~~l~~~~~~~~~~QviitTHs~~~~~~~~d-~~~~~v~~~~~  360 (369)
T g1ii8.1         282 GERIALGLAFRLAMSLYLAGEISLLILDEPTPYLDEERRRKLITIMERYLKKIPQVILVSHDEELKDAAD-HVIRISLEN  360 (369)
T ss_dssp             HHHHHHHHHHHHHHHHHHHSSCSEEEEECCSSSSCSHHHHHHHHHHHHHTGGGSSEEEEEESCGGGGGTSS-EEEEEEECS
T ss_pred             HHHHHHHHHHHHHHHHHhcCCCCCEEEEECCCCCCCHHHHHHHHHHHHHHHhcCCEEEEEechHHHHHhCC-EEEEEEEeC


Q ss_pred             Ccc
Q gi|389815356|r  233 GER  235 (288)
Q Consensus       235 ge~  237 (291)
                    |++
T Consensus       361 g~s  363 (369)
T g1ii8.1         361 GSS  363 (369)
T ss_dssp             SSE
T ss_pred             CEE


No 2
>g1f2t.1 c.37.1.12 (A:,B:) Rad50 {Archaeon Pyrococcus furiosus [TaxId: 2261]}
Probab=96.35  E-value=0.045  Score=47.57  Aligned_cols=59  Identities=14%  Similarity=0.197  Sum_probs=0.0

Q ss_pred             CcccEEEEECChhhcCHHHHHHHHHHHHhhc---eeEEEEeCCHHHHHhcCccceeeee-ecCCcc
Q gi|389815356|r  175 KATNVCFYDFPENELNKSELKELFQFATSQA---CTMICLTTSPQVINQVGLSNVHLI-KRTGER  235 (288)
Q Consensus       177 ~keNvl~~lFPE~hLg~sd~K~f~~~~~~~~~~---~t~Ii~T~hP~~I~~i~~~nv~Li-K~nge~  237 (291)
                    .+..++++|=|++||.....+++.++++++.   .+.|+.|.+|.+++.+  |+|..+ +.+|++
T Consensus       224 ~~~~llllDEp~~~Ld~~~~~~~l~~~l~~~~~~~~~qviv~TH~~~~~~~~~~--D~ii~l~~~~g~~  286 (292)
T g1f2t.1         224 GEISLLILDEPTPYLDEERRRKLITIMERYLKKIPQVILVSHDEELKDAA--DHVIRISLENGSS  286 (292)
T ss_dssp             SSCSEEEEESCSCTTCHHHHHHHHHHHHHTGGGSSEEEEEESCGGGGGGC--SEEEEEEETTEE
T ss_pred             CCCCEEEEeCCccccCCHHHHHHHHHHHHHHHhCCCEEEEEeecHHHHHhC--CEEEEEEecCCEE


No 3
>cd03275 ABC_SMC1_euk Eukaryotic SMC1 proteins; SMC proteins are large (approximately 110 to 170 kDa), and each is arranged into five recognizable
domains.  Amino-acid sequence homology of SMC proteins between species is largely confined to the amino- and carboxy-terminal globular domains. The
amino-terminal domain contains a 'Walker A' nucleotide-binding domain (GxxGxGKS/T, in the single-letter amino-acid code), which by mutational
studies has been shown to be essential in several proteins.  The carboxy-terminal domain contains a sequence (the DA-box) that resembles a 'Walker
B' motif, and a motif with homology to the signature sequence of the ATP-binding cassette (ABC) family of ATPases.  The sequence homology within
the carboxy-terminal domain is relatively high within the SMC1-SMC4 group, whereas SMC5 and SMC6 show some divergence in both of these sequences.
In eukaryotic cells, the proteins are found as heterodimers of SMC1 paired with SMC3, SMC2 with SMC4, and SMC5 with SMC6 (for
Probab=96.25  E-value=0.32  Score=45.21  Aligned_cols=123  Identities=11%  Similarity=0.083  Sum_probs=0.0

Q ss_pred             HHHHHHHHHhceE-ecCCeEeecccchhhHHHHHhhcCcchhhccccchhHHHHHHHHHhhhhhhccCCcccEEEEECCh
Q gi|389815356|r  108 QGELEKAFEQCVV-EFEGMELQPSTELISFEQLIKIAPIEVRTVSDEPMGAYQYQSFLLKSWLHLVHNKATNVCFYDFPE  186 (288)
Q Consensus       110 e~~le~~~~sl~i-~~~e~yq~~~qss~i~l~kLlk~f~~dfhn~lddln~~~qrk~Fid~~~~~L~h~~keNvl~~lFPE  188 (291)
                    ++.++.+++.|++ -+...-+-+|...    .++.++.|+...+..-.-...-||+..-....+-+-......++++|=|-
T Consensus       111 ~~~~~~~~l~~~gi~~~~~~~~~i~Qg~v---~~~i~~~~p~~~~~~~~~~~~~LS~G~k~rl~la~al~~~~~~~p~~lllDEPt  187 (247)
T cd03275         111 LKEYNEELEKINILVKARNFLVFQGDV---ESIASKNPPGKRFRDMDNLSGGEKTMAALALLFAIHSYQPAPFFVLDEVD  187 (247)
T ss_pred             HHHHHHHHHhCCCCCCCeEEEECCch---hhhhhccCcchhhhHHHcCHHHHHHHHHHHHHHHHhccCCCCCEEEEeccc


Q ss_pred             hhcCHHHHHHHHHHHHhh---ceeEEEEeCCHHHHHhcCccceeeeeecCCcc
Q gi|389815356|r  187 NELNKSELKELFQFATSQ---ACTMICLTTSPQVINQVGLSNVHLIKRTGER  235 (288)
Q Consensus       189 ~hLg~sd~K~f~~~~~~~~---~~t~Ii~T~hP~~I~~i~~~nv~LiK~nge~  237 (291)
                    +.|.......+.++++++   +.++|++|.+|.++  -..|.+..+.+++++
T Consensus       188 ~~LD~~~~~~~l~~~i~~~~~~~g~~vi~isH~~~~~~---~~~d~i~~~~~~~~~  237 (247)
T cd03275         188 AALDNTNVGKVASYIREQAGPNFQFIVISLKEEFF--SKADALVGVYRDQEC  237 (247)
T ss_pred             ccCCHHHHHHHHHHHHhccCCcEEEEEEECCHHHH--hhCCeEEEEEecCCC


No 4
>3qkt_A DNA double-strand break repair RAD50 ATPase; RECA-like fold, coiled-coils, ATP binding, DNA bindi MRE11, replication; HET: DNA ANP; 1.90A
{Pyrococcus furiosus} PDB: 3qku_A* 1ii8_A 3qks_B* 3qkr_B* 1ii8_B
Probab=95.57  E-value=0.34  Score=43.27  Aligned_cols=76  Identities=12%  Similarity=0.135  Sum_probs=0.0

Q ss_pred             HHHHHHHHHhhhhhhhccCCcccEEEEECChhhcCHHHHHHHHHHHHhh---ceeEEEEeCCHHHHHhcCccceeeeeecC
Q gi|389815356|r  156 GAYQYQSFLLKSWLHLVHNKATNVCFYDFPENELNKSELKELFQFATSQ---ACTMICLTTSPQVINQVGLSNVHLIKRT  232 (288)
Q Consensus       158 ~~~qrk~Fid~~~~~L~h~~keNvl~~lFPE~hLg~sd~K~f~~~~~~~~---~~t~Ii~T~hP~~I~~i~~~nv~LiK~n  234 (291)
                    |--+.-+....-++...-+....++++|=||+||-....+++..+++++   +...|+.|.+|.+++.+  ++|..+...
T Consensus       252 G~~~~~~la~~la~~~~~~~~~~~~llliDEpe~~Lhp~~~~~~l~~~L~~l~~~~~~QiiitTHs~~l~~~~~--d~i~~v~~e  329 (339)
T 3qkt_A         252 GERIALGLAFRLAMSLYLAGEISLLILDEPTPYLDEERRRKLITIMERYLKKIPQVILVSHDEELKDAA--DHVIRISLE  329 (339)
T ss_dssp             HHHHHHHHHHHHHHHHHHTTTTCEEEEECCCTTCCHHHHHHHHHHHHHTGGGSSEEEEEESCGGGGGGC--SEEEEEEE
T ss_pred             chHHHHHHHHHHHHHHHhCCCCCEEEEECCCCCCCCHHHHHHHHHHHHHHHhCCCEEEEEEccHHHHHhC--CEEEEEec


Q ss_pred             C
Q gi|389815356|r  233 G  233 (288)
Q Consensus       235 g  235 (291)
                    +
T Consensus       330 ~  330 (339)
T 3qkt_A         330 N  330 (339)
T ss_dssp             T
T ss_pred             C
```

```
No 5
>PRK11147 ABC transporter ATPase component; Reviewed
Probab=95.53  E-value=0.31  Score=51.69  Aligned_cols=93  Identities=16%  Similarity=0.246  Sum_probs=0.0

Q ss_pred             HHHHHhhcCc---chhhcccchhHHHHHHHHHhhhhhhhccCCcccEEEEECChhhcCHHHHHHHHHHHHhhceeEEEEe
Q gi|389815356|r  136 FEQLIKIAPI---EVRTVSDEPMGAYQYQSFLLKSWLHLVHNKATNVCFYDFPENELNKSELKELFQFATSQACTMICLT  212 (288)
Q Consensus       138 l~kLlk~f~~---dfhn~lddln~~~qrk~Fid~~~~~L~h~~keNvl~~lFPE~hLg~sd~K~f~~~~~~~~~~t~Ii~T  214 (291)
                      .+.+|+.|++   +.+....+|-+=+|++.-|-...+.     ..||+++|=|-+||...-...+.++++.++.|+|++|
T Consensus       420 ~~~~l~~~~l~~~~~~~~~~~~~LSgGekqRl~la~al~~-----~p~lLlLDEPt~~LD~~~~~~l~~~l~~~~~tvi~vS  494 (635)
T PRK11147        420 VLGYLQDFLFHPKRAMTPVKALSGGERNRLLLARLFLK-----PSNLLILDEPTNDLDVETLELLEELLDSYQGTVLLVS  494 (635)
T ss_pred             HHHHHHhcCCCHHHHhChhhhCCHHHHHHHHHHHHHhc-----CCCEEEEcCCCCCCCHHHHHHHHHHHHhCCCeEEEEE


Q ss_pred             CCHHHHHhcCcceeeeeecCCc
Q gi|389815356|r  213 TSPQVINQVGLSNVHLIKRTGE  234 (288)
Q Consensus       215 ~hP~~I~~i~~~nv~LiK~nge  236 (291)
                      .+++++.++ .+.+..++.+|+
T Consensus       495 Hd~~~~~~~-~d~i~~l~~~g~  515 (635)
T PRK11147        495 HDRQFVDNT-VTECWIFEGNGK  515 (635)
T ss_pred             CCHHHHHHh-cCEEEEEeCCCe


No 6
>3auy_A DNA double-strand break repair RAD50 ATPase; DNA repair, ABC transporter ATPase domain-like; HET: DNA ADP; 2.70A {Methanocaldococcus
jannaschii} PDB: 3aux_A* 3av0_B*
Probab=95.51  E-value=1.2  Score=38.93  Aligned_cols=153  Identities=13%  Similarity=0.067  Sum_probs=0.0

Q ss_pred             hhhhhHHHHHHHHHHHHHHHhcCchHHHHHHHHHHHHHHHHHhceEecCCeEeecccchhhHHHHHhhcCcchhhcccch
Q gi|389815356|r   75 EFFKDKHHAKQLIELMKFELENSPELLEVFTVLQGELEKAFEQCVVEFEGMELQPSTELISFEQLIKIAPIEVRTVSDEP  154 (288)
Q Consensus        77 ~~f~~K~~~e~ly~~fe~qi~n~ettn~~f~k~e~~le~~~~sl~i~~e~yq~~~qss~i~l~kLlk~f~~dfhn~lddl  156 (291)
                      ..-..........................-..++..+...+.........-++...-+       +.....+.....+.+
T Consensus       209 ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~-------~~~~~~~~~~~~~~~  281 (371)
T 3auy_A         209 KLTKFVEYLDKVRRIFGRNGFQAYLREKYVPLIQKYLNEAFSEFDLPYSFVELTKDFE-------VRVHAPNGVLTIDNL  281 (371)
T ss_dssp             HHHHHHHHHHHHHHHHSTTTHHHHHHHHHHHHHHHHHHHHHHHTTCSCSCEEECTTCC-------EEEEETTEEECGGGS
T ss_pred             HHHHHHHHHHHHHHHHHhhHHHHHHHHHHHHHHHHHHHHHHHhhhccccccccccccc-------eeeeccccccccccc


Q ss_pred             hHHHHHHHHHh--hhhhhhccCCcccEEEEECChhhcCHHHHHHHHHHHHh--hceeEEEEeCCHHHHHhcCcceeeeee
Q gi|389815356|r  155 MGAYQYQSFLL--KSWLHLVHNKATNVCFYDFPENELNKSELKELFQFATS--QACTMICLTTSPQVINQVGLSNVHLIK  230 (288)
Q Consensus       157 n~~~qrk~Fid--~~~~~L~h~~keNvl~~lFPE~hLg~sd~K~f~~~~~~--~~~t~Ii~T~hP~~I~~i~~~nv~LiK  232 (291)
                      .+-.|+...+.  .+-.......+..+++++=||+||...-.+++..+++.  .+...|+.|.||++.+.. ..-+.+-|
T Consensus       282 s~G~~~~~~la~~la~~~~~~~~~~~~~~~~lliDEpe~~L~~~~~~~~l~~~l~~l~~~~QiivtTHs~~~~~~~~-d~~i~i~k  360 (371)
T 3auy_A         282 SGGEQIAVALSLRLAIANALIGNRVECIILDEPTVYLDENRRAKLAEIFRKVKSIPQMIIITHHRELEDVA-DVIINVKK  360 (371)
T ss_dssp             CHHHHHHHHHHHHHHHHHHHHSSCCSEEEEESTTTTCCHHHHHHHHHHHHHHCCSCSEEEEEESCGGGGGGC-SEEEEEEE
T ss_pred             cchHHHHHHHHHHHHHHHHHhCCCCCeEEEECCcccCCHHHHHHHHHHHHHhccCCEEEEEEECcHHHHHhC-CcEEEEEE


Q ss_pred             cCCcc
Q gi|389815356|r  231 RTGER  235 (288)
Q Consensus       233 ~nge~  237 (291)
                      .+|.+
T Consensus       361 ~~~~s  365 (371)
T 3auy_A         361 DGNVS  365 (371)
T ss_dssp             SSSCE
T ss_pred             CCCEE


No 7
>4i99_A Chromosome partition protein SMC; winged-helix domain and SMC head domain, chromosome condensa SCPB, DNA binding protein; 2.30A {Pyrococcus
furiosus}
Probab=95.47  E-value=0.33  Score=44.85  Aligned_cols=82  Identities=13%  Similarity=0.182  Sum_probs=0.0

Q ss_pred             chhHHHHHHHHHhhhhhhhccCCcccEEEEECChhhcCHHHHHHHHHHHHh--hceeEEEEeCCHHHHHhcCcceeeeee
Q gi|389815356|r  153 EPMGAYQYQSFLLKSWLHLVHNKATNVCFYDFPENELNKSELKELFQFATS--QACTMICLTTSPQVINQVGLSNVHLIK  230 (288)
Q Consensus       155 dln~~~qrk~Fid~~~~~L~h~~keNvl~~lFPE~hLg~sd~K~f~~~~~~--~~~t~Ii~T~hP~~I~~i~~~nv~LiK  232 (291)
                      +..+.-++......-++.+....+..+++++=||+||...-++++..+++.  .++-.|+.|.+|++++..+ .-+.+.+
T Consensus       244 ~~~s~g~~~~~~l~l~~~~~~~~~~~~~i~liDEpe~~Lhp~~~~~l~~~L~~~~~~~QviitTHsp~~l~~~d-~l~~v~~  322 (354)
T 4i99_A         244 EAMSGGEKALTALAFVFAIQKFKPAPFYLFDEIDAHLDDANVKRVADLIKESSKESQFIVITLRDVMMANAD-KIIGVSM  322 (354)
T ss_dssp             GGSCHHHHHHHHHHHHHHHHHHSCCSEEEEESTTTTCCHHHHHHHHHHHHHHHTTSEEEEECCCHHHHTTCS-EEEEEEC
T ss_pred             cccchhHHHHHHHHHHHHHHhhcCcchhhhccHHhhcCHHHHHHHHHHHHhCcCCeEEEEECCHHHHHhcc-cEEEEEE


Q ss_pred             cCCcc
Q gi|389815356|r  231 RTGER  235 (288)
Q Consensus       233 ~nge~  237 (291)
                      .+|.+
T Consensus       323 ~~~~s  327 (354)
T 4i99_A         323 RDGVS  327 (354)
T ss_dssp             SSSCC
T ss_pred             ECCce


No 8
>SUPFAM0046693 c.37.1 P-loop containing nucleoside triphosphate hydrolases (52540) SCOP seed sequence: d1f2t.1.
Probab=95.46  E-value=0.58  Score=35.81  Aligned_cols=73  Identities=16%  Similarity=0.195  Sum_probs=0.0

Q ss_pred             HHHHHHhhhhhh--hccCCcccEEEEECChhhcCHHHHHHHHHHHHhh---ceeEEEEeCCHHHHHhcCcceeeeeecCC
Q gi|389815356|r  159 QYQSFLLKSWLH--LVHNKATNVCFYDFPENELNKSELKELFQFATSQ---ACTMICLTTSPQVINQVGLSNVHLIKRTG  233 (288)
Q Consensus       161 qrk~Fid~~~~~--L~h~~keNvl~~lFPE~hLg~sd~K~f~~~~~~~---~~t~Ii~T~hP~~I~~i~~~nv~LiK~ng  235 (291)
                      ||+......++.  ........++++|=|.++|.......+.++++++   +.|.|++|..|+++  --.+.+..+...+
T Consensus       207 ~~~~~~~a~~l~~~~~~~~~~~~~~l~lDEp~~~LD~~~~~~~~~~~~l~~~~~~~~~~~~~~~H~~~~~--~~~d~~~~~~~g~  284 (293)
T SUPFAM0046693   207 ERQRVALARALALALLLLSDPPLLLLDEPTAGLDPETREELLELLKELAKEGKTVIVVTHDLELL--ALADRIIVLDDGR  284 (293)
T ss_pred             HHHHHHHHHHHHHHHHHhccCCCeEEEeCCccccCCHHHHHHHHHHHHHhcCCEEEEEEccHHHH--hhcCEEEEEeCCe
```

```
No 9
>cd03293 ABC_NrtD_SsuB_transporters NrtD and SsuB are the ATP-binding subunits of the bacterial ABC-type nitrate and sulfonate transport systems,
respectively.  ABC transporters are a large family of proteins involved in the transport of a wide variety of different compounds, like sugars,
ions, peptides, and more complex organic molecules.  The nucleotide binding domain shows the highest similarity between all members of the family.
ABC transporters are a subset of nucleotide hydrolases that contain a signature motif, Q-loop, and H-loop/switch region, in addition to, the Walker
A motif/P-loop and Walker B motif commonly found in a number of ATP- and GTP-binding and hydrolyzing proteins.
Probab=95.32  E-value=0.75  Score=41.53  Aligned_cols=93  Identities=14%  Similarity=0.162  Sum_probs=0.0

Q ss_pred             HHHHHhhcCcc--hhhccccchhHHHHHHHHHhhhhhhhccCCcccEEEEECChhhcCHHHHHHHHHHHHhh----ceeEE
Q gi|389815356|r  136 FEQLIKIAPIE--VRTVSDEPMGAYQYQSFLLKSWLHLVHNKATNVCFYDFPENELNKSELKELFQFATSQ----ACTMI  209 (288)
Q Consensus       138 l~kLlk~f~~d--fhn~lddln~~~qrk~Fid~~~~~~L~h~~keNvl~~lFPE~hLg~sd~K~f~~~~~~----~~t~I  211 (291)
                      ..++++.++++  .+....++-+-++++..|-...++     ...++++|=|-+.|...-.+++.++++..    +.|+|
T Consensus       112 ~~~~l~~~~l~~~~~~~~~~~~LSgG~~qrl~la~al~~-----~p~lllLDEPt~~LD~~~~~~~~~~l~~~~~~~~~~tii  186 (220)
T cd03293         112 AEELLELVGLSGFENAYPHQLSGGMRQRVALARALAV-----DPDVLLLDEPFSALDALTREQLQEELLDIWRETGKTVL  186 (220)
T ss_pred             HHHHHHHcCChhhhhCCccccCCHHHHHHHHHHHHHHc-----CCCEEEECCCCCCCHHHHHHHHHHHHHHHHHHcCCEEE


Q ss_pred             EEeCCHHHHHhcCcceeeeeeec-CCc
Q gi|389815356|r  210 CLTTSPQVINQVGLSNVHLIKR-TGE  234 (288)
Q Consensus       212 i~T~hP~~I~~i~~~nv~LiK~-nge  236 (291)
                      ++|.+|..+..+- +.+.+++. +|+
T Consensus       187 ~sH~~~~~~~~~~~-d~i~~l~~~~G~  211 (220)
T cd03293         187 VTHDIDEAVFLA-DRVVVLSARPGR  211 (220)
T ss_pred             EEecCHHHHHHhC-CEEEEEECCCCE


No 10
>d1e69a_ c.37.1.12 (A:) Smc head domain {Thermotoga maritima [TaxId: 2336]}
Probab=95.25  E-value=0.36  Score=43.76  Aligned_cols=76  Identities=13%  Similarity=0.133  Sum_probs=0.0

Q ss_pred             HHHHHHhhhhhhhccCCccccEEEEECChhhcCHHHHHHHHHHHHhhc--eeEEEEeCCHHHHHhcCcceeeeeeecCCcc
Q gi|389815356|r  159 QYQSFLLKSWLHLVHNKATNVCFYDFPENELNKSELKELFQFATSQA--CTMICLTTSPQVINQVGLSNVHLIKRTGER  235 (288)
Q Consensus       161 qrk~Fid~~~~~~L~h~~keNvl~~lFPE~hLg~sd~K~f~~~~~~~~~--~t~Ii~T~hP~~I~~i~~~nv~LiK~nge~  237 (291)
                      |||.+....-+-........|+++|=|+++|.....+.+..++++..  +++|+.|.+|+++...... ..+++.+|.+
T Consensus       224 ~~~~i~la~~la~~~~~~~~~vlllDEp~~~Ld~~~~~~~l~~~l~~~~~~~~qviittH~~~~~~~~~~~~~~-~~~~~~~~~~~~~~  301 (308)
T d1e69a_         224 EKALVGLALLFALMEIKPSPFYVLDEVDSPLDDYNAERFKRLLKENSKHTQFIVITHNKIVMEAADLL-HGVTMVNGVS  301 (308)
T ss_dssp             HHHHHHHHHHHHHHTTTSCCSEEEEESCCSSCCHHHHHHHHHHHHHHHHTTTSEEEEECCCTTGGGGCSEE-EEEEESSSCE
T ss_pred             hHHHHHHHHHHHHHhcccCCCEEEEeCCccccCHHHHHHHHHHHHHHhCcCCeEEEEECCHHHHhhchhE-EEEEEeCCEE


No 11
>TIGR02211 LolD_lipo_ex lipoprotein releasing system, ATP-binding protein. This model represents LolD, a member of the ABC transporter family
(pfam00005). LolD is involved in localization of lipoproteins in some bacteria. It works with a transmembrane protein LolC, which in some species
is a paralogous pair LolC and LolE. Depending on whether the residue immediately following the new, modified N-terminal Cys residue, the nascent
lipoprotein may be carried further by LolA and LolB to the outer membrane, or remain at the inner membrane. The top scoring proteins excluded by
this model include homologs from the archaeal genus Methanosarcina.
Probab=95.22  E-value=0.82  Score=41.08  Aligned_cols=94  Identities=13%  Similarity=0.282  Sum_probs=0.0

Q ss_pred             HHHHHhhcCcc--hhhccccchhHHHHHHHHHhhhhhhhccCCcccEEEEECChhhcCHHHHHHHHHHHHhh----ceeEE
Q gi|389815356|r  136 FEQLIKIAPIE--VRTVSDEPMGAYQYQSFLLKSWLHLVHNKATNVCFYDFPENELNKSELKELFQFATSQ----ACTMI  209 (288)
Q Consensus       138 l~kLlk~f~~d--fhn~lddln~~~qrk~Fid~~~~~~L~h~~keNvl~~lFPE~hLg~sd~K~f~~~~~~----~~t~I  211 (291)
                      ..++++.++++  .+.....+-+-++++.-|-..+++     ...|+++|=|-+.|...-..++.+.++..    +.|+|
T Consensus       122 ~~~~l~~~~l~~~~~~~~~~~~LS~G~~qrv~laral~~-----~p~illlDEPt~~LD~~~~~~~l~~~l~~~~~~~~~~tii  196 (221)
T TIGR02211       122 AYEMLEKVGLEHRINHRPSELSGGERQRVAIARALVN-----QPSLVLADEPTGNLDNNNAKIIFDLMLELNRELNTSFL  196 (221)
T ss_pred             HHHHHHHcCChhhhhCChhhCCHHHHHHHHHHHHHHhC-----CCCEEEEeCCCCcCCHHHHHHHHHHHHHHHHHhcCCEEE


Q ss_pred             EEeCCHHHHHhcCcceeeeeeecCCcccc
Q gi|389815356|r  210 CLTTSPQVINQVGLSNVHLIKRTGERYS  237 (288)
Q Consensus       212 i~T~hP~~I~~i~~~nv~LiK~nge~~~  239 (291)
                      ++|.++..+..+  +.+.+++ +|+-++
T Consensus       197 ~~tH~~~~~~~~~~--d~v~~l~-~G~i~~  221 (221)
T TIGR02211       197 VVTHDLELAKKL--DRVLEMK-DGQLFN  221 (221)
T ss_pred             EEeCCHHHHhhc--CEEEEEe-CCEecC


No 12
>SUPFAM0050337 c.37.1 P-loop containing nucleoside triphosphate hydrolases (52540) SCOP seed sequence: d1w1wa_.
Probab=95.16  E-value=0.74  Score=38.14  Aligned_cols=74  Identities=18%  Similarity=0.277  Sum_probs=0.0

Q ss_pred             HHHHHHhhhhhhhccCCccccEEEEECChhhcCHHHHHHHHHHHHhh---ceeEEEEeCCHHHHHhcCcceeeeeeecCC
Q gi|389815356|r  159 QYQSFLLKSWLHLVHNKATNVCFYDFPENELNKSELKELFQFATSQ---ACTMICLTTSPQVINQVGLSNVHLIKRTG  233 (288)
Q Consensus       161 qrk~Fid~~~~~~L~h~~keNvl~~lFPE~hLg~sd~K~f~~~~~~---~~t~Ii~T~hP~~I~~i~~~nv~LiK~ng  235 (291)
                      |++.......+.+..-...+++++|=|.++|...-...+.+++.++   +.+.|++|..|.++..+ .+.+.+.+..+
T Consensus       337 e~~~~~la~~~~~~~~~~~~~~~~~lDE~~~~Ld~~~~~~~~~~l~~~~~~~~~~~~~~~~~~h~~~~~~~~~-~~~~~~~~~~~g~  413 (427)
T SUPFAM0050337   337 EKQRLALALALALAELLPPPLLLLDEPTAGLDPENRERLLELLRELASEGGQVIVVTHDPELAARL-ADRILVLKDGG  413 (427)
T ss_pred             HHHHHHHHHHHHHhccCCCeEEEecCcccCCHHHHHHHHHHHHHHHhcCCEEEEEEccHHHHhh-ccEEEEEeCCc


No 13
>3zgx_A SMC, chromosome partition protein SMC; cell cycle; 3.40A {Bacillus subtilis}
Probab=95.04  E-value=0.42  Score=42.72  Aligned_cols=77  Identities=10%  Similarity=0.156  Sum_probs=0.0

Q ss_pred             HHHHHHhhhhhhhccCCccccEEEEECChhhcCHHHHHHHHHHHHhh--ceeEEEEeCCHHHHHhcCcceeeeeeecCCcccc
Q gi|389815356|r  159 QYQSFLLKSWLHLVHNKATNVCFYDFPENELNKSELKELFQFATSQ--ACTMICLTTSPQVINQVGLSNVHLIKRTGERY  236 (288)
Q Consensus       161 qrk~Fid~~~~~~L~h~~keNvl~~lFPE~hLg~sd~K~f~~~~~~--~~t~Ii~T~hP~~I~~i~~~nv~LiK~nge~~  238 (291)
                      +++.....-++-+......-++++|=||+||...-++++.++++.+  ++-.||.|.+|.++...  +.+..++.+.+-+
T Consensus       333 ek~~~~lal~~a~~~~~~~~~~~~~~ilDEpe~~Lhp~~~~~~l~~~l~~~s~~~QiiitTHsp~~~~~~~~--d~~~~v~~~~~~g~  410 (426)
T 3zgx_A          333 ERALTAIALLFSILKVRPVPFCVLDEVEAALDEANVFRFAQYLKKYSSDTQFIVITHRKGTMEEA--DVLYGVTMQESGV  410 (426)
T ss_dssp             HHHHHHHHHHHHHHHHSCCSCEEEESTTTTSCTHHHHHHHHHHHHCSSSSCEEEECSCHHHHHHC--SCEEEEEC----C
T ss_pred             hHHHHHHHHHHHHHhcCCCCEEEEECCCCCCCHHHHHHHHHHHHHHhCCCCEEEEEeCcHHHHHhc--CcEEEEEEcCCCc


Q ss_pred             c
Q gi|389815356|r  237 S  237 (288)
Q Consensus       239 ~  239 (291)
                      +
T Consensus       411 s  411 (426)
T 3zgx_A          411 S  411 (426)
T ss_dssp             B
T ss_pred             E


No 14
>1w1w_A Structural maintenance of chromosome 1; cohesin, chromosome segregation, cell adhesion, kleisin, MIT cell cycle; HET: AGS; 2.90A
{Saccharomyces cerevisiae} SCOP: c.37.1.12
Probab=94.95  E-value=0.79  Score=41.46  Aligned_cols=77  Identities=8%  Similarity=0.119  Sum_probs=0.0

Q ss_pred             HHHHHHhhhhhhhccCCccccEEEEECChhhcCHHHHHHHHHHHHhh---ceeEEEEeCCHHHHHhcCcceee-eeecCCc
Q gi|389815356|r  159 QYQSFLLKSWLHLVHNKATNVCFYDFPENELNKSELKELFQFATSQ---ACTMICLTTSPQVINQVGLSNVH-LIKRTGE  234 (288)
Q Consensus       161 qrk~Fid~~~~~~L~h~~keNvl~~lFPE~hLg~sd~K~f~~~~~~---~~t~Ii~T~hP~~I~~i~~~nv~-LiK~nge  236 (291)
                      +|......-.+.+.......++++|=|++||...-..++.+++...   ++-.||.|.+|.++  -..|++. +.|..+.
T Consensus       338 e~~~~~lal~la~~~~~~~~i~llDEpd~~Ld~~~~~~~l~~~l~~~~~~~~~~Q~IvtTH~~~~~~--~~ad~~~~v~~~~~~~
T 1w1w_A          338 EKTVAALALLFAINSYQPSPFFVLDEVDAALDITNVQRIAAYIRRHRNPDLQFIVISLKNTMF--EKSDALVGVYRQQQE  415 (430)
T ss_dssp             HHHHHHHHHHHHHHTSSCCSEEEESSTTTTCCHHHHHHHHHHHHHHHCBTTBEEEEECSCHHHH--TTCSEEEEEEEETTT
T ss_pred             HHHHHHHHHHHHHHhcCCCCEEEEECCCCCCHHHHHHHHHHHHHHHhCCCCEEEEEcCCHHHH--HhcCcEEEEEEEcCC
```

```
Q ss_pred             ccc
Q gi|389815356|r  235 RYS  237 (288)
Q Consensus       237 ~~~  239 (291)
                      .++
T Consensus       416 ~~s  418 (430)
T 1w1w_A          416 NSS  418 (430)
T ss_dssp             TEE
T ss_pred             Cce


No 15
>PRK10247 putative ABC transporter ATP-binding protein YbbL; Provisional
Probab=94.94  E-value=1  Score=41.20  Aligned_cols=94  Identities=13%  Similarity=0.127  Sum_probs=0.0

Q ss_pred             HHHHHhhcCcc---hhhccccchhHHHHHHHHhhhhhhhccCCcccEEEEECChhhcCHHHHHHHHHHHHhh----ceeE
Q gi|389815356|r  136 FEQLIKIAPIE---VRTVSDEPMGAYQYQSFLLKSWLHLVHNKATNVCFYDFPENELNKSELKELFQFATSQ----ACTM  208 (288)
Q Consensus       138 l~kLlk~f~~d---fhn~lddln~~~qrk~Fid~~~~~L~h~~keNvl~~lFPE~hLg~sd~K~f~~~~~~~~----~~t~  210 (291)
                      .+++++.++++   .+.....+-+-++|+.=|-..+++     ...|+++|=|-+.|...-...+.+++.+.    +.|+
T Consensus       117 ~~~~l~~~~l~~~~~~~~~~~~~LS~G~~qrv~laral~~-----~p~llllDEPt~~LD~~~~~~~l~~~l~~~~~~~~~~~tv  191 (225)
T PRK10247        117 FLDDLERFALPDTILTKNIAELSGGEKQRISLIRNLQF-----MPKVLLLDEITSALDESNKHNVNEIIHRYVREQNIAV  191 (225)
T ss_pred             HHHHHHHcCCChHHhcCCcccCCHHHHHHHHHHHHHHhc-----CCCEEEEeCCcccCCHHHHHHHHHHHHHHHHHHhcCCEE


Q ss_pred             EEEeCCHHHHHhcCcceeeeeecCCccc
Q gi|389815356|r  209 ICLTTSPQVINQVGLSNVHLIKRTGERY  236 (288)
Q Consensus       211 Ii~T~hP~~I~~i~~~nv~LiK~nge~~  238 (291)
                      ||+|.++..+. . .+.+..+..++..+
T Consensus       192 ii~sh~~~~~~~-~-~d~i~~l~~~~~~~~  217 (225)
T PRK10247        192 LWVTHDKDEIN-H-ADKVITLQPHAGEM  217 (225)
T ss_pred             EEEECChHHHH-h-CCEEEEEecccchH
```

HHPRED hits for Csn2 short

```
 No Hit                             Prob E-value P-value  Score    SS Cols Query HMM  Template HMM
  1 2yl5_A Beta-N-acetylhexosamini  68.5      45  0.0003   36.0   6.5   70  141-210     3-112 (442)
  2 d1t5ia_ c.37.1.19 (A:) Spliceo  66.9      64 0.00043   29.8   6.3   45  163-212    13-57  (168)
  3 d1s2ma2 c.37.1.19 (A:252-422)   64.2      67 0.00045   29.6   5.9   45  163-212    18-62  (171)
  4 d2gjxa1 c.1.8.6 (A:167-528) be  51.3      81 0.00054   33.3   4.7   61  150-210     5-88  (362)
  5 d2rb4a1 c.37.1.19 (A:307-474)   51.2 1.3E+02 0.00085   28.1   5.4   48  160-212    15-62  (168)
  6 cd06564 GH20_DspB_LnbB-like Gl  50.5 1.2E+02 0.00079   31.6   5.6   61  150-210     4-99  (326)
  7 d1oywa3 c.37.1.19 (A:207-406)   49.2 2.2E+02  0.0015   26.8   6.7   44  164-212    17-60  (200)
  8 d1fuka_ c.37.1.19 (A:) Initiat  48.9 2.7E+02  0.0018   25.9   7.1   44  164-212    14-57  (162)
  9 2hjv_A ATP-dependent RNA helic  48.3 2.4E+02  0.0016   25.6   6.6   45  163-212    21-65  (163)
 10 PF14258 DUF4350:  Domain of un  47.5 2.1E+02  0.0014   22.7   5.4   59  142-212    10-69  (70)
 11 cd02742 GH20_hexosaminidase Be  47.1 1.9E+02  0.0013   29.9   6.4   61  150-210     3-89  (303)
 12 SUPFAM0048920 c.1.8 (Trans)gly  46.0 2.4E+02  0.0016   26.4   6.4   61  150-210     5-107 (443)
 13 cd06565 GH20_GcnA-like Glycosy  45.7 2.2E+02  0.0015   29.6   6.6   61  150-210     3-77  (301)
 14 1t5i_A C_terminal domain of A   45.2 2.5E+02  0.0017   25.9   6.3   45  163-212    17-61  (172)
 15 PF09664 DUF2399:  Protein of u  44.3 1.1E+02 0.00077   28.6   4.1   31   22-64     20-50  (152)


No 1
>2yl5_A Beta-N-acetylhexosaminidase; hydrolase; 2.15A {Streptococcus pneumoniae} PDB: 2yla_A* 2yl9_A*
Probab=68.49  E-value=45  Score=36.01  Aligned_cols=70  Identities=16%  Similarity=0.270  Sum_probs=0.0

Q ss_pred              cHHHHHHhcCcEecccCCCcHHHHHHHHHHHHHHhcCCCEEEEEE------------------------------------
Q gi|339625078|r  126 ELERLLHSQNIVVEKGNNSTIFGKIEEVIHVMGMLDESRYLFLT------------------------------------  169 (224)
Q Consensus       141 d~~~llK~~~ikie~~~~~~~~eki~~~iki~~~l~~kkllvfv------------------------------------  184 (251)
                      +.++|+|.-|+.+|+...=-+.+-|..+|+.+....-..+=.=.
T Consensus         3 ~~~~l~~~RG~mLDvaR~f~~~~~ik~~ID~mA~~KlN~lH~HltDdgwriei~~~pl~~~g~~~~~~~~~~~~~~~g~~~   82 (442)
T 2yl5_A           3 HMEKLAKNKVISIDAGRKYFTLNQLKRIVDKASELGYSDVHLLLGNDGLRFLLDDMTITANGKTYASDDVKKAIIEGTKA   82 (442)
T ss_dssp              CHHHHHSEEEEEEETTTSCCCHHHHHHHHHHHHHHTCCEEEEEEEESSSBCBCCSCCCEEETTEEECHHHHHHHHHHHHH
T ss_pred              ccchhhhccceEEecCCCCCCHHHHHHHHHHHHHHhCCcEEEEEEcCCCcceeccCCcceeceeeccccccccccCCccc


Q ss_pred              ----ChHHhCCHHHHHHHHHHHHHHhCCeEE
Q gi|339625078|r  170 ----NASLYCKLSDLNQLHECLLAEGVNLI  195 (224)
Q Consensus       185 ----Nl~~YLt~~el~~L~~~i~~~~l~vL  210 (251)
                          .-+.|.|.+|+.++.+|++..++.|+
T Consensus        83 ~~~~~~~~~YT~~di~eIv~YA~~rgI~VI  112 (442)
T 2yl5_A          83 YYDDPNGTALTQAEVTELIEYAKSKDIGLI  112 (442)
T ss_dssp              HCCCTTCSCBCHHHHHHHHHHHHTTTCEEE
T ss_pred              cccCCCCCCcCHHHHHHHHHHHHHcCCeee


No 2
>d1t5ia_ c.37.1.19 (A:) Spliceosome RNA helicase BAT1 (UAP56) {Human (Homo sapiens) [TaxId: 9606]}
Probab=66.89  E-value=64  Score=29.83  Aligned_cols=45  Identities=13%  Similarity=0.269  Sum_probs=0.0

Q ss_pred              HHHHHHHHHHHHhcCCCEEEEEEChHHhCCHHHHHHHHHHHHHHhCCeEEEE
Q gi|339625078|r  148 GKIEEVIHVMGMLDESRYLFLTNASLYCKLSDLNQLHECLLAEGVNLISI  197 (224)
Q Consensus       163 eki~~~iki~~~l~~kkllvfvNl~~YLt~~el~~L~~~i~~~~l~vL~i  212 (251)
                      +|+-.++++.......|++||+|     +.+....|.+.+...+.++..+
T Consensus        13 ~K~~~L~~ll~~~~~~k~iIF~~-----~~~~~~~~~l~~~L~~~~~~~~~i   57 (168)
T d1t5ia_          13 EKNRKLFDLLDVLEFNQVVIFVK-----SVQRCIALAQLLVEQNFPAIAI   57 (168)
T ss_dssp              GHHHHHHHHHHHSCCSSEEEECS-----SHHHHHHHHHHHHHTTCCEEEE
T ss_pred              HHHHHHHHHHhCCCCeEEEEEe-----eeecchhhhhhccccccccc


No 3
>d1s2ma2 c.37.1.19 (A:252-422) Putative ATP-dependent RNA helicase DHH1 {Baker's yeast (Saccharomyces cerevisiae) [TaxId: 4932]}
Probab=64.23  E-value=67  Score=29.63  Aligned_cols=45  Identities=11%  Similarity=0.121  Sum_probs=0.0

Q ss_pred              HHHHHHHHHHHHhcCCCEEEEEEChHHhCCHHHHHHHHHHHHHHhCCeEEEE
Q gi|339625078|r  148 GKIEEVIHVMGMLDESRYLFLTNASLYCKLSDLNQLHECLLAEGVNLISI  197 (224)
Q Consensus       163 eki~~~iki~~~l~~kkllvfvNl~~YLt~~el~~L~~~i~~~~l~vL~i  212 (251)
                      +|+....++.......|++||+|     |......+.+++...+.++.++
T Consensus        18 ~K~~~L~~ll~~~~~~k~iVF~~-----~~~~~~~~~l~~~L~~~g~~~~~~   62 (171)
T d1s2ma2          18 QKLHCLNTLFSKLQINQAIIFCN-----STNRVELLAKKITDLGYSCYYS   62 (171)
T ss_dssp              GHHHHHHHHHHHSCCSEEEEECS-----SHHHHHHHHHHHHHHHTCCEEEE
T ss_pred              HHHHHHHHHHhCCCCCceEEEEe---------eeehhhHhHHhhcccccccc


No 4
>d2gjxa1 c.1.8.6 (A:167-528) beta-hexosaminidase A {Human (Homo sapiens) [TaxId: 9606]}
Probab=51.25  E-value=81  Score=33.35  Aligned_cols=61  Identities=13%  Similarity=0.296  Sum_probs=0.0

Q ss_pred              CcEecccCCCcHHHHHHHHHHHHHHhcCCCEEE-----------------EEChHHh------CCHHHHHHHHHHHHHHhC
Q gi|339625078|r  135 NIVVEKGNNSTIFGKIEEVIHVMGMLDESRYLF-----------------LTNASLY------CKLSDLNQLHECLLAEG  191 (224)
Q Consensus       150 ~ikie~~~~~~~~eki~~~iki~~~l~~kkllv-----------------fvNl~~Y------Lt~~el~~L~~~i~~~  206 (251)
                      |+.+|+...-=+.+-|..+|+.++.+.-..+-.                 ++..+.|      +|.+|+.+|.+|++..+
T Consensus         5 G~mlD~aR~f~~~~~~ik~~Id~ma~~KlN~lhlHl~D~~~~~rle~~~~~P~lt~~ga~~~~~~~~~yT~~di~~lv~yA~~~rg   84 (362)
T d2gjxa1           5 GLLLDTSRHYLPLSSILDTLDVMAYNKLNVFHWHLVDDPSFPYESFTFPELMRKGSYNPVTHIYTAQDVKEVIEYARLRG   84 (362)
T ss_dssp              EEEEETTTSCCCHHHHHHHHHHHHHHTTCCEEEEECCCSSCCCBCCSSCTHHHHHSSCTTTSCBCHHHHHHHHHHHHHHTT
T ss_pred              ceEEecCCCCcCHHHHHHHHHHHHHcCCcEEEEEeecccCceeeccCCcccccCCcCCCCccCHHHHHHHHHHHHcC


Q ss_pred              CeEE
Q gi|339625078|r  192 VNLI  195 (224)
Q Consensus       207 l~vL  210 (251)
                      +.|+
T Consensus        85 I~VI   88 (362)
T d2gjxa1          85 IRVL   88 (362)
T ss_dssp              CEEE
T ss_pred              Ceee


No 5
>d2rb4a1 c.37.1.19 (A:307-474) ATP-dependent RNA helicase DDX25 {Human (Homo sapiens) [TaxId: 9606]}
Probab=51.19  E-value=1.3e+02  Score=28.07  Aligned_cols=48  Identities=8%  Similarity=0.167  Sum_probs=0.0

Q ss_pred              cHHHHHHHHHHHHHHhcCCCEEEEEEChHHhCCHHHHHHHHHHHHHHhCCeEEEE
Q gi|339625078|r  145 TIFGKIEEVIHVMGMLDESRYLFLTNASLYCKLSDLNQLHECLLAEGVNLISI  197 (224)
Q Consensus       160 ~~~eki~~~iki~~~l~~kkllvfvNl~~YLt~~el~~L~~~i~~~~l~vL~i  212 (251)
                      +.-+|+..+++++......+++||+|     +.+....|.+.+...+.++.++
T Consensus        15 ~~~~K~~~L~~ll~~~~~~~~~~~~lIF~-----~~~~~~~~~l~~~l~~~~~~~~~~~~   62 (168)
T d2rb4a1          15 HRKDKYQALCNIYGSITIGQAIIFCQ-----TRRNAKWLTVEMIQDGHQVSLL   62 (168)
T ss_dssp              SHHHHHHHHHHHHTTSCCSEEEEECS-----CHHHHHHHHHHHHHTTTCCEEEE
T ss_pred              CHHHHHHHHHHHHhCCCCcEEEEcC-----HHHHHHHHHHHHHhcCCcceec
```

```
No 6
>cd06564 GH20_DspB_LnbB-like Glycosyl hydrolase family 20 (GH20) catalytic domain of dispersin B (DspB), lacto-N-biosidase (LnbB) and related
proteins. Dispersin B is a soluble beta-N-acetylglucosamidase found in bacteria that hydrolyzes the beta-1,6-linkages of PGA (poly-beta-(1,6)-N-
acetylglucosamine), a major component of the extracellular polysaccharide matrix. Lacto-N-biosidase hydrolyzes lacto-N-biose (LNB) type I
oligosaccharides at the nonreducing terminus to produce lacto-N-biose as part of the GNB/LNB (galacto-N-biose/lacto-N-biose I) degradation pathway.
The lacto-N-biosidase from Bifidobacterium bifidum has this GH20 domain, a carbohydrate binding module 32, and a bacterial immunoglobulin-like
domain 2, as well as a YSIRK signal peptide and a G5 membrane anchor at the N and C termini, respectively. The GH20 hexosaminidases are thought to
act via a catalytic mechanism in which the catalytic nucleophile is not provided by solvent or the enzyme, but by the substrate itself.
Probab=50.45  E-value=1.2e+02  Score=31.57  Aligned_cols=61  Identities=18%  Similarity=0.315  Sum_probs=0.0

Q ss_pred             CcEeccccCCCcHHHHHHHHHHHHHhcCCCEEE-----------------------------------EEChHHhCCHHH
Q gi|339625078|r  135 NIVVEKGNNSTIFGKIEEVIHVMGMLDESRYLF-----------------------------------LTNASLYCKLSD  179 (224)
Q Consensus       150 ~ikie~~~~~~~~~eki~~~iki~~~l~~kkllv-----------------------------------fvNl~~YLt~~e  194 (251)
                     |+.+|+...-=+.+-|..+|+.++.+.-..+-.                                   -..-..|+|.+|
T Consensus         4 G~mLD~aR~~~~~~~ik~~id~ma~~K~N~lhlHltD~~~~~~~~~p~l~~~~~~~~~~~~~~~~~~~~~~~~~~YT~~d   83 (326)
T cd06564           4 GFMLDVGRKYYSMDFLKDIIKTMSWYKMNDLQLHLNDNLIFNLDDMSTTVNNATYASDDVKSGNNYYNLTANDGYYTKEE   83 (326)
T ss_pred             eeEEEccCCCCCHHHHHHHHHHHHHcCCceEEEeecCCcccccCCCchhhhhhhhcccccccccccCCCCCCCCcccHHH


Q ss_pred             HHHHHHHHHHhCCeEE
Q gi|339625078|r  180 LNQLHECLLAEGVNLI  195 (224)
Q Consensus       195 l~~L~~~i~~~~l~vL  210 (251)
                     +.+|.+|++..++.|+
T Consensus        84 i~eiv~yA~~rgI~vI   99 (326)
T cd06564          84 FKELIAYAKDRGVNII   99 (326)
T ss_pred             HHHHHHHHHcCCeEe


No 7
>d1oywa3 c.37.1.19 (A:207-406) RecQ helicase domain {Escherichia coli [TaxId: 562]}
Probab=49.19  E-value=2.2e+02  Score=26.83  Aligned_cols=44  Identities=7%  Similarity=0.167  Sum_probs=0.0

Q ss_pred             HHHHHHHHHHHhcCCCEEEEEChHHhCCHHHHHHHHHHHHHHhCCeEEEE
Q gi|339625078|r  149 KIEEVIHVMGMLDESRYLFLTNASLYCKLSDLNQLHECLLAEGVNLISI  197 (224)
Q Consensus       164 ki~~~iki~~~l~~kkllvfvNl~~YLt~~el~~L~~~i~~~~~l~vL~i  212 (251)
                     ++..+++++.....++++||+|     |..+...+.+++...+..+..+
T Consensus        17 k~~~l~~~l~~~~~~~~~~lIF~~-----s~~~~~~l~~~L~~~~~~~~~~~~~~   60 (200)
T d1oywa3          17 PLDQLMRYVQEQRGKSGIIYCN-----SRAKVEDTAARLQSKGISAAAY   60 (200)
T ss_dssp             HHHHHHHHHHHTTTCCEEEECS-----SHHHHHHHHHHHHHHTTCCEEEE
T ss_pred             HHHHHHHHHHhcCCCcEEEEEC-----cHHHHHHHHHHHHHHCCCCEEEE


No 8
>d1fuka_ c.37.1.19 (A:) Initiation factor 4a {Baker's yeast (Saccharomyces cerevisiae) [TaxId: 4932]}
Probab=48.91  E-value=2.7e+02  Score=25.86  Aligned_cols=44  Identities=14%  Similarity=0.210  Sum_probs=0.0

Q ss_pred             HHHHHHHHHHHhcCCCEEEEEChHHhCCHHHHHHHHHHHHHHhCCeEEEE
Q gi|339625078|r  149 KIEEVIHVMGMLDESRYLFLTNASLYCKLSDLNQLHECLLAEGVNLISI  197 (224)
Q Consensus       164 ki~~~iki~~~l~~kkllvfvNl~~YLt~~el~~L~~~i~~~~~l~vL~i  212 (251)
                     |+....+++.....+|++||+|     |.....++.+.....+.++..+
T Consensus        14 K~~~L~~ll~~~~~~~k~iIF~~-----s~~~~~~l~~~L~~~~~~~~~~~~~~   57 (162)
T d1fuka_          14 KYECLTDLYDSISVTQAVIFCN-----TRRKVEELTTKLRNDKFTVSAI   57 (162)
T ss_dssp             HHHHHHHHHHHTTCSCEEEEES-----SHHHHHHHHHHHHHHTTCCEEEE
T ss_pred             HHHHHHHHHHhCCCCcEEEEEE-----EEchHHHHHHHHhhcCceEEEe


No 9
>2hjv_A ATP-dependent RNA helicase DBPA; parallel alpha-beta, hydrolase; 1.95A {Bacillus subtilis}
Probab=48.29  E-value=2.4e+02  Score=25.60  Aligned_cols=45  Identities=18%  Similarity=0.144  Sum_probs=0.0

Q ss_pred             HHHHHHHHHHHHhcCCCEEEEEChHHhCCHHHHHHHHHHHHHHhCCeEEEE
Q gi|339625078|r  148 GKIEEVIHVMGMLDESRYLFLTNASLYCKLSDLNQLHECLLAEGVNLISI  197 (224)
Q Consensus       163 eki~~~iki~~~l~~kkllvfvNl~~YLt~~el~~L~~~i~~~~~l~vL~i  212 (251)
                     +|+..+.++.......+++||+|     |.++...+.+.+...++++..+
T Consensus        21 ~K~~~L~~ll~~~~~~~~~~~~~lVF~~-----~~~~~~~~~l~~~L~~~~~~~~~~~~   65 (163)
T 2hjv_A          21 NKFSLLKDVLMTENPDSCIIFCR-----TKEHVNQLTDELDDLGYPCDKI   65 (163)
T ss_dssp             GHHHHHHHHHHHHCCSSEEEECS-----SHHHHHHHHHHHHHHTTCCEEEE
T ss_pred             HHHHHHHHHHHhcCCCcEEEEEC-----CHHHHHHHHHHHHHHcCCcEEEE


No 10
>PF14258 DUF4350:  Domain of unknown function (DUF4350)
Probab=47.48  E-value=2.1e+02  Score=22.67  Aligned_cols=59  Identities=22%  Similarity=0.297  Sum_probs=0.0

Q ss_pred             HHHHHHhcCcEeccccCCCcHHHHHHHHHHHHHhcCCCEEEEEChHHhCCH-HHHHHHHHHHHHhCCeEEEE
Q gi|339625078|r  127 LERLLHSQNIVVEKGNNSTIFGKIEEVIHVMGMLDESRYLFLTNASLYCKL-SDLNQLHECLLAEGVNLISI  197 (224)
Q Consensus       142 ~~~llK~~~ikie~~~~~~~~~eki~~~iki~~~l~~kkllvfvNl~~YLt~-~el~~L~~~i~~~~~l~vL~i  212 (251)
                     +.++++..|+++..-..  +++.+..         ....+++++-...++. +|+++|.+++. .+=+|++.
T Consensus        10 ~~~~L~~~g~~v~~~~~--~~~~l~~---------~~~tll~i~~~~~~~~~~~~~~~~~l~~~v~-~G~~lvl~   69 (70)
T PF14258_consen   10 LYQLLEEQGVKVERWRK--PYEALEA---------DDGTLLVIGPDLRLSEPEEAEALLEWVE-AGNTLVLA   69 (70)
T ss_pred             HHHHHHHCCCeeEEecc--cHHHhCC---------CCCEEEEEeCCCCCCchHHHHHHHHHHH-cCCEEEEe


No 11
>cd02742 GH20_hexosaminidase Beta-N-acetylhexosaminidases of glycosyl hydrolase family 20 (GH20) catalyze the removal of beta-1,4-linked N-acetyl-
D-hexosamine residues from the non-reducing ends of N-acetyl-beta-D-hexosaminides including N-acetylglucosides and N-acetylgalactosides.  These
enzymes are broadly distributed in microorganisms, plants and animals, and play roles in various key physiological and pathological processes.
These processes include cell structural integrity, energy storage, cellular signaling, fertilization, pathogen defense, viral penetration, the
development of carcinomas, inflammatory events and lysosomal storage disorders. The GH20 enzymes include the eukaryotic beta-N-
acetylhexosaminidases A and B, the bacterial chitobiases, dispersin B, and lacto-N-biosidase.  The GH20 hexosaminidases are thought to act via a
catalytic mechanism in which the catalytic nucleophile is not provided by the solvent or the enzyme, but by the substrate itself.
Probab=47.13  E-value=1.9e+02  Score=29.93  Aligned_cols=61  Identities=18%  Similarity=0.322  Sum_probs=0.0

Q ss_pred             CcEeccccCCCcHHHHHHHHHHHHHhcCCCE-EEEE-------------------------ChHHhCCHHHHHHHHHHHH
Q gi|339625078|r  135 NIVVEKGNNSTIFGKIEEVIHVMGMLDESRY-LFLT-------------------------NASLYCKLSDLNQLHECLL  188 (224)
Q Consensus       150 ~ikie~~~~~~~~~eki~~~iki~~~l~~kkl-lvfv-------------------------Nl~~YLt~~el~~L~~~i~  203 (251)
                     |+.+|+...=-+.+-|..+|+.+....-..+ +=++                         --+.++|.+|+.+|.+|++
T Consensus         3 G~mlD~aR~~~~~~~~~~lk~~id~ma~~K~N~lhlHl~D~~~~~le~~~~p~l~~~g~~~~~~~~~~~~~~yT~~di~elv~yA~   82 (303)
T cd02742           3 GIMLDVSRHFLSVESIKRTIDVLARYKINTFHWHLTDDQAWRIESKKFPELAEKGGQINPRSPGGFYTYAQLKDIIEYAA   82 (303)
T ss_pred             ceeeccccCcCHHHHHHHHHHHHHhCCcEEEEeeecCCCceEeeCccchhhhhcccccCCCCCCeECHHHHHHHHHHHH


Q ss_pred             HhCCeEE
Q gi|339625078|r  189 AEGVNLI  195 (224)
Q Consensus       204 ~~~l~vL  210 (251)
                     ..++.|+
T Consensus        83 ~rgI~vi   89 (303)
T cd02742          83 ARGIEVI   89 (303)
T ss_pred             HcCCEEE
```

```
No 12
>SUPFAM0048920 c.1.8 (Trans)glycosidases (51445) SCOP seed sequence: d1qbaa3; InterPro: IPR017853.
Probab=46.02  E-value=2.4e+02  Score=26.43  Aligned_cols=61  Identities=10%  Similarity=0.224  Sum_probs=0.0

Q ss_pred             CcEecccCCCcHHHHHHHHHHHHHhcCCCEEEEEE------------------------------------------ChH
Q gi|339625078|r  135 NIVVEKGNNSTIFGKIEEVIHVMGMLDESRYLFLT------------------------------------------NAS  172 (224)
Q Consensus       150 ~ikie~~~~~~~~~eki~~~iki~~~l~~kkllvfv------------------------------------------Nl~  187 (251)
                      |+.+|+...--..+.|...|+.+.......+-.-.                                          .-.
T Consensus         5 G~~~D~~R~~~~~~~~~~~~~~~~~~~~k~N~l~~hl~D~~~~~~~~~~~~~P~l~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~   84 (443)
T SUPFAM0048920     5 GLMLDVARHFFSVETIKRLIDAMALYKLNVLHLHLTDDQGWRLEIKSYPELTEVGAYRGSDLVETALLLPQLGSLPDSGG   84 (443)
T ss_pred             ccccccCCCcCCHHHHHHHHHHHHHHhCCCEEEEEeecCCCceeecCCchhhhccceecccccchheeccccccccccCCC


Q ss_pred             HhCCHHHHHHHHHHHHHhCCeEE
Q gi|339625078|r  173 LYCKLSDLNQLHECLLAEGVNLI  195 (224)
Q Consensus       188 ~YLt~~el~~L~~~i~~~~l~vL  210 (251)
                      .++|.+|+.+|.+|++..++.|+
T Consensus        85 ~~yt~~~~~~~~~~yA~~~~i~v~  107 (443)
T SUPFAM0048920    85 GFYTQEDIKEIVAYAAERGIEVI  107 (443)
T ss_pred             CccCHHHHHHHHHHHHhCCcEEE


No 13
>cd06565 GH20_GcnA-like Glycosyl hydrolase family 20 (GH20) catalytic domain of N-acetyl-beta-D-glucosaminidase (GcnA, also known as BhsA) and
related proteins. GcnA  is an exoglucosidase which cleaves N-acetyl-beta-D-galactosamine (NAG) and N-acetyl-beta-D-galactosamine residues from 4-
methylumbelliferylated (4MU) substrates, as well as cleaving NAG from chito-oligosaccharides (i.e. NAG polymers).  In contrast, sulfated forms of
the substrate are unable to be cleaved and act instead as mild competitive inhibitors. Additionally, the enzyme is known to be poisoned by several
first-row transition metals as well as by mercury.  GcnA forms a homodimer with subunits comprised of three domains, an N-terminal zincin-like
domain, this central catalytic GH20 domain, and a C-terminal alpha helical domain.  The GH20 hexosaminidases are thought to act via a catalytic
mechanism in which the catalytic nucleophile is not provided by solvent or the enzyme, but by the substrate itself.
Probab=45.73  E-value=2.2e+02  Score=29.59  Aligned_cols=61  Identities=7%  Similarity=0.221  Sum_probs=0.0

Q ss_pred             CcEecccC-CCcHHHHHHHHHHHHHhcCCCEEEEEECh-------------HHhCCHHHHHHHHHHHHHhCCeEE
Q gi|339625078|r  135 NIVVEKGN-NSTIFGKIEEVIHVMGMLDESRYLFLTNA-------------SLYCKLSDLNQLHECLLAEGVNLI  195 (224)
Q Consensus       150 ~ikie~~~-~~~~~eki~~~iki~~~l~~kkllvfvNl-------------~~YLt~~el~~L~~~i~~~~l~vL  210 (251)
                      |+.+|+.. .--..+-|..+|+.+..+.-..+..=..=             +.++|.+|+.+|.+|++..++.|+
T Consensus         3 G~mlD~~R~~~~~~~~~lk~~id~ma~~k~N~l~lhl~D~f~~~~~p~~~~~~~~~~yT~~ei~ei~~yA~~~gI~vI   77 (301)
T cd06565           3 GVHLDLKRNAVPKVSYLKKLLRLLALLGANGLLLYYEDTFPYEGEPEVGRMRGAYTKEEIREIDDYAAELGIEVI   77 (301)
T ss_pred             eEEEecCCCCCCCHHHHHHHHHHHHHcCCCEEEEEEEecceecCCCccccccCCCCcCHHHHHHHHHHHHHcCCEEE


No 14
>1t5i_A C_terminal domain of A probable ATP-dependent RNA helicase; RECA-like fold, PRE-mRNA processing protein; 1.90A {Homo sapiens} SCOP:
c.37.1.19
Probab=45.25  E-value=2.5e+02  Score=25.94  Aligned_cols=45  Identities=13%  Similarity=0.269  Sum_probs=0.0

Q ss_pred             HHHHHHHHHHHhcCCCEEEEEEChHHhCCHHHHHHHHHHHHHhCCeEEEE
Q gi|339625078|r  148 GKIEEVIHVMGMLDESRYLFLTNASLYCKLSDLNQLHECLLAEGVNLISI  197 (224)
Q Consensus       163 eki~~~iki~~~l~~kkllvfvNl~~YLt~~el~~L~~~i~~~~l~vL~i  212 (251)
                      +|+..+.++.......+++||+|     |..+...+.+.+...++++..+
T Consensus        17 ~K~~~L~~ll~~~~~~~~~~lVF~~-----~~~~~~~~l~~~L~~~~~~~~~~~~~   61 (172)
T 1t5i_A           17 EKNRKLFDLLDVLEFNQVVIFVK-----SVQRCIALAQLLVEQNFPAIAI   61 (172)
T ss_dssp             GHHHHHHHHHHHSCCSSEEEECS-----SHHHHHHHHHHHHHTTCCEEEE
T ss_pred             HHHHHHHHHHHhCCCCcEEEEEC-----CHHHHHHHHHHHHhcCCCEEEE


No 15
>PF09664 DUF2399:  Protein of unknown function C-terminus (DUF2399);  InterPro: IPR024465 This domain is found in archaeal, bacterial and
eukaryotic proteins. Its function is unknown.
Probab=44.26  E-value=1.1e+02  Score=28.56  Aligned_cols=31  Identities=13%  Similarity=0.228  Sum_probs=0.0

Q ss_pred             EEEEECCHHHHHHHHHHHHhcccccccCccccccceEEEcCCC
Q gi|339625078|r   21 NLITVANSELYLRLIAKLS----NDIEDD-------IIYSYDN   52 (224)
Q Consensus        22 tvL~ien~~~f~~li~~l~~~~~~~~~~~~~~~~~~~~~~~~l~d~~   64 (251)
                      +++++|||.||+.++..+     +...-+       +++.++.
T Consensus        20 ~V~VvENp~Vf~~~~~~~~~-----~~~~~~p-------LVCt~G~   50 (152)
T PF09664_consen   20 RVYVVENPAVFSALADEL-----GASCPP-------LVCTSGQ   50 (152)
T ss_pred             EEEEEecHHHHHHHHHhc-----CCCCCe-------EEEcCCc
```

C

Proteins with structure most similar to *S. thermophilus* Csn2 structure (3zth)

	Target	Q-score	P-score	Z-score	Seq%
1w1w:A	*S. cerevisiae* Smc1	0.1447	-0	6.767	11.73
4i99:A	*P. furiosus* Smc	0.1404	0.6379	8.12	11.9
1vpl:A	*T. maritima* ABC transporter ATP-binding protein	0.1374	-0	5.211	13.38
2onk:A	*A. fulgidus* Molybdate/tungstate ABC transporter, ATP-binding protein	0.1346	-0	5.001	7.273
4g1u:D	*Y. pestis* Hemin import ATP-binding protein HmuV	0.1308	1.04e-09	5.957	13.69
4hlu:C	*T. maritima* Energy-coupling factor transporter ATP-binding protein EcfA	0.1288	-0	5.253	11.52
2olk:A	*G. stearothermophilus* Amino acid ABC transporter	0.1271	-0	5.316	10.19
1g6h:A	*M. jannaschii* high-affinity branched-chain amino acid transport ATP-binding protein	0.1266	-0	5.862	10.69
4fi3:C	*E. coli* Vitamin B12 import ATP-binding protein BtuD	0.1253	-0	5.085	10.78
2fgk:B	*E. coli* Alpha-hemolysin translocation ATP-binding protein hlyB	0.1231	-0	4.874	12.26
2yz2:A	*T. maritima* Putative ABC transporter ATP-binding protein TM_0222	0.1219	-0	5.316	0.1562
1l2t:A	*M. jannaschii* Hypothetical ABC transporter ATP-binding protein MJ0796	0.1216	0.001364	5.357	0.1241

Proteins with structure most similar to *S. pyogenes* Csn2 structure (3toc)

	Target	Q-score	P-score	Z-score	Seq%
2onk:A	*A. fulgidus* Molybdate/tungstate ABC transporter, ATP-binding protein	0.1986	0.05449	4.77	9.028
2pcj:A	*A. aeolicus* Lipoprotein-releasing system ATP-binding protein lolD	0.1959	0.00532	4.49	7.692
3gfo:A	*C. perfringens* Cobalt import ATP-binding protein cbiO 1	0.1766	0.1979	5.312	14.39
1ji0:A	*T. maritima* ABC transporter TM_1139	0.1728	0.859	5.647	15.15
3n70:D	*E. coli* P-loop NTPase domain of the Sigma-54 transport activator	0.172	0.09933	3.876	6.522
2fgj:A	*E. coli* Alpha-hemolysin translocation ATP-binding protein hlyB	0.1675	0.3295	5.372	8.759
2olk:B	*G. stearothermophilus* Amino acid ABC transporter ArtP	0.1624	-0	4.349	4.511
3vx4:A	*S. mutans* Putative ABC transporter, ATP-binding protein ComA	0.1614	0.493	5.853	6.923
2ihy:A	*S. aureus* ATPase subunit of ABC transporter	0.1602	-0	4.327	11.03
1g9x:C	*M. jannaschii* High-affinity branched-chain amino acid transport ATP-binding protein	0.1597	-0	4.683	10.08
1l7v:C	*E. coli* Vitamin B12 transport system permease protein Btuc	0.1591	8.58e-10	4.305	8.209

```
3zth:Ass          ------EE-EE-------------------------EE-EE------EEEEE--HHHHHHHHHHHHHHHH--HH---------HHHHHH-------------------------------------------------------------
3zth:A (1)        ---X--KF-FV-QH--P---------Y--K----ERIE-LNI-G--AITQIVGQNNELKYYTWQILSWYF--GG--KK----YSSEDLS----------IFDYEEP---------TILDEAREIVKRSSY--------------------

1w1w:Ass          --EE--EE-EEE----------------------EE-E-EE------EEEEE--------HHHHHHHHHH--HHH-------HH--------------------HHH-----EEEEEEE---------EEE--EEEEEEEEEE----E--
1w1w:A (2)        GRLV--GL-ELSNF--K---------SY-RG---VT-K-VGFGES-NFTSIIGP-NGS--GKSNMMDAIS--FVLGV-----LK--------------------DLIYRGPQSAYVKAF---------YQKGNKLVELMRIISRNGDT--
4i99:A (2)        --PYI-EKLE--LKGFK---------S--YGNKKV--VIPFS-K--GFTAIVGA-NGS--GKSNIGDAILFVLG--GLSAKAXRASRISDLIFAGSKNEPPAKYAE---------VAIY---------FNNEDRGFPIDEDEVVIRRRVY
1vpl:A (2)        ---GAVVV-KDLRK--RIG-----KKEILK----GI-S-FEI-EEGEIFGLIGP-NGA--GKTTTLRIIS--T---L-----I----------------------KPS------SGIVT---------VFGKNV----------------
2onk:A (1)        --MF--LK-VRAEK--RLG-----NFR--L----NV-D-FEM-GR-DYCVLLGP-TGA--GKSVFLELIA--G---I-----V----------------------KPD------RGEVR---------LNGADIT---------------
4g1u:D (10)       --ALL-EA-SHLHY--HVQ-----QQALIN----DV-S-LHI-ASGEMVAIIGP-NGA--GKSTLLRLLT--G---Y-----L----------------------SPS------HGECH---------LLGQNLN---------------
4hlu:C (999)      GSGRI-EL-NSVSF--RYN-----GDYVLK----DV-N-AEF-ETGKIYVVVGK-NGS--GKTTLLKILA--G---L-----L----------------------AA-------AGEIF---------LDGSPAD---------------
2olk:A (1)        ----MIDV-HQLKK--SFG-----SLEVLK----GI-N-VHI-REGEVVVVIGP-SGS--GKSTFLRCLN--L---L-----E----------------------DFD------EGEII---------IDGINLKA--------------
1g6h:A (4)        -TMEILRT-ENIVK--YFG-----EFKALD----GV-S-ISV-NKGDVTLIIGP-NGS--GKSTLINVIT--G---F-----L----------------------KAD------EGRVY---------FENKDIT---------------
4fi3:C (2)        --SIVMQL-QDVAE--S---------TRLG----PL-S-GEV-RAGEILHLVGP-NGA--GKSTLLARMA--G---M-----T----------------------SG-------KGSIQ---------FAGQPLE---------------
2fgk:B (467)      ---DI-TF-RNIRF--RYKP---DSPVILD----NI-N-LSI-KQGEVIGIVGR-SGS--GKSTLTKLIQ--R---F-----Y----------------------IPE------NGQVL---------IDGHDLA---------------
2yz2:A (1)        --MRI-EV-VNVSH--IFHRGTPLEKKALE----NV-S-LVI-NEGECLLVAGN-TGS--GKSTLLQIVA--G---L-----I----------------------EPT------SGDVL---------YDGERK----------------
1l2t:A (1)        ---MI-KL-KNVTK--TYKMG-EEIIYALK----NV-N-LNI-KEGEFVSIMGP-SGS--GKSTMLNIIG--C---L-----D----------------------KPT------EGEVY---------IDNIKTNDL-------------


3toc:Ass          --------------------EEE--------------EEEEEE-HH-H-HHHHHHHHHHH--------E--EE----------------------E-EEE-----------HHHHHHHHHHHHHHHHH-HHHHHHHHHHHHHHHHHHHHH
3toc:A (1)        -------------------XNLNFSLLDEPI-PLRG-GTILVLEDV-C-VFSKIVQYCYQYES---ELK--FFXKT------IKES--------EI-XLV-T-DILGFDVNSSTILKLIHADLESQFNEKPEVKSXIDKLVATITELIVF

2pcj:Ass          -----EEEEEEEEEEEE--EEEEEEEE-EE--EE---EEEEE-E-----HHHHHHHHHH-------EEEEEE--------HHHHHHHHH---HHHE-EEE------------HHHHHHHHHHHHH--HHHHH------------------
2pcj:A (2)        ----AEILRAENIKKVIRGYEILKGIS-LS--VKKGEFVSII-GAS-GSGKSTLLYILG-LLDAPTEGKVFLEGKEVDYTNEKELSLLR---NRKL-GFV-FQFH-YLIPELTALENVIVPMLKMGKPKKEA------------------
2onk:A (1)        --------MFLKVRAEKRLGNFRLNVD-FE--MGRD-YCVLL-GPT-GAGKSVFLELIA-GIVKPDRGEVRLNGAD-----ITPLPPE----RRGI-GFV-PQDY-ALFPHLSVYRNIAYGLRNVERVERD-------------------
3gfo:A (2)        --EDYILKVEELNYNYSDGTHALKGIN-MN--IKRGEVTAIL-GGN-GVGKSTLFQNFN-GILKPSSGRILFDNKPID-YSRKGIMKL----RESI-GIV-FQDPDNQLFSASVYQDVSFGAVNMKLPEDEI------------------
1ji0:A (1)        --VSDIVLEVQSLHVYYGAIHAIKGID-LK--VPRGQIVTLI-GAN-GAGKTTTLSAIA-GLVRAQKGKIIFNGQDIT-NKPAHVINR-----XGI-ALV-PEGR-RIFPELTVYENLXXGAYNRKDKEGIK------------------
3n70:D (158)      -------------------------RL-QQLSE-TD-IAVWL-YGAPGTGRXTGARYLH-Q-------F--GRN---------AQ----------GEFVY-R-E--LT------------------------------------------
2fgj:A (467)      ----DITFRNIRFRYKPDSPVILDNIN-LS--IKQGEVIGIV-GRS-GSGKSTLTKLIQ-RFYIPENGQVLIDGHDL---ALADPNWL----RRQV-GVV-LQD--NVLLNRSIIDNISLANPGMSVEKVIYAA----------------
2olk:B (1)        ------MIDVHQLKKSFGSLEVLKGIN-VH--IREGEVVVVI-GPS-GSGKSTFLRCLN-LLEDFDEGEIIIDGINLK-AKDTNLNKV----REEV-GMV-FQR--FNLFPHMTVLNNITLAPMKVRKWPREK-----------------
3vx4:A (522)      SFLDGDISFENLSYKYGFGRDTLSDIN-LS--IKKGSKVSLV-GAS-GSGKTTLAKLIV-NFYEPNKGIVRINGNDL---KVIDKTAL----RRHI-SYLPQ-Q--AYVFSGSIMDNLVLGAKEGTSQEDIIRA----------------
2ihy:A (1)        -----XLIQLDQIGRXKQGKTILKKIS-WQ--IAKGDKWILY-GLN-GAGKTTLLNILN-AYEPATSGTVNLFGKX-----PGYSAETV---RQHI-GFV-S-HSLLEKFQEGERVIDVVISGAIDDEIR--------------------
1g9x:C (2004)     ---TMEILRTENIVKYFGEFKALDGVS-IS--VCKGDVTLII-GPN-GSGKSTLINVIT-GFLKADEGRVYFENKD-----ITNKEPAELYHYGIV-RTF-Q-T--PQPLKEMTVLENLLIGEINPGESPLNSLFYKKW-----------
1l7v:C (2)        --------SIVXQLQDVAESTRLGPLS-GE--VRAGEILHLV-GPN-GAGKSTLLARXA-GXTS-GKGSIQFAGQPL----EAWSATKL---ALHR-AYLSQ-Q--QTPPFATPVWHYLTLHQHDKTRTE--------------------
```

```
3zth:Ass        -----------------------E----------EEE-E--HH--HH---------------------H---------------------------------------HHHH-----HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH
3zth:A          -----------------------H----------YID-I-SSF--KD---------------------L---------------------------------------LEQXEYKKGTLAQGYLRKIVNQVDIVGHLEKINEQVELIEEA

1w1w:Ass        -----EEEE--EEEHHHHHHHHHHHHH-----------------HH----------------------H---------------------------------------HH----------------------------------------
1w1w:A          -----SYKIDGKTVSYKDYSIFLENENILIKAKNFLV-FQGDV-EQ----------------------I---------------------------------------AA----------------------------------------
4i99:A          PDGRSSYWLNGRRATRSEILDILTAAXIS-PDGYNI--VLQGDITK----------------------FI-------------------------------------KX-----------------------------------------
1vpl:A          -------------------VEEPHEVR----KLISYLPE-EA---G----------------------A---------------------------------------Y-----------------------------------------
2onk:A          ------------------PLPPERRG-------IGFV-P-QDY--A----------------------L---------------------------------------F-----------------------------------------
4g1u:D          ------------------SWQPKALAR-----TRAVM-R-QY---S----------------------E---------------------------------------LAFPFSVSEVIQMGRAPYGGSQD-------------------
4hlu:C          ------------------PFLLRKN--------VGYV-F-QNP--SS---------------------Q---------------------------------------IIGATVEEDVAFSLEIMGLDESE-------------------
2olk:A          -----------------KDTNLNKVRE-----EVGMV-F-QR---F----------------------N---------------------------------------LF----------------------------------------
1g6h:A          ------------------NKEPAELYH----YGIVRT-F-QT---PQPLKEMTVLENLLIGEICPGESPLNSLFYKKWIPKEEEMVE---KAFKILEFLKLSHLYDRKA-----------------------------------------
4fi3:C          ------------------AWSATKLAL-----HRAYLSQ-QQ---T----------------------PPFATPVWHYLTLHQHDKTRTELLNDVAGALALDDKLGRST-----------------------------------------
2fgk:B          ------------------LADPNWLRR-----QVGVV-L-QD---N----------------------V---------------------------------------LLNRSIIDNISLANPGMSVEKVIYA-----------------
2yz2:A          --------------------KGYEIRR-----NIGIA-F-QYP-ED----------------------Q---------------------------------------F-----------------------------------------
1l2t:A          -----------------DDDELTKIRR----DKIGFV-F-QQF--N----------------------L---------------------------------------I-----------------------------------------
```

```
3toc:Ass        HHHHH---------HHHHHHHHHH---------H--HHHHHHHHHHHHH------------EEEEE-H-HHHH----HHHHHHHHHHHHHHH-------EEEEE--------------------------EEEE--EEE-----------
3toc:A          ECLENELDLEYDEITILELIKSLGVKVETQSD-T--IFEKCLEILQIFKYLTK------KKLLIFVNS-GAFL----TKDEVASLQEYISLTN---L-TVLFLEPREL---Y-D-----F--P-------QYILDLITKNX---------

2pcj:Ass        -------HHHHHHHHHHH--------HHHHHH-HHHHHHHHHHHHH---------------EEEEE-------HHHHHHHHHHHHHHHHHH-------EEEEE-----HHHH-H-----H---------EEEEEE--EEEEEEE------
2pcj:A          -------KERGEYLLSELGLGDKLSRKPYELS-GGEQQRVAIARALA---NE-------PILLFADEP-TGNLDSANTKRVMDIFLKINEG-G---T-SIVMVT---HEREL-A-----E--LTH----RTLEMKDGKVVGEITRV----
2onk:A          --------RRVREMAEKLGIAHLLDRKPARLS-GGERQRVALARALV---IQ-------PRLLLLDEP-LSAVDLKTKGVLMEELRFVQREFD---V-PILHVT---HDLIE-AA----M--LAD----EVAVMLNGRIVEKGKLKELFS
3gfo:A          -------RKRVDNALKRTGIEHLKDKPTHCLS-FGQKKRVAIAGVLV---ME-------PKVLILDEP-TAGLDPMGVSEIMKLLVEMQKELG---I-TIIIATH--DIDIV-P-----L--YCD----NVFVMKEGRVILQGNPKEVFA
1ji0:A          -------RDLEWIFSLFPRLKERLKQLGGTLS-GGEQQXLAIGRALX---SR-------PKLLXXDEP-SLGLAPILVSEVFEVIQKINQE-G---T-TILLVE---QNALGAL-----K--VAH----YGYVLETGQIVLEGKASELLD
3n70:D          ------------------------------PDNAP-QLNDFIALA------Q-------G-GTLVLSH-PEHL----TREQQYHLVQLQSQEH---RPFRLIGI---GDTSL-VEIAELYYCFA-----XTQ------------------
2fgj:A          ----KLAGAHDFISELREGYNTIVGEQGAGLSGGQ-RQRIAIARALV---NN-------PKILIFDEATSALD----YESEHVIMRNMHKICKG--R-TVIIIA---A---R-L-----S--TVKN-ADRIIVMEKGKIVEQGKHKELLS
2olk:B          ------AEAKAMELLDKVGLKDKAHAYPDSLS-GGQAQRVAIARALA---ME-------PKIMLFDEP-TSALDPEMVGEVLSVMKQLANE-G---M-TMVVVT---H---E-M-----G--FAREVGDRVLFMDGGYIIEEGKPEDLFD
3vx4:A          ----CEIAEIRSDIEQMPQGYQTELSDGAGISGGQ-KQRIALARALL---TQ-------APVLILDAA-TSSL---DILTEKKIISNLLQM-TE--K-TIIFVA---H---R-L-----S--ISQR-TDEVIVMDQGKIVEQGTHKELLA
2ihy:A          --------NEAHQLLKLVGXSAKAQQYIGYLS-TGEKQRVXIARALX---GQ-------PQVLILDEP-AAGLDFIARESLLSILDSLSDSYPT--L-AXIYVTH--FI-EE-I-----T--ANF---SKILLLKDGQSIQQGAVEDILT
1g9x:C          IPKEEEMVEKAFKILEFLKLSHLYDRKAGELS-GGQMKLVEIGRALM---TN-------PKMIVMDEP-IAGV---APGLAHDIFNHVLEL-KAKGI-TFLIIEH--R---L-D-----I--VLNY-IDHLYVMFNGQIIAEGRGEEEIK
1l7v:C          ---------LLNDVAGALALDDKLGRSTNQLSGGE-WQRVRLAAVVL---QITPQANPAGQLLLLDEP-XNSL----DVAQQSALDKILSALCQQGL-AIVXSS---H---D-L-----N--HTLRHAHRAWLLKGGKXLASGRREEVLT
```

```
3zth:Ass            HHH------------------HHHHHHHHH----EEE--EEEHHH----------------------------------------------------------HH--------------------HHHHHHHHHHHHHHHHHH-------
3zth:A              XNRHINLNCGQVEYHLENLPLTLDQLLTKNFSPFFAIENKNLSFE----------------------------------------------------------WV--------------------SNIDKLSLFLEXLDHLLSQ---TTE

1w1w:Ass            -------------------------------------------HHHHHHHHH-HHHHHHHH---------------------------HHHHHHHH--EEEE--EEEE--------HHHHHHHHHHHHHHHHHHHHHHH-----------
1w1w:A              -------------------------------------------QSPVELSRMFTFDYVSDH---------------------------LDAIYRELTGNASLTKYHATPPLKRFK-DMEYLSGGEKTVAALALLFAINSYQP-------S
4i99:A              -------------------------------------------SPLERRLLIDDISGIKEKKNVFXRTFEAISRNFSEIFAKLSPGGSARLILENPEDPFSGGLEIEAKPAGKDVKRIEAXSGGEKALTALAFVFAIQKF-KP------A
1vpl:A              -------------------------------------------RNMQGIEYLRFVAGFYASSSSE------------------IEEMVERATEIAGLGEKIKDRVST-----------------YSKGMVRKLLIARALMVN-------P
2onk:A              -------------------------------------------PHLSVYRNIAYGLRNVERVER---------------------DRRVREMAEKLGIAHLLDRKPA----------------RLSGGERQRVALARALVIQ-------P
4g1u:D              ----------------------RQALQQVMAQTDCLALAQRDYR------------------------------------------------------------VL-------------------SGGEQQRVQLARVLAQL-------W
4hlu:C              ---------------------MRKRIKKVLELVGLSGLAAADPL------------------------------------------------------------NL-------------------SGGQKQRLAIASMLARD-------T
2olk:A              -------------------------------------------PHMTVLNNITLAPMKVRKWPRE------------------KAEAKAMELLDKVGLKDKAHAYPD----------------SLSGGQAQRVAIARALAME-------P
1g6h:A              -------------------------------------------G------------------------------------------------------------EL-------------------SGGQMKLVEIGRALMTN-------P
4fi3:C              -------------------------------------------N------------------------------------------------------------QL-------------------SGGEWQRVRLAAVVLQITPQANPAG
2fgk:B              -------------------AKLAGAHDFISELREGYNTIVGEQGA-----------------------------------------------------------GL-------------------SGGQRQRIAIARALVNN-------P
2yz2:A              -------------------------------------------FAERVFDEVAFAVKNFYPDRDP------------------VPLVKKAMEFVGLDFDSFKDRVPF----------------FLSGGEKRRVAIASVIVHE-------P
1l2t:A              -------------------------------------------PLLTALENVELPLIFKYRGAMSG----------------EERRKRALECLKMAELEERFANHKPN---------------QLSGGQQQRVAIARALANN-------P
```

```
3zth:Ass     HHH------------------HHHHHHHHH----EEE--EEEHHH----------------------------------------------------------HH--------------------HHHHHHHHHHHHHHHHHH-------
3zth:A       XNRHINLNCGQVEYHLENLPLTLDQLLTKNFSPFFAIENKNLSFE----------------------------------------------------------WV--------------------SNIDKLSLFLEXLDHLLSQ---TTE

1w1w:Ass     -------------------------------------------HHHHHHHHH-HHHHHHHH---------------------------HHHHHHHH--EEEE--EEEE--------HHHHHHHHHHHHHHHHHHHHHHH-----------
1w1w:A       -------------------------------------------QSPVELSRMFTFDYVSDH---------------------------LDAIYRELTGNASLTKYHATPPLKRFK-DMEYLSGGEKTVAALALLFAINSYQP-------S
4i99:A       -------------------------------------------SPLERRLLIDDISGIKEKKNVFXRTFEAISRNFSEIFAKLSPGGSARLILENPEDPFSGGLEIEAKPAGKDVKRIEAXSGGEKALTALAFVFAIQKF-KP------A
1vpl:A       -------------------------------------------RNMQGIEYLRFVAGFYASSSSE------------------IEEMVERATEIAGLGEKIKDRVST-----------------YSKGMVRKLLIARALMVN-------P
2onk:A       -------------------------------------------PHLSVYRNIAYGLRNVERVER---------------------DRRVREMAEKLGIAHLLDRKPA----------------RLSGGERQRVALARALVIQ-------P
4g1u:D       ----------------------RQALQQVMAQTDCLALAQRDYR------------------------------------------------------------VL-------------------SGGEQQRVQLARVLAQL-------W
4hlu:C       ---------------------MRKRIKKVLELVGLSGLAAADPL------------------------------------------------------------NL-------------------SGGQKQRLAIASMLARD-------T
2olk:A       -------------------------------------------PHMTVLNNITLAPMKVRKWPRE------------------KAEAKAMELLDKVGLKDKAHAYPD----------------SLSGGQAQRVAIARALAME-------P
1g6h:A       -------------------------------------------G------------------------------------------------------------EL-------------------SGGQMKLVEIGRALMTN-------P
4fi3:C       -------------------------------------------N------------------------------------------------------------QL-------------------SGGEWQRVRLAAVVLQITPQANPAG
2fgk:B       -------------------AKLAGAHDFISELREGYNTIVGEQGA-----------------------------------------------------------GL-------------------SGGQRQRIAIARALVNN-------P
2yz2:A       -------------------------------------------FAERVFDEVAFAVKNFYPDRDP------------------VPLVKKAMEFVGLDFDSFKDRVPF----------------FLSGGEKRRVAIASVIVHE-------P
1l2t:A       -------------------------------------------PLLTALENVELPLIFKYRGAMSG----------------EERRKRALECLKMAELEERFANHKPN---------------QLSGGQQQRVAIARALANN-------P
```

A

```
312126587|Caldicellulosiruptor hydrothermalis 108      -------------------------------MRFVPVVDKDGNPLMPTKPS-RARRWIKSGKATPFWSK---GIFCVRLNI-------------------------EPSARYKQPVALGI
434408285|Stanieria cyanosphaera PCC 7437              -----------------------MYIQLGLIAMYVPVVDKNQIPLMPTTST-RAKRWIKSGKATGFWKR---GVFCVRLNI-------------------------EPSNRNYQPIACGV
414076241|Anabaena sp. 90                              --------------------------------MRVPVISVDNTPLMPTKPS-RARRWIKCGKAIGKFDKL--GIFYVQLLA-------------------------ESSNTETQEIVLGL
428299778|Calothrix sp. PCC 6303                       --------------------------------MRVPVISKDNLPLMLTKPS-RARRWIKEGKAIGKFNKL--GIFYVKLLT-------------------------EASDEKIQEVVIGL
 17233043|Nostoc sp. PCC 7120                          ----------------------------------------------------------------------------MQLTG-------------------------EPSGRATQDIVIGI
166363864|Microcystis aeruginosa NIES-843              ----------------MLRKGTLTLLLRIISMARVPVISKDGKSLMPTKPS-RARRWIKEGKAIGKFNDL--GIFYVQLID-------------------------EPSDSKTQPIAIGI
166367659|Microcystis aeruginosa NIES-843              -------------------------------MARVPVISKDGKSLMPTKPS-RARRWIKEGKAIGKFNDL--GIFYVQLID-------------------------EPSDSKTQPIAIGI
307150343|Cyanothece sp. PCC 7822                      -------------------------------MERVPVISKEGKPLMPTKSS-RARKWIKDGKAVGKFNDL--NQFYVQLAV-------------------------EPSDSKIQLISIGI
307153352|Cyanothece sp. PCC 7822                      -------------------------------MLRVPVISKEGKPLMPTKPS-RARKWIKDGKAVGKFNDL--NQFYVQLIV-------------------------NSSDNKTQTISIGI
428306661|Crinalium epipsammum PCC 9333                -------------------------------MLRVPVISKNNLPLMPTKSS-RARRWIEEGKAVGALDDL--GQFYVQLIE-------------------------SASDNQTQPIAVGI
307150912|Cyanothece sp. PCC 7822                      -------------------------------MLRVPVIAPDKTPLMPTKCS-RARRWIKEGKAIGKFNDL--GQFYVQLIE-------------------------QPSDNKTQPITIGI
428775289|Halothece sp. PCC 7418                       --------------------------------------MSNLRLMVRSGST-HDKGCVREERSEV-------KVSCCVLKT---------------------------------SLEGDL
428779516|Dactylococcopsis salina PCC 8305             -------------------------------MYRVPVRSKNGKPLMPTKAS-RARRWLKEGKAVIYKNDL--HSFAVQLTI--------------------------DSGEETQPIAVGI
428778599|Dactylococcopsis salina PCC 8305             -------------------------------MYRVPVISKDGKPLMPTKSS-RARRWLNEGKAVVYKNDL--HSFAVQLTI--------------------------DSGKETQPIAVGI
428781696|Dactylococcopsis salina PCC 8305             -------------------------------MYRVPVISKEGRPLMPTKTS-RARRWLKEGKAVIYKNDL--HSFAIQLTV--------------------------DSGEEIQPIAVGV
428215045|Oscillatoria acuminata PCC 6304              ------------MLDSVGEETITPIGGLSQTMLRVPVLSPDGKPLMPTQPS-RARRWIRDGLAVGKWSDL--GMFYVQRVE-------------------------QTSGEETQPIAVGV
428313425|Microcoleus sp. PCC 7113                     ---------------------------------------------MPTKPS-RARRWIQDGLAVKKWSDL--GVFYIQLIN-------------------------QPSGEKTQPIAIGI
158340109|Acaryochloris marina MBIC11017               ------------------------------MILRVPVISPEGQPLMPTKPS-RARRMVRDGKAVGKWNDA--GVYYIQLTA-------------------------DPSDTQTQPIHLGC
428775002|Halothece sp. PCC 7418                       -------------------------------MNRIPVVSKDGQPLMPTKPS-RARKWIESGRAISKWSDL--GLYYVQLLQ-------------------------EPSGTETQPVVAGL
307150700|Cyanothece sp. PCC 7822                      -------------------------------MLRVPVIAPNKTPLMPTKCS-RARRWIKEGKALGRFNDL--GQFYLQLIE-------------------------QPSDNKTQPIAIGI

312126663|Caldicellulosiruptor hydrothermalis 108      --------------------------------MLLFTIDRNGKPGHPTRRFDMVRKLVKRRRAKIIGGGASGKPPVVMFLD----------------------REFDYSKTVSRKLIIVL
312621424|Caldicellulosiruptor kronotskyensis 2002     --------------------------------MVIFTVDKHGRPGHPTRRFDMVRKLVKQGRAKIIGGGTSGKPPVVMFLD----------------------REFDYTKTIERRLFVVL
222530402|Caldicellulosiruptor bescii DSM 6725         --------------------------------MVIFTVDKHGRPGHPTRRFDMVRKLVKQGRAKIIGGGASGKPPVVMFLD----------------------REFDYSKTIERRLFVVL
222530431|Caldicellulosiruptor bescii DSM 6725         --------------------------------MVIFTVDKHGRPGHPTRRFDMVRKLVKQGRAKIIGGGASGKPPVVMFLD----------------------REFDYSKTIERRLFVVL

307591386|Cyanothece sp. PCC 7822                      ------------------------------MSNFVFVLDAKKNPLSPCHPS-VARKLLKQGKAAVLRR----YPFTIILKK--------------------------ECQKPTETIKLKL
302391131|Acetohalobium arabaticum DSM 5501            -----------------------------MAINYAFVLDNEGKMLDPTKAK-KAWYLIRQGKAELVEE----YPLIIQLNR-----------------------TVPSSQINDDEIVLGI
435854579|Halobacteroides halobius DSM 5150            ----------------------------MAQLQYALVIDNEGKILDPTKVK-KAWYLIRKGRVKLVEE----YPLIIQLKK-----------------------KIPKEEVNNDKIVLGI
435852864|Halobacteroides halobius DSM 5150            -----------------------------MPNKYAFVLDSKGKLLDPTKSK-KAWYLIRKGKASLVEE----YPLIIKLKR-----------------------EVPKDQVNSDKLILGI
435854272|Halobacteroides halobius DSM 5150            -----------------------------MPNKYAFVLDSKGKLLDPTKSK-KAWYLIRKGKASLVEE----YPLIIKLKR-----------------------EVPKDQVNSDKLILGI
313123202|Lactobacillus delbrueckii  ND02              --------------------------MKHENANRVFLLNKDGKPLMPCCPR-KARLLLKSGKAFVVKK----YPFTIQLKY--------------------------GSYGYKQKVSLGV
77362205|Pseudoalteromonas haloplanktis TAC125         --------------------------------MLAYVLNKHGKPLMPCKAA-KARKLLKAGEAKVVKR----TPFTIKLLH--------------------------GSSGHTQHVVAGM
298675374|Methanohalobium evestigatum Z-7303           --------------------------------MHVYVLSKNKKPLMPTRPA-KARRLLKSGKAKVVRR----TPFSIKLLY--------------------------ETTEYTQPVIAGM
298674707|Methanohalobium evestigatum Z-7303           --------------------------------MRVFVLSKNKKPLMPTRPA-KARRLLESGKAKVVRR----SPFTIKLLY--------------------------DTAEYTQPVIAGM
298675090|Methanohalobium evestigatum Z-7303           --------------------------------MRVFVISKNKKPLMPTRPA-KARRLLKSGKAKVVRR----TPFTIKLLY--------------------------ETTEYTQPVVAGM
169834784|Clostridium botulinum A3 str. Loch Maree     ------------------------------MPLMVYVLNKEGKPLMPTNRYAKIRVLLKSKKAVVVNS----KPFTIKLSY--------------------------NTKDFVQPITLGI
229587454|Clostridium botulinum Ba4 str. 657           -------------------------MIKRGVPLMVYVLSKKGKPLMPTNRYAKVRVLLKFKKAVVVSS----KPFTIKLLY--------------------------NTKDFVQPITLGI

297585104|Bacillus selenitireducens MLS10              --------------------------------MFVYVLNKNGQPLMPCKPR-KARLLLNEKKAKVVKQ----TPFTIQLLY--------------------------GSSGYKQPVSLGV
188587532|Natranaerobius thermophilus JW-NM-WN-LF      ---------------------------------MVYVLAKNGKPLMPTKKHGKVRHLLKDGKAKIVKR----TPFTIQLLY--------------------------ESDNYTQPVNLGV
312793946|Caldicellulosiruptor kristjanssonii 177R1B   ---------------------------------MVYVISKDGKPLMPTKRHGKVRRLLKQGLAKVVRR----EPFTIQLLY--------------------------DTTTYTQPVTVGV
344996431|Caldicellulosiruptor lactoaceticus 6A        ---------------------------------MVYVISKDGKPLMPTKRHGKVRRLLKQGLAKVVRR----EPFTIQLLY--------------------------DTTTYTQPVTVGV
217077332|Thermosipho africanus TCF52B                 ---------------------------------MVYVISKHGKTLMPTKRHGKVRRLLKQGLAKVVRR----GPFTIQLLY--------------------------DTTTYTQPLTVGI
375009168|Geobacillus thermoleovorans CCB US3 UF5      --------------------------------MFVYIINKHGKPLMPCSPR-KARLLLKQKKAKVVRR----TPFTIQLLY--------------------------GSSGYKQPVSLGV
407477159|Exiguobacterium antarcticum B7               --------------------------------MFVFVLNHHGEPLMPCKPS-KARKLLHNSKAKLVKQ----TPFTIQLFY--------------------------GSSGYKQEISLGV
172057416|Exiguobacterium sibiricum 255-15             --------------------------------MFVFVLNYHGEPLMPCKPS-KARKLLNNGKAKLIKQ----TPFTIQLLF--------------------------GSSGYKQEISLGV
169823684|Finegoldia magna ATCC 29328                  --------------------------------MVVYVLNIDNKPLMPTTRFGKVRRLLRNKKANVVKR----TPFTIKLLY--------------------------RTDNNVQDITLGV
160902522|Petrotoga mobilis SJ95                       --------------------------------MLVYVLNKYGKPLMPCKPS-KARKLLKEGKAKVVRK----EPFTIQLLY--------------------------GSSGYKQPITLGV
160902616|Petrotoga mobilis SJ95                       --------------------------------MLVYVLNKHGEPLMPCKSS-KARKLLKDGKAKVVRK----EPFTIQLLY--------------------------GSSGYKQPITLGV

333978211|Desulfotomaculum kuznetsovii DSM 6115        ------------------------MAPAFAGHSKVFVVDAEGKPLLPCHPA-RARKLLKVGKVEVLRG----SPFTIRLKY--------------------------PVENPVGSLRAKV
260892456|Ammonifex degensii KC4                       ------------------------MAPTFVGYPKVFVVDAEGKPLLPCHPA-RARKLLKSGKAEVLRG----SPFTIKLKY--------------------------PVSTPVGSLRAKV
260893631|Ammonifex degensii KC4                       ------------------------MAPTFVGYPKVFVVDAEGKPLLPCHPA-RARKLLKSGKAEVLRG----SPFTIKLKY--------------------------PVSNPVGSLRAKV
302391857|Acetohalobium arabaticum DSM 5501            ---------------------------------MVFVLNKDKTPLNPCHNG-KAKWLLDNNYAVVHKI----QPFTIRLKK-------------------------KVTNPNLKDYTLKI
428778981|Dactylococcopsis salina PCC 8305             ------------------------------MQNYIFQLDSNKQPLDMIHPA-KARRLQSQGKAATFRM----YPYTIIRNI-------------------------SVKSPNTKSYILKI
260892735|Ammonifex degensii KC4                       ---------------------------------MVFVLDRKGRPLMPCSER-RARKLLEQGRAVVVRL----EPFVIRLKD------------------------RVAEEASLQPLRLKI
260893993|Ammonifex degensii KC4                       ---------------------------------MVFVLDRKGRPLMPCSER-RARKLLEQGRAVVVRL----EPFVIRLKD------------------------RVAEEASLQPLRLKI
433655673|Thermoanaerobacterium M0795                  MWQTQKAEITLARRDYLLIALLVLYRTPKGEFYMVFVLDKHKKPLMPCTEK-RARLLLESGRAVIHKI----NPFVIRLKD------------------------VTVDNCIIQPCRIKI
332982963|Mahella australiensis 50-1 BON               ---------------------------------MVFVLDKHKKPLMPCSEK-RARLLLERSRAVVHKK----EPFTIRLKD------------------------RTEEESQLQSLRLKL
332983053|Mahella australiensis 50-1 BON               ---------------------------------MVFVLDKHKKPLMPCSEK-RARLLLERSRAVVHKK----EPFTIRLKD------------------------RTEEESQLQSLRLKL
222530385|Caldicellulosiruptor bescii DSM 6725         ---------------------------------MVFVLNRDKTPLAPCHEA-VARKLLKHGKAVIHRI----YPFTIRLKE------------------------QKDTSTFKPNYRLKI
298674346|Methanohalobium evestigatum Z-7303           ---------------------------------MVFVLNKNKKPLTPCHPA-KARILLKAGKATIHKK----YPFTIRLKE-------------------------LKSDDVKDDFRLKI
298675173|Methanohalobium evestigatum Z-7303           ---------------------------------MVFVLNKNKKFLTPCHPA-KARILLKARKATIHKK----YPFTIRLKE-------------------------LKSDDIKDDFRLKI
292490518|Nitrosococcus halophilus Nc 4                --------------------------------MGVFVLDKNRQPLMPCQPA-RARKLLREGKAAVFRC----YPFTIILKE--------------------------RAGGTTQPISLTV
428778759|Dactylococcopsis salina PCC 8305             ------------------------------MSNYVFVIDQKKTPLKPVHPK-RARQLLSSGKAAVFRL----FPFTLILKK-------------------------EVPNPDIHPLTLKI
300114477|Nitrosococcus watsonii C-113                 -------------------------------MNSVFVLDANKTPLMPCRPS-RARRLLRDRKAAVYRM----QPFTLMLKY--------------------------RVEASPQPIEFKV
292491781|Nitrosococcus halophilus Nc 4                -------------------------------MNRVFVLDTDRKPLMPCSPA-RARRLLRDGKAAVYRM----QPFTIILKY--------------------------RVDPNPQPVEFKV
292491869|Nitrosococcus halophilus Nc 4                -------------------------------MNRVFVLDTDRKPLMPCSPA-RARRLLRDGKAGVYRM----QPFTIILKY--------------------------RVDPNPQPVEFKV
292491893|Nitrosococcus halophilus Nc 4                -------------------------------MNRVFVLDTDRKPLMPCSPA-RARRLLRDGKAAVYRM----QPFTIILKY--------------------------RVDPNPQPVEFKV
292491036|Nitrosococcus halophilus Nc 4                -------------------------------MNSVFVLDTDKTPLMPCRPS-RARRLLRDGRAAVFRM----QPFTVILKY--------------------------RVVPTPQPVEFKV
292490884|Nitrosococcus halophilus Nc 4                -------------------------------MNRVFVLDTDKTALMPCRPS-RARRLLRDGKAGVYRM----QPFTIILKY--------------------------RVDSNPQPVEFRV
292491492|Nitrosococcus halophilus Nc 4                -------------------------------MNRVFVLDTDKTALMPCRPS-RARRLLRDGKAGVYRM----QPFTIILKY--------------------------RVDSNPQPVEFRV
297563003|Nocardiopsis dassonvillei DSM 43111          ---------MPPGAGEQAREDRPPKTTPREGHPYVFVPDRHGTPLQPTHPA-RARRLLARGRAIAARH----TPFVIRLKD------------------------RTTAESGVDGVETGI
374989417|Streptomyces bingchenggensis BCW-1           ----------------------------------MFVLDKHGRPLQPCTPA-RARKLLKKGRAVVHRH----TPFAIRLKD------------------------RTIAESEVDGVELGI
374989654|Streptomyces bingchenggensis BCW-1           ----------------------------------MFVLDKHGQPLQPCTPA-SARKLLKKGRAVVHRH----TPFVIRLKD------------------------RTIAESEVDGVELGI
271963428|Streptosporangium roseum DSM 43021           ----------------------------------MFVLDKHGQPLQPTSPA-YARKLLRSGRAVVHRH----TPFVIRLKD------------------------RTAADSTVEGMHVGI
271964894|Streptosporangium roseum DSM 43021           ----------------------------------------------------------------MVRH----TPFVIRLKD------------------------RTVAGSTMQGVELGI
258511478|Alicyclobacillus acidocaldarius DSM 446      -----------------------------MRQNRVLVLDTHRRPLMPCHPA-RARQLLKAGRASVFRR----YPFTIILHD--------------------------RNGGDVQPMQLKL
160902264|Petrotoga mobilis SJ95                       --------------------------MEKESVQRVFVLDKNKQPLMPCHPA-RARELLKKGKAAVFRY----HPFTIILKD--------------------------REGGDTQPIQVKI
121604978|Polaromonas naphthalenivorans CJ2            ---------------------------------------------MPCSEK-RARKLLESGRARVHKL----FPFGIRLVD------------------------RLVEDSVFQPLRLSL
347754630|Chloracidobacterium thermophilum B           ---------------------------------------------MPCSEK-RARLLLERGRARVHRL----VPFTIRLLD------------------------REQEDCALQPVRLKL
218667603|Acidithiobacillus ferrooxidans ATCC 23270    --------------------------------MAVFVLDRRKRPLMPCSEK-RARLLLERGRARVHRM----VPFTIRLVD------------------------RLQADSVLQPLRLKL
218667415|Acidithiobacillus ferrooxidans ATCC 23270    --------------------------------MAVLVLDKRKKPLMPCSEK-RARLLLERGRARVHRM----VPFTIRLVD------------------------RLQEDSTLQPVRLKL
344201059|Acidithiobacillus ferrivorans SS3            --------------------------------MAVFVLDKRKKPLMPCSEK-RARILLERGRARVHRM----VPFTIRLVD------------------------RLQDDSVLQPVRVAA
344201127|Acidithiobacillus ferrivorans SS3            --------------------------------MAVLVLDKRKKPLMPCSEK-RARILLERGRARVHRM----VPFTIRLVD------------------------RRVENSVLQPVRVAA
288940385|Allochromatium vinosum DSM 180               --------------------------------MAVFVLDKQKHPLMPCTEK-RARLLLERGRAVVVRL----APFTIRLKD--------------------------RIGGALQPLRLKL
288941921|Allochromatium vinosum DSM 180               --------------------------------MAVFVLDKQKHPLMPCTEK-RARLLLERGRAVVVRL----APFTLRLKD--------------------------RIGGAVQPLRLKL
288940786|Allochromatium vinosum DSM 180               --------------------------------MAVFVLDKQKHPLMPCTEK-RARLLLERGRAVVVRL----APFTLRLKD--------------------------RIGGAVQPLRLKL
288941171|Allochromatium vinosum DSM 180               --------------------------------MAVFVLDKQKHPLMPCTEK-RARLLLERGRAVVVRL----APFTIRLKD--------------------------RIGGAVQPLRLKL
288941306|Allochromatium vinosum DSM 180               --------------------------------MAVFVLDKQKHPLMPCTEK-RARLLLERGRAVVVRL----APFTLRLKD--------------------------RIGGAVQPLRLKL
121582913|Polaromonas naphthalenivorans CJ2            --------------------------------MAVFVLDRSARPLMPCTEK-RARLLLERGRARVHKV----MPFTIRVVD------------------------LEAEDCEFQALRVKL
121604604|Polaromonas naphthalenivorans CJ2            ---------------MCRKACVTRPVRAEGFEMAVFVLDRSGRPLMPCTEK-RARLLLERGRARVHKM----LPFTIRVVD------------------------RFAESSEFQALRIKL
91791204|Polaromonas sp. JS666                         ---------------------------------------------MPCTEK-RARLLLARGRARVHRL----VPLVIRLVD------------------------RQAAACDFQPLRIKL
91791283|Polaromonas sp. JS666                         ---------------------------------------------MPCTEK-RARLLLARGRARVHRL----TPMVIRLVD------------------------RQAAACDFQPLRLKL
121582916|Polaromonas naphthalenivorans CJ2            --------------------------------MAVFVLDRSGQPVMPCSEK-RARLLLQSKRARVHRV----MPFTIRLID------------------------RSQADCLLQPLRLKL
121582983|Polaromonas naphthalenivorans CJ2            --------------------------------MAVFVLDRGGQPVMPCSEK-RARLLLQSKRARVHRV----MPFTIRLID------------------------RSQTDCLLQPLRLKL
428304158|Crinalium epipsammum PCC 9333                --------------------------------MRVFVLDKNLQPLDPCHPA-RARELMNKGRAALFKR----YPFTIVLQD------------------------RSVEESVTHPHRIKI
428204921|Chroococcidiopsis thermalis PCC 7203         ------------------------------MQNYVFVIDQNKQPLNLIPPA-RARELLTKQKAAVFRM----YPFTIILKH-------------------------AVANPLPKPLTIKI
307153539|Cyanothece sp. PCC 7822                      -------------------------------MSKVFVLAPNKKPLNPVHSG-QARRLLNQGKAAIFRR----IPFSIILKK--------------------------EVNNSANPLRVKI
428216193|Oscillatoria acuminata PCC 6304              --------------------------------MHVFVLDTNKKPLSPCHPC-KARQLLKSGRASVFRR----YPFTIILHD------------------------TKALDSTVGETRLKI
414079584|Anabaena sp. 90                              ------------------------------MSNFVLVLDTNKKPLTPIHPG-DARFLLNQQKAAVFRR----FPFTIILKE-------------------------PKSHVPTQPIELKI
172055147|Cyanothece sp. ATCC 51142                    --------------------------------MFVFVLSKDKRPLSPCHQA-RARKLLKEGKAAVYRK----YPFTIILKE-------------------------EVNPSKPQPCRIKL
428316510|Oscillatoria nigro-viridis PCC 7112          ------------------------------MSNFVFVLDTNKQPLEPVAPG-QARRLLKLTQAAVYLR----YPFTIILKY-------------------------AVDHPKTQPAQLKI
428314951|Oscillatoria nigro-viridis PCC 7112          ------------------------------MSNFVFVVDTNGQPLNPVPPG-QARRLLKLQKAAIYRR----YPLTIVLKY-------------------------TVSAPQIQPHQLKI
307591386|Cyanothece sp. PCC 7822                      ------------------------------MSNFVFVLDAKKNPLSPCHPS-VARKLLKQGKAAVLRR----YPFTIILKK--------------------------ECQKPTETIKLKL
75812341|Anabaena variabilis ATCC 29413                -------------------------------MCKVFVINKEKRPLNPIHPA-QARQLLRSGKAAVFKR----FPFTIIVKE-------------------------SNVNTNVSPLRLKL
17233293|Nostoc sp. PCC 7120                           -------------------------------MCKVFVIDNEKRSLNPIHPA-QARQLLKRGKAAVFRR----FPFTIVVKE-------------------------SYTDATVSPLRLKF
75812782|Anabaena variabilis ATCC 29413                -------------------------------MSKVFVLDTEKRPLDPIHSA-QARQLLRNKKAAVFRR----FPFTIILKE-------------------------SRADASVSDLRIKL
434407609|Cylindrospermum stagnale PCC 7417            -------------------------------MSKIFVIDTDRQPLNPIHPA-QARQLLRNKKAAIFRR----FPFTLILKE-------------------------SRTNSPSSPLRLKI
414078405|Anabaena sp. 90                              -------------------------------MSKTLVLDTNKRPLDPIHPA-QARQLLRNKKAAVFRQ----FPFTIILKE-------------------------SRPDSPVSPLRLKI
440685156|Anabaena cylindrica PCC 7122                 -------------------------------MSKVFVLDTNFIPLNPIHSA-QARQLLRNKKAAIFRQ----FPFTIILKE-------------------------SRPDSPVSSLRLKI
428315656|Oscillatoria nigro-viridis PCC 7112          -----------------------------MVLNNLGVQECQSRMMRKCHVR-----FVGERDTATYAL----LPDLKV-----------------------------EVTAPQKPITLKL
428204926|Chroococcidiopsis thermalis PCC 7203         ------------------------------MSNQVLVLDTNRKPLTPCTPG-TARSLLKAGKAAVFRR----FPFSIILKK--------------------------EVKDVPQPLTLKL
218440244|Cyanothece sp. PCC 7424                      ------------------------------MSNFVFVLDTNKRRLNPVHPR-EARLMLNQGKAAVFRK----YPFTIILKE-------------------------SCPDVLVQDLELLL
428775092|Halothece sp. PCC 7418                       ------------------------------MSNFVFVLDTNKQPLDPVHPG-KARRLLSRGNAAVFRR----YPFTIILKE-------------------------ACPDIPMQDLELKL
414075475|Anabaena sp. 90                              ------------------------------MSNYALVLDTNKQPLSPCHPS-VARKLLDSGKAALFRM----YPFTIILNK-------------------------SCSDIVNQAIELKI
428305120|Crinalium epipsammum PCC 9333                ------------------------------MSNFILVLDTNRKPLNPCTPG-IARGLLKSGKARVFRR----FPFVIILNK--------------------------AVLSSPAQMQLKL
427713881|Synechococcus sp. PCC 6312                   ------------------------------MSNFIFVLDTQKRPLDPVHPG-KARHLLNQGKAAVYRR----FPFTLILKE-------------------------AHPDVPVQDLELKL

S. pyogenes Cas9                                                                                                                                                                   (7) KKYSIGL
S. pyogenes Cas9 domains                                                                                                                                                                rrrrrrr
A. naeslundii Cas9                                                                                                                                                                 (12)   LRVGI
A. naeslundii Cas9 domains                                                                                                                                                                rrrrr
```

Supplementary Figure S8. Distant Cas9 homologs. Compare with Figure 2. (**A**) Multiple sequence alignment of Cas9 distant homologs (See Supplementary Material and Methods). Sequences are denoted by Genbank identifiers and the complete organism name. Mapped sequences and domains of *A. naeslundii* and *S. pyogenes* Cas9 are shown in the last rows of the alignment with catalytic residues highlighted in yellow. For Cas9 domain description, see Supplementary Figure S4. Mapping was done manually according to HHPRED hits for Cas9 homologs shown in (**B**).
(**B**) HHPRED hits for Cas9 distant homologs. Top ten hits and selected hits to Cas9, RuvC and transposase structures are shown.

```
312126587|Caldicellulosiruptor hydrothermalis 108     DPGSKMEGYSVKS---------KA-----------------HTYLNIQ-AYA--------VTHVKSAVET--------------RRNMRRARRY------RK---------TRR------
434408285|Stanieria cyanosphaera PCC 7437             DPGSKREGYTVKS---------KA-----------------HQYLNIQ-ATT--------VDWVNEHLKT--------------RREMRRARRF------RK---------TPC------
414076241|Anabaena sp. 90                             DPGKLFSGIAVQS---------KK-----------------FTLQMLHLVLP--------FKTVKDRMDQ--------------RSMMRRNRRG--RRINRK---------LSF----NK
428299778|Calothrix sp. PCC 6303                      DPGKMFSGVAVQS---------KK-----------------YTLQMLHLVLP--------FKTVKNRMEQ--------------RSIMRRGRRG--RRINRK---------LSF----KK
 17233043|Nostoc sp. PCC 7120                         DPGKKFSGIGVQS---------AK-----------------FTLYTAHLILP--------FQTVRDRMDA--------------RRLMRRGRRG--RRIDRK---------VNF----LK
166363864|Microcystis aeruginosa NIES-843             DPGKLFSGIGVQS---------SL-----------------FTLWKAHLELP--------FKRVKERLDN--------------RRLMRRGRRK--RRINRQ---------LSF----NL
166367659|Microcystis aeruginosa NIES-843             DPGQLFSGIGVQS---------SL-----------------FTLWKAHLELP--------FKRVKERMDN--------------RRLMRRGRRK--RRINRQ---------LSF----NL
307150343|Cyanothece sp. PCC 7822                     DPGKLFSGIGVQS---------SS-----------------YSLWTGHLELP--------FKTVKKRMED--------------RRMMRRARRG--RRINRN---------LPF----EL
307153352|Cyanothece sp. PCC 7822                     DPGKLFSGIGVQS---------SS-----------------YSLWAGHLELP--------FKTIKKRMED--------------RRMMRRTRRG--RRINRH---------LPF----EL
428306661|Crinalium epipsammum PCC 9333               DPGKLFTGIGVQS---------SR-----------------FTLWTAHLELP--------FKRVKERMDN--------------RRLMRRGRRG--RRINRQ---------LPF----NC
307150912|Cyanothece sp. PCC 7822                     DPGKLFSGIGVQS---------SH-----------------FSLWKAHLELP--------FKRVRERMDN--------------RRLMRRSRRG--RRINGK---------ISF----ES
428775289|Halothece sp. PCC 7418                      QAEFNYSGIGVQS---------SK-----------------TTLFTAHLELP--------FQRVKDRMET--------------RAMMRRNRRG--RRINRK---------LPF----DQ
428779516|Dactylococcopsis salina PCC 8305            DPGKHYSGIGVQS---------SF-----------------MTLFTAHLELP--------FQRVKDRMET--------------RAMMRRNRRG--RRIDRK---------LPF----NQ
428778599|Dactylococcopsis salina PCC 8305            DPGKHYSGIGVQS---------SK-----------------ATLFTAHLELP--------FQRVKNRMET--------------RAMMRRSRRG--RRINRK---------LPF----DQ
428781696|Dactylococcopsis salina PCC 8305            DPGKHYSGIGVQS---------SQ-----------------ATLFTAHLELP--------FQRVKDRMET--------------RAMMRRGRRG--RRINRK---------LPF----DQ
428215045|Oscillatoria acuminata PCC 6304             DPGKLYSGIGVQS---------AK-----------------ATLFMAHLILP--------FQTVKDRMEQ--------------RRMMRRNRRG--RRINRQ---------LEY----SK
428313425|Microcoleus sp. PCC 7113                    DPGKHYSGIGVQS---------SK-----------------VTLFMAHLVLP--------FQSIKDRMEQ--------------RRMMRRGRRG--RRINRK---------APY----NQ
158340109|Acaryochloris marina MBIC11017              DPGKSYSGIAVQS---------AH-----------------FTLYTAHLVLP--------FQRIKERLGAAVIKNGKVIKNVRGRALQRRVRRG--RRINRK---------VPF----HL
428775002|Halothece sp. PCC 7418                      DPGKSYSGVGVQS---------AK-----------------CTLLQLHLILP--------FGRVRSRMDQ--------------RRLLRRSRRS--RRINRD---------VPF----KL
307150700|Cyanothece sp. PCC 7822                     DPGKFFSGIGVQS---------SL-----------------FTLWKAHLELP--------FQRVRERIDN--------------RGLMRRSRRG--RRINRK---------ISF----ES

312126663|Caldicellulosiruptor hydrothermalis 108     DPGYHCIGFVVC-EL------KGGKL---------------IVYCTGILVSR--------IPEIKELITE--------------RKAHRRNRRYYSRCKKKRLSARQNRVLLKF------
312621424|Caldicellulosiruptor kronotskyensis 2002    DPGYHHIGFAVC-EL------RWGVL---------------IVYCIGVLETR--------IPEIKDLMTK--------------RRGYRRNRRYHSRCRKKRMSKRQSRVLTKF------
222530402|Caldicellulosiruptor bescii DSM 6725        DPGYHHIGFAVC-EL------RWGVL---------------IVYCIGVLETR--------IPEIKDLMTK--------------RRGYRRNRRYHSRCRKKRMSKRHSRVLTKF------
222530431|Caldicellulosiruptor bescii DSM 6725        DPGYHHIGFAVC-EL------RWGVL---------------IVYCIGVLETR--------IPEIKDLMTK--------------RRGYRRNRRYHSRCRKKRMSKRHSRVLTKF------

307591386|Cyanothece sp. PCC 7822                     DPGSKTTGIALVQQD--------------------------KLIWGAELIHR--------GQQIKDNLLT--------------RRQIRRSRRN------RK---------TRY------
302391131|Acetohalobium arabaticum DSM 5501           DDGAKYVGVALVQDC------KTKN----------------KVLFKAKMKQR---------QDVSAKMEE--------------RNDYRGYRRS------HK----------RY------
435854579|Halobacteroides halobius DSM 5150           DDGYDKVGFALIQKC------QSKN----------------KVLFKATMKQR---------QEVSKLMTQ--------------RREYRRYRRQ------HK----------RY------
435852864|Halobacteroides halobius DSM 5150           DDGTKKVGFALVQKC------QTKN----------------KVLFKAVMEQR---------QDVSKKMEE--------------RRGYRRYRRS------HK----------RY------
435854272|Halobacteroides halobius DSM 5150           DDGAKKVGFALVQKC------QTKN----------------KVLFKAVMEQR---------QDASKKMEE--------------RRGYRRYRRS------HK----------RY------
313123202|Lactobacillus delbrueckii  ND02             DTGQRHIGFAIVSQN--------------------------KVLHQSEVELR---------QDVHKNLYT--------------RKIYRRSKRN------RK---------TRY------
77362205|Pseudoalteromonas haloplanktis TAC125        DTGSKHVGCAAIANG--------------------------KVIYQSEVALR---------TDVSKKMQQ--------------RAMYRRTRRN------RK---------LRY------
298675374|Methanohalobium evestigatum Z-7303          DSGSKYIGCAAVAND--------------------------EVMYQSEIKLR---------DDVSKKMKD--------------RADHRMNRRY------RK---------TRY------
298674707|Methanohalobium evestigatum Z-7303          DSGSKHIGCAAVAND--------------------------EVLYQSEVELR---------DDVSKKVQQ--------------RASYRRNRRS------RK---------TRY------
298675090|Methanohalobium evestigatum Z-7303          DSGSKHIGCAAVVND--------------------------EVLYQSEVELR---------DDVSRKMKD--------------RADYRRNRRS------RK---------TRY------
169834784|Clostridium botulinum A3 str. Loch Maree    DSGYLNIGFSAITKE--------------------------NELISGEVKLL---------KGMSERLKE--------------KAMYRRQRRS-------R---------LRY------
229587454|Clostridium botulinum Ba4 str. 657          DSGYLNIGFSAITKE--------------------------KELISGEVKLL---------KCMSERLKE--------------KAMYRRQRRS-------R---------LRY------

297585104|Bacillus selenitireducens MLS10             DAGTKHIGFSAPTEK--------------------------DVLFEGDVQLR---------TDIQDKLAT--------------RRQYRRSRRS------RK---------TRY------
188587532|Natranaerobius thermophilus JW-NM-WN-LF     DAGTKHVGLSATTGN--------------------------EVLFEGEAQLR---------TNIQELLST--------------RREARRARRN------RK---------TRY------
312793946|Caldicellulosiruptor kristjanssonii 177R1B  DIGSKVVGVSAITEK--------------------------QELFSSEVELR---------QDIRKLLLE--------------RREHRRFRRY------RK---------RRY------
344996431|Caldicellulosiruptor lactoaceticus 6A       DIGSKVVGVSAITEK--------------------------QELFSSEVELR---------QDIRKLLLE--------------RREHRRFRRY------RK---------RRY------
217077332|Thermosipho africanus TCF52B                DIGSKTVGISAITEK--------------------------QELLSMEVELR---------QDIKRLLLE--------------RREYRRNRRY------RK---------RRY------
375009168|Geobacillus thermoleovorans CCB US3 UF5     DMGTRHVGISATTKK--------------------------DVLFEAEAQLR---------TDIVELLAI--------------RRQFRRSRRN------RK---------TRY------
407477159|Exiguobacterium antarcticum B7              DAGTKHIGFSATTST--------------------------KVLLEGEVQLR---------TDIQDLLAT--------------RRAMRSARRS------RK---------TRY------
172057416|Exiguobacterium sibiricum 255-15            DAGTKHIGFSATTST--------------------------KVLLEGEVQLR---------TDIQDLLAT--------------RRAMRSARRS------RK---------TRY------
169823684|Finegoldia magna ATCC 29328                 DSGSKYIGLSASTEN--------------------------KEIFSAEYELR---------NDIVKKLSS--------------RRECRRTRRN-------R---------LRH------
160902522|Petrotoga mobilis SJ95                      DAGSKSIGLSATTEN--------------------------KELFAAEVEIR---------DDIPKLISQ--------------KRQYRKGRRF------RK---------TRY------
160902616|Petrotoga mobilis SJ95                      DAGSKTVGLSATTKK--------------------------KELFASEVEIR---------DDITKLISQ--------------KRQYRRDRRL------RK---------TRY------

333978211|Desulfotomaculum kuznetsovii DSM 6115       DDGSKWVGIALVNEH------TG------------------EVVFRGTLIQR--------GDVVR-LLTL--------------RREYRRNRRY------RL---------VRH------
260892456|Ammonifex degensii KC4                      DDGSRYVGIALVNEH------TG------------------EVVFRGVLVQR--------GDVVR-LLTL--------------RREYRRGRRY------RI---------VRH------
260893631|Ammonifex degensii KC4                      DDGSRYVGIALVNEH------TG------------------EVVFRGVLVQR--------GDVVR-LLTL--------------RREYRRGRRY------RI---------VRH------
302391857|Acetohalobium arabaticum DSM 5501           DPGSKTTGLAINNGS--------------------------QVAFLANLHHR--------ADTIKQNLQT--------------RREHRRFRRS-------Q---------LKY------
428778981|Dactylococcopsis salina PCC 8305            DPGATWTGFAIQCGE--------------------------EIIFRMELKHR--------GGLIKKSLEQ--------------RADFRKGRRS------RN---------LRY------
260892735|Ammonifex degensii KC4                      DPGYEVTGISVVRVA------GGKE----------------VVVYFAEVHHR---------TDVPGKLLS--------------RRQARRSRRS------RK---------TRY------
260893993|Ammonifex degensii KC4                      DPGYEVTGISVVRVA------GGKE----------------VVVFFAEVHHR---------TDVPGKLLS--------------RRQARRSRRS------RK---------TRY------
433655673|Thermoanaerobacterium
thermosaccharolyticum M0795                           DPGSKVTGIAILQGE--------------------------KVLLLVELYHK---------QGIKKSLDD--------------RRNHRRFRRN-------K---------LRY------
332982963|Mahella australiensis 50-1 BON              DPGSKATGVAVLADD------T-------------------KAIWLGEIHHK---------SNIKKKLKD--------------RRALRSSRRN------RK---------TRY------
332983053|Mahella australiensis 50-1 BON              DPGSKATGAAVLADD------T-------------------KAIWLGEIHHK---------SNIKKKLKD--------------RRALRSSRRN------RK---------TRY------
222530385|Caldicellulosiruptor bescii DSM 6725        DYGSRCTGIVILKNN------C-------------------EVVFMMKLYHR---------TEIKENMDR--------------RRSLRRSRRN------RK---------TRY------
298674346|Methanohalobium evestigatum Z-7303          DYGSRYTGLAILKNN------K-------------------DVIWLDQVHHR---------TDIKKKLDD--------------RRNFRRRRRS------QN---------LRY------
298675173|Methanohalobium evestigatum Z-7303          DYGSKHTGLAILKNN------K-------------------DVIWLAQIHHR---------TDIKKKLDD--------------RRSFRRRRRS------KN---------LRY------
292490518|Nitrosococcus halophilus Nc 4               DPGSKTTGISLNAQF------NSGT----------------VCIFAMQLEHR--------GRQIKESLDS--------------RRGVRRSRRN------RN---------TRY------
428778759|Dactylococcopsis salina PCC 8305            DPGSKVTGFALVTDQ------G-------------------EVVWRMQLEHR--------GGIIKKRLEQ--------------RRALRRKRRS-------R---------LRY------
300114477|Nitrosococcus watsonii C-113                DPGSKTTGLALVGKF------QAQGR---------------VVLWAGNLTHR--------GQAIRDRLER--------------RRALRRGRRY------RN---------TRY------
292491781|Nitrosococcus halophilus Nc 4               DPGSKITGLALVGNF------PQQGR---------------VVLWAANLTHR--------GYAIRERLAS--------------RRSLRRGRRG------RK---------TRY------
292491869|Nitrosococcus halophilus Nc 4               DPGSKITGLALVGNF------PQQGR---------------VVLWAANLTHR--------GYAIRERLAS--------------RRSLRRGRRG------RK---------TRY------
292491893|Nitrosococcus halophilus Nc 4               DPGSKITGLALVGNF------PQQGR---------------VVLWAANLTHR--------GYAIRERLAS--------------RRSLRRGRRG------RK---------TRY------
292491036|Nitrosococcus halophilus Nc 4               DPGSKTTGMALVGHF------SVQGR---------------VVLWAANLMHR--------GQSVRDRLER--------------RRLLRRGRRG------RK---------TRY------
292490884|Nitrosococcus halophilus Nc 4               DPGSKTTGLALVGNF------PQQGR---------------VVLWAANLEHR--------GQSVRDRLEK--------------RRSLRRGRRG------RK---------TRY------
292491492|Nitrosococcus halophilus Nc 4               DPGSKTTGLALVGNF------PQQGR---------------VVLWAANLEHR--------GQSVRDRLEK--------------RRSLRRGRRG------RK---------TRY------
297563003|Nocardiopsis dassonvillei DSM 43111         DPGSKRAGTAVFTDR------RGER----------------RGRCAIRPDHR--------GGQIAKRMRQ--------------RAAYRKGRRS------RN---------LRY------
374989417|Streptomyces bingchenggensis BCW-1          DPGSKHTGIATFTAQ------AGER----------------RGRYAVHLDHR--------GAVIRKKRGQ--------------RAAYRRRRRT------SN---------LRY------
374989654|Streptomyces bingchenggensis BCW-1          DPGSRHTGIAAFTAR------AGER----------------RGRYAVQLDHR--------GAVIRKRMGQ--------------RAAYRRRRRT------SH---------LRY------
271963428|Streptosporangium roseum DSM 43021          DPGSKHTGIAVFTEH------GGSR----------------TGRYALQLDHR--------GAAIRDKLAS--------------RAALRRGRRS------RN---------LRY------
271964894|Streptosporangium roseum DSM 43021          DPGSKHTGIAAFSER------GGSR----------------IGLYALQLDHR--------GGQIRDKLAS--------------RAALRRGRRS------RN---------LRY------
258511478|Alicyclobacillus acidocaldarius DSM 446     APGSKTTGVALVADF------QRGK----------------TVVWAAEIQHR--------GDQIRQALLR--------------RRMLRRGRRY------RK---------TRY------
160902264|Petrotoga mobilis SJ95                      DPGSKITGVTLVGDF------KNGK----------------KVIWGAEIHHR--------GQSIKKALDT--------------RRGVRRSRRN------RK---------IRY------
121604978|Polaromonas naphthalenivorans CJ2           DPGSKTTGLAVCRVA------ETIDVTTGEIQPTM------HIQFLMELMHR--------GATIKKSLGA--------------RSAMRRRRRG-------K---------LRY------
347754630|Chloracidobacterium thermophilum B          DPGSRATGVALVRES------QEVGADTGEVQRSA------HVLFLAELVHR--------GHAIRAALRQ--------------RAAFRRRRRC------AN---------LRH------
218667603|Acidithiobacillus ferrooxidans ATCC 23270   DPGSKTTGLALVREK------ETVDVSTGEVLRAI------TILMLLELKHR--------GQAIRDALAS--------------RGAFRRRRRG-------Q---------LRY------
218667415|Acidithiobacillus ferrooxidans ATCC 23270   DPGSKTTGMALVREQ------ESVDEDTGEIQRKA------IVLMLLELQHR--------GYAIRDALTQ--------------RRAFRRRRRG-------N---------LRY------
344201059|Acidithiobacillus ferrivorans SS3           APGSKTTGLALVRDA------ETVDAITGKVTRTV------VVLMLLELQHR--------GHAIRDALTQ--------------RRAFRRRRRG-------K---------LRY------
344201127|Acidithiobacillus ferrivorans SS3           APGSKTTGLALVRDA------ETVDAITGEVTLTV------VVLMLLELKHR--------GHAIRDALTQ--------------RGGHRRFRRS-------K---------LRY------
288940385|Allochromatium vinosum DSM 180              DPGSRVTGLALVRES------ETCDADTGAVERLE------HGLWFGELAHR--------GQAIRETLGQ--------------RRHLRRARRS------RK---------TRY------
288941921|Allochromatium vinosum DSM 180              DPGSRVTGLALVRES------ETGDADTGAVERLE------HGLWFGELVHR--------GQAIREALTL--------------RRHYRRARRS------RK---------THY------
288940786|Allochromatium vinosum DSM 180              DPGSRVTGLAIVRES------ETCDPETGAIERLE------HGLWFGELAHR--------GQAIREALTL--------------RRHYRRARRS------RK---------TRY------
288941171|Allochromatium vinosum DSM 180              DPGSRVTGLAIVRES------ETGDADTGAVERLE------HGLWLGELAHR--------GQAIREALTL--------------RRHYRRARRS------RK---------TRY------
288941306|Allochromatium vinosum DSM 180              DPGSRVTGLAIVRES------ETGDADTGAVERLE------HGLWLGELAHR--------GQAIREALTL--------------RRHYRRARRS------RK---------TRY------
121582913|Polaromonas naphthalenivorans CJ2           DPGSKTTGIALVRET------AE--------KNGV------AVINLFELMHR--------GKQISEALTA--------------RRGHRRMRRS-------H---------LRY------
121604604|Polaromonas naphthalenivorans CJ2           DPGSKTTGIALVREA------ERGAV---------------AVLNLFELAHR--------GRQISKALTA--------------RRGHRRLRRG-------K---------LRY------
91791204|Polaromonas sp. JS666                        DPGSKTTGVALVRDV------ESVDASTGEIHSGV------AVINLLDLVHR--------GRQISEALTA--------------RRQMRRRRRS-------N---------LRC------
91791283|Polaromonas sp. JS666                        DPGSKTTGVALVRDV------ESIEVATGEIHCGV------AVLNLMELVHR--------GRHISEALTA--------------RRQMRRRRRS-------N---------LRY------
121582916|Polaromonas naphthalenivorans CJ2           DPGSRATGLALVRDI------ETIEPAMGEVTRGA------AVVSLLELEHR--------GRQISEALTA--------------RRQMRRRRRN-------Q---------LRY------
121582983|Polaromonas naphthalenivorans CJ2           DPGSRVTGLAVVREK------PNTPGAKNAPAFEI------AVINLFELVHR--------GRQISEALTA--------------RRQMRRRRRG-------N---------LRY------
428304158|Crinalium epipsammum PCC 9333               DPGSKVTGIAVVTEE------TG------------------RVTSALLISHR--------GQQIKNALES--------------RRALRRGRRN------RK---------TRY------
428204921|Chroococcidiopsis thermalis PCC 7203        DPGSKVTGIAIFDGE--------------------------NVIWVAELEHR--------GGQIKNALLS--------------RRSLRRSRRN------RK---------TRY------
307153539|Cyanothece sp. PCC 7822                     DPGAKTTGLSLVNDQ------TG------------------EVIWAGELTHR--------GFAIRESLTS--------------RRQLRRSRRG------RK---------TRY------
428216193|Oscillatoria acuminata PCC 6304             DPGSQITGLAICQEN--------------------------RVIWAAELVHR--------GSAIKDALLS--------------RRQLRRGRRN------RK---------TRY------
414079584|Anabaena sp. 90                             DPGSKTTGFALVQNN--------------------------KVIWGMELQHR--------GLAIKESLET--------------RKGVRRGRRS------RH---------TRY------
172055147|Cyanothece sp. ATCC 51142                   DPGAKVTGIALIQGD--------------------------RMIFGAELEHR--------GFRIKDSLES--------------RRQLRRGRRG------RK---------TRY------
428316510|Oscillatoria nigro-viridis PCC 7112         DPGSKTTGLAIVQGN--------------------------QVIWGAELTHR--------GQQIKNDLES--------------RRAIRRNRRN------RK---------TRY------
428314951|Oscillatoria nigro-viridis PCC 7112         DPGAKVSGLAIVRDD--------------------------KVIWGAELTHR--------GQQIKHDLAS--------------RRALRSHRRN------RK---------TRY------
307591386|Cyanothece sp. PCC 7822                     DPGSKTTGIALVQQD--------------------------KLIWGAELIHR--------GQQIKDNLLT--------------RRQIRRSRRN------RK---------TRY------
75812341|Anabaena variabilis ATCC 29413               DPGAKTTGIAIINDA------ND------------------EVVFAAELKHR--------GFAIRDAITS--------------RRQLRRSRRN------RK---------TRY------
17233293|Nostoc sp. PCC 7120                          DPGAKYTGIALVNDV------NG------------------EVVFAAELKHR--------GFAIRDALTS--------------RRQLRRSRRS------RK---------TRY------
75812782|Anabaena variabilis ATCC 29413               DPGAKITGIALVNDS------TG------------------EVVFAADLKHR--------GFAIRDALIS--------------RRQLRRTRRN------RK---------TRY------
434407609|Cylindrospermum stagnale PCC 7417           DPGAKHTGIALVNDS------TG------------------EVVFAADLEHR--------GFAIRDSLTT--------------RRQLRRSRRN------RK---------TRY------
414078405|Anabaena sp. 90                             DPGAKFTGMALVNDS------TG------------------EVVFAAELKHR--------GFVIRDALTS--------------RRQLRRSRRA------RK---------TRY------
440685156|Anabaena cylindrica PCC 7122                DPGAKFTGMALVNDS------TG------------------EVVFAAELKHR--------GFAIRDALTS--------------RRQLRRSRRA------RK---------TRY------
428315656|Oscillatoria nigro-viridis PCC 7112         DPGSTTTGIALVQGE--------------------------EVIFGAQLAHR--------GQSIKASLES--------------RRSLRRGRRN------RH---------TRY------
428204926|Chroococcidiopsis thermalis PCC 7203        DPGSKTTGIALLKGS--------------------------QLLWVAELIHR--------GAAIKDAILS--------------RRQLRRGRRN------RK---------TRY------
218440244|Cyanothece sp. PCC 7424                     DPGSKVTGIAIKQGN--------------------------KVLFGAEVQHR--------GQQIRDALLS--------------RRQLRRGRRN------RK---------TRY------
428775092|Halothece sp. PCC 7418                      DPGSRVTGIAIVQGT--------------------------KVIFGAELEHR--------GQQIKDALLS--------------RRQLRRSRRN------RK---------TRY------
414075475|Anabaena sp. 90                             DPGSKTTGIALVQGE--------------------------KIIFGAELSHR--------GSAIKASLES--------------RRSLRRGRRN------RH---------TRY------
428305120|Crinalium epipsammum PCC 9333               DPGSKTTGIALLQGE--------------------------QVVFGAELTHR--------GQQIKNALES--------------RRALRRGRRS------RK---------TRY------
427713881|Synechococcus sp. PCC 6312                  DPGSKVTGIAIKQGN--------------------------KIIFGAELQHR--------GQQIKDALLS--------------RRQLRRGRRN------RK---------TRY------

S. pyogenes Cas9                                      DIGTNSVGWAVITDE------YKVPSKKFKVLGNTDRHSIKKNLIGALLFDS--------GET----AEA--------------TRLKRTARRR------YT---------RR-------
S. pyogenes Cas9 domains                              rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr|RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
A. naeslundii Cas9                                    DVGTHSVGLATLRVDDHGTPI--------------------ELLSALSHIHDSGVGKEGKKDHDTRKKLS--------------GIARRARRLL------HH---------RRT------
A. naeslundii Cas9 domains                            rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr|RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR
```

```
312126587|Caldicellulosiruptor hydrothermalis 108        ---RPVRFNNR-KNKD---------------------------------LPPSTKARWQWKLRIVKWLTRIF---PITD----FVVEDIKAVTKLGQ-------------------KRW-
434408285|Stanieria cyanosphaera PCC 7437                ---RQPRANRLINKFC---------------------------------LPPSTKSRWQWKLRLCSWLSKMF---PITT----FVIEDIKAKT---------------------WGGKW-
414076241|Anabaena sp. 90                                RNHRQARFDNR-RGSK---------------------------------LPPSIRANKDLEYRVISLLLQLY---PIKT----VVIEEV--------------------------EARG-
428299778|Calothrix sp. PCC 6303                         RSHRQARFDNR-KNKK---------------------------------LPPSIRANKDLEFRVINLLCELY---PVST----IVIEEV--------------------------EAKG-
 17233043|Nostoc sp. PCC 7120                            RAHRQKRFDNR-RQGK---------------------------------LPPSIRANRQLELRIVSELCKIY---PISE----IRYEYVRADVDLTS---------------GRKKARS-
166363864|Microcystis aeruginosa NIES-843                RAHRQKRFSNR-KQGK---------------------------------LAPSIRANRQLELRVVSELTKIY---PITD----IYFEYVKADVDLTS---------------GRKGAKS-
166367659|Microcystis aeruginosa NIES-843                RAHRQQRFSNR-KQGK---------------------------------LAPSIRANRQLELRVVSELTKIY---PITD----IYFEYVKADVDLTS---------------GRKGAKS-
307150343|Cyanothece sp. PCC 7822                        RAHREKRFNNR-KQGK---------------------------------LPPSIRANRQLELRVIKELCKVF---PITD----IYFEYVKVDVDLTS---------------GRKKARS-
307153352|Cyanothece sp. PCC 7822                        RAHRQKRFNNR-KQGK-------------------------LPGVRNEFRTPSIRANRQLELRVVKELCKVF---PITD----IYFEYVKADVDLTS---------------GRKKACS-
428306661|Crinalium epipsammum PCC 9333                  RAHRQKRFSNR-KRSK---------------------------------LAPSIKANRQLEIRVVSELCKIY---PITD----IYFEYIKADVDLTS---------------GRKAAKS-
307150912|Cyanothece sp. PCC 7822                        RAHRQKRFSNR-RQGK---------------------------------LSPSIKANRQLELRVVSELSKIY---PITD----IYFEYVKADVDLTS---------------GRKKARA-
428775289|Halothece sp. PCC 7418                         RSHRQKRFSNR-RGNK---------------------------------LPPSIKANRQLEYRVVKELSQLF---PITS----IIYEIV--------------------------KAKG-
428779516|Dactylococcopsis salina PCC 8305               RSHRQKRFNNR-RCHK---------------------------------LPPSIKANRQLEYRIVKELSQLF---PISS----IIYEIV--------------------------KAKG-
428778599|Dactylococcopsis salina PCC 8305               RAHRQKRFNNR-CGHK---------------------------------LPPSIKANRQLEYRVVKELSQLF---PISS----IIYEIV--------------------------KARG-
428781696|Dactylococcopsis salina PCC 8305               RAHRQKRFDNR-RGHK---------------------------------LPPSIKANRQLEYRVVKELSQLF---PISS----IIYEIL--------------------------KAKG-
428215045|Oscillatoria acuminata PCC 6304                RAHRQQRFDNR-RRNK---------------------------------LPPSVRANRQFELRVVTELCRLF---PVSK----IVYESV--------------------------KARG-
428313425|Microcoleus sp. PCC 7113                       RAHRQKRFANR-RGNK---------------------------------LPPSVRANRQLELRVATELCQIF---PISQ----VVYEYIKADVDLTS---------------GRKKARS-
158340109|Acaryochloris marina MBIC11017                 RAHRQKRFNNR-RQPK---------------------------------LPPSIKASRQMELRVIKELSKIF---PITS----IVYEVVKADVDRTS---------------GRKGAKS-
428775002|Halothece sp. PCC 7418                         RNHRQKRFDNR-RGKK---------------------------------LAPSIRASRQLEIRVVKEIASIY---PITA----IGYEKVRADVDLTS---------------GRKGARS-
307150700|Cyanothece sp. PCC 7822                        RAHRQKRFSNR-RQGK---------------------------------LAPSIKANRQLELRIISELSKIY---PITD----IYFEYVKADVDLTS---------------GRKKARA-

312126663|Caldicellulosiruptor hydrothermalis 108        ---KVPRTVR--DRDR---------------------------------LNATLRHGIETHLNLYKKLLKFF---PFPAEQVVFVMEDNIFDVRTMT----------------WGKTDG-
312621424|Caldicellulosiruptor kronotskyensis 2002       ---KAPRNVRTKDRTN-----------------------------------ATLRHGIETHLNLYKKLLKFF---PFPVEQFVFVMEDNIFDVRTMT----------------WGKTDG-
222530402|Caldicellulosiruptor bescii DSM 6725           ---KAPRNVRTKDRTN-----------------------------------ATLRHGIETHLNLYKKLLKFF---PFPAEQVVFVMEDNIFDVRTMT----------------WGKTYG-
222530431|Caldicellulosiruptor bescii DSM 6725           ---KAPRNVRTKDRTN-----------------------------------ATLRHGIETHLNLYKKLLKFF---PFPAEQVVFVMEDNIFDVRTMT----------------WGKTYG-

307591386|Cyanothece sp. PCC 7822                        ---RQARFLNR-TRLK---------------------------GW----LPPSLRHRVETTMTWVKRICKFV---HVTN----ISVELVKFDTQALD----------------NPEISG-
302391131|Acetohalobium arabaticum DSM 5501              ---RPKRFDNRFASKR--------------------------KGR----LPPSIKQRKQAILRVVKHLSKHI---RLDK----IILEDVSIDIRKLT---------------EGEELYS-
435854579|Halobacteroides halobius DSM 5150              ---RAKRFDNRSSSKR--------------------------KGR----IAPSIKQKKQAILRVVNYLNKFI---RIDK----IVLEDVAIDIRKAV---------------EGKELYS-
435852864|Halobacteroides halobius DSM 5150              ---RPARFDNRSSSKR--------------------------KGR----IPPSILQKKQAILRVVNKLKKYI---RIDK----IVLEDVSIDIRKLT---------------EGRELYN-
435854272|Halobacteroides halobius DSM 5150              ---RPARFDNRSSSKR--------------------------KGR----IPPSILQKKQAILRVVNKLKKYI---RIDK----IVLEDVSIDIRKLT---------------EGRELYN-
313123202|Lactobacillus delbrueckii  ND02                ---RQARFLNRVHGKR-------------------------DGLW----LPPSVKSKVSHNIAWIKRYLAVL---PNPD----LHIEVGKFDMAKMV----------------NPDISG-
77362205|Pseudoalteromonas haloplanktis TAC125           ---RPARWLNR-ASMR-----------KE--------------GR----LAPSLKSKLDSHLREKKQVEAIL---PISH----WKVETASFDIHKIT----------------NPIVTS-
298675374|Methanohalobium evestigatum Z-7303             ---RKPRFNNR-GNSK-----------KE--------------NR----IAPSIKSKVQSHERERRFVESIL---PVKK----WIVETTSFDIHKIT----------------NPNVSG-
298674707|Methanohalobium evestigatum Z-7303             ---RKPRFNNR-GNSK-----------KD--------------GR----IAPSIKSKVQSHERERRFVESIL---PVKK----WIVETASFDIHKIT----------------NPNVSG-
298675090|Methanohalobium evestigatum Z-7303             ---RKPRFDNR-GNSK-----------KE--------------NR----IAPSIKNKVQSHDRERRFVESIL---PVKK----WIVETASFDIHKIT----------------NPNVSG-
169834784|Clostridium botulinum A3 str. Loch Maree       ---RKPRFNNR-KRSK---------------------------EW----LAPSIQHKYDSHIRFIDKLKKLL---PITN----IIIEVANFDTHKLK----------------NPSIEG-
229587454|Clostridium botulinum Ba4 str. 657             ---RKPRFDNR-KRSE---------------------------NW----LAPSIQHKYDSHIRFIDKLKKLL---PITN----IIIEVANFDTHKLK----------------NPSIEG-

297585104|Bacillus selenitireducens MLS10                ---RQPRFDNR-KKEK---------------------------GW----LAPSVQHKVDSHIRFVEKVHALL---PITT----ITVEVAQFDIQKIK----------------KPEIEG-
188587532|Natranaerobius thermophilus JW-NM-WN-LF        ---RKPRFNNR-KKPN---------------------------GW----LAPSIQNKVDTHLKLVDMVCSIL---PVSN----IKVEVAQFDIHKIK----------------NPEIEG-
312793946|Caldicellulosiruptor kristjanssonii 177R1B     ---RKPRFLNR-RRPE---------------------------GW----LAPSVQWKVDAHIRLVNFVAKIL---PVTK----VVVEVAPFDIHKIV----------------NPTVEG-
344996431|Caldicellulosiruptor lactoaceticus 6A          ---RKPRFLNR-RRPE---------------------------GW----LAPSVQWKVDAHIRLVNFVAKIL---PVTK----VVVEVAPFDIHKIV----------------NPTVEG-
217077332|Thermosipho africanus TCF52B                   ---RKPRFLNR-RRYE---------------------------GW----LAPSVQWKVDAHIRLVNLIAKIL---PVTK----IVIEIAPFDTHKIV----------------NPDVEG-
375009168|Geobacillus thermoleovorans CCB US3 UF5        ---REARFLNR-RKPE---------------------------GW----LPPSIQHKINSHIKLIDLVHNIL---PVTS----VTIEVAAFDTQKLK----------------NLNIRG-
407477159|Exiguobacterium antarcticum B7                ---RQARFLNR-KKPK---------------------------GW----LAPSIQHKVDAHLKLIRMIHRLL---PIKH----LTIEVAQFDIQKIK----------------NPDISG-
172057416|Exiguobacterium sibiricum 255-15              ---RQARFLNR-KKPN---------------------------GW----LAPSVQHKVDSHLKLIRMIHRLL---PIKH----LTIEVAQFDIQKIK----------------SPDISG-
169823684|Finegoldia magna ATCC 29328                   ---RKARFNNR-VSSK-----------KK--------------GW----LAPSIRHKVNSHLKIISDVCSII---PIKK----IIVEVASFDMQKIK----------------NPDISG-
160902522|Petrotoga mobilis SJ95                        ---RKPRFLNR-VHSR-----------KKQKQADATQPYGRQGNW----LAPSVEHKIKSHIKLIDLVHSIL---PITT----IVVEVASFDIQKIK----------------NPNIEG-
160902616|Petrotoga mobilis SJ95                        ---RKPRFLNR-VHSK-----------NK--------------GW----IAPSVEHKIKSHIKLIDLVYRLL---PITK----IVVEVASFDIQKIK----------------NPNIEG-

333978211|Desulfotomaculum kuznetsovii DSM 6115          ---RECRNHNR-KQLA---------------------------PF------PSIRQKKEAVCRVLTDLAKIA---PISG----VDVELVAAGVK-------------------NP-----
260892456|Ammonifex degensii KC4                         ---RPCRNRAR-QQVV---------------------------PF------PSVRQKKEAIYRVLADLAKIA---PISG----VDVELVSSGVK-------------------NP-----
260893631|Ammonifex degensii KC4                         ---RPCRNRAR-QQVV---------------------------PF------PSVRQKKEAIYRVLADLAKIA---PISG----VDVELVSAGVK-------------------NP-----
302391857|Acetohalobium arabaticum DSM 5501              ---REKRFDNR-KRGD---------------------------NW----LPPSVDSIVNNIKTWVQRLKKLC---PISK----VVVEMVRFDTQKMQ----------------NPEIEG-
428778981|Dactylococcopsis salina PCC 8305               ---RKKRF-NR-KKPE---------------------------GW----LAPSLRHRLQTVETWVKKFLKLC---PITC----IEIEQVRFDLQKLE----------------NPEISG-
260892735|Ammonifex degensii KC4                         ---REPRFDNR-RKPE---------------------------GW----LPPSAEARVSQVFSVVGKLVRWL---PVTE----IVVESAKFDTQKLQ----------------NPEVSG-
260893993|Ammonifex degensii KC4                         ---REPRFDNR-GRPE---------------------------GW----LPPGAVARVNQVLSVVGKLVRWL---PVTE----IVVESAKFDTQKLQ----------------NPEVSG-
433655673|Thermoanaerobacterium
thermosaccharolyticum M0795                              ---REPRFDNR-KREK---------------------------GW----LPLSLEARVQQVINTVKKLAKYI---PIDN----ISIEYVKFDTQLMQ----------------NPEISG-
332982963|Mahella australiensis 50-1 BON                 ---RQARFNNR-HPAKCVACGKNAKHGSRYCRPCQKARNYVDNGYREKRLPPSLQARVNQTINCVTKLCKLL---PITA----ISTEHVKFDTQLMQ----------------NPDIND-
332983053|Mahella australiensis 50-1 BON                 ---RQARFNNR-HPAKCVACGKNAKHGSRYCRPCQKARNYVDNGYREKRLPPSLQARVNQTINCVTKLCKLL---PITA----ISTEHVKFDTQLMQ----------------NPDIND-
222530385|Caldicellulosiruptor bescii DSM 6725           ---RKARFSNR-RRDE---------------------------NW----LPPTLLSRVRNIETWVKRLCKLC---PVTA----ISYENVKFDTQKLR----------------NPEISG-
298674346|Methanohalobium evestigatum Z-7303             ---RKPRFDNR-KKPE---------------------------GW----LPPSLQSRVDNIQTWITRLRELV---PINC----LSYENVNFDTQLMQ----------------NPEIKG-
298675173|Methanohalobium evestigatum Z-7303             ---RKPRFDNR-KKPK---------------------------GW----LPPSLQSRVDNIQTWVRKLMKLI---PINS----ISYENIKFDTQLMQ----------------NPEIKG-
292490518|Nitrosococcus halophilus Nc 4                  ---RQPRFLNR-SRPK---------------------------GW----LPPSLLSRVDNTETWAERLVRLA---PICS----ANVEIARFDLQLLE----------------NPDITG-
428778759|Dactylococcopsis salina PCC 8305               ---RQPRFLNR-KRQE---------------------------GW----LPPSLMHRVQTTETWAKRLIKFC---PVQE----IWIEKVKFDTQAME----------------NPEIDG-
300114477|Nitrosococcus watsonii C-113                   ---RAPRFLNR-TKPK---------------------------GW----LPPSLRSRVENVSTWLNRLLDRA---PISE----CHIETVRFDIQKIQ----------------NPEISG-
292491781|Nitrosococcus halophilus Nc 4                  ---RAPRFLNR-TKPK---------------------------GW----LPPSLNSRVENVSSWFNRLLDRV---PISE----CHIETVRFDTQKIQ----------------NPEING-
292491869|Nitrosococcus halophilus Nc 4                  ---RAPRFLNR-TKPK---------------------------GW----LPPSLNSRVENVSSWFNRLLDRV---PISE----CHIETVRFDTQKIQ----------------NPEING-
292491893|Nitrosococcus halophilus Nc 4                  ---RAPRFLNR-TKPK---------------------------GW----LPPSLNSRVENVSSWFNRLLDRV---PISE----CHIETVRFDTQKIQ----------------NPEING-
292491036|Nitrosococcus halophilus Nc 4                  ---RAPRFLNR-TRPQ---------------------------GW----LAPSLRSRVENVSTWFNRLLDRA---PISA----CHIETVRFDLQKIQ----------------NPEISG-
292490884|Nitrosococcus halophilus Nc 4                  ---RAPRFLNR-TRPA---------------------------GW----LPPSLRSRVENVSTWFNRLLDRA---PLSE----CHIETVRFDMQKIQ----------------NPEISG-
292491492|Nitrosococcus halophilus Nc 4                  ---RAPRFLNR-TRPA---------------------------GW----LPPSLRSRVENVSTWFNRLLDRA---PLSE----CHIETVRFDMQKIQ----------------NPEISG-
297563003|Nocardiopsis dassonvillei DSM 43111            ---RTARFDNR-TRPR---------------------------GW----PPPSLRHRVDTTLAWVGRPARWA---PVRA----VRVERVAFDVHALG---------------AGRPAEG-
374989417|Streptomyces bingchenggensis BCW-1             ---RAPRFNNR-TRPD---------------------------GW----LAPSLQHRVTTTVSWVRRLARWA---PVRA----VHVERVAFDTHAIS---------------AGRPLEG-
374989654|Streptomyces bingchenggensis BCW-1             ---RAPRFNNR-ARPD---------------------------GW----LAPSLRHRVATAVSWVTRLSRWA---PVRA----VHVERVAFDTHAIS---------------AGRPLEG-
271963428|Streptosporangium roseum DSM 43021             ---RAPRFNNR-TKPK---------------------------GW----LAPSLKHRMDTTMSWADRLRRWA---PVSA----VHVERVAFDTHALS---------------AGRPLEG-
271964894|Streptosporangium roseum DSM 43021             ---RAPRFNNR-TRPQ---------------------------GW----IAPSLRHRVDGTVSWVSRLSRWA---PVTA----VHVERVAFDTHLLS---------------AGRPLEG-
258511478|Alicyclobacillus acidocaldarius DSM 446        ---RKPRFDNR-RRLE---------------------------GW----LPPSLMSRVHNIETWVARLRRWA---PITH----LSMQLVSFDTQKLQ----------------YPEISG-
160902264|Petrotoga mobilis SJ95                        ---RIARFDNR-KRSK---------------------------GW----LPPSLISRVENILTWIKRIRRFS---PITG----ISLELVRFDTQKLQ----------------DPEING-
121604978|Polaromonas naphthalenivorans CJ2              ---RAPRFSNR-PKPQ---------------------------GW----LAPSLRHRVETTMTQVKRLCRLA---PITH----LAQELVRFDMQKMQ----------------NPEISG-
347754630|Chloracidobacterium thermophilum B             ---RQSRFDNR-TRRN---------------------------GW----LPPSLQHRVDTITSWVRRLCRWA---PVSS----LSQELVRFDTQALQ----------------NPEISG-
218667603|Acidithiobacillus ferrooxidans ATCC 23270      ---RPARFDNR-TKSE---------------------------DW----LAPSLQHRVDTVLAWVNRLMRLA---PVTA----LSQELVRFDNQTMQ----------------NPEISG-
218667415|Acidithiobacillus ferrooxidans ATCC 23270      ---RPARFDNR-AKPE---------------------------GR----LAPSLQHRVDTTMAWVQRLLRWA---PVSA----LSTMLHRFDTQALQ----------------NPEISG-
344201059|Acidithiobacillus ferrivorans SS3             ---RPARFDNR-TKPA---------------------------GW----LAPSLRHRVDAAIAWIARLRRLA---PVSA----LSTMLHRFDTQLLQ----------------NPEISG-
344201127|Acidithiobacillus ferrivorans SS3             ---RPARFNNR-TKPK---------------------------GW----LAPSLQHRVDTTLAWVARLRRWA---PVTS----LSTMLHRFDTQLLQ----------------NPEISG-
288940385|Allochromatium vinosum DSM 180                ---RAARFLNR-TRRT---------------------------GW----LPPSVQHRVESTVNWVKRLRRLA---PITA----LSQELVRFDTQALQ----------------NPEISG-
288941921|Allochromatium vinosum DSM 180                ---RAPRVLNR-PRRE---------------------------GW----LPPSLQHRLDTTLNWVERLRRLA---PITA----LSQELVRFDTQALQ----------------NPEISG-
288940786|Allochromatium vinosum DSM 180                ---RAPRFLNR-TRRE---------------------------GW----LPPSLQHRLDTTLNWVARLRRLA---PITA----LSQELVRFDTQALQ----------------NPEISG-
288941171|Allochromatium vinosum DSM 180                ---RAPRFLNR-PRRE---------------------------GW----LPPSLQHRLDTTLNWVARLRRLA---PITA----LSQELVRFDTQALQ----------------NPEISG-
288941306|Allochromatium vinosum DSM 180                ---RAPRFLNR-PRRE---------------------------GW----LPPSLQHRLDTTLNWVARLRRLA---PITA----LSQELVRFDTQALQ----------------NPEISG-
121582913|Polaromonas naphthalenivorans CJ2              ---RAPRFLNRGNKKS---------------------------GW----LAPSLQHRVDTTLAWVNRLERLA---PVSA----ISTELVRFDMQALE----------------NPDIEG-
121604604|Polaromonas naphthalenivorans CJ2              ---RAPRFLNRDNKAK---------------------------GW----LAPSLQHRVDTTLAWVNRLEQLA---PVSA----ISTEMVRFDMQALE----------------NPEIEG-
91791204|Polaromonas sp. JS666                          ---RAPRFLNRGNKKS---------------------------GW----IAPSLQHRVDTAMAWVRRIQRWA---PVRA----ISSELVRFDMQALQ----------------NPEISG-
91791283|Polaromonas sp. JS666                          ---RAPRFLNRGNKKS---------------------------GW----IAPSLQHRVDTSIAWVKRIQHWA---PVRA----ISSERVRFDMQALQ----------------NPEISG-
121582916|Polaromonas naphthalenivorans CJ2              ---RAPRFLNRGNKQK---------------------------GW----LAPSLQHRVDTTAAWVARIQRWA---PVTA----LSSELVRFDMQQLQ----------------NTEIEG-
121582983|Polaromonas naphthalenivorans CJ2              ---RAPRGLNRGNKQK---------------------------GW----LAPSLQHRVDTTLAWVARLRRLA---PITA----LSSELVRFDMQQLA----------------NPQIEG-
428304158|Crinalium epipsammum PCC 9333                  ---RQPRFSHR-TRKP---------------------------GW----LAPSLFSRIANIETWVRRIRKLC---PITA----ISTELVRFDFQQME----------------NPEISG-
428204921|Chroococcidiopsis thermalis PCC 7203           ---RPARFDNR-KRKE---------------------------GW----LPPSLMHRVLTVETWVKRLCRYA---PITQ----VVMELVRFDTQKMQ----------------NPEIDG-
307153539|Cyanothece sp. PCC 7822                        ---RQPRFNNR-KRVD---------------------------KS----LPPSLMSRVHNILTWVRKLTRLA---IITA----ISSELVRFDTQALE----------------NPEIRG-
428216193|Oscillatoria acuminata PCC 6304                ---RKPRFLNR-TRKP---------------------------GW----LPPSLESRISNVLTWVNRLLLVC---HVTA----ISQELVRFDTQKLQ----------------NPEVTG-
414079584|Anabaena sp. 90                                ---RQARFLNR-TKPQ---------------------------GW----LAPSLSHRVLTINTWVKRLCNFA---PITD----IVQELARFDLQQLE----------------NPEISG-
172055147|Cyanothece sp. ATCC 51142                      ---RKPRFLNR-TRPK---------------------------GW----LPPSLQSRVENINTWVNRLRKLC---PVDS----ISVELVRFDTQQLE----------------NPEISG-
428316510|Oscillatoria nigro-viridis PCC 7112            ---RKPRFLNR-TRAP---------------------------GW----LPPSLESRVKNIVTWVNRIRRYV---PVTG----ISQELVLFDLLAMQ----------------NPEISG-
428314951|Oscillatoria nigro-viridis PCC 7112            ---RQPRFLNR-TRKS---------------------------GW----LPPSLESRVENILTWVNRIRRYV---PITG----ISQELVKFDTQAMQ----------------NPEISG-
307591386|Cyanothece sp. PCC 7822                        ---RQARFLNR-TRLK---------------------------GW----LPPSLRHRVETTMTWVKRICKFV---HVTN----ISVELVKFDTQALD----------------NPEISG-
75812341|Anabaena variabilis ATCC 29413                  ---RQPRFLNR-TKPE---------------------------GW----LAPSLQSRVENIKTWVSKLRKVA---PITA----ISQELVKFDMQLIR----------------RPDIQG-
17233293|Nostoc sp. PCC 7120                            ---RQPRFFNR-TRPQ---------------------------GW----LAPSLQSRVENIKTWVNRLRKVA---PITA----ISQELVKFDTQLIR----------------HPDIQG-
75812782|Anabaena variabilis ATCC 29413                  ---RKPRFLNR-TRSE---------------------------GW----LAPSLMSRVHNVETWVNRLRKFA---PITA----ISTELVKFDMQLMR----------------NPEIEG-
434407609|Cylindrospermum stagnale PCC 7417              ---RQPRFMNR-NRPQ---------------------------GW----LAPSLQSRVENIKTWVNKLVKFA---PIEA----ISQELVRFDMQLIS----------------NPDIQG-
414078405|Anabaena sp. 90                                ---RQPRFLNR-TRPN---------------------------GW----LAPSLQSRVENIKTWVKKLRKFA---PIEA----ISQELVRFDMQLMA----------------NPDIQG-
440685156|Anabaena cylindrica PCC 7122                   ---RQPRFLNR-TRPE---------------------------GW----LAPSLQSRIENIKTWVKKLRQFA---PIGA----ISQELVRFDMQLMA----------------NPDLQG-
428315656|Oscillatoria nigro-viridis PCC 7112            ---RKARFLNR-TRSQ---------------------------GW----LAPSLQHRVETILTWVTRLMKVA---PISS----IAQELVRFDLQQLE----------------NPEISG-
428204926|Chroococcidiopsis thermalis PCC 7203           ---REARFLNR-TRPQ---------------------------CW----LAPSLQHRVDTTLTWVKRLTRYT---PVAE----VRQELVKFDLQLMQ----------------NTEISG-
218440244|Cyanothece sp. PCC 7424                        ---RQPRFLNR-TRPS---------------------------GW----LAPSLQHRVLTTLTWVRKFIRLA---PIAS----LAQELVRFDLQLMQ----------------NSEVSG-
428775092|Halothece sp. PCC 7418                         ---RQARFLNR-TRPK---------------------------GW----LPPSLQHRVDTTMTWVNRFRKLA---PIAR----ISQELVRFDLQLMQ----------------NPEISG-
414075475|Anabaena sp. 90                                ---RQARFLNR-TRTE---------------------------NW----LAPSLQHRVETTLTWVNKLIKFI---PITR----ISQELVRFDLQQID----------------NPEISG-
428305120|Crinalium epipsammum PCC 9333                  ---RKPRFLNR-TRQC---------------------------GW----LAPSLQHRVETILTWVNRLRRYA---PITN----ISTELVRFDMLALD----------------NPEISG-
427713881|Synechococcus sp. PCC 6312                     ---RQPRFLNR-NRPD---------------------------GW----LAPSLKHRVDTVLTWVNRLRNLA---PLGS----IAQELVRFDLQKLQ----------------NPEISG-

S. pyogenes Cas9                                                                              (732)  A----GSPAIKKGILQTVKVVDELVKVMGRHKPEN----IVIEMAR-ENQTTQKGQKNSRERMKRIEEGIKELGSQ
S. pyogenes Cas9 domains                                                                             rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr|lllllll|NNNNNNNNNNNNNNNNNNNNNN
A. naeslundii Cas9                                       ---QLQQLDEV-LRDL                              (476) PSVDRTLKIVGRYLSAVESMWG--TPEV----IHVEHVR----DGF----------------TSERMA-
A. naeslundii Cas9 domains                               RRRRR|hhhhhhhhhh                                   rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr|NNNNNNNNNNNNNNNNNNNNNN
```

```
312126587|Caldicellulosiruptor hydrothermalis 108      --NESF-----------------------SPLE-TGK-----------------------------------------------------------------------------------
434408285|Stanieria cyanosphaera PCC 7437              --GKSF-----------------------SPLQ-VGK-----------------------------------------------------------------------------------
414076241|Anabaena sp. 90                              --NKGF-----------------------SPVM-VGQ-----------------------------------------------------------------------------------
428299778|Calothrix sp. PCC 6303                       --SKSF-----------------------SPVM-VGQ-----------------------------------------------------------------------------------
 17233043|Nostoc sp. PCC 7120                          --GKGF-----------------------SAVM-VGQ-----------------------------------------------------------------------------------
166363864|Microcystis aeruginosa NIES-843              --GKGF-----------------------SSVM-VGQ-----------------------------------------------------------------------------------
166367659|Microcystis aeruginosa NIES-843              --GKGF-----------------------SSVM-VGQ-----------------------------------------------------------------------------------
307150343|Cyanothece sp. PCC 7822                      --GKGF-----------------------SPVM-VGQ-----------------------------------------------------------------------------------
307153352|Cyanothece sp. PCC 7822                      --GKGF-----------------------SPVM-VGQ-----------------------------------------------------------------------------------
428306661|Crinalium epipsammum PCC 9333                --GSGF-----------------------SAVM-VGQ-----------------------------------------------------------------------------------
307150912|Cyanothece sp. PCC 7822                      --GKGF-----------------------SAVM-VGQ-----------------------------------------------------------------------------------
428775289|Halothece sp. PCC 7418                       --DKGF-----------------------SPVM-VSQ-----------------------------------------------------------------------------------
428779516|Dactylococcopsis salina PCC 8305             --DKGF-----------------------SPVM-VSQ-----------------------------------------------------------------------------------
428778599|Dactylococcopsis salina PCC 8305             --DQGF-----------------------SPVM-VGQ-----------------------------------------------------------------------------------
428781696|Dactylococcopsis salina PCC 8305             --DKGF-----------------------SPVM-VGQ-----------------------------------------------------------------------------------
428215045|Oscillatoria acuminata PCC 6304              --SKSF-----------------------SPVM-VGQ-----------------------------------------------------------------------------------
428313425|Microcoleus sp. PCC 7113                     --GNGF-----------------------SPVM-VGQ-----------------------------------------------------------------------------------
158340109|Acaryochloris marina MBIC11017               --GKGF-----------------------SPVM-TAQ-----------------------------------------------------------------------------------
428775002|Halothece sp. PCC 7418                       --GKGF-----------------------SPVM-IGQ-----------------------------------------------------------------------------------
307150700|Cyanothece sp. PCC 7822                      --GKGF-----------------------SAVM-VGQ-----------------------------------------------------------------------------------

312126663|Caldicellulosiruptor hydrothermalis 108      --TGYQ-----------------------KSPR-VP-----------A-EKKCI--ICG---SKEK--L-----QKHHLIQRKYGGTDI-RENVVYLCRDCH--EDVHA-----------
312621424|Caldicellulosiruptor kronotskyensis 2002     --TGYQ-----------------------KSPR-VP-----------A-EKKCI--LCG---TEDN--L-----QKHHLIQRKCGGTDV-QENVVYLCRDCH--EDVHA-----------
222530402|Caldicellulosiruptor bescii DSM 6725         --TGYQ-----------------------KSPR-VP-----------A-EKKCI--ICG---TEDN--L-----QKHHLIQRKCGGTDV-QENLVYLCRDCH--EDVHA-----------
222530431|Caldicellulosiruptor bescii DSM 6725         --TGYQ-----------------------KSPR-VP-----------A-EKKCI--ICG---TEDN--L-----QKHHLIQRKCGGTDV-QENLVYLCRDCH--EDVHA-----------

307591386|Cyanothece sp. PCC 7822                      --KEYQ-----------------------QGEL-FGY-EIREYLLEKW-GRRCV--YCG--IKDVP--L-----EVEHILAKSKGGSDR-CSNLTISCRICN--QLKGN---QDIKDFLS
302391131|Acetohalobium arabaticum DSM 5501            --WQYQ-----------------------ESNR-LDE-NLRKATLKRD-DYTCQ--LCE--VEDTM--L-----HAHHIIPVRDGGADS-IYNLITLCADCH--NNKVD-----NNEYKY
435854579|Halobacteroides halobius DSM 5150            --WEYQ-----------------------ESNR-LDE-NLRKATLYRD-DCAFQ--ICG--ATDTM--L-----EAHHIIPQRDSGADS-IYNLITLCKDCH--KDKVD-----GSEYKY
435852864|Halobacteroides halobius DSM 5150            --WEYQ-----------------------ESNR-LDE-NLRKATLYRD-DCTCQ--LCG--TTETM--L-----HAHHIMPRRDGGADS-IYNLITLCKACH--KDKVD-----NNEYQY
435854272|Halobacteroides halobius DSM 5150            --WEYQ-----------------------ESNR-LDE-NLRKATLYRD-DCTCQ--LCG--ATETM--L-----HAHHIMPRRDGGADS-IYNLITLCKACH--KDKVD-----GREYQY
313123202|Lactobacillus delbrueckii  ND02              --KQYQ-----------------------EGSL-KDWKNYEYYVLARD-EYTCQ--LCHKHGEGVK--L-----VVHHIVYRSQGGTDR-VDNLITLCTNCH--TTKNH----QPGGKLY
 77362205|Pseudoalteromonas haloplanktis TAC125        --AGYQ-----------------------QGAL-LGYYNVKAYVLQRD-NYKCQ--SGRKTKHNAK--L-----HVHHKVFRSQGGTDA-LSNLIALCETCH--NDLHV-----------
298675374|Methanohalobium evestigatum Z-7303           --TGYQ-----------------------NGDL-KGYYNIRIYVLHRD-DYTCQ--KCKNKSKDPK--L-----ECHHIIFRSKGGTDR-PNNLITLCETCH--SKLHK------GEFSI
298674707|Methanohalobium evestigatum Z-7303           --TDYQ-----------------------NGDL-KGYYNIRTYVLHRD-NYTCQ--KCKNRSKDPK--L-----ECHHIVFRSKGGTNR-PDNLITLCETCH--GKLHN------GKFTI
298675090|Methanohalobium evestigatum Z-7303           --TGYQ-----------------------NGDL-KGYYNIRTYVLHRD-NYTCQ--KCKNRSKDPK--L-----ECHHIVFRSNGGTDR-PDNLITLCETCH--GKLHD------GEFTI
169834784|Clostridium botulinum A3 str. Loch Maree     --EAYQ-----------------------NGEQ-KDFFNLREYILHRD-GHKCQNPNCKSKGAESI--L-----QVHHIGFWKQDRSDR-PGNLITLCNKCH--TPKNH----KENGFLY
229587454|Clostridium botulinum Ba4 str. 657           --TAYQ-----------------------NGEQ-KDFFNLREYILYRD-GHKCQNPNCKNKSPEPI--L-----QVHHIGFWKQDRSDR-PGNLITLCNKCH--TPKNH----KGKGFLY

297585104|Bacillus selenitireducens MLS10              --EGYQ-----------------------QGDQ-FGFWNVREYVFFRD-NHRCQ--HCKGKSKDKI--L-----NVHHIESRKTGG-DS-PDNLLTLCETCH--KKIHR-----------
188587532|Natranaerobius thermophilus JW-NM-WN-LF      --EQYQ-----------------------QGEQ-LGFWNVREYVLYRD-NHTCQ--YCRGKSKDSI--L-----NVHHIRSRQSHG-DR-PGNLVTLCETCH--TKLHK-----------
312793946|Caldicellulosiruptor kristjanssonii 177R1B   --TKYQ-----------------------EGPQ-KDFWDVREYCLWRA-GHKSE--LSG---KKGV--L-----EIHHIIPKSQGGTNN-PSNLIVLTVEEHKAIHEGK-----------
344996431|Caldicellulosiruptor lactoaceticus 6A        --TKYQ-----------------------EGPQ-KDFWDVREYCLWRA-GHKSE--LSG---KKGV--L-----EIHHIIPKSQGGTNN-PSNLIVLTVEEHKAIHEGK-----------
217077332|Thermosipho africanus TCF52B                 --IQYQ-----------------------EGPQ-KSFSDVREYCLWRA-GYKSE--LSG---KKGI--L-----EVHHIIPRSKGGTDN-PSNLIVLTAEEHKALHEGK-----------
375009168|Geobacillus thermoleovorans CCB US3 UF5      --VEYQ-----------------------QGEQ-MGFWNVREYVLYRD-NHTCQ--YCKGKKKDPV--L-----QVHHIESRKTGG-DS-PDNLITLCKTCH--HEIHE-----------
407477159|Exiguobacterium antarcticum B7                --DLYQ-----------------------KGDQ-LGFWNVREYVFFRD-KHICQ--HCKGKSKDKI--L-----NVHHIESRRTGG-DS-PDNLITLCETCH--KKIHK-----------
172057416|Exiguobacterium sibiricum 255-15             --DLYR-----------------------KGDQ-LGFWNVREYVFFRD-KHMCQ--HCKGKSKDKI--L-----NVHHIESRRTGG-DA-PDNLITLCETCH--KKIHK-----------
169823684|Finegoldia magna ATCC 29328                  --VEYQ-----------------------QGEQ-LGFWNIREYVLFRD-NHTCQ--HCKGKSKDPI--L-----NVHHIESRQTGG-NS-PNNLITLCESCH--NAYHK-----------
160902522|Petrotoga mobilis SJ95                       --KQYQ-----------------------QGEQ-LDFWNVREYVLWRD-NYTCQ--RCKDKSKDKR--L-----NVHHIESRQIGG-NA-PSNLITLCETCH--KDYHD-----------
160902616|Petrotoga mobilis SJ95                       --KQYQ-----------------------QGEQ-LDFWNVREYVLWRD-NYTCQ--HCKGKSKDKR--L-----NVHHIESRQIGG-NA-PSNLITLCETCH--RDYHN-----------

333978211|Desulfotomaculum kuznetsovii DSM 6115        -------------------------------AL-PGE-GAREKVLNRD-G-ACV--LCG---SEEK--L-----QRHHLVPRSKGGSDT-PMNQVVLCEKCH--RKIHA-----------
260892456|Ammonifex degensii KC4                       -------------------------------AL-PGK-STREKVLNRD-G-ACV--LCG---STRR--L-----QRHHLVPRSKGGTDT-PENQVLLCAECH--RKLHA-----------
260893631|Ammonifex degensii KC4                       -------------------------------AL-PGK-STREKVLNRD-G-ACV--LCG---STRR--L-----QRHHLVPRSKGGTDT-PENQVLLCAECH--RELHA-----------
302391857|Acetohalobium arabaticum DSM 5501            --IEYQ-----------------------QGTL-QGY-NVREYLLYKH-NHTCQ--YCGGESGDPV--L-----EVEHKIPSSRGGTDS-IKNLTLSCKTCN--QDKDN---RTLKEWLQ
428778981|Dactylococcopsis salina PCC 8305             --VQYQ-----------------------QGEL-QGY-EIREYLLEKW-GRQCV--YCE--KQDTP--L-----QIEHIKPKSKGGSNR-IGNLTLSCPECN--QNKGN---QDVEDFLS
260892735|Ammonifex degensii KC4                       --VEYQ-----------------------QGEL-QGY-EVREYLLEKY-GRRCA--YCG--KENVP--L-----EVDHVVPKSRGGTDR-VSNLTLACRECN--QAKGN---KLPQEWLE
260893993|Ammonifex degensii KC4                       --VEYQ-----------------------RGEL-QGY-EVREYLLEKY-GRKCA--YCG--KENVP--L-----EVDHVVPKSRGGTDR-VSNLTLACRECN--RAKGN---KLPQEWLE
433655673|Thermoanaerobacterium
thermosaccharolyticum M0795                            --VEYQ-----------------------QGEL-QGY-EVREYLLVKW-GRKCA--YCG--KENVP--L-----EVEHIVPKSRGGTDR-VSNLTIACHECN--QKKGN---MTAEEF--
332982963|Mahella australiensis 50-1 BON               --VEYQ-----------------------HGEL-FGY-EVKEYLLEKW-GRKCT--YCG--KENVP--L-----EVEHIIPKSRGGTDR-ISNLTIACHKCN--QKKGN---MTAEEF--
332983053|Mahella australiensis 50-1 BON               --VEYQ-----------------------HGEL-FGY-EVKEYLLEKW-GRKCA--YCG--KENVP--L-----EVEHIIPKSRGGTDR-ISNLTIACHKCN--QKKGN---MTAEEF--
222530385|Caldicellulosiruptor bescii DSM 6725         --IEYQ-----------------------HGTL-QGY-EVKEYLLEKF-NWRCV--YCG--ATGVP--L-----EVEHVIPKSRGGTDR-VDNLVIACHGCN--QKKGN---KTAEEF--
298674346|Methanohalobium evestigatum Z-7303           --VEYQ-----------------------QGTL-QGY-NVREYLLEKF-QRTCA--YCG--KQDVQ--L-----EIEHIVPRSRGGSNR-VSNLTIACHDCN--QRKSN---LTAEEF--
298675173|Methanohalobium evestigatum Z-7303           --VEYQ-----------------------QGTL-KGY-EVREYLLEKF-QRTCA--YCD--KQNVQ--L-----EVEHIVPRSRGGSDR-VSNLTIACHDCN--QRKSN---LTAEEF--
292490518|Nitrosococcus halophilus Nc 4                --IEYQ-----------------------QGTL-YGF-ELREYLLARH-QHTCA--YCHGLSKDSI--L-----EIEHILPKSLGGTDR-VGNLVVACRTCN--QHKNNLHPQAWADACR
428778759|Dactylococcopsis salina PCC 8305             --KEYQ-----------------------QGTL-QGY-SVREYLLEKW-DRKCA--YCK--KENVP--L-----QVEHIHSKSKGGSNQ-VSNLCLACERCN--QKKGD---KSIKDFLK
300114477|Nitrosococcus watsonii C-113                 --VEYQ-----------------------QGEL-LGY-EVREYLLEKW-GRKCA--YCN--KQDIP--L-----EVEHITPRCRGGSHR-VSNLTLACTPCN--QRKNH---QTAAEF--
292491781|Nitrosococcus halophilus Nc 4                --VEYQ-----------------------QGEL-MGY-EVREYLLEKW-DRKCA--YCN--KKDIP--L-----EVEHIIPRSRGGSNR-VSNLTLARAPCN--KKKNS---KTAAEF--
292491869|Nitrosococcus halophilus Nc 4                --VEYQ-----------------------QGEL-MGY-EVREYLLEKW-DRKCA--YCN--KKDIP--L-----EVEHIIPRSRGGSNR-VSNLTLARAPCN--KKKNS---KTAAEF--
292491893|Nitrosococcus halophilus Nc 4                --VEYQ-----------------------QGEL-MGY-EVREYLLEKW-DRKCA--YCN--KKDIP--L-----EVEHIIPRSRGGSNR-VSNLTLARAPCN--KKKNS---KTAAEF--
292491036|Nitrosococcus halophilus Nc 4                --VEYQ-----------------------QGEL-QGY-EVREYVLEKW-GRKCA--YCH--KTNIP--L-----EIEHLVPRSRGGSHR-VSNLTLACTPCN--QKKNS---KTAAEF--
292490884|Nitrosococcus halophilus Nc 4                --VEYQ-----------------------QGEL-MGY-EVREYLLEKW-GRKCA--YCS--NKDVP--L-----EVEHIVPRSRGGSHR-VSNLTLVCTPCN--QKKNS---KTATEF--
292491492|Nitrosococcus halophilus Nc 4                --VEYQ-----------------------QGEL-MGY-EVREYLLEKW-GRKCA--YCS--NKDVP--L-----EVEHIVPRSRGGSHR-VSNLTLACTPCN--QKKNS---KTATEF--
297563003|Nocardiopsis dassonvillei DSM 43111           --VEYR-----------------------HGTL-HGY-EVREYLLAKW-GRACA--YCG--ATGTP--L-----NIDHVRPRSRGGTDR-VADLALACVPCN--RAKAD---QPVEEFVT
374989417|Streptomyces bingchenggensis BCW-1           --AEYQ-----------------------HGTL-HGT-EIREYLLAKF-GRACV--YCG--IAGVP--L-----NIDHVHPRSRGGSDR-ISNLVIACIPCN--ESKSH---RPVEEFVR
374989654|Streptomyces bingchenggensis BCW-1           --AEYQ-----------------------HGTL-HGT-EIREYLLAKF-GRACV--YCN--ISGVP--L-----NIDHVHPRSRGGSDR-ISNLVTACIPCN--ETKSN---RPVEEFVR
271963428|Streptosporangium roseum DSM 43021           --AAYQ-----------------------QGTL-VGY-EVREYLLAKW-GRACA--YCG--ISRTP--L-----NIDHIHPRSRGGCDR-ISNLTVACIPCN--QTKNA---TPVEEFLK
271964894|Streptosporangium roseum DSM 43021           --VEYR-----------------------YGTL-HGY-EVREYLLAKW-GRACA--YCG--ASGVP--L-----NLDHIHPRSRGGSNR-ISNLCVACVGCN--QAKNA---TPIEEFLT
258511478|Alicyclobacillus acidocaldarius DSM 446      --EEYQ-----------------------QGTL-YGY-EVREYLLEKW-GRKCA--YCG--AENVP--L-----EIDHVIPRSRGGSDR-VSNLVLACHRCN--QAKAN---RPVEEFLA
160902264|Petrotoga mobilis SJ95                       --IEYQ-----------------------RGTL-YGY-EIKEYLLEKW-GRKCV--YCG--KENVP--L-----EIEHIVPKSKGGSDR-ISNLTLACHECN--QKKGN---QSIEEFLT
121604978|Polaromonas naphthalenivorans CJ2            --AEYQ-----------------------QGTL-QGY-EVREYLLEKF-NRTCA--YCD--AQNVP--L-----QMEHIHPKAAGGTNR-ISNLALACQACN--QKKGV---QDIKDFLL
347754630|Chloracidobacterium thermophilum B           --IEYQ-----------------------QGTL-AGC-EVREYLLEKW-GRKCA--YCD--AENVP--L-----GIDHIHPRSKGGSDR-VSNLTLACHPCN--RRKGA---QGIEDFLA
218667603|Acidithiobacillus ferrooxidans ATCC 23270    --TEYQ-----------------------QGTL-FGY-EVREYLLEKW-GRECA--YCG--AEHTP--L-----EIDHIHPRSKGGSDR-VSNLTIACHDCN--QAKGN---MPIEQFLA
218667415|Acidithiobacillus ferrooxidans ATCC 23270    --IEYQ-----------------------RGEL-FGY-EVREYLLEKW-GRKCA--YCD--AQNTP--L-----TIDHIHPRSAGGSDR-VSNLTLACFPCN--QRKSN---RDVREFLA
344201059|Acidithiobacillus ferrivorans SS3            --IEYQ-----------------------QGEL-QGY-EVREYLLEKW-GRKCV--YCD--AEHTP--L-----TIDHIHPKSAGGSDR-VSNLTLACFPCN--QRKSN---QDVRAFLA
344201127|Acidithiobacillus ferrivorans SS3            --VEYQ-----------------------QGEL-QGY-EVREYLLEKW-GRKCA--YCD--AEHTP--L-----TIDHIHPKSTGGSDR-VSNLTLACLPCN--QRKSN---QDVRAFLA
288940385|Allochromatium vinosum DSM 180               --VEYQ-----------------------QGEL-AGY-EVREYLLEKW-HRTCA--YCS--ATGVP--L-----EIEHIVPRSRGGSDR-VSNLTLACRACN--QRKGN---QSIEDFLK
288941921|Allochromatium vinosum DSM 180               --VEYQ-----------------------QGEL-AGY-EVREYLLEMW-HRSCA--YCG--ATNVP--L-----EIEHIVPRAKGGSDR-ASNLTLACRPCN--QRKGS---RSIEDFLQ
288940786|Allochromatium vinosum DSM 180               --VEYQ-----------------------QGEL-AGY-EVREYLLETW-HRTCA--YCG--ATNVP--L-----EIEHIVPRARGGSDR-VSNLTLACRPCN--QRKGN---RSIEDFLK
288941171|Allochromatium vinosum DSM 180               --VEYQ-----------------------QGEL-AGY-EVREYLLETW-HRTCA--YCG--ATHVP--L-----EIEHIVPRARGGSDR-VSNLTLACRACN--QRKAN---QSIEDFLK
288941306|Allochromatium vinosum DSM 180               --VEYQ-----------------------QGEL-AGY-EVREYLLETW-HRTCA--YCG--ATHVP--L-----EIEHIVPRARGGSDR-VSNLTLACRACN--QRKAN---QSIEDFLK
121582913|Polaromonas naphthalenivorans CJ2            --AQYQ-----------------------QGTL-AGY-EVREYLLEKW-GRTCA--YCG--ARNVP--L-----QIEHIHPKSKSGSDR-IGNLTLACQCCN--QKKAA---LPVEVYLA
121604604|Polaromonas naphthalenivorans CJ2            --AQYQ-----------------------QGTL-AGY-EVREYLLEKW-GRTCA--YCG--AKNVP--L-----QIEHIHPKSKSGSDR-IGNLTLACQCCN--QKKAA---LPVEVFLA
91791204|Polaromonas sp. JS666                         --VEYQ-----------------------QGTL-FGY-ELREYLLEKW-GRQCA--YCD--AKDTP--L-----QIEHIRPKALGGSNR-ASNLTLACQCCN--QKKAA---RSIEEFLA
91791283|Polaromonas sp. JS666                         --VEYQ-----------------------QGTL-YGY-ELREYLLEKW-GRQCA--YCD--AKDTP--L-----QIEHIRPKALGGSNR-ASNLTLACQCCN--QKKAA---RSIEDFLA
121582916|Polaromonas naphthalenivorans CJ2            --AEYS-----------------------QGTL-AGY-EVREYLLEKW-KRTCA--YCD--AQNTP--L-----QIEHIEPRARGGSHR-ISNLCLACQPCN--QKKAA---RTLQDFLK
121582983|Polaromonas naphthalenivorans CJ2            --AEYQ-----------------------QGTL-AGY-EVREYLLEKW-SRTCA--YCD--AQNTP--L-----QIEHIEPRARGGSHR-ISNLCLACGPCN--QKKAA---RTLQDFLK
428304158|Crinalium epipsammum PCC 9333                --VEYQ-----------------------RGEL-YGF-EVKEYLLQKW-KRKCA--YCG--IEDVP--F-----EIEHIIAKSKGGSNR-VSNLCLSCHSCN--QAKGN---QDIRDFLK
428204921|Chroococcidiopsis thermalis PCC 7203         --VEYQ-----------------------QGEL-AGY-EVREYLLEKW-GHNCV--YCD--KQDVP--L-----QIEHIYPRAKGGSNR-VSNLTLSCEKCN--SLKGI---LLIDEFLK
307153539|Cyanothece sp. PCC 7822                      --TEYQ-----------------------QGEL-AGY-EVREYLLEKF-NRKCI--YCS--QTDTR--L-----EIEHLTPRSKGGSNR-VSNLGIACQPCN--QKKGN---RDVLSFLK
428216193|Oscillatoria acuminata PCC 6304              --VEYQ-----------------------QGEL-FSY-EVREYLLTKW-GRKCA--YCS--AENVP--L-----EVEHIHPKSKGGSDR-VSNLTLACHPCN--QAKGN---RNIREFLS
414079584|Anabaena sp. 90                              --FEYQ-----------------------QGEL-QGY-EVREYLLNKW-NRKCA--YCT--AENVP--L-----QVEHIKPKAKGGTNR-ISNLCLACEKCN--IKKGT---QDIEKFLV
172055147|Cyanothece sp. ATCC 51142                    --VEYQ-----------------------QGTL-CGY-EVREYLLEKY-HRTCQ--YCG--AKEVP--F-----EIEHIVPKSKGGSDR-VSNLTLSCHKCN--QKKGN---RSVEEFLK
428316510|Oscillatoria nigro-viridis PCC 7112           --KEYQ-----------------------QGEL-AGY-EVREYLLAKW-GRKCA--YCG--VENVP--F-----EVEHIHPKSKGGSDR-VSNLTLACRGCN--QAKGN---QDIRDFLS
428314951|Oscillatoria nigro-viridis PCC 7112           --IQYQ-----------------------QGEL-AGY-EIREYLLEKW-GRKCA--YCN--AQNIP--L-----EIEHIHPRSKGGSSR-VSNLTIACRHCN--EAKGN---QELKDFLA
307591386|Cyanothece sp. PCC 7822                      --KEYQ-----------------------QGEL-FGY-EIREYLLEKW-GRRCV--YCG--IKDVP--L-----EVEHILAKSKGGSDR-CSNLTISCRICN--QLKGN---QDIKDFLS
75812341|Anabaena variabilis ATCC 29413                --KEYQ-----------------------QGTL-AGF-ETRQYLLEKW-NRECA--YCG--IKDVP--L-----QIEHIHPKSKGGSNS-ITNLTLSCEKCN--LRKGN---QDIRDYLE
17233293|Nostoc sp. PCC 7120                           --KEYQ-----------------------QGTL-AGF-ETRQYFLEKW-NRECA--YCG--IKDVP--L-----QIEHIHPKSKGGSNS-ITNLTLSCEKCN--LKKGT---QDIKNFLK
75812782|Anabaena variabilis ATCC 29413                --KEYQ-----------------------QGTL-AGY-ETREFLLEKW-NRQCA--YCG--IKDIP--L-----QIEHIHSRSKGGSNS-ITNLTLSCEKCN--VKKGT---KDIKDFLK
434407609|Cylindrospermum stagnale PCC 7417            --KEYQ-----------------------QGTL-AGY-ETREFLLEKW-NRQCA--YCG--VKDVP--L-----QIEHIHPRARGGSNS-ITNLTLSCEKCN--RKKGT---KDIKEFLQ
414078405|Anabaena sp. 90                              --KEYQ-----------------------QGTL-AGY-ETKEYLLEKW-NRQCA--YCG--VKDVP--F-----QVEHIHPRAKGGSNS-ITNLTLSCEKCN--TKKGT---KDIKDFLK
440685156|Anabaena cylindrica PCC 7122                 --REYQ-----------------------QGTL-AGY-ETREYLLEKW-DRQCA--YCG--VKDVP--F-----QVEHIHPRAKGGSDS-ITNLTLSCEKCN--TKKGT---KDIKDFLK
428315656|Oscillatoria nigro-viridis PCC 7112           --LEYQ-----------------------QGVL-CGY-EVREYLLNKW-ERKCA--YCG--VTDTP--L-----QVEHIHPKAKGGSNR-ISNLCLACEKCN--VKKGT---QDIEQFLS
428204926|Chroococcidiopsis thermalis PCC 7203         --IEYQ-----------------------QGEL-QGY-EVREYLLEKF-NRQCV--YCG--AKDMP--L-----EVEHIKPRSRGGSDR-VSNLVLACHSCN--QSKGN---QDIKDFLS
218440244|Cyanothece sp. PCC 7424                      --VEYQ-----------------------RGEL-QGY-EVREYLLEKW-NRRCA--YCS--VENVP--F-----EVEHIQPKSKGGSER-VSNLTIACHDCN--QSKGN---QDIRDFLS
428775092|Halothece sp. PCC 7418                       --IEYQ-----------------------QGEL-QGY-EVREYLLEKW-NRKCA--YCG--AKNVA--L-----EVDHIHPKSKGGTDQ-VSNLTLACHSCN--QSKSN---RDIRDFLS
414075475|Anabaena sp. 90                              --IEYQ-----------------------QGTL-SGY-EVREYLLEKW-QRKCA--YCG--IENVP--L-----QVEHIHPQARGGTNR-ISNLCLACEKCN--IKKGT---QDIKDFLA
428305120|Crinalium epipsammum PCC 9333                --IEYQ-----------------------QGTL-FGY-EVREYLLAKW-DRKCA--YCS--AENVP--L-----QIEHIHPKAKGGTNR-ISNLCVACEPCN--IKKGT---QDVEVFLT
427713881|Synechococcus sp. PCC 6312                   --IEYQ-----------------------QGEL-QGY-EVREYLLEKW-GRKCT--YCG--TKDVP--L-----EVEHIHPRSRGGTDR-VSNLTMACHACN--QSKGN---QDIRDFLF

S. pyogenes Cas9                                       ILKEHP-----------------------VENT-QLQ-NEKLYLYYLQ-NGRDM--YVD--QELDI--NRLSDYDVDHIVPQSFLKDDS-IDNKVLTRSDKN--RGKSD---NVPSEEVV
S. pyogenes Cas9 domains                               NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
A. naeslundii Cas9                                     --DERDKANRRRYNDNQEAMKKIQRDYGKEGYI-SRG-DIVRLDALELQGCACL--YCG--TTIGYHTC-----QLDHIVPQAGPGSNNRRGNLVAVCERCN--RSKSN----TPFAVWA
A. naeslundii Cas9 domains                             NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
```

```
312126587|Caldicellulosiruptor hydrothermalis 108      -----------------------------------------------------------------------NW----------------------FYA-ELSRL---G---KVHIKEGRE
434408285|Stanieria cyanosphaera PCC 7437              -----------------------------------------------------------------------DW----------------------FYS-ELAKL---A---EVHTLQGME
414076241|Anabaena sp. 90                              -----------------------------------------------------------------------RY----------------------QIS-RLSKL---T---NVVLKKGWE
428299778|Calothrix sp. PCC 6303                       -----------------------------------------------------------------------RF----------------------QIN-RLSEI---A---SIKLKKGWE
 17233043|Nostoc sp. PCC 7120                          -----------------------------------------------------------------------KW----------------------MLK-QLEQF---A---PVAKVEGYQ
166363864|Microcystis aeruginosa NIES-843              -----------------------------------------------------------------------KW----------------------AIE-QLSQL---A---TVHTRFGWQ
166367659|Microcystis aeruginosa NIES-843              -----------------------------------------------------------------------KW----------------------AIE-QLSQL---A---TVHTRFGWQ
307150343|Cyanothece sp. PCC 7822                      -----------------------------------------------------------------------KW----------------------MLE-QLRQL---A---NVHTRYGWQ
307153352|Cyanothece sp. PCC 7822                      -----------------------------------------------------------------------KW----------------------MLE-QLRQL---G---NVDTRYGWQ
428306661|Crinalium epipsammum PCC 9333                -----------------------------------------------------------------------KW----------------------AIE-QLSKF---A---RVHTRFGWQ
307150912|Cyanothece sp. PCC 7822                      -----------------------------------------------------------------------KW----------------------MLE-QLKQF---G---DVHTRLGWQ
428775289|Halothece sp. PCC 7418                       -----------------------------------------------------------------------KV----------------------MTEQWLPNI---A---PVTAKYGWE
428779516|Dactylococcopsis salina PCC 8305             -----------------------------------------------------------------------KV----------------------MTEQWLPQI---A---PVTTKYGWE
428778599|Dactylococcopsis salina PCC 8305             -----------------------------------------------------------------------KV----------------------ITEQWLPSI---A---PVTIKHGWE
428781696|Dactylococcopsis salina PCC 8305             -----------------------------------------------------------------------KV----------------------MTEQWLPSI---A---PVTTKFGWE
428215045|Oscillatoria acuminata PCC 6304              -----------------------------------------------------------------------KV----------------------MLG-WLSKL---A---PIETQLGYQ
428313425|Microcoleus sp. PCC 7113                     -----------------------------------------------------------------------KW----------------------MLE-HLSKL---A---PTSTLYGWE
158340109|Acaryochloris marina MBIC11017               -----------------------------------------------------------------------YW----------------------AME-QLEAI---A---PVIKRHGWQ
428775002|Halothece sp. PCC 7418                       -----------------------------------------------------------------------QF----------------------CIS-ELEKI---A---PVYVREGWQ
307150700|Cyanothece sp. PCC 7822                      -----------------------------------------------------------------------KW----------------------MLK-QLKQF---G---NVHTRLGWQ

312126663|Caldicellulosiruptor hydrothermalis 108      ---------------------------------------------------------GRVYIP-VEGVRQWRA----------------------LGTVNA----------IIGRLREI-
312621424|Caldicellulosiruptor kronotskyensis 2002     ---------------------------------------------------------GRVYIP-VDGVKQWRA----------------------LGTVNA----------IIGRLREI-
222530402|Caldicellulosiruptor bescii DSM 6725         ---------------------------------------------------------GRVYIP-VEGVRQWRA----------------------LGTMNA----------IIGQLREI-
222530431|Caldicellulosiruptor bescii DSM 6725         ---------------------------------------------------------GRVYIP-VEGVRQWRA----------------------LGTMNA----------IIGQLREI-

307591386|Cyanothece sp. PCC 7822                      NQP----------------------------------S-------------------LLEKIL-KQSKQSLKN----------------------VAAVNTTRW---A---LFNKLKETG
302391131|Acetohalobium arabaticum DSM 5501            KDQ------------------------------------------------------LLTKI--DGKELNLKP----------------------AMHVMQGKS---W---LREELSTMA
435854579|Halobacteroides halobius DSM 5150            KD-------------------------------------------------------EFLNII-DGKELNIKP----------------------AMHVMQGKT---W---LREKLSKIV
435852864|Halobacteroides halobius DSM 5150            KDQ------------------------------------------------------FLAIID-SKELSDLKS----------------------ASHVMQGKT---W---LRDKLSKIA
435854272|Halobacteroides halobius DSM 5150            KDQ------------------------------------------------------FLDIIN-SKELSDLKS----------------------ASHVMQGKT---W---LRDKLSKIA
313123202|Lactobacillus delbrueckii  ND02              KW-------------------------------------------------------MKAK---KKVTKQLKG----------------------ETFMNILRK---R---IMTTFPEAS
77362205|Pseudoalteromonas haloplanktis TAC125         ----------------------------------------------------------------GKFELKVKR----------------------SKTKHATEI---G---ILKAQLKQC
298675374|Methanohalobium evestigatum Z-7303           NKN----------------------------------P------------------------------SNTKH----------------------PTEMGIVRS---Q---IRKLN----
298674707|Methanohalobium evestigatum Z-7303           NKK----------------------------------P------------------------------SNTKH----------------------PTEIGIVRS---Q---IRKLN----
298675090|Methanohalobium evestigatum Z-7303           NKK----------------------------------P------------------------------SNTKH----------------------PTEMGIVRS---Q---IRKLN----
169834784|Clostridium botulinum A3 str. Loch Maree     GW--------------------------------------------------------------QPKIKSFKE----------------------ATFMSMIRW---R---LVNTLGCKH
229587454|Clostridium botulinum Ba4 str. 657           GW--------------------------------------------------------------HPKVKSFKE----------------------ATFMSMIRW---R---LVNTLGCKH

297585104|Bacillus selenitireducens MLS10              ------------------------------------------------EGLEH--------LF-QRKKKPFRD----------------------ASQMAVMRW---F---IYNGLKARY
188587532|Natranaerobius thermophilus JW-NM-WN-LF      ------------------------------------------------ESLEN--------YF-HPKDKGFKD----------------------ESQFTILRW---F---IYNGIKEKY
312793946|Caldicellulosiruptor kristjanssonii 177R1B   ------------------------------------------------IKIPR--------SK-LEKVKIFKD----------------------ASHVSTIGW---Y---IVNRLKYQY
344996431|Caldicellulosiruptor lactoaceticus 6A        ------------------------------------------------IKIPR--------SK-LEKVKIFKD----------------------ASHVSTIGW---Y---IVNRLKYQY
217077332|Thermosipho africanus TCF52B                 ------------------------------------------------IKIPR--------SK-LEKVRILKD----------------------ASHVSTIGW---H---IVNRLKQQY
375009168|Geobacillus thermoleovorans CCB US3 UF5      ------------------------------------------------KGLEH--------IF-QRKRPPMRD----------------------ASQMTAMRW---A---MFSRIREKY
407477159|Exiguobacterium antarcticum B7               ------------------------------------------------EKLEH--------LF-QRKSRSFRD----------------------ASQMTVMRW---F---IYNAVKETY
172057416|Exiguobacterium sibiricum 255-15             ------------------------------------------------ENLEH--------LF-QRKSRSFRD----------------------ASQMTVMRW---F---IYNAVKETY
169823684|Finegoldia magna ATCC 29328                  -------------------------------------------------GKIK--------LN-IKRGKSLKD----------------------SAFMGIMRW---A---LYNRLKEIY
160902522|Petrotoga mobilis SJ95                       -------------------------------------------------GKIK--------LD-FKRGQKFKD----------------------ATSMNIMRW---T---LYNRLKDLY
160902616|Petrotoga mobilis SJ95                       -------------------------------------------------GKIK--------LN-FKRGQKFKD----------------------ATFMGIMRW---T---LYNRLKDLY

333978211|Desulfotomaculum kuznetsovii DSM 6115        -------------------------------------------------GEVS--------LD-REGWTFAWV----------------------AHAVLGKAY-------LLALLSRFG
260892456|Ammonifex degensii KC4                       -------------------------------------------------GGAT--------LD-RRGKTFAWV----------------------AHAVLGKAY-------LLALLSRFG
260893631|Ammonifex degensii KC4                       -------------------------------------------------GEAA--------LD-RRGKTFAWV----------------------AHAVLGKAY-------LLALLSRFG
302391857|Acetohalobium arabaticum DSM 5501            DLK----------------------------------EQDFRK------KLPKVRYKNVKQILESKVDISLKD----------------------AGRVNSYRY---K---LLDQLKSIC
428778981|Dactylococcopsis salina PCC 8305             GKP----------------------------------D-----------LLKR--------VL-AEAPKPLDS----------------------GSAVNSTRY---A---IVEMAKNLC
260892735|Ammonifex degensii KC4                       ELK----------------------------------DSGRAVDRKRAENVPK--------VL-ARLKEPLKA----------------------AACMNATRY---V---LVERLRALG
260893993|Ammonifex degensii KC4                       ELK----------------------------------NSGRAVDRKRAENIPK--------VL-ARLKEPLKA----------------------AACMNATRY---V---LVERLRALG
433655673|Thermoanaerobacterium
thermosaccharolyticum M0795                            -------------------------------------------------GYPD--------IQ-KQAKQPLKD----------------------AAMINATRW---K---TYNLLKEMY
332982963|Mahella australiensis 50-1 BON               -------------------------------------------------GYPE--------IQ-KQAKQPLKD----------------------AAMMNATRW---A---LYNNLKKFG
332983053|Mahella australiensis 50-1 BON               -------------------------------------------------GYPE--------IQ-KQAKQPLKD----------------------AAMMNATRW---A---LYDNLKKFG
222530385|Caldicellulosiruptor bescii DSM 6725         -------------------------------------------------GYPE--------IQ-KLVKAPLRD----------------------CALVNATRW---R---IYEVLKNTG
298674346|Methanohalobium evestigatum Z-7303           -------------------------------------------------GHPE--------VY-EKAKKPMKD----------------------AAVINATKD---K---VLDVLKNTG
298675173|Methanohalobium evestigatum Z-7303           -------------------------------------------------GYPE--------VH-EKARKSLKD----------------------AAVINATKW---K---VLDVLKNTG
292490518|Nitrosococcus halophilus Nc 4                KRK----------------------------------N-----------KLEQQRSLAMRAIL-KGHRPTLKD----------------------AAAVNATRY---A---VGRVIKQLI
428778759|Dactylococcopsis salina PCC 8305             KKP----------------------------------T-----------ILES--------VL-KQAKSPLKD----------------------ASAVNATRN---K---IVEVISQHS
300114477|Nitrosococcus watsonii C-113                 -------------------------------------------------GYPK--------IQ-AKAKLPLRD----------------------A-AVNATRY---A---IGRAIQSLS
292491781|Nitrosococcus halophilus Nc 4                -------------------------------------------------GYTQ--------IQ-SKSKLPLKD----------------------AAAVNATRY---A---IGCTIQSVG
292491869|Nitrosococcus halophilus Nc 4                -------------------------------------------------GYTQ--------IQ-SKSKLPLKD----------------------AAAVNATRY---A---IGCTIQSVG
292491893|Nitrosococcus halophilus Nc 4                -------------------------------------------------GYTQ--------IQ-SKSKLPLKD----------------------AAAVNATRY---A---IGCTIQSVG
292491036|Nitrosococcus halophilus Nc 4                -------------------------------------------------GYPH--------LQ-TKANQPLKD----------------------AAAVNATRY---A---IGRAIQSIG
292490884|Nitrosococcus halophilus Nc 4                -------------------------------------------------GYSQ--------IQ-SKANQPLKD----------------------AAAVNATRY---A---VGRAIQSVG
292491492|Nitrosococcus halophilus Nc 4                -------------------------------------------------GYSQ--------IQ-SKANQPLKD----------------------AAAVNATRY---A---VGRAIQSVG
297563003|Nocardiopsis dassonvillei DSM 43111          NP-----------------------------------R-----------ALAG--------IS-AQTKAPLRD----------------------AAAVDATRW---A---LWRSLDEC-
374989417|Streptomyces bingchenggensis BCW-1           DK-----------------------------------S-----------KLAK--------IL-AQAKAPLRD----------------------AAAVQSTRW---A---LWRALDQM-
374989654|Streptomyces bingchenggensis BCW-1           DK-----------------------------------T-----------KLAK--------ML-SRAKAPLRD----------------------AAAVQSTRW---A---LWRALDQM-
271963428|Streptosporangium roseum DSM 43021           AKP----------------------------------T-----------LLAK--------IL-KQAKAPLRD----------------------AAAVNATRW---A---LWRALEATG
271964894|Streptosporangium roseum DSM 43021           DRP----------------------------------V-----------VLVK--------IL-QQSKAPLRD----------------------AAAVNATRW---A---LWRALTATG
258511478|Alicyclobacillus acidocaldarius DSM 446      HDQ----------------------------------E-----------RLHR--------IQ-AQLKLSLRD----------------------AAAMNATRW---A---VFRRLKDTG
160902264|Petrotoga mobilis SJ95                       NNP----------------------------------E-----------RLKQ--------IK-SESKRPLKD----------------------TAALNATRW---Y---IFNQLRGNS
121604978|Polaromonas naphthalenivorans CJ2            KDP----------------------------------K-----------RLAK--------VL-KQAKAPLLD----------------------AAAVNTTRW---T---LFGALKQTG
347754630|Chloracidobacterium thermophilum B            KDP----------------------------------K-----------RLAR--------IE-AQRKAPLRD----------------------AAAVNSTRW---E---LWRRLTATA
218667603|Acidithiobacillus ferrooxidans ATCC 23270    KFP----------------------------------E-----------RIRK--------IL-AQAKAPLRD----------------------AAAVNSTRW---A---LFQRLKATN
218667415|Acidithiobacillus ferrooxidans ATCC 23270    HDP----------------------------------K-----------RLTR--------IE-ASRKAPLKD----------------------TAAVNSTRW---A---LWRQLVATG
344201059|Acidithiobacillus ferrivorans SS3            HDQ----------------------------------K-----------RLAR--------IE-SQRKAPLRD----------------------AAAVNSSRW---A---LLQRLEETG
344201127|Acidithiobacillus ferrivorans SS3            HDT----------------------------------K-----------RLAH--------IE-AVRKAPLKD----------------------AAVINSTRW---A---LWQRLVDTG
288940385|Allochromatium vinosum DSM 180               RQP----------------------------------A-----------LLRQ--------IQ-AQAQAPLQD----------------------AAAVNATRW---A---LFAALKAQG
288941921|Allochromatium vinosum DSM 180               RQP----------------------------------V-----------RLHQ--------IQ-AQARAPLQD----------------------AAAVNATRW---A---LLNALQTTG
288940786|Allochromatium vinosum DSM 180               RQP----------------------------------A-----------LLHR--------IQ-AQAKAPLQD----------------------AAAVNATRW---A---LLNALQTTG
288941171|Allochromatium vinosum DSM 180               RQP----------------------------------A-----------LLHR--------IQ-AQAQAPLQD----------------------AAAVNATRW---A---LLNALQTTG
288941306|Allochromatium vinosum DSM 180               RQP----------------------------------A-----------LLHR--------IQ-AQAQAPLQD----------------------AAAVNATRW---A---LLNALQTTG
121582913|Polaromonas naphthalenivorans CJ2            RKP----------------------------------E-----------LFKR--------IL-GQAKRPLKD----------------------AAAVNSTRW---A---LVNALKTTG
121604604|Polaromonas naphthalenivorans CJ2            RKP----------------------------------E-----------LLKR--------IL-GQAKRPLKD----------------------AAAVNSTRW---A---LVNALKTTG
91791204|Polaromonas sp. JS666                         KDK----------------------------------K-----------RLTG--------IL-AQAQKPLRD----------------------AAAVNVTRW---A---LANALKATG
91791283|Polaromonas sp. JS666                         RDK----------------------------------K-----------RLDR--------IL-AQATKPLRD----------------------AAAVNATRW---T---LANALKATG
121582916|Polaromonas naphthalenivorans CJ2            KDP----------------------------------K-----------RLAR--------IL-AQAQRPLRD----------------------AAAVNVTRW---A---LANALKTTG
121582983|Polaromonas naphthalenivorans CJ2            KDP----------------------------------A-----------RLAR--------IL-AQAKKPLRD----------------------AAAVNATRW---A---LAEALKATS
428304158|Crinalium epipsammum PCC 9333                GKP----------------------------------D-----------LQLR--------IL-SQAKTPLKD----------------------AAAVNATRW---K---LYRKLQLTG
428204921|Chroococcidiopsis thermalis PCC 7203         KDL----------------------------------T-----------RLER--------IK-RQSKTPLKD----------------------AAAVNSTRW---K---LYNTLKEM-
307153539|Cyanothece sp. PCC 7822                      GKQ----------------------------------D-----------LAKK--------IL-DQAKKPLAD----------------------TAAVNATRW---E---LFNQLKQFN
428216193|Oscillatoria acuminata PCC 6304              AKS----------------------------------D-----------ILNR--------IL-KQSKTPLKD----------------------AAAVNATRW---K---LYQQLKETG
414079584|Anabaena sp. 90                              KKP----------------------------------E-----------LLKQ--------IL-SQAQRTLKD----------------------ASAVNSTRW---A---LFNKLKETG
172055147|Cyanothece sp. ATCC 51142                    GKP----------------------------------E-----------VLKR--------VK-ANAKKPLRD----------------------AAAVNAARW---A---VFNRLKASG
428316510|Oscillatoria nigro-viridis PCC 7112          KKP----------------------------------D-----------ILSR--------IL-KQASQPLKD----------------------ANAVNSTRW---A---LFQQLKQTG
428314951|Oscillatoria nigro-viridis PCC 7112          QQP----------------------------------E-----------LFDR--------IA-KQAKQPLKD----------------------AAAVNSTRW---Q---LFNRLKETS
307591386|Cyanothece sp. PCC 7822                      NQP----------------------------------S-----------LLEK--------IL-KQSKQSLKN----------------------VAAVNTTRW---A---LFNKLKETG
75812341|Anabaena variabilis ATCC 29413                KDK----------------------------------P-----------RLEK--------IL-ALTKKPLAD----------------------AAAVNATRF---A---LLYALKSTG
17233293|Nostoc sp. PCC 7120                           KDK----------------------------------S-----------RLEK--------IL-ALTKKSLAD----------------------AAAVNTTRF---A---LLEVLKSTG
75812782|Anabaena variabilis ATCC 29413                KDP----------------------------------T-----------RLQK--------IL-AQAKKPLAD----------------------AAAVNATRY---K---LLEVLKSTD
434407609|Cylindrospermum stagnale PCC 7417            KDP----------------------------------S-----------RLEK--------PI-KQVKRPLSD----------------------AAAVNTTRL---A---LLFVLKSTG
414078405|Anabaena sp. 90                              KDP----------------------------------T-----------RLQK--------IL-KQAKRPLAD----------------------AAAVNATRW---A---LFRVLKDTG
440685156|Anabaena cylindrica PCC 7122                 KDP----------------------------------S-----------RLEK--------IL-KQAKRPLAD----------------------AASVNTTRL---A---LLKVLKATG
428315656|Oscillatoria nigro-viridis PCC 7112          KKP----------------------------------D-----------VLKK--------IL-SQAKRPLKD----------------------AAVVNSTRW---A---LFNRLGMTG
428204926|Chroococcidiopsis thermalis PCC 7203         NKP----------------------------------D-----------VLKR--------VI-SQAKAPLKD----------------------ATAVNVTRW---A---LFNALKELG
218440244|Cyanothece sp. PCC 7424                      SKS----------------------------------D-----------LLNH--------VL-KQAKSPLKD----------------------AAAVNSTRW---A---LFNALKATG
428775092|Halothece sp. PCC 7418                       SKP----------------------------------D-----------VLNR--------IL-KQSQLPLKD----------------------AAAVNSTQW---A---LFNALKQTG
414075475|Anabaena sp. 90                              KKP----------------------------------D-----------LLKC--------IL-AQAKRPLKD----------------------AAAVNSTRW---A---LFSRLQETG
428305120|Crinalium epipsammum PCC 9333                KKP----------------------------------D-----------VLKR--------IL-AQAKRPLKD----------------------ASAVNSTRW---A---LMNRLKQTG
427713881|Synechococcus sp. PCC 6312                   GKP----------------------------------E-----------LLQR--------IL-RKAKAPLKD----------------------AAAVNSTRW---A---LFNALKATG

S. pyogenes Cas9                                       KKM----------------------------------K-----------NYWR--------QL-LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKH---VAQILDSRM
S. pyogenes Cas9 domains                               NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN|ll|rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
A. naeslundii Cas9                                     QKCGIPHVGVKEAIGRVRGWRKQTPNTSSEDLTRLKKE-----------VIAR--------LR-RTQEDPEID----------------------ERSMESVAW---MANELHHRIAAAY
A. naeslundii Cas9 domains                             NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN|rrrrrrrrrrrrrrrrrrrrr
```

```
312126587|Caldicellulosiruptor hydrothermalis 108      TKQLRE--------------SYGLKKTS---KKLEETFE-----------------SHCVDAW-CLAN-------------------WCTGGHVQP-----------DNTELLVIIPLRF
434408285|Stanieria cyanosphaera PCC 7437             TKQIRD--------------ELGLKK-T--KNKLAETFE-----------------AHCVDSF-ALAY-------------------SVVGGNSLP-----------KNTNLLIVVPLRF
414076241|Anabaena sp. 90                              TSNLRK--------------YLGLHKEK--SDKSLQIPE-----------------THAVDAV-ALASSEFVKYQTWEGAK--NHGASWIGGVDITESQF------------TIVRRPPI
428299778|Calothrix sp. PCC 6303                       TSNLRK--------------HLGMHKEK--TDKSLQIPE-----------------THAVDAI-TLASSEFVKYKSFEGKN--TRGASWVGNVSITESQF------------TIVRRPPI
 17233043|Nostoc sp. PCC 7120                          TASTRK--------------YLGLTKNS--SDKSKSEFN-----------------THAVDGV-AIAATAFVEYRQYHTVK--VNGGDWFGNVVITFALF------------KLIRRPPY
166363864|Microcystis aeruginosa NIES-843              TSNLRK--------------YLGLEKS---KNKAEQSPE-----------------SHANDGI-ALACFQFLDYWPFHNYN--GHGYDWKGSVKVTNAPF------------AVIKRPPI
166367659|Microcystis aeruginosa NIES-843              TSNLRK--------------YLGLEKS---KNKAEQSPE-----------------SHANDGI-ALACFQFLDYWPFHNYN--GHGYDWKGSVKVTNAPF------------AVIKRPPI
307150343|Cyanothece sp. PCC 7822                      TSNLRQ--------------HLGLEKS---KDKAKQSKE-----------------SHANDGL-TLACFQFLDYKSFQTVN--SQGHHWVGSVAITSCPF------------AVIKRPPF
307153352|Cyanothece sp. PCC 7822                      TSNLRQ--------------HLGLEKS---KDKAKQSKE-----------------SHGLDGL-TLACFQFLDYKSFQTVN--SQGHHWVGSVTITNCPF------------AVIKRPPF
428306661|Crinalium epipsammum PCC 9333                TSNLRK--------------HLGLEKS---KNKAEKSPK-----------------SHANDGI-ALACFRFLDYLPFHTAN--SHGHEWQGKVTLTSAIF------------VVVKRPPI
307150912|Cyanothece sp. PCC 7822                      TSNLRK--------------HLGLEKS---KDKAKRSPE-----------------SHANDGI-TLACFRFLDYREFHTHN--SRGHDWIGKVQLTPSPF------------AVVKRPPI
428775289|Halothece sp. PCC 7418                       TSKRRE--------------LLGLEKD---KQKSNQTPA-----------------THAVDGV-ALASHEFLKYKFLKGDQGW-----FEGSVEITNAPF------------SVIKRPPV
428779516|Dactylococcopsis salina PCC 8305             TSARRE--------------MLGLEKD---REKSNQTPA-----------------THAVDGV-ALASYQFLKYKRIKGYKGW-----FEGSVEITNAPF------------STIKRPPV
428778599|Dactylococcopsis salina PCC 8305             TAKRRD--------------TLGLEKD---RNKSSQTPA-----------------THAVDGV-ALASYQFLKYKYLKGDKGW-----FEGSVEITNAPF------------SVIKRPPI
428781696|Dactylococcopsis salina PCC 8305             TARRRD--------------SLGLEKD---REKSNQNPA-----------------THAVDGV-ALASHDFLKYKYLKGDKGW-----FEGSVEITNAPF------------SVIKRPPV
428215045|Oscillatoria acuminata PCC 6304              TASLRT--------------QLGLVKS---QDKSSKTPE-----------------SHAVDGV-TLACSEFIKYKPFQTAK--THGHTWKGSVQLTHAAF------------VVIRRPPI
428313425|Microcoleus sp. PCC 7113                     TAAIRQ--------------QIGLVKEK--KDKSNQSPE-----------------THAVDGI-SLASSQFIQYRQLKGKQGW-----WDGFVSITPAPF------------VVIRRPPV
158340109|Acaryochloris marina MBIC11017               KDGNGTS--------QIRK-HLGLIKC---KDKKAQAPE-----------------AHAVDGI-ALAASQFTRYQASQSDSMD-----WVGSIDVTPCQF------------RVITRPAY
428775002|Halothece sp. PCC 7418                       KDGNGTS--------QLRT-ALGLVKDK--QNKSEAKPE-----------------THSVDGV-ALACGYFIQFRRFHSSCGHGKT--WKGKVTTTNSQF--RIITRP-GAVKRGKEYGV
307150700|Cyanothece sp. PCC 7822                      TCNLRK--------------HLGLEKS---KDKAKQSPE-----------------SHANDGI-ALACFHFLDYREFHTHN--SRGHDWIGKVYLTPSPF------------VVVKRPPI

312126663|Caldicellulosiruptor hydrothermalis 108      ---------------PWLKFVPASDVAQMRKRL--GLEK-----------------GHANDA--LATAAVFC--NCTETDRTHMIELTLVKFRRHNRARIHAVRD---------------
312621424|Caldicellulosiruptor kronotskyensis 2002     ---------------PWLKFVPASDVVQMRKRL--GLEK-----------------GHANDA--LATAAVFC--NCTETDRTHMIELNMVKFRRHNRARIHAVRD---------------
222530402|Caldicellulosiruptor bescii DSM 6725         ---------------PWLKFVPASDAAQMRKKL--GLKK-----------------GHANDA--LATAAVFC--SCTEADRTHMIELTLVKFRRHNRARIHAVRD---------------
222530431|Caldicellulosiruptor bescii DSM 6725         ---------------PWLKFVPASDAAQMRKKL--GLKK-----------------GHANDA--LATAAVFC--SCTEADRTHMIELTLVKFRRHNRARIHAVRD---------------

307591386|Cyanothece sp. PCC 7822                      ---------------LSIETGTGGRTKYNRCRL--NLEK-----------------RHFIDA--GCVGNLES---LKLLTRQ---PLLIKATGHGNRQMCGINKY---------------
302391131|Acetohalobium arabaticum DSM 5501            ----------------KLEITTGGDTANRRIDY--NIEK-----------------THSNDA--ICITGLLP---IDRLDIK---GYFIKPLRKKRASKIDSLGD---------------
435854579|Halobacteroides halobius DSM 5150            ----------------NLEITTGGDTANKRIDY--DIKK-----------------SHSNDS--ICITNLLP---VDNLDVK---EYFIKPLRKKSKAKIKELKG---------------
435852864|Halobacteroides halobius DSM 5150            ----------------QLEITSGGNTANKRIDY--EIEK-----------------SHSNDA--ICTTGLLP---VDNIDDIK--EYYIKPLRKKSKAKIKELKC---------------
435854272|Halobacteroides halobius DSM 5150            ----------------QLEITSGGNTANKRIDY--EIEK-----------------SHSNDA--ICTTGLLP---VDNIDDIK--EYYIKPLRKKSKAKIKELKC---------------
313123202|Lactobacillus delbrueckii  ND02              -------------------FQYGSQTYVDRKNL--LLPK-----------------GHFMDA--IAISGIKP---VGQMPET---VTMISQFRKKKRSLHEATAR---------------
77362205|Pseudoalteromonas haloplanktis TAC125        ---------------WDFEETFGYVTKFKREQV-LGLPK-----------------THYYDAVSICCEDGEL----VQQGKH---VLRKRHVASGDYQQTK-------------------
298675374|Methanohalobium evestigatum Z-7303           ---------------WKFQETFGYITKYKREQK-LKLEK-----------------THYNDAVAICCDENNP---NLQLSNV---VYYKKHVNKGDYQQTK-------------------
298674707|Methanohalobium evestigatum Z-7303           ---------------WKFQETFGYITKYKREQI-LGLEK-----------------THYNDAVAICCDENNP---NLQLSNV---VYYKKHVSSGDYQQTK-------------------
298675090|Methanohalobium evestigatum Z-7303           ---------------WKFQETFGYITKYKREQK-LNLEK-----------------THYNDAVAICCDENNP---NIQLSDV---VYYKKHVSKGDYQQTK-------------------
169834784|Clostridium botulinum A3 str. Loch Maree     --------------------TYGYITKSNRIAL--KLEK-----------------THYNDA--FCIAGSSN----QNRVKP----LIFEQIKRNSRSLEKFYDA---------------
229587454|Clostridium botulinum Ba4 str. 657           --------------------TYGYITKHNRIAL--NLEK-----------------TNYNDA--FCIAGGSN----QNRVKP----LIFEQIKRNSRSLEKFYDA---------------

297585104|Bacillus selenitireducens MLS10              ---------------PDAKLTYGYLTKNTRIEH--SLEK-----------------SHAVDA--RCISGHPT---AKPSEAT----YQYKQVRKNNRQLHKTTIL---------------
188587532|Natranaerobius thermophilus JW-NM-WN-LF      ---------------PDVNLTFGYITKNTRIKN--NLPK-----------------SHAVDA--RCISGNPL---AKPPVNT----LYIKQVRKNNRQLHKFNPS---------------
312793946|Caldicellulosiruptor kristjanssonii 177R1B   ----------------EVEITYGSITKAKRVEM--GLER-----------------THRNDAF-VIAGGNKNV--NRVPEW-----YFGKFFRRQNRSLHKANPI---------------
344996431|Caldicellulosiruptor lactoaceticus 6A        ----------------EVEITYGSITKAKRVEM--GLER-----------------THRNDAF-VIAGGNKNV--NRVPEW-----YFGKFFRRQNRSLHKANPI---------------
217077332|Thermosipho africanus TCF52B                 ----------------DVGITFGSITKSKRIEM--GLEK-----------------THRNDA--FVIADGNK---NINRTTK---WYFGKFFRRQNRSIHKANPI---------------
375009168|Geobacillus thermoleovorans CCB US3 UF5      ---------------SHANITYGYITKCIRIAN--GLSK-----------------SHMVDA--RCISGNPL---AKPSGTV----YLLKFVRKNNRQLHKATIL---------------
407477159|Exiguobacterium antarcticum B7               ---------------PYVKLTYGFLTKNTRIEN--GLEK-----------------RHAVDA--RCISGNPL---AEAPKVS----YLFRQVRANNRQLHKMTIG---------------
172057416|Exiguobacterium sibiricum 255-15             ---------------PYVKLTYGFLTKNRRIQN--GLEK-----------------RHAVDA--RCISGNPL---AEEPKVS----YLFRQVRANNRQLHKMTVG---------------
169823684|Finegoldia magna ATCC 29328                  ---------------DDVSLTYGYLTKNKRIEN--NLPK-----------------EHKIDA--FCIANNLS---AKLSDTT----FYFKKVRCHNRQIHKANYL---------------
160902522|Petrotoga mobilis SJ95                       ---------------SNVHLTYGYITKNTRIQN--SLTK-----------------THRIDA--LCISGHPQ---AKQLDYY----YHIKEVRRHNRQIHKAKIL---------------
160902616|Petrotoga mobilis SJ95                       ---------------SNVHLTYGYITKNTRIQN--NLPK-----------------THRIDA--LCISGHPQ---AKQLDYY----YHIKEVRRHNRQIHKANIL---------------

333978211|Desulfotomaculum kuznetsovii DSM 6115        ----------------KVRMCEGWRTKEWREGL--GLAK-----------------SHANDA--ACLFPPGG-------------PLKIFGPEYLIYPLRRRKWE---------------
260892456|Ammonifex degensii KC4                       ----------------EVRVAEGWQTKEWRESL--GLPK-----------------SHTNDA--LSLFLPGG-------------PLKVFGPEYLIWPLRRRRWE---------------
260893631|Ammonifex degensii KC4                       ----------------EVRVAEGWQTKEWRESL--GLPK-----------------SHTNDA--LSLFLPGG-------------PLKVFGPEYLIWPLRRRKWK---------------
302391857|Acetohalobium arabaticum DSM 5501            ---------------SNIETSTGAKTKYNRNQV-AELEK-----------------THYFDA--LCVGTAKQ-EYSFQKGFK---VLSIKATGRGTRQRTLLDKY---------------
428778981|Dactylococcopsis salina PCC 8305             ---------------DHVKCWTGGRTKYNRVNQ--ELPK-----------------SHSIDA--ACVGESGS--KIKLLVSQ---PLIVESKGHGTKQARRANQQ---------------
260892735|Ammonifex degensii KC4                       ---------------PPVLTATTAQTKYNRARF--NLPK-----------------THYFDA--CCVGEVSG--ELEVAAGY---VQVFRAVGRGTRRVANLDRH---------------
260893993|Ammonifex degensii KC4                       ---------------PPVLTATAAQTKYNRARF--NLPK-----------------THYFDA--CCVGEVSG--ELEVAAGY---VQVFRAVGRGTRQMANLDRH---------------
433655673|Thermoanaerobacterium
thermosaccharolyticum M0795                            ----------------PVECGTGALTKKNRIDR--NLPK-----------------EHYCDA--CCVGQSTP-KEFVFKADY---ILEFHAKGRGTRQRTLLNKY---------------
332982963|Mahella australiensis 50-1 BON               ---------------LPVECGTGAMTKKQRLAH--KLPK-----------------THYYDA--CFVGASTS-DEITISTEY---VSIWSAIGRGNRQMCRTDKY---------------
332983053|Mahella australiensis 50-1 BON               ---------------LPVECGTGAMTKKQRLAH--KLPK-----------------THYYDA--CCVGASTS-DEITISTEY---VSIWSAIGRGTRQMCRTDKY---------------
222530385|Caldicellulosiruptor bescii DSM 6725         ---------------LPVECGSGALTKMNRIKL--GLPK-----------------DHHFDA--ICVGYSTP---NRIWLKT-KTVLHVIAKGRGTRQIAILDRY---------------
298674346|Methanohalobium evestigatum Z-7303           ---------------LPIECGTGALTKMNRIKL--NLPK-----------------DHHFDA--CCVGESTP-DELRFKTNS---VLHIHAKGRGSYKRCKLDKY---------------
298675173|Methanohalobium evestigatum Z-7303           ---------------LPVECGTGALTKMNRIKL--NLPK-----------------DHHFDA--CCVGESTP-DELRFKTNS---VLHIHAKGRGSYKRCKLDKY---------------
292490518|Nitrosococcus halophilus Nc 4                ---------------PDIRFCSGGQTKRNRTLQ--GYPK-----------------AHWIDA--ACVGEQGQ--QVRLECSG---YLQVSAKGHGARQQCRMDKY---------------
428778759|Dactylococcopsis salina PCC 8305             ----------------SVKTATGAQTKMNRVRL--GLPK-----------------EHCLDA--ACVGEVET---LTILSHQ---PLMVTATGTGGRQKCQTNKY---------------
300114477|Nitrosococcus watsonii C-113                 ---------------LPTSFWSGGRTKKNRLSQ--GYPK-----------------DHWIDA--VCVGENGA-KVTITEDVR---ALHIKAMGRGARQVVRTDKH---------------
292491781|Nitrosococcus halophilus Nc 4                ---------------LPTSFWSGGRTKKNRISQ--GYTK-----------------DHWIDA--ACVGESAG-QVTITEGYR---ALHIKATGRGTRQVVRTDKY---------------
292491869|Nitrosococcus halophilus Nc 4                ---------------LPTSFWSGGRTKKNRISQ--GYTK-----------------DHWIDA--ACVGESAE-QVTITEGYR---ALHIKATGRGTRQVVRTDKY---------------
292491893|Nitrosococcus halophilus Nc 4                ---------------LPTSFWSGGRTKKNRISQ--GYTK-----------------DHWIDA--ACVGESAE-QVTITEGYR---ALHIKATGRGTRQVVRTDKY---------------
292491036|Nitrosococcus halophilus Nc 4                ---------------LPTSFWSGGRTKKNRLSQ--GYAK-----------------DHWIDA--ACVGESGE-QVTLAEGDR---PLQIKATGRGTRQVVRTDKY---------------
292490884|Nitrosococcus halophilus Nc 4                ---------------LPTSFWSGGRTKKNRISQ--SYGK-----------------DHWIDA--ACVGEIGE-QVTLTEGYR---ALQIKATGRGTRQVARTDKY---------------
292491492|Nitrosococcus halophilus Nc 4                ---------------LPTSFWSGGRTKKNRISQ--SYGK-----------------DHWIDA--ACVGEIGE-QVTLTEGYR---ALQIKATGRGTRQVARTDKY---------------
297563003|Nocardiopsis dassonvillei DSM 43111          ---------------PPTRVGSGGRTKWNRTRD--HLPE-----------------SHTLDA--LAVGRVEA---VTTAVST---VLVVGAAGRGPYARTRADRH---------------
374989417|Streptomyces bingchenggensis BCW-1           ---------------LPTHIASGGRTKWNRTRC--GLPK-----------------SHTLDA--LNVGKLDT---VTETANP---ILVAGCSGRGQHQRTSPDRH---------------
374989654|Streptomyces bingchenggensis BCW-1           ---------------LPPHIASGGRTKWNRTRC--GLPK-----------------SHTLDA--LNVGKLDT---VTETVTP---ILVAGCSGRGQHQRTSPDRH---------------
271963428|Streptosporangium roseum DSM 43021           ---------------LPVATASGARTKWNRSRT--GAAK-----------------SHALDA--LHVGHLET---VAGWPST---VLVVTATGRGSYARTTPDRY---------------
271964894|Streptosporangium roseum DSM 43021           ---------------LPVATASGGRTKWNRSRT--GAAK-----------------SHTLDA--LHVGALDH---VTGWPSM---VLVIAATGRGTYARTRADRY---------------
258511478|Alicyclobacillus acidocaldarius DSM 446      ---------------LPVEAGSGGRTKYNRFSQ--GYPK-----------------AKWIDA--ACVGESGQ-CVRLDTQMQ---VLTIVAKGHGKRQRCITDKY---------------
160902264|Petrotoga mobilis SJ95                       LTAGKEE--------LPIEVGTGGRTKYNRETQ--NYPK-----------------KHWIDA--ACVGESGQ-NVQIEPDMQ---VLEIKAMGHGMRRMCFVDKY---------------
121604978|Polaromonas naphthalenivorans CJ2            ---------------LPVETGTGGQTKYNRCRL--SIPK-----------------THALDA--ACVGQVQS---VNNTNAP---MLRVICTGRVSRSKTRLDKY---------------
347754630|Chloracidobacterium thermophilum B           ---------------LPVEAGTGGRTKWNRHRL--NLPK-----------------AHALDA--ACVGEVHA---VTNWRVP---TLNIQCTGRGSYQRTRLTKH---------------
218667603|Acidithiobacillus ferrooxidans ATCC 23270    ---------------LPVEVASGGRTKWNRHRL--QVPK-----------------EHCLDA--ACVGHVDT---IEHWNQS---VLSIKATGRGSYQRTRLTRH---------------
218667415|Acidithiobacillus ferrooxidans ATCC 23270    ---------------LDVEVGTGGRTKWNRSRL--QIPK-----------------EHCLDA--ACVGHVDG---LEHWQQP---VLGIKATGRGSYQRTRLTKH---------------
344201059|Acidithiobacillus ferrivorans SS3            ---------------LDVEVGSGGGTKWNRKRF--SVPK-----------------AHCLDA--ACTGHVDA---IQDRQQP---VLSIKATGRGRYKRTQLTKY---------------
344201127|Acidithiobacillus ferrivorans SS3            ---------------MDVEAASGGRTKWNRTRF--SMPK-----------------AHCLDA--ACVGHVDA---IDAWRQP---VLSVKSTGRGSYQRTRLTKH---------------
288940385|Allochromatium vinosum DSM 180               ---------------LPVETGSGGRTKFNRSRL--NLPK-----------------THALDA--SCVGAVDQ---VRDWNRP---VLAIRATGRGTYSRTRLDRF---------------
288941921|Allochromatium vinosum DSM 180               ---------------LAVETGSGGRTKFNRTRL--DIPK-----------------THALDA--ACVGAVDQ---VRDWNRP---VLAIRATGRGAYRRTRTFNN---------------
288940786|Allochromatium vinosum DSM 180               ---------------LAVETGSGGRTKFNRTRL--DIPK-----------------THALDA--ACVGAVDQ---VRDWNRP---VLAICATGRGAYSRTRTFNN---------------
288941171|Allochromatium vinosum DSM 180               ---------------LAVETGSGGRTKFNRTRL--DIPK-----------------THALDA--ACVGAVDQ---VRDWNRP---VLAIRATGRGAYRRTRTFNN---------------
288941306|Allochromatium vinosum DSM 180               ---------------LAVETGSGGRTKFNRTRL--DIPK-----------------THALDA--ACVGAVDQ---VRDWNRP---VLAIRATGRGAYRRTRTFNN---------------
121582913|Polaromonas naphthalenivorans CJ2            ---------------LVVETASGGRTKFNRIQL--GIPK-----------------THALDA--ACVGYVNS---ITGWRKP---TLCLKAMGRGSYQRTRLDKY---------------
121604604|Polaromonas naphthalenivorans CJ2            ---------------LVVETASGGRTKFNRMAL--GIPK-----------------THALDA--ACVGHVNA---MTGWHKP---TLRLKAMGRGSYQRTRLDKY---------------
91791204|Polaromonas sp. JS666                         ---------------LPVELASGGRTKFNRCTL--GVPK-----------------THALDA--ACVGQVRF---IQDWSRP---TLTLKAMGRGSYQRTRLDRF---------------
91791283|Polaromonas sp. JS666                         ---------------LPVELASGGQTKFNRCRL--GVPK-----------------THALDA--ACVGQVRF---IQDWSKP---TLTLKAMGRGSYQRTRLDKF---------------
121582916|Polaromonas naphthalenivorans CJ2            ---------------FPLELASGGRTKFNRCTL--DVPK-----------------THALDA--ACVGQVEA---ITGWQQPAFYTLTIKAMGRGSYQRTRLDAY---------------
121582983|Polaromonas naphthalenivorans CJ2            ---------------LPLELASGGQTKFNRCTL--GMPK-----------------THALDA--ACVGRVDA---IAGWQNHASATLTIKAMGRGSYQRTRLDGF---------------
428304158|Crinalium epipsammum PCC 9333                ---------------LPVEVSSGGRTKFNRTKC--GIEK-----------------THWTDA--ACVGASTP-EKLLINGIK---PLTIKAKGHGIRQRCRPDKY---------------
428204921|Chroococcidiopsis thermalis PCC 7203         ---------------LPITTGTGGQTKFNRIRF--GLKK-----------------QHWIDA--VCVGNIEA---IKLLTQQ---PLRIKCTGWGTRKMCSTDEY---------------
307153539|Cyanothece sp. PCC 7822                      ---------------LPVEVGTGGLTKFNRCRQ--NLDK-----------------THWLDA--ACVGQSTP-EQLIIKGVR---PLLIKANGHGTRQMCGTNKY---------------
428216193|Oscillatoria acuminata PCC 6304              ---------------LTVETGTGALTKYNRTRL--NLPK-----------------THWFDA--ACVGQSTP-SSLKIQVLK---PLQIQATGRGNRQRVLPNKY---------------
414079584|Anabaena sp. 90                              ---------------LPITTGSGGLTKFNRTRL--GLPK-----------------THWIDA--ACVGKVET---LKILTTK---ILTVKSTGHSCRRFCRINKF---------------
172055147|Cyanothece sp. ATCC 51142                    ---------------LEVEAGSGGLTKFNRTKH--ELPK-----------------THWLDA--ACVGKSTP-SKLRMLVQK---PLKIKATGHGTRQMCRTNKY---------------
428316510|Oscillatoria nigro-viridis PCC 7112          ---------------LAIEVSTGGKTKYNRSRL--GLPK-----------------THWLDA--ACVGNVEV---LQVFTKQ---PLLIAAKGWGSRQMCTTNKY---------------
428314951|Oscillatoria nigro-viridis PCC 7112          ---------------LPVEIGTGGQTKYNRTRL--ELPK-----------------THWLDA--ACVGLVSQ---LKILTAG---PLLIQASGWGSRQICQPNKY---------------
307591386|Cyanothece sp. PCC 7822                      ---------------LSIETGTGGRTKYNRCRL--NLEK-----------------RHFIDA--GCVGNLES---LKLLTRQ---PLLIKATGHGNRQMCGINKY---------------
75812341|Anabaena variabilis ATCC 29413                ---------------LPIETGSGGLTKFNRTQQ--ALNK-----------------SHWLDA--CCVGASTP--ILKIKGIK---PLLITANGHGTRQSCRTDKY---------------
17233293|Nostoc sp. PCC 7120                           ---------------LPVEIGSGGLTKFNRTQQ--GLNK-----------------SHWLDA--CCVGASTP--ILKIKGIR---LLLITANGHGTRQSCRTDKY---------------
75812782|Anabaena variabilis ATCC 29413                ---------------LPVECGSGGLTKFNRTNQ--QLQK-----------------THWLDA--ACVGQSTP--ILIIKGIK---PLLITANGHGTRQMCRTDKF---------------
434407609|Cylindrospermum stagnale PCC 7417            ---------------LPVEIGSGGLTKFNRSQL--KLEK-----------------SHWLDA--ACVGKSTP--VLKVKGIK---DLLIKANGHGTRQSCRTDKF---------------
414078405|Anabaena sp. 90                              ---------------LPVETGSGGLTKFNRSQQ--NLAK-----------------THWIDA--ACVGQSTP--ILTIKGIK---PLLIAANGHGSRQSCRTDKF---------------
440685156|Anabaena cylindrica PCC 7122                 ---------------LPVETGSGGLTKFNRSQQ--KLEK-----------------THWIDA--ACVGQLTP--ILNIKGVK---PLLITANGHGTRQSCRTDKY---------------
428315656|Oscillatoria nigro-viridis PCC 7112          ---------------LPVSTGSGGLTKFNRTRL--QLPK-----------------SHWLDA--ACVGRVEE---LKVLTSQ---PLLIKAAGHGTRQMCGTDRF---------------
428204926|Chroococcidiopsis thermalis PCC 7203         ---------------LPVVTSSGGKTKFNRSRL--GLPK-----------------THFFDA--ACVGETLE---LQVCTRE---PLLIKTTGHGTRQMCGTDKF---------------
218440244|Cyanothece sp. PCC 7424                      ---------------LPVTTGTGGQTKFNRTRL--NLPK-----------------AHWTDA--ACVGRIDS---LLILTSK---PLLIAAKGHGTRQMCRTDKY---------------
428775092|Halothece sp. PCC 7418                       ---------------LPVATGTGGQTKFNRTRL--DLPK-----------------THWLDA--ASVGEIDS---LKVLTTK---PLLIAAKGHGTRQMCGTDKF---------------
414075475|Anabaena sp. 90                              ---------------LPVFTGSGGQTKFNRTRL--NFPK-----------------THWLDA--ACVGQIGE---LEILTNK---PLLIKATGHGTRQMCRTDKF---------------
428305120|Crinalium epipsammum PCC 9333                ---------------LPVLAASGGKTKFNRCRL--LLPK-----------------THWLDA--ACVGEVEQ---LEVKTTQ---PLLIKATGHGTRQMCGTDRF---------------
427713881|Synechococcus sp. PCC 6312                   ---------------LPLTTGTGGQTKFNRTRL--DLPK-----------------AHWLDA--ACVGQVDA---LKVLTSK---PLLITAKGHGTRQMCGTDKF---------------

S. pyogenes Cas9                                       NTKYDENDKLIRE--VKVITLKSKLVSDFRKDF--QFYK----------VREINNYHHAHDA--YLNAVVGT
S. pyogenes Cas9 domains                               rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
A. naeslundii Cas9                                     -------------PETTVMVYRGSITAAARKAA--GIDSRINLIGEKGRKDRIDRRHHAVDA--SVVALMEA
A. naeslundii Cas9 domains                             rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr
```

```
312126587|Caldicellulosiruptor hydrothermalis 108      HRRQLHMLQFDKGGIRRRYGGTR-S-LGFKRGSYVKHPKYG---------------------------------------------------------------LCYVGGHMNG------
434408285|Stanieria cyanosphaera PCC 7437              HRRQLHRLEHAPGAIRSPYGGTM-S-AGFKRGSLVKHPKYN---------------------------------------------------------------LCFVGGSSKG------
414076241|Anabaena sp. 90                              SRRQLHLMTFSEGGNRRKYGGTT-TRHGFRKGDFVEATQGS------------------------------------------------------------KTFFGWVSGD---------
428299778|Calothrix sp. PCC 6303                       SRRQLHLMVPTKGGNRRKYGGTT-TRHGFRKGDYVKATQGK------------------------------------------------------------KTFFGWVSGD---------
 17233043|Nostoc sp. PCC 7120                          SRRQLHLMVPAKSGVRRKYGGST-TRHGLRKGDLVTS-PKG---------------------------------------------------------------IGYVSGD---------
166363864|Microcystis aeruginosa NIES-843              SRRQLHLMVFSKGGKRRKYGGST-TGHGFRKGDLVSS-PKG---------------------------------------------------------------IGYVSGD---------
166367659|Microcystis aeruginosa NIES-843              SRRQLHLMVFSKGGKRRKYGGST-TRHGFRKGDLVSS-PKG---------------------------------------------------------------IGYISGD---------
307150343|Cyanothece sp. PCC 7822                      SRRQLHLMLPSQGGLRRKYGGTT-TDFDLIKGDLVDS-PKG---------------------------------------------------------------IGYVSGQ---------
307153352|Cyanothece sp. PCC 7822                      SRRQLHLMLKSKGGIRRQYGGTI-TDFDLRKGDLVNC-PKG---------------------------------------------------------------IDYVSGQ---------
428306661|Crinalium epipsammum PCC 9333                SRRQLHLMLPYQGGIRRKYGGTT-TEFSLRKGDLVNS-SKG---------------------------------------------------------------IGYVSGQ---------
307150912|Cyanothece sp. PCC 7822                      SRRQLHLMLPSKGGIRRKYGGTT-TDFDLRKGDLVNS-PKG---------------------------------------------------------------IGYVSGQ---------
428775289|Halothece sp. PCC 7418                       SRRQLHLMIPAKGGIRRKYGGTV-TRHGFRKGDYVEVERKG------------------------------------------------------------IKYRGWVSGD---------
428779516|Dactylococcopsis salina PCC 8305             SRRQLHLMIPAKGGVRRKYGGTV-TRHDFRKGDYVETERKG------------------------------------------------------------IKYRGWVSGE---------
428778599|Dactylococcopsis salina PCC 8305             SRRQLHLMLPAKDAVRRKYGGTV-TRHNVRKGDYVETERKG------------------------------------------------------------VKFRGWVSGD---------
428781696|Dactylococcopsis salina PCC 8305             SRRQLHLMLPAKGGVRRKYGGTV-TRHGFRKGDYVETERKG------------------------------------------------------------VKFRGWVGGD---------
428215045|Oscillatoria acuminata PCC 6304              SRRQLHLMVPSKGGVRRKYGGTM-TRHGVRKGDLVRASMDR------------------------------------------------------------RVVMGWVSGD---------
428313425|Microcoleus sp. PCC 7113                     SRRQLHLMIPAKGGKRRKYGGTT-TRHDFRKGDLVKAEMAG------------------------------------------------------------RVCVGWVSGD---------
158340109|Acaryochloris marina MBIC11017               FRRALHFDNPGKGGTHKRKGGTV-TPWVFRSGDLVVGEKAG------------------------------------------------------------KVFRGWVGGY---------
428775002|Halothece sp. PCC 7418                       FRRQLHFEIPGQNGVRKRKGGTI-TPWGFRVGDLVRATKGK------------------------------------------------------------TESIGYIGGY---------
307150700|Cyanothece sp. PCC 7822                      SRRQLHLMVPSKGGIRRKYGGTT-TDFGLRKGDLVNS-PKG---------------------------------------------------------------IGYVSGQ---------

312126663|Caldicellulosiruptor hydrothermalis 108      --RLYKVNGKIVAKNRRKRTDQKEPSFADISPLSPEIQRKLK--------------------------VYPGTKILNPLRKETPTI--AGDVWIYE------PTGKRFVATGVVSQK--Y
312621424|Caldicellulosiruptor kronotskyensis 2002     --RLYKVDGKIVAKNRRKRTDQKEPSFADISPLPPEIQRKLK--------------------------VYPGTKILNPLRKEMPTI--AGEVWVHE------PTGKRFVTTGVVSQK--Y
222530402|Caldicellulosiruptor bescii DSM 6725         --RLYKVDGKIVAKNRRKRTDQKEPSFADISPLPPEIQRKLK--------------------------VYPGTKILNPLRKEMPTI--AGDVWIHE------PTGKRFVTTGVVSQK--Y
222530431|Caldicellulosiruptor bescii DSM 6725         --RLYKVDGKIVAKNRRKRTDQKEPSFADISPLPPEIQRKLK--------------------------VYPGTKILNPLRKETPTI--AGDVWIHE------PTGKRFVTTGVVSQK--Y

307591386|Cyanothece sp. PCC 7822                      --------GFP-----IRHRSRNKFYKG----------------------------------------FQTGAQVKAVVTKGKK-----VGVYLGR------V-LCR--ASGRF-----D
302391131|Acetohalobium arabaticum DSM 5501            --------------------------------------------------------------------FKHRDVIRYIKRDGTT--------YLGY------ITALR-VKNNKYNSKVCN
435854579|Halobacteroides halobius DSM 5150            --------------------------------------------------------------------FKHRDIVRYTKRNGET--------YDGY------ITSLR-IKNNKYKSKVCN
435852864|Halobacteroides halobius DSM 5150            --------------------------------------------------------------------FRQRDLVKYTKRNGET--------YTGY------ITSLR-IKNNKYNSKVCN
435854272|Halobacteroides halobius DSM 5150            --------------------------------------------------------------------FRQRDLVKYTKRNGET--------YTGY------ITSLR-IKNNKYNSKVCN
313123202|Lactobacillus delbrueckii  ND02              ------KGRKQPNMSSKRNEKNTKYACG----------------------------------------LWLNDYVRVI-----------GNHVKGY------VKGFK--SNGYYV----Y
77362205|Pseudoalteromonas haloplanktis TAC125        --------GGR-----SEKRIPVGKLFG----------------------------------------LKKHDFVQTPQG---------TGFVKGK----------R--SSGYF-----A
298675374|Methanohalobium evestigatum Z-7303           --------GVK-----SEKKIPTGKLHG----------------------------------------LKKFDFIETPYV---------SGFVKGK----------R--TSGNF-----S
298674707|Methanohalobium evestigatum Z-7303           --------GIR-----SDKKIPTGKLHG----------------------------------------LRKFDFIKTPDV---------SGFVKGK----------R--TSGNF-----S
298675090|Methanohalobium evestigatum Z-7303           --------GVR-----SEKKIPTGKLHG----------------------------------------LRKFDFIQTPDI---------SGFVKGK----------R--TSGNF-----S
169834784|Clostridium botulinum A3 str. Loch Maree     -------KVID-----IRTNTKVSGAELFNGRRTRNKNLNSENLRKYRGVKVSKGQRRIRTKRYF---YQPGDLVKYEDK---------VYTVKGT----------Q--NKGKY-----I
229587454|Clostridium botulinum Ba4 str. 657           -------KVID-----IRTNTKVSGAELFNGRRTRNKSFNSENLRKYRGIKISKGQRRIRTKRYF---YQPGDLVKYEDK---------IYTVKGT----------Q--NKGKY-----I

297585104|Bacillus selenitireducens MLS10              -------KGGI-----RKANKAERFVKG----------------------------------------FQLFDKVLCEGQ---------PCFIFGR----------R--KTGSF-----D
188587532|Natranaerobius thermophilus JW-NM-WN-LF      -------KGGI-----RKANKAPYLVKG----------------------------------------FRLFDKVLYNNR---------KCFIFGR----------R--KSGYF-----D
312793946|Caldicellulosiruptor kristjanssonii 177R1B   -------KGG------KRPVNTVKEVHG----------------------------------------FRRFDKIEYQGG---------TYIIFGL----------R--SSGYF-----A
344996431|Caldicellulosiruptor lactoaceticus 6A        -------KGG------KRPVNTVKEVHG----------------------------------------FRRFDKIEYQGG---------TYIIFGL----------R--SSGYF-----A
217077332|Thermosipho africanus TCF52B                 -------KGG------IRPSNTVKEVRG----------------------------------------FRRFDKVRYKNQ---------IGIIHGL----------R--SSGYF-----D
375009168|Geobacillus thermoleovorans CCB US3 UF5      -------KGGK-----RKSNKAPRFVKG----------------------------------------FQLFDKVVYERK---------ECFIFGR----------R--SSGYF-----D
407477159|Exiguobacterium antarcticum B7                -------KKGK-----RKANKAERLVHG----------------------------------------YQLFDKVQYEEQ---------TCFVFGR----------R--KNGYF-----D
172057416|Exiguobacterium sibiricum 255-15              -------KKGK-----RKANKAERLVHG----------------------------------------YQLFDKVQYEEQ---------TCFVFGR----------R--KNGYF-----D
169823684|Finegoldia magna ATCC 29328                  -------KGGI-----KKNNQASYLVKG----------------------------------------FRLFDRVKYQDK---------EYFIFGR----------R--SSGQF-----H
160902522|Petrotoga mobilis SJ95                       -------KGGI-----RKLNQAPYLVHG----------------------------------------FRLFDKVKYKGI---------ECFIFGR----------R--STGYF-----D
160902616|Petrotoga mobilis SJ95                       -------KGGT-----RKLNQAPYLVHG----------------------------------------FRLFDKVRYNGI---------ECFIFGR----------R--STGYF-----D

333978211|Desulfotomaculum kuznetsovii DSM 6115        --------GNP--------TKTCEERSG----------------------------------------FQHWDIVRA-ERAGRR--------VFGC------V-RSL--KARAM-----T
260892456|Ammonifex degensii KC4                       --------GNP--------TKTCEEKNG----------------------------------------FRHWDVVRA-VRAGKV--------VFGC------V-RSL--KARAM-----T
260893631|Ammonifex degensii KC4                       --------NNP--------TKTCEEKNG----------------------------------------FRHWDVVRA-VRAGKV--------VFGC------V-RSL--KARAM-----T
302391857|Acetohalobium arabaticum DSM 5501            --------GFP-----RAYRSRNKYVDN----------------------------------------FKTGDLVKAVIPKGKN-----KGVYFAR------V-STR--KSGYF-----R
428778981|Dactylococcopsis salina PCC 8305             --------GFPALNKDNQPIKPKTKYTH----------------------------------------CEAGDMIKFRLEKNRKTV--RAGIYKGR------V-KTP--TPKGF-----E
260892735|Ammonifex degensii KC4                       --------GFP-----RGHRAREKVHFG----------------------------------------FATGDLVVAEVPNGKY-----AGRYVGY------V-AVR--RSGYF-----D
260893993|Ammonifex degensii KC4                       --------GFP-----RGHRARGKIHSG----------------------------------------FMTGDLAVAEVPAGRY-----AGRYVGY------V-AVR--RSGYF-----D
433655673|Thermoanaerobacterium
thermosaccharolyticum M0795                            --------GFP-----RAYLSRQKEYFN----------------------------------------FQNGDYVKAEVPKGKY-----KGQYTGY------V-AVR--KSGYF-----D
332982963|Mahella australiensis 50-1 BON               --------GFP-----KGYRKNRKQYFG----------------------------------------FQTGDIVKANIPNGKY-----QGNFTGR------V-AVR--SSGYF-----D
332983053|Mahella australiensis 50-1 BON               --------GFP-----KGYRKNRKQYFG----------------------------------------FQTGDIVKANIPNGKY-----QGNFTGR------V-AVR--SSGYF-----D
222530385|Caldicellulosiruptor bescii DSM 6725         --------GFP-----RGHRTRKKFFYG----------------------------------------FQTGDMVKVVVPKGKY-----KGTWVGT------V-SCR--NSGYF-----D
298674346|Methanohalobium evestigatum Z-7303           --------GFP-----RKPAPRIKYIFG----------------------------------------FQSGDIIKAIVEKGKY-----KGTWKGA------V-ACR--SNGYF-----D
298675173|Methanohalobium evestigatum Z-7303           --------GFP-----RKPAPRIKYIFG----------------------------------------FQSGDIIKAIVENGKY-----KGTWKGA------V-TCR--SKGYF-----D
292490518|Nitrosococcus halophilus Nc 4               --------GFP-----RTSAKAARTVQG----------------------------------------FRTGDIVEARVPTGKK-----AGHYIGR------V-AVR--SSGFF-----N
428778759|Dactylococcopsis salina PCC 8305             --------GYP-----IKHRP-LRPIQG----------------------------------------FQTGDIVQVDIPKGKN-----QGQWTGR------L-CPY--SDGNC-----E
300114477|Nitrosococcus watsonii C-113                 --------GFP-----RGKAGRCKRVKG----------------------------------------FQTGDLAKLIQPKGKY-----AGIHIGM------LAGIR--ARGTF-----D
292491781|Nitrosococcus halophilus Nc 4               --------GFP-----RGKAGRCRRVKG----------------------------------------FQTGDLVRPSQPQGKY-----AGDHVGI------LASIR--ASGTF-----D
292491869|Nitrosococcus halophilus Nc 4               --------GFP-----RGKAGRCRRVKG----------------------------------------FQTGDLVRLSQPQGKY-----AGDHVGI------LASIR--ASGTF-----D
292491893|Nitrosococcus halophilus Nc 4               --------GFP-----RGKAGRCRRVKG----------------------------------------FQTGDLVRLSQPQGKY-----AGDHVGI------LASIR--ASGTF-----D
292491036|Nitrosococcus halophilus Nc 4               --------GFP-----RGKAGRCKRVKG----------------------------------------FQTGDRVRLIQPKGKY-----EGAHTGI------LAGIR--ANGVF-----D
292490884|Nitrosococcus halophilus Nc 4               --------GFP-----RGKAGRCKRVRG----------------------------------------FQTGDLVRLIQPKGKY-----GGTHTGI------LAGIR--ANGAF-----D
292491492|Nitrosococcus halophilus Nc 4               --------GFP-----RGKAGRCKRVRG----------------------------------------FQTGDLVRLIQPKGKY-----GGTHTGI------LAGIR--ANGAF-----D
297563003|Nocardiopsis dassonvillei DSM 43111         --------GFP-----RLRLPRRKRFFG----------------------------------------FQTGDLARAAVPSDMN-----AGTHTGR------V-AVR--SSGSH-----T
374989417|Streptomyces bingchenggensis BCW-1          --------GFP-----RRTSPRIKTVHG----------------------------------------YATGDLVTASVPKGKY-----TGTHAGR------V-MIR--TTGRF-----D
374989654|Streptomyces bingchenggensis BCW-1          --------GFP-----RRTSPRIKTVHG----------------------------------------YATGDLVTASVPKGKY-----TGTHTGR------V-MIR--TTGKF-----D
271963428|Streptosporangium roseum DSM 43021          --------GFP-----RLVRPRIKLHHG----------------------------------------FQTGGLVRANVPAGKK-----AGVHLGR------V-LVR--SSGSF-----D
271964894|Streptosporangium roseum DSM 43021          --------GFP-----RLALPRTKQHHG----------------------------------------FQTGDLVRAVVPTGKK-----AGVHTGR------V-AVR--STGNF-----N
258511478|Alicyclobacillus acidocaldarius DSM 446      --------GFP-----RSHAPSSRSYMG----------------------------------------FRTGDLVQAHVPRGKY-----AGTHVGR------I-VIR--HRPSF-----Q
160902264|Petrotoga mobilis SJ95                       --------GFP-----KKYRPKERTYMG----------------------------------------YKTGDIVLAVIPKGKN-----MGIHIGR------I-AIR--HRPSF-----L
121604978|Polaromonas naphthalenivorans CJ2            --------GFP-----RTYLTHQKTAFG----------------------------------------FKTGDMVIATVPKGVK-----KGIHKGR------V-AIR--LTGSF-----N
347754630|Chloracidobacterium thermophilum B           --------GFP-----RGDLTRSKSAFG----------------------------------------FQTGDLVRVVVTTGKK-----AGNYLGR------V-AIR--ASGSF-----N
218667603|Acidithiobacillus ferrooxidans ATCC 23270    --------GFP-----RGYLTRSKSAFG----------------------------------------FQTGDMVKAVVTKGKK-----VGTYLGR------V-AIR--ASGSF-----N
218667415|Acidithiobacillus ferrooxidans ATCC 23270    --------GFP-----RGYLTRSKSAFG----------------------------------------FQTGDMVKAVVTKGKK-----VGTYLGR------V-AIR--ASGSF-----N
344201059|Acidithiobacillus ferrivorans SS3            --------GFP-----RGYVTRQKRHFG----------------------------------------FQTGDMVRAVVTRGKK-----AGTYVGR------V-VIR--ASGSF-----S
344201127|Acidithiobacillus ferrivorans SS3            --------GFP-----RGYLTRTKSAFG----------------------------------------FQTGDMVRAVVTTGKK-----KGEYLGR------V-AIR--ASGSF-----N
288940385|Allochromatium vinosum DSM 180               --------GFP-----RGYLIREKRVHG----------------------------------------FQTGDWVRAEVPAGKR-----AGVHVGR------V-AVR--RTGAF-----N
288941921|Allochromatium vinosum DSM 180               --------GFP-----RGYLMRHKRVHG----------------------------------------FQTGDWVSAVVPTGKK-----AGVHIGR------V-AVR--QTGSF-----N
288940786|Allochromatium vinosum DSM 180               --------GFP-----RGYLMRHKRVHG----------------------------------------FQTGDWVRAEVPKGKK-----AGVHVGR------V-AVR--QTGSF-----N
288941171|Allochromatium vinosum DSM 180               --------GFP-----RGYLMRHKRVHG----------------------------------------FQTGDWVRAEVPKGKK-----AGVHVGR------V-AVR--QTGSF-----N
288941306|Allochromatium vinosum DSM 180               --------GFP-----RGYLMRHKRVHG----------------------------------------FQTGDWVRAEVPKGKK-----AGVHVGR------V-AVR--QTGSF-----N
121582913|Polaromonas naphthalenivorans CJ2            --------GCA-----RGYLTRSKSIQG----------------------------------------FQTGDMVKAEVTQGKK-----TGIYTGR------V-AVR--ASGSF-----N
121604604|Polaromonas naphthalenivorans CJ2            --------GSA-----RGYLTRSKSIQG----------------------------------------FQTGDMVKAEVTQGKK-----TGIHTGR------V-AVR--ASGSF-----N
91791204|Polaromonas sp. JS666                         --------GFP-----RGYLMRGKRVQG----------------------------------------FGTGDMVRADVPKGVK-----AGVHVGR------V-AVR--ASGSF-----N
91791283|Polaromonas sp. JS666                         --------GFP-----RGYLMRSKRVHG----------------------------------------FGTGDMVRADVTKGVK-----AGHHAGR------V-AVR--ASGSF-----N
121582916|Polaromonas naphthalenivorans CJ2            --------GFP-----RGYLMRAKSVHG----------------------------------------FQTGDRVKAVVPQGKK-----VGTHVGR------V-AIR--KTGSF-----N
121582983|Polaromonas naphthalenivorans CJ2            --------GFP-----RGYLMREKSVHG----------------------------------------FQTGDLVKAVVPKGKK-----AGTHVGR------V-AIR--KTGSF-----N
428304158|Crinalium epipsammum PCC 9333                --------GFP-----KAHAPSAKFFMG----------------------------------------FTTGDMVKANIPTGKF-----AGKYVGR------I-AIR--FRPKF-----V
428204921|Chroococcidiopsis thermalis PCC 7203         --------GFP-----IRHRDSNQVHFG----------------------------------------FRTGNIVKAIVTSGKK-----VGEYVGR------V-LCR--KTGSF-----D
307153539|Cyanothece sp. PCC 7822                      --------GFP-----IRHRSRIKIHKG----------------------------------------FQTGDIVKAIVTSGKK-----IGIHAGR------V-LCR--ASGSF-----D
428216193|Oscillatoria acuminata PCC 6304              --------GFA-----RGHRPRQRYFFG----------------------------------------FKTGDIVIADIPKGKK-----AGRYLGR------L-ACR--TSGRF-----N
414079584|Anabaena sp. 90                              --------GFP-------CTEPKKIFTH----------------------------------------VSTGDFVKATLHKDRKNI--TSGKYVSR------V-KTP--TKNGC-----E
172055147|Cyanothece sp. ATCC 51142                    --------GFP-----TRYVPHKKDVKG----------------------------------------FQTGDIVKAVVTKGKK-----TGIYVGR------V-AVR--STGSF-----N
428316510|Oscillatoria nigro-viridis PCC 7112          --------GFP-----VKHRTRCKTFFG----------------------------------------FGTGDMVSAILLFGKF-----AGTHIGR------L-TVR--ESGVF-----E
428314951|Oscillatoria nigro-viridis PCC 7112          --------GFP-----IRHKTRCKEFFG----------------------------------------FKTGDMVQATLPTGKF-----AGTHRGR------L-IVR--ASGVF-----E
307591386|Cyanothece sp. PCC 7822                      --------GFP-----IRHRSRNKFYKG----------------------------------------FQTGAQVKAVVTKGKK-----VGVYLGR------V-LCR--ASGRF-----D
75812341|Anabaena variabilis ATCC 29413                --------GLP-----VRHCSRTKFHFG----------------------------------------FQTGDIVKAVVTKGKK-----VGTYVGR------I-ATR--ATGSF-----N
17233293|Nostoc sp. PCC 7120                           --------GFP-----IRHCSRTKFHFG----------------------------------------FQTGDIVKAVVTKGKK-----IGTYIGR------I-ATR--ATGSF-----N
75812782|Anabaena variabilis ATCC 29413                --------GFP-----NRYVPKLKFIKG----------------------------------------FQTGDIVKAVVTNGKK-----IGEYIGR------V-AVR--STGSF-----N
434407609|Cylindrospermum stagnale PCC 7417            --------GFP-----SRYVPRFKYVLG----------------------------------------FQTGDIVKAIVTRGKK-----IGEYIGR------I-AVR--TSGSF-----N
414078405|Anabaena sp. 90                              --------GFP-----SRYVPRFKFVKG----------------------------------------FQTGDIVKAVVTAGKK-----VGEYFGR------V-AIR--TSGTF-----N
440685156|Anabaena cylindrica PCC 7122                 --------GFP-----SRYVPRFKFVNG----------------------------------------FQTGDIVKAVVTKGKK-----IGAYIGR------V-AVR--TSGSF-----N
428315656|Oscillatoria nigro-viridis PCC 7112          --------GFP-----IRHRSKKQIHLG----------------------------------------FQTGDIVGATVTAGKK-----VGSYVGR------I-LCR--ASGSF-----D
428204926|Chroococcidiopsis thermalis PCC 7203         --------GFP-----TRHRSKAIIHKG----------------------------------------FKTGDMIKAVVTTGKK-----AGNYVGR------V-LCR--KSGSF-----D
218440244|Cyanothece sp. PCC 7424                      --------GFP-----SRYVPRNKFVKG----------------------------------------FQTGDIVKAIVEKGKK-----IGQYVGR------V-AVR--STGSF-----N
428775092|Halothece sp. PCC 7418                       --------GFP-----KRHRSRVQTHKG----------------------------------------FQTGDIITATVTKGKK-----IGSYVGR------V-LCR--ASGSF-----D
414075475|Anabaena sp. 90                              --------GFP-----SRYVPRIKFVKG----------------------------------------FQTGDIVKAIVTTGKK-----IGKYIGR------V-AVR--ATGSF-----N
428305120|Crinalium epipsammum PCC 9333                --------GFP-----TRHRKTIQIHKG----------------------------------------FQTGDIVKAAVREGKK-----FGTYVGR------V-LCR--ATGSF-----D
427713881|Synechococcus sp. PCC 6312                   --------GFS-----TRHKSRVQIHKG----------------------------------------FQTGDIVRATVTAGKK-----IGFYVGR------V-LCR--ASGSF-----D
```

```
312126587|Caldicellulosiruptor hydrothermalis 108      ----KLSLHDINTGK-RFTQQANPDDC-KFICFASWRY--KYKEL------------------------------------------------------------------
434408285|Stanieria cyanosphaera PCC 7437              ----RVSLHSVETGK-RLCQNAKPQDI-KFLTYNTWRT--KYAN-------------------------------------------------------------------
414076241|Anabaena sp. 90                              ----TEKAVSVSDANWKRLGQCTVKKV-KLILRSTSLI-ITAVKTARVASLSALK--------------------------------------------------------
428299778|Calothrix sp. PCC 6303                       ----TETKVSVSDADWKRLGQCTAKKV-QLVTRSTGLI-IKATKVVSPSETLCERVASLSGLKTLSFPHPARSL-------------------------------------
 17233043|Nostoc sp. PCC 7120                          ----TEKQVSVSDANWKRLGQIASSKV-QLISRSNGLI-VA----------------------------------------------------------------------
166363864|Microcystis aeruginosa NIES-843              ----TEKQLSVSDTNGQRLGQIAVSKI-QLIRRSNGLI-VSH---------------------------------------------------------------------
166367659|Microcystis aeruginosa NIES-843              ----TEKQLSVSDTNGQRLGQIAVSKI-QLIRRSNGLI-VSR---------------------------------------------------------------------
307150343|Cyanothece sp. PCC 7822                      ----TKKQISVSDANWKRLGQIAFSKV-TLIRRSTGLI-VSY---------------------------------------------------------------------
307153352|Cyanothece sp. PCC 7822                      ----TKKQISVSDANWKRLGQIAVSKV-TLIRRSNGLI-VNC---------------------------------------------------------------------
428306661|Crinalium epipsammum PCC 9333                ----TKKQISVSDANWKRLGQIASSKV-TLIRRSTGLI-VS----------------------------------------------------------------------
307150912|Cyanothece sp. PCC 7822                      ----TKKQISVSDANWKRLGQIAVSKV-TLIKRSNGLI-VTA---------------------------------------------------------------------
428775289|Halothece sp. PCC 7418                       ----TKTQISVSDFGWKRVGQFSAKKT-QLLKRSTNLI-VNTSNGRRAFLRGVASGVSPDVYTPRLPLNQRL---------------------------------------
428779516|Dactylococcopsis salina PCC 8305             ----TKNQVSVSGFDWKRLGQFGAKKT-QLLKRNTNLI-VNTQTGGRAFLSR-----------------------------------------------------------
428778599|Dactylococcopsis salina PCC 8305             ----TKTQVSVSDFDWKRLGQFSAKNT-QVLKRSTNLIVNTLNGGRAFLSH------------------------------------------------------------
428781696|Dactylococcopsis salina PCC 8305             ----TKTQVSVSDFDWKRLGQFSAKKT-QLLKRNTNLI-VNTQTGGRAFLSQ-----------------------------------------------------------
428215045|Oscillatoria acuminata PCC 6304              ----TARQVSVSNCDWKRLGQFTASKV-QLIHRATGLL-VSRLHYLLVNGPSIA---------------------------------------------------------
428313425|Microcoleus sp. PCC 7113                     ----TQKQVSVSDLNWKRLGQFTASKV-QLLCRATGLL-VSPLRYCQLLGY------------------------------------------------------------
158340109|Acaryochloris marina MBIC11017               TK--KAKSVSLYDHNWRRIGKFVVSKC-TLLRRSNKLC-VA----------------------------------------------------------------------
428775002|Halothece sp. PCC 7418                       SE--VNKVMSLYDWQWKRIGQFSVSKT-TLIRRSNGLC-VA----------------------------------------------------------------------
307150700|Cyanothece sp. PCC 7822                      ----TKKQISVSDANWKRLGQIASSKV-ILIQRSTGLI-VSH---------------------------------------------------------------------

312126663|Caldicellulosiruptor hydrothermalis 108      LYSPQ----LEKIV---GKTYVKPEEC-KRVLRNEGIV-VMYNSLYHN---------------------------------------------------------------
312621424|Caldicellulosiruptor kronotskyensis 2002     LYSPQ----LEKIV---GKTYVKPEEC-KRILRNEGIV-VMYNSLYHS---------------------------------------------------------------
222530402|Caldicellulosiruptor bescii DSM 6725         LYSPQ----LKKIV---GKMYVQPEEC-RQVLHNEGMV-VMYNSLYHS---------------------------------------------------------------
222530431|Caldicellulosiruptor bescii DSM 6725         LYSPQ----LKKIV---GKMYVQPEEC-RQVLHNEGMV-VMYNSLYHS---------------------------------------------------------------

307591386|Cyanothece sp. PCC 7822                      IASHQ-----GRTT---GITY---KFC-TAIQKKDGYN-YEF---------------------------------------------------------------------
302391131|Acetohalobium arabaticum DSM 5501            FTTLD-----GEKTFR-GYGL---TNL-TLVNRPQGLM-II----------------------------------------------------------------------
435854579|Halobacteroides halobius DSM 5150            FTTLN-----GEKTFR-GYGL---TNL-TLINRPKGLM-II----------------------------------------------------------------------
435852864|Halobacteroides halobius DSM 5150            FSTLK-----GKIFR--GYGF---RNL-TLLNRPKGLM-IV----------------------------------------------------------------------
435854272|Halobacteroides halobius DSM 5150            FSTLK-----GKIFR--GYGF---RNL-TLLNRPKGLM-IV----------------------------------------------------------------------
313123202|Lactobacillus delbrueckii  ND02              LTDGL-----GNYVLNNGKNYINGQQC-RLIMHNGNWQ-KAEQKLSLYEFK------------------------------------------------------------
77362205|Pseudoalteromonas haloplanktis TAC125         LETIL-----GDKVHASANIK---KNTVRISARTTTLT-QLMEAAIPLCTKVPSILAVN-----------------------------------------------------
298675374|Methanohalobium evestigatum Z-7303           IMDIF-----GNTSKKNPNVK---KNCRRISARCTTLT-QQIQI-------------------------------------------------------------------
298674707|Methanohalobium evestigatum Z-7303           IMDIF-----GNTFKKNPNVK---KNCRRISARCTTLT-QQIQI-------------------------------------------------------------------
298675090|Methanohalobium evestigatum Z-7303           IMDIF-----GNTFK---------KNCRRISARCTTLT-QQIQI-------------------------------------------------------------------
169834784|Clostridium botulinum A3 str. Loch Maree     ALKEL-----KKVPK---------VELLTPYKFRKGLV-CV
229587454|Clostridium botulinum Ba4 str. 657           ALKEL-----KKVPK---------VELLAPYKFRKGLV-CV

297585104|Bacillus selenitireducens MLS10              LRLLD-----GTVISR-GKSY---KKI-ALKEKATSWL-VERSETVHIPPHE-----------------------------------------------------------
188587532|Natranaerobius thermophilus JW-NM-WN-LF      LRLLD-----GTKIHA-SASH---KRL-DLIESATSML-KEVS--------------------------------------------------------------------
312793946|Caldicellulosiruptor kristjanssonii 177R1B   TGTLS-----GEKISD-SVKY---TCL-RLVEKAKTLM-FERREQRIFLHLAEDGVSCAVL--------------------------------------------------
344996431|Caldicellulosiruptor lactoaceticus 6A        TGTLS-----GEKISD-SVKY---TCL-RLVEKAKTLM-FERREQRIFLHLAEDGVSCAVL--------------------------------------------------
217077332|Thermosipho africanus TCF52B                 IRTLS-----GEKIHS-SAKW---SNL-KLLEKAKTLM-LESRNQLIPLHLNGVSCAGS----------------------------------------------------
375009168|Geobacillus thermoleovorans CCB US3 UF5      LRLLD-----GTKVHA-SASW---KKL-KRVEHASTLL-IERRKGDSSPTFALA---------------------------------------------------------
407477159|Exiguobacterium antarcticum B7               LRTIA-----GFSIHK-SASY---KKL-TLLEKAKTWL-VDIQSEGGEGRSQP----------------------------------------------------------
172057416|Exiguobacterium sibiricum 255-15             LRTLA-----GFSIHK-SASY---KKL-ILLEKAKTWL-VDIQPEGGEGRSHP----------------------------------------------------------
169823684|Finegoldia magna ATCC 29328                  IRDLS-----GNKVNKGSISC---KKL-EFIEPRKTVL-VERR--------------------------------------------------------------------
160902522|Petrotoga mobilis SJ95                       IRKLD-----GTVIHR-SAKS---KDL-ILVSKAKTLL-WERRENSAFLSPLPYLP-------------------------------------------------------
160902616|Petrotoga mobilis SJ95                       IRKLD-----GTVIHR-SAKS---KDL-ILVSKAKTLL-WERRENSAFLSPLPYLP-------------------------------------------------------

333978211|Desulfotomaculum kuznetsovii DSM 6115        LRTAG-----DGNF---EVSY---SKA-KLLHRPRGLA-YVPVW-------------------------------------------------------------------
260892456|Ammonifex degensii KC4                       LRTAK-----DDNF---EVSY---SKA-RLLHRPRGLA-YMPVW-------------------------------------------------------------------
260893631|Ammonifex degensii KC4                       LRTAK-----DDNF---EVSY---SKA-RLLHRPRGLA-YMPA--------------------------------------------------------------------
302391857|Acetohalobium arabaticum DSM 5501            LDCFN----GSKVD---GVNS---KYL-NLLQRGAGYS-YSFEDIA-----------------------------------------------------------------
428778981|Dactylococcopsis salina PCC 8305             VKIDG-----CRVS----QKM---EYLIRFMHRNDGYA-YSF---------------------------------------------------------------------
260892735|Ammonifex degensii KC4                       LKDLS---GRRVCQ---GIAF---HHF-KLVQRFDGWR-YEKFKVAALSSPHLKVGASSAA--------------------------------------------------
260893993|Ammonifex degensii KC4                       LKDLS---GRRVCQ---GIAF---RHF-KLVQRFDGWR-YEKFKIAALSSPHLKVGASSAA--------------------------------------------------
433655673|Thermoanaerobacterium
thermosaccharolyticum M0795                            IRDAN---GKNIAQ---GISY---QYC-HIIQRYDGYK-YGRRKRHILLHPSDAMDGVSCA--------------------------------------------------
332982963|Mahella australiensis 50-1 BON               IKDST---GNCLCQ---GVSW---KYM-QVLQRNSGWQ-YDKERRNPLLSQLK----------------------------------------------------------
332983053|Mahella australiensis 50-1 BON               IKDST---GNCLCQ---GVSW---KYM-QVLQRNSGWQ-YDKERRNPLLSQLK----------------------------------------------------------
222530385|Caldicellulosiruptor bescii DSM 6725         IKDKT---GKRIVQ---SISY---KHC-KIIQRFDGYC-YELERIRISGTFPLQPVEVGASMCQ-----------------------------------------------
298674346|Methanohalobium evestigatum Z-7303           IKNGT----KKIAQ---GINH---KYF-KLVQRFDGYT-YQLDRLNLNKNTNTD--GRNSSHC------------------------------------------------
298675173|Methanohalobium evestigatum Z-7303           IKNGT----KRIAQ---GINH---KYF-KLVQRFDGYT-YQLDRLNLNSNTNTD--GRNSSHC------------------------------------------------
292490518|Nitrosococcus halophilus Nc 4                ITSQR-----IVTQ---GISY---QYC-RTLHRCDGYG-YGYVIDTQIASTQQEDAGEAA---------------------------------------------------
428778759|Dactylococcopsis salina PCC 8305             ICPKG-----RKRL---GTKL---SYVSRVVHQKDSYK-YRHLS-------------------------------------------------------------------
300114477|Nitrosococcus watsonii C-113                 IKAKA-----GKIS----ANW---KNF-QLIQRGDGYD-YGLLTV------------------------------------------------------------------
292491781|Nitrosococcus halophilus Nc 4                VKAKA-----GKIS----ANW---KNF-KLIQRGDGYD-YSLSAV------------------------------------------------------------------
292491869|Nitrosococcus halophilus Nc 4                VKAKA-----GKIS----ANW---INF-KLIQRGDGYD-YSLSAV------------------------------------------------------------------
292491893|Nitrosococcus halophilus Nc 4                VKAKA-----GKIS----ANW---INF-KLIQRGDGYD-YSLSAV------------------------------------------------------------------
292491036|Nitrosococcus halophilus Nc 4                IKAKA-----GKIS----ANW---KNF-KLIQRGEGYD-YGLSAV------------------------------------------------------------------
292490884|Nitrosococcus halophilus Nc 4                IKAKA-----RKIS----ANW---KHF-KLIQRGDGYD-YGLSAV------------------------------------------------------------------
292491492|Nitrosococcus halophilus Nc 4                IKAKA-----RKIS----ANW---KHF-KLIQRGDGYD-YGLSAV------------------------------------------------------------------
297563003|Nocardiopsis dassonvillei DSM 43111          LQTSY-----GPIK----TSW---KNL-CLLQRADGYG-YTTQKEAGAPSTA-----------------------------------------------------------
374989417|Streptomyces bingchenggensis BCW-1           IRTPH-----GRAA---GIQH---TNI-RLLQRGDGYT-YTHQPEPTRA--------------------------------------------------------------
374989654|Streptomyces bingchenggensis BCW-1           IRTPH-----GRAA---GIRH---TNI-RLLQRGDGYT-YTHQPEQACA--------------------------------------------------------------
271963428|Streptosporangium roseum DSM 43021           ITTRH-----GRVA---GISH---RHV-RHLQRADGYG-YTTRKEADRND-------------------------------------------------------------
271964894|Streptosporangium roseum DSM 43021           IRTRH-----GSVR---GISH---RHV-RLLQRADGYG-YTTHPEARNRAAFPPPPEGGGIHAGGN---------------------------------------------
258511478|Alicyclobacillus acidocaldarius DSM 446      LNGFN-------------VHP---KYL-KVLQRGDGYG-YAME--------------------------------------------------------------------
160902264|Petrotoga mobilis SJ95                       LNGVG------------DVHP---KYL-TLLQKNDGYG-YQIS--------------------------------------------------------------------
121604978|Polaromonas naphthalenivorans CJ2            IQTGI--AGAATVQ---GISY---KHC-RITQRADGYG-YSHRAQTKKECGNKECALHTALSIPGLKAEVSRAI-------------------------------------
347754630|Chloracidobacterium thermophilum B           IQSAT-----GLVQ---GIHH---RFC-TLLQRADGYG-YSWTRIAI----------------------------------------------------------------
218667603|Acidithiobacillus ferrooxidans ATCC 23270    IQTAT-----GLVQ---GIHH---RFC-TRIQCADGYG-YSWTKIASTQGEAGMRAAHAAALSLPGLNPGVSRANG-----------------------------------
218667415|Acidithiobacillus ferrooxidans ATCC 23270    IQTGN-----GLVQ---HIHY---RFC-KLVQRGDGYG-YHWSLLHPALNHGIAEAGRN----------------------------------------------------
344201059|Acidithiobacillus ferrivorans SS3            INAAS-----GLVS---GIHH---RFF-SLIQRADGYG-YSLTKIASTQGDAGRGRSVAAALSLPGMNARVSRANG-----------------------------------
344201127|Acidithiobacillus ferrivorans SS3            IQTGN-----GLVQ---HVHH---RFC-KLVQRADGYG-YAFTTRPQRQGLPHGLIWSKGS--------------------------------------------------
288940385|Allochromatium vinosum DSM 180               IQTQD-----ATVQ---GISY---RHC-RVLQRADGYG-YAFQSKPDAEKARRAA--------------------------------------------------------
288941921|Allochromatium vinosum DSM 180               IQAPA-----GTVQ---GIAY---RHC-CLIQRADGYG-YAFQPKPDTEEARRAA--------------------------------------------------------
288940786|Allochromatium vinosum DSM 180               IQAQG-----GTVQ---GVSY---RHC-RVLQRADGYG-YAFQPKPDTEKARRAA--------------------------------------------------------
288941171|Allochromatium vinosum DSM 180               IQAQG-----GTVQ---GVSY---RHC-RVLQRADGYG-YAFQPKPDAEKARRAA--------------------------------------------------------
288941306|Allochromatium vinosum DSM 180               IQAQG-----GTVQ---GVSY---RHC-RVLQRADGYG-YAFQPKPDAEKARRAA--------------------------------------------------------
121582913|Polaromonas naphthalenivorans CJ2            IQTRH-----GLVQ---GVSH---RYC-TVVQRGDGYG-YSLVAKTDATGTSPKQGDALHPALYLPGLKAGVSRAN-----------------------------------
121604604|Polaromonas naphthalenivorans CJ2            IQTRH-----GLVQ---GVSH---RYC-TVVQRGDGYG-YQLVAKTDATGTPPKQGDASHPALYLPGLKAEVSRANR----------------------------------
91791204|Polaromonas sp. JS666                         IQTHQ-GGISGVVQ---GISH---KHC-RVTQRNDGYG-YFFNRADHTGREQVWPKASDAAHPALYLPAMNGRVSRAI---------------------------------
91791283|Polaromonas sp. JS666                         IQTRR-NGISGVIQ---GISH---KHC-RVIQRNDGYG-YFFNRTDHTGREQVRPKASDAAHPALYLPAMNGRVSRAI---------------------------------
121582916|Polaromonas naphthalenivorans CJ2            ITTPA-----GVVQ---GISH---KHC-RIVQRNDGYG-YFFHRADLTQDANR----------------------------------------------------------
121582983|Polaromonas naphthalenivorans CJ2            ITTAR-----GVVQ---GIGH---KNC-RVVQRNDGYG-YFFNRAQHTGCEQAGPGATDALHPALYLPGLNPGVSRAS---------------------------------
428304158|Crinalium epipsammum PCC 9333                LQLPT-----QKFD----VHP---KYL-KTIHHSDGYE-YQS---------------------------------------------------------------------
428204921|Chroococcidiopsis thermalis PCC 7203         IATKT-----GRVT---GISH---RFC-TSIHKKDGYS-YVF---------------------------------------------------------------------
307153539|Cyanothece sp. PCC 7822                      IASKS-----GRVT---GISH---KNC-RVVHKNDGYA-YEF---------------------------------------------------------------------
428216193|Oscillatoria acuminata PCC 6304              IKTVE-----LTVQ---GISH---KYC-QITHHGDGYL-YGL---------------------------------------------------------------------
414079584|Anabaena sp. 90                              IVING-----FRVE---FSTM---KDI-TKVHCSDGYS-YV----------------------------------------------------------------------
172055147|Cyanothece sp. ATCC 51142                    ISTAK-----GVIQ---GINH---KYC-RQIHKKDGYS-YAT---------------------------------------------------------------------
428316510|Oscillatoria nigro-viridis PCC 7112          MRTPL-----GKVS---PVRH---KYC-KSIHRNDGYM-YAFSTNVRPMNQRLEA--------------------------------------------------------
428314951|Oscillatoria nigro-viridis PCC 7112          MMSPT-----GKVS---PVRH---KYC-RAIHHNDGYT-YTLSTCVQ----------------------------------------------------------------
307591386|Cyanothece sp. PCC 7822                      IASHQ-----GRTT---GITY---KFC-TAIQKKDGYN-YEF---------------------------------------------------------------------
75812341|Anabaena variabilis ATCC 29413                ISTAK-----GLIQ---GISH---KYC-KTIHKKDGYA-YGI---------------------------------------------------------------------
17233293|Nostoc sp. PCC 7120                           ISTVQ-----GLIQ---GISH---KYC-KTIHKKDGYA-YAI---------------------------------------------------------------------
75812782|Anabaena variabilis ATCC 29413                ISAQR-----GLIQ---GINY---KFC-KSIHKKDGYS-YSS---------------------------------------------------------------------
434407609|Cylindrospermum stagnale PCC 7417            ISTQS-----CLVQ---GISY---KAC-KHIHKKDGYE-YGKSRI------------------------------------------------------------------
414078405|Anabaena sp. 90                              ISTNS-----GLVQ---GISH---KYC-LATHKKDGYS-YAN---------------------------------------------------------------------
440685156|Anabaena cylindrica PCC 7122                 ISTVS-----GLIQ---GISY---KYC-KSIHKKDGYE-YGK---------------------------------------------------------------------
428315656|Oscillatoria nigro-viridis PCC 7112          IATMK-----GRVA---GISH---KYC-KPIHKKDGYS-YGF---------------------------------------------------------------------
428204926|Chroococcidiopsis thermalis PCC 7203         IATKS-----GRVT---GISY---KCC-QPIQRKDGYS-YAI---------------------------------------------------------------------
218440244|Cyanothece sp. PCC 7424                      ISTNN-----GLVQ---GISH---KYC-SIIHRKDGYN-YGFCHDAEFIQRVVLPPRVKTSRFPYSRRFLDEKTRRSH---------------------------------
428775092|Halothece sp. PCC 7418                       VTTQS-----GRVS---GISH---RYC-QPIHKRDGYA-YAFSHTIGD---------------------------------------------------------------
414075475|Anabaena sp. 90                              ISASK------LVQ---GISY---KYC-KIIHRKDGYD-YAF---------------------------------------------------------------------
428305120|Crinalium epipsammum PCC 9333                IATKS-----GRVA---GISH---KFC-THIHRKDGYN-YAY---------------------------------------------------------------------
427713881|Synechococcus sp. PCC 6312                   IATAT-----GRVA---GISH---KYC-KAMHKKDGYA-YVF---------------------------------------------------------------------
```

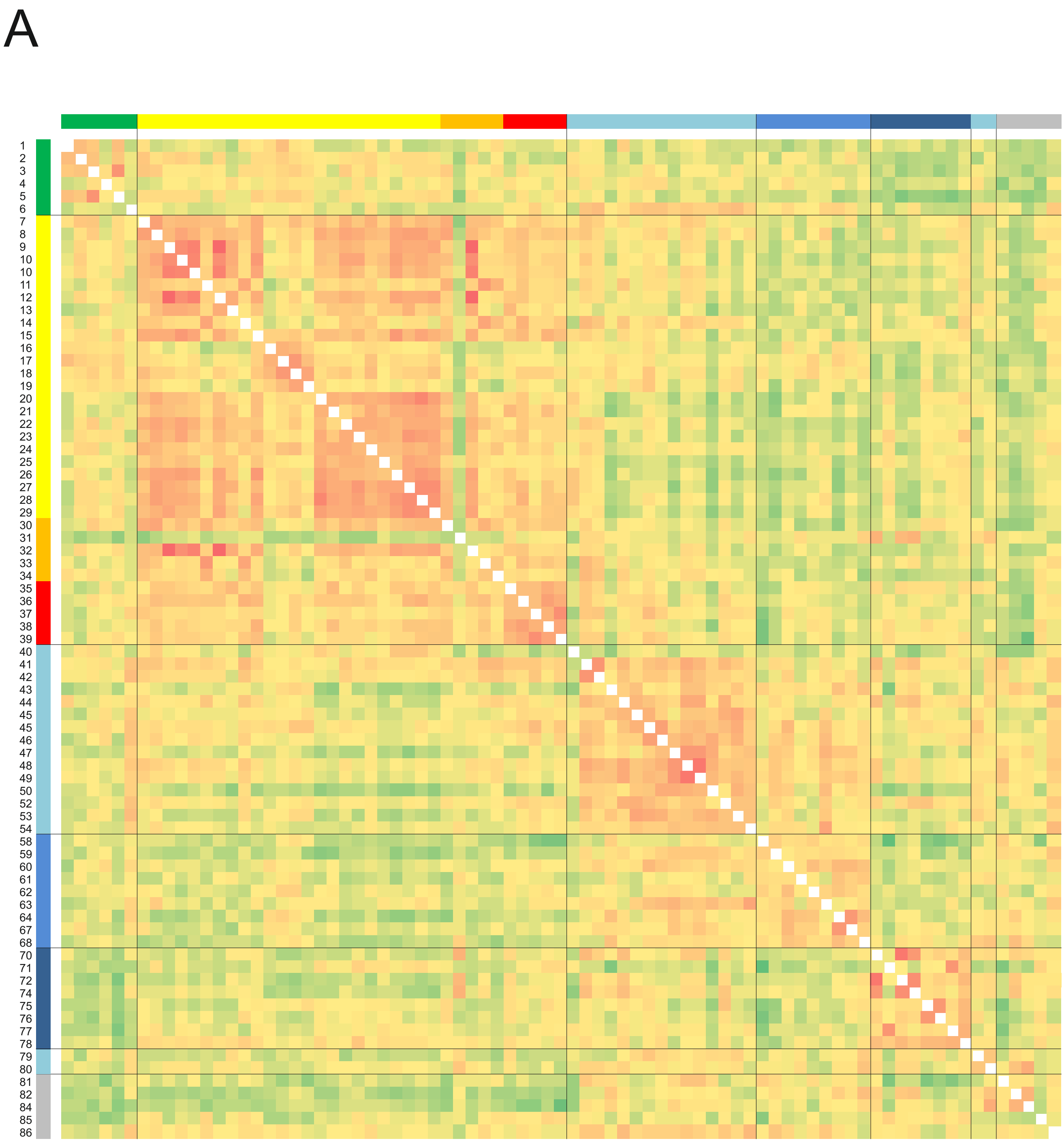
B

```
No Hit                             Prob E-value P-value  Score    SS Cols Query HMM  Template HMM
 1 PF14239 RRXRR:  RRXRR protein  100.0 1.6E-79 2.7E-84  564.5  20.9  174    5-178     1-176 (176)
 2 4oge_A HNH endonuclease domain  99.7 6.1E-19 1.1E-23  202.5   4.6  189  128-321   476-748 (1101)
 3 4cmp_A CAS9, crispr-associated  99.7 6.7E-18 1.2E-22  197.5   6.4  196  125-321   732-999 (1372)
 4 COG3513 Predicted CRISPR-assoc  99.5 3.6E-14 6.2E-19  154.1   6.0  138  184-321   561-748 (1088)
 5 2qgp_A HNH endonuclease; Q39X4  99.4 1.8E-13 3.2E-18  119.1   4.5   60  186-245    27-88  (112)
 6 PF13395 HNH_4:  HNH endonuclea  99.2 4.6E-12 7.9E-17   96.8   2.4   46  197-242     1-53  (54)
 7 PF01844 HNH:  HNH endonuclease  99.1 2.7E-11 4.7E-16   88.5   1.9   44  197-240     1-47  (47)
 8 COG1403 McrA Restriction endon  98.8 5.4E-09 9.4E-14   91.5   3.3   61  184-245    59-120 (146)
 9 PF14279 HNH_5:  HNH endonuclea  97.1 0.00038 6.7E-09   56.6   2.0   39  197-238     1-41  (71)
10 PF13391 HNH_2:  HNH endonuclea  97.0  0.0006   1E-08   53.1   2.5   44  197-240     1-61  (66)

13 1e7l_A GP49, recombination end  91.2    0.29   5E-06   46.0   3.7   53  184-238    10-66  (157)

16 1hjr_A Holliday junction resol  87.7     1.5 2.7E-05   41.0   5.7   41   56-96      2-46  (158)
17 PF02075 RuvC:  Crossover junct  87.4     1.4 2.4E-05   40.7   5.3   40   57-96      1-45  (149)
18 4ep4_A Crossover junction endo  87.0     1.3 2.3E-05   41.8   5.0   41   56-96      2-50  (166)

22 4oge_A HNH endonuclease domain  80.5     1.3 2.3E-05   52.9   2.6   74   56-129    12-99  (1101)

37 4cmp_A CAS9, crispr-associated  58.7     5.3 9.2E-05   49.2   1.0   53   54-110     7-79  (1372)

44 COG0675 Transposase and inacti  50.6      75  0.0013   31.6   7.5  152   40-233   161-328 (364)

57 COG3513 Predicted CRISPR-assoc  36.7      30 0.00052   40.7   2.4   63   55-117     4-80  (1088)

75 PF01385 OrfB_IS605:  Probable   29.0      39 0.00068   31.9   1.6  110   40-156   106-221 (227)


  No 1
>PF14239 RRXRR:  RRXRR protein
  Probab=100.00  E-value=1.6e-79  Score=564.55  Aligned_cols=174  Identities=61%  Similarity=1.015  Sum_probs=0.0

  Q ss_pred                EEEECCCCCcCCCCCHHHHHHHHHcCCeEEEecCCEEEEEEEeccCCCccCCEEEEEcCCCCcEEEEEEECCEEEEEEEEE
  Q Thu_Feb_20_16:    5 IFVLDTQKRPLDPVHPGKARHLLNQGKAAVYRRFPFTLILKEAHPDVPVQDLELKLDPGSKVTGIAIKQGNKIIFGAELQ   84 (425)
  Q Consensus         5 VfVld~~gkPLmP~~p~rAR~lLk~gkA~v~~~~pFtI~L~~~~~~~~~~~qpi~lgiDpG~k~~G~av~~~~~~l~~~el~   84 (425)
                        |||||+||+|||||+|+|||+|||+|||+|++++||||||++++++.++|||+|||||||++|||||++++++||++||+
  T Consensus         1 VfVld~~gkPLmP~~p~rAR~LLk~GkA~V~r~~PFtI~L~~~~~~~~~~~qpi~lgiDpGsk~tGiav~~~~~vl~~~el~   80 (176)
  T PF14239_consen    1 VFVLDKNGKPLMPCHPARARKLLKQGKAKVVRRYPFTIQLKYEPSGSYTQPIRLGIDPGSKTTGIAVVSEKKVLWAAELE   80 (176)
  T ss_pred                CeEECCCCCcCCCCCHHHHHHHHHCCCEEEEeCCCeEEEEEeccCCCccccCEEEEECCCCCeEEEEEEeCCEEEEEeEee

  Q ss_pred                eCchHHHHHH-HHHHHHHhhhc-cccCccCCchhhcccCCCCCCChhHHHHHHHHHHHHHHHHhCCcchheeeeeeeeh
  Q Thu_Feb_20_16:   85 HRGQQIKDAL-LSRRQLRRGRR-NRKTRYRQPRFLNRNRPDGWLAPSLKHRVDTVLTWVNRLRNLAPLGSIAQELVRFDL  162 (425)
  Q Consensus        85 ~r~~~vk~~l-~~Rr~~RR~RR-~rk~ryR~~RF~NR~r~~gwL~PSi~~~~~~~~~~~~~v~~l~~~~Pi~~i~~E~v~fD~  162 (425)
                        ||+++|+++| ++||++||+|| |||+|||+|||+||++++|||||||+||+|+|++||++|++|+||++|++|+|+||+
  T Consensus        81 ~r~~~i~~~lr~~RR~~RR~RR~~rk~RyR~~RF~NR~r~~gwL~PSl~~rv~~~l~~v~~L~~~~PIt~i~~E~v~FD~  160 (176)
  T PF14239_consen   81 HRGQDIKDRLRTQRRAYRRGRRYNRKTRYRKARFDNRKRPKGWLPPSLRHRVDTHLRWVKRLCKLLPITAIVVELVKFDT  160 (176)
  T ss_pred                ccHHHHHHHHHHHHHHHhhhcccccccccccccccCCCCCCcCcCHHHHHHHHHHHHHHHHHHhCCcceEEEEEEEeeh

  Q ss_pred                hhhcCCcccccceeccc
  Q Thu_Feb_20_16:  163 QKLQNPEISGIEYQQG  178 (425)
  Q Consensus       163 qk~~np~i~G~~YQqg  178 (425)
                        |+|+||||+|+|||||
  T Consensus       161 q~l~np~i~G~~YQqG  176 (176)
  T PF14239_consen  161 QKLQNPEISGVEYQQG  176 (176)
  T ss_pred                hhccCcCccccccCCC


  No 2
>4oge_A HNH endonuclease domain protein; crispr-CAS, CAS9, RUVC, RNA-guided DNA endonuclease, cytoplasmic,
  hydrolase; HET: SPD; 2.20A {Actinomyces naeslundii} PDB:  4ogc _A*
  Probab=99.75  E-value=6.1e-19  Score=202.49  Aligned_cols=189  Identities=24%  Similarity=0.347  Sum_probs=0.0

  Q ss_pred                hhHHHHHHHHHHHHHHHHHhC-Ccchheeeeeeeehhhhccccccceec-----------------------cccccch
  Q Thu_Feb_20_16:  128 PSLKHRVDTVLTWVNRLRNLA-PLGSIAQELVRFDLQKLQNPEISGIEYQ-----------------------QGELQGY  183 (425)
  Q Consensus       128 PSi~~~~~~~~~~~~~v~~l~~~~~-Pi~~i~~E~v~fD~qk~~np~i~G~~YQ-----------------------qg~l~g~  183 (425)
                        |....-+...+++|+.|.+-+ +...|+||+++    -+++.+=.-.+.+                       .+....-
  T Consensus       476 P~V~RaL~Q~rKVVNaIikkyG~P~~I~IElAR----El~~s~ker~~i~k~~kkn~kene~~~kel~e~~~~~~~~~~~~~~  551 (1101)
  T 4oge_A          476 PSVDRTLKIVGRYLSAVESMWGTPEVIHVEHVR----DGFTSERMADERDKANRRRYNDNQEAMKKIQRDYGKEGYISRG  551 (1101)
  T ss_dssp                HHHHHHHHHHHHHHHHHHHHHCCCSEEEEECCS----SCSSCHHHHHHHHHHHHHHHHHHHHHHHHHHHHHCCCSCCCHH
  T ss_pred                HHHHHHHHHHHHHHHHHHHHHhCCchhhhHHhh----hhcccHHHHHHHHHHHHHHHHHHHHHHHHHHHHhccccCCChh

  Q ss_pred                hHHHHHHHHh-CCcccccCCcCCc--eeEeeeccccCCCCcc-hhhhhhhhhhhhhhhccCCHHHHhcccH---------
  Q Thu_Feb_20_16:  184 EVREYLLEKW-GRKCTYCGTKDVP--LEVEHIHPRSRGGTDR-VSNLTMACHACNQSKGNQDIRDFLFGKP---------  250 (425)
  Q Consensus       184 ~vrey~l~k~-~~~C~yCg~~~~~--Lev~HI~prs~gGsd~-~sNl~l~C~~Cn~~K~n~~~~~~fl~~~~~---------  250 (425)
                        ++..|-|+.. +..|.|||++...  +|||||+|+|..++|+ .+|++|+|..||+.|+| +|.+|+....
  T Consensus       552 di~K~~Lw~eQ~g~C~YsG~~I~~~~~~eIDHIiP~S~~~dDS~~~NkVLv~~~~N~~Kgn-tP~e~f~~~~~~~~~~~~~~~~  630 (1101)
  T 4oge_A          552 DIVRLDALELQGCACLYCGTTIGYHTCQLDHIVPQAGPGSNNRRGNLVAVCERCNRSKSN-TPFAVWAQKCGIPHVGVKE  630 (1101)
  T ss_dssp                HHHHHHHHHHTTSBCTTTCCBCCTTTCEEEESSCBSSSSCCCSGGGEEEECHHHHHHHTT-SCHHHHHHHHCCTTCSHHH
  T ss_pred                HHHHHHHHHHcCCcCCCCCCcCChhheeeeeeeeeccCCCCCcccccceEEcHHHHHhCCC-ChHHHHhccccccchhhHHH

  Q ss_pred                -------------------------HHHHHHHHhccCCchhHHHHHHHHHH---HHHHHHhhC--CCeEEeechhhhhhhh
  Q Thu_Feb_20_16:  251 -------------------------ELLQRILRKAKAPLKDAAAVNSTRWA---LFNALKATG--LPLTTGTGGQTKFNR  300 (425)
  Q Consensus       251 -------------------------~~l~~i~~~~k~~lkda~~vn~tR~~---l~~~Lk~~~--~~V~~~~G~~Tk~~R  300 (425)
                                                 ..+.++.......=-.+..+|+|||+   +.+.|++.+  ..|.++.|.+|...|
  T Consensus       631 ~~~~~~~~w~~~~~~~~~~~~~~~~~~kk~~~~~~~~~~~~~~~~~Fi~R~L~dTryisk~~~~~~L~~~~~~~~~~V~~v~G~lTs~lR  710 (1101)
  T 4oge_A          631 AIGRVRGWRKQTPNTSSEDLTRLKKEVIARLRRTQEDPEIDERSMESVAWMANELHHRIAAAYPETTVMVYRGSITAAAR  710 (1101)
  T ss_dssp                HHHHHHTCCCCC---CHHHHHHHHHHHHHHHHCCSCBCCCCTTTTBCCCHHHHHHHHHHHHHSTTSEEEEECTTHHHHHH
  T ss_pred                HHHHHHHHHHHHHHHhhhhhHHHHHHHHhhccccccccchhcccHHHHHHHHHHHHHHHHHCCCCeEEEECcHHHHHHH

  Q ss_pred                hhcCCcc-----------------ccccccceeeccccc
  Q Thu_Feb_20_16:  301 TRLDLPK-----------------AHWLDAACVGQVDA  321 (425)
  Q Consensus       301 ~~~gL~K-----------------~H~~DA~cig~~~~  321 (425)
                        ..|||.+                 |||+||++||+.++
  T Consensus       711 ~~wgl~~~~~~~~~~~~~~~~~K~R~d~rHHAiDA~VIA~~~~  748 (1101)
  T 4oge_A          711 KAAGIDSRINLIGEKGRKDRIDRRHHAVDASVVALMEA  748 (1101)
  T ss_dssp                HHHTCTTTSCCSSCCCCSSCCBTTHHHHHHHHHHHCCH
  T ss_pred                HHcCCchhhhhhccccccCCCCCCchHHHHHHHHHhCCH
```

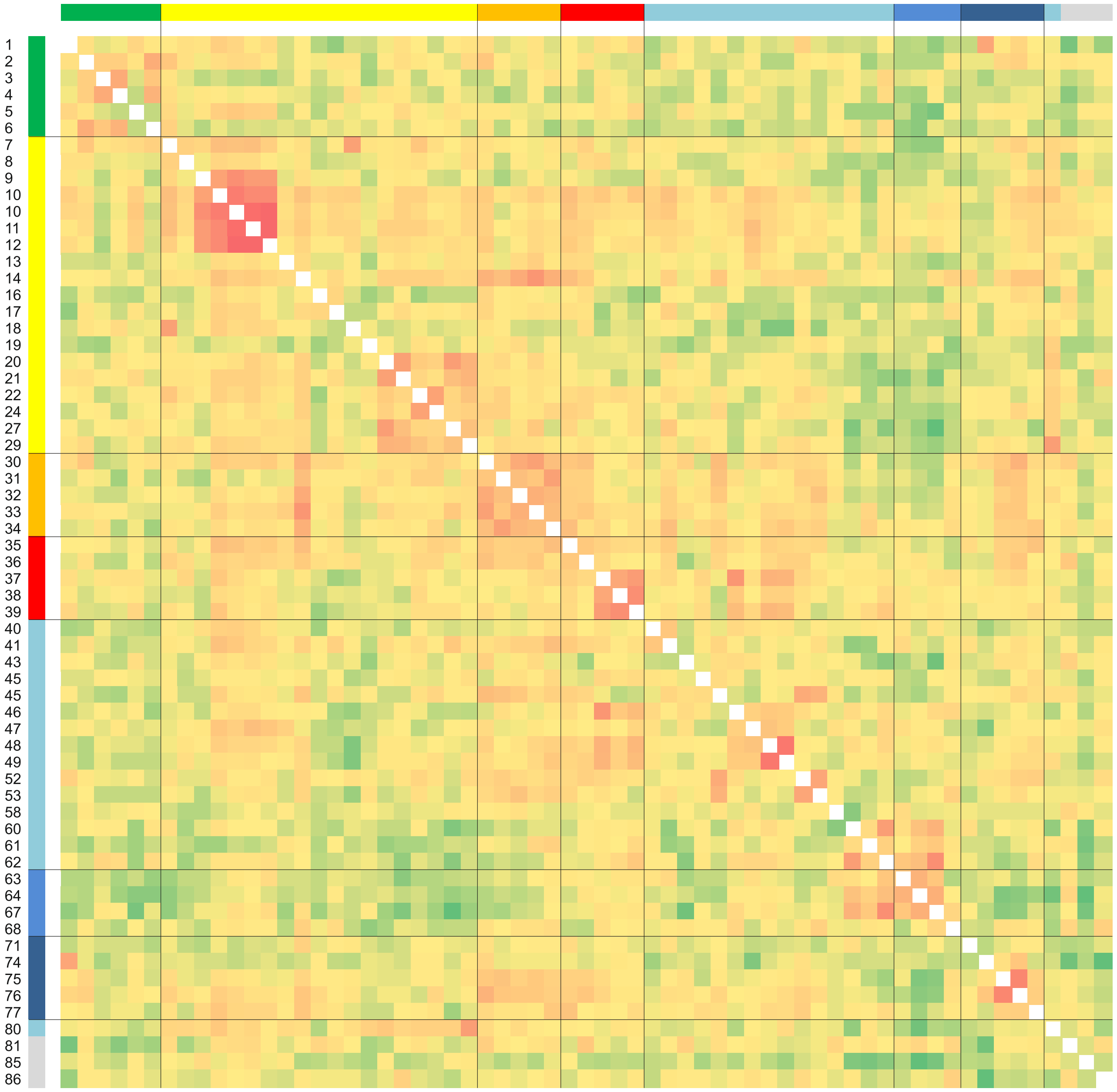
```
  No 3
>4cmp_A CAS9, crispr-associated endonuclease CAS9/CSN1; hydrolase, DNAse, RNA-guided, immunity, crRNA, genome
  editin; 2.62A {Streptococcus pyogenes} PDB:  4cmq _A
  Probab=99.71  E-value=6.7e-18  Score=197.53  Aligned_cols=196  Identities=20%  Similarity=0.121  Sum_probs=0.0

  Q ss_pred             CCChhHHHHHHHHHHHHHHHHHhC---Ccchheeeeeeeehhhhc----------------CCcccc---ceeccccccc
  Q Thu_Feb_20_16:  125 WLAPSLKHRVDTVLTWVNRLRNLA---PLGSIAQELVRFDLQKLQ----------------NPEISG---IEYQQGELQG  182 (425)
  Q Consensus       125 wL~PSi~~~~~~~~~~v~~l~~~~---Pi~~i~~E~v~fD~qk~~----------------np~i~G---~~YQqg~l~g  182 (425)
                        ...|+++.-+...+++|+.|.+.+   +...|+||+++ |.+.-.                +-++..   .++..-.-.-
  T Consensus       732 ~~nP~V~r~l~Q~~kVVn~Iik~~G~~~P~~I~IElAR-El~~s~k~rk~~~~~~~~~~k~~k~~~~~~~~e~~~~~~~~  810 (1372)
  T 4cmp_A          732 AGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMAR-ENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQL  810 (1372)
  T ss_dssp             SSCHHHHHHHHHHHHHHHHHHHHTTTCCCSEEEEEEC-----------CHHHHHHHHHHHHHTTC-------CCCCGGGG
  T ss_pred             cCChHHHHHHHHHHHHHHHHHHHhCCCCCCeEEEEecC-CCCccHhHHHHHHHHHHHHHHHHHHHHHHHHHHHhcCCCcchh

  Q ss_pred             hhHHHHHHHHhCCcccccCCcCCce-----eEeeeccccCCCCcchhhhhhhhhhhhhhhccCCHHHHhcccHHHHHHHH
  Q Thu_Feb_20_16:  183 YEVREYLLEKWGRKCTYCGTKDVPL-----EVEHIHPRSRGGTDRVSNLTMACHACNQSKGNQDIRDFLFGKPELLQRIL  257 (425)
  Q Consensus       183 ~~vrey~l~k~~~~C~yCg~~~~~L-----ev~HI~prs~gGsd~~sNl~l~C~~Cn~~K~n~~~~~fl~~~~~~l~~i~  257 (425)
                        -..|-||-+..+++|+|||+....-     |||||+|+|..++||.+|+||+|.+||+.|+|++|.+|+..+.......+
  T Consensus       811 ~~~K~~Lw~eQ~gkc~YTG~~I~i~~L~~ydIDHIiP~S~~~DDS~~NKVLv~~~~N~~K~n~tP~e~~~~~~~~~~w~~l  890 (1372)
  T 4cmp_A          811 QNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQL  890 (1372)
  T ss_dssp             GSHHHHHHHHTTSBCTTTSCBCCHHHHHHHCEEESSCTTTCCCCSGGGEEECC--CCTTTCSSSSSCTTGGGGHHHHHHH
  T ss_pred             HHHHHHHHHHcCCCCCCCCCcCChhhccCcCccceeecccCCCCccchhhhccHHHHHhcCCCChHHHHHHHHHHHHHHH

  Q ss_pred             HhccCCchh----------------------HHHHHHHHH---HHHHHHHHhC-------------CCeEEeechhhhhh
  Q Thu_Feb_20_16:  258 RKAKAPLKD----------------------AAAVNSTRW---ALFNALKATG-------------LPLTTGTGGQTKFN  299 (425)
  Q Consensus       258 ~~~k~~lkd----------------------a~~vn~tR~---~l~~~Lk~~~-------------~~V~~~~G~~Tk~~  299 (425)
                        ...+.--..                      +..+|.|||   .+.+.|++..             ..|.++.|.+|...
  T Consensus       891 ~~~~~is~~K~~~L~~~~~~~~~~~e~~~~~Fi~RqL~DTryisK~~~~~L~~~~~~~~~~~~~~~~~~~~~Vv~vkG~~Ts~l  970 (1372)
  T 4cmp_A          891 LNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDF  970 (1372)
  T ss_dssp             HHTTSSCHHHHHHHHC-----CCHHHHHHHHHHHHCCCCHHHHHHHHHHHHHHCCCBCTTSCBCCCCEEEECCHHHHHHH
  T ss_pred             HHhhhcCHHHHHHHHhhhhcccccchhhcchhhccCHHHHHHHHHHHHHHHHHHhhhcccccccCCCCCEEEEECCHhhHHH

  Q ss_pred             hhhcCCcc-------cccccceeeccccc
  Q Thu_Feb_20_16:  300 RTRLDLPK-------AHWLDAACVGQVDA  321 (425)
  Q Consensus       300 R~~~gL~K-------~H~~DA~cig~~~~  321 (425)
                        |..|||+|       |||+||++|++.+.
  T Consensus       971 R~~wgl~K~r~~n~~HHA~DA~lia~~~~~  999 (1372)
  T 4cmp_A          971 RKDFQFYKVREINNYHHAHDAYLNAVVGT  999 (1372)
  T ss_dssp             HHHTTCCCCTTTCSHHHHHHHHHHHHHHH
  T ss_pred             HHHcCCCccCCCCcchhHHHHHHHHhccH

  No 4
>COG3513 Predicted CRISPR-associated nuclease, contains McrA/HNH-nuclease and RuvC-like nuclease domain [Defense
  mechanisms]
  Probab=99.48  E-value=3.6e-14  Score=154.15  Aligned_cols=138  Identities=28%  Similarity=0.304  Sum_probs=0.0

  Q ss_pred             hHHHHHHHHhCCccccCCCc--------CCceeEeeeccccCCCCcchhhhhhhhhhhhhhhccCCHHHHhccc-HHHHH
  Q Thu_Feb_20_16:  184 EVREYLLEKWGRKCTYCGTK--------DVPLEVEHIHPRSRGGTDRVSNLTMACHACNQSKGNQDIRDFLFGK-PELLQ  254 (425)
  Q Consensus       184 ~vrey~l~k~~~~C~yCg~~--------~~~Lev~HI~prs~gGsd~~sNl~l~C~~Cn~~K~n~~~~~fl~~~-~~~l~  254 (425)
                        +.+=||.+-.+.+|+|.|++        ...+|||||+|.|+..+||++|++|+.++.||.|+|++|.|++.++ ...|+
  T Consensus       561 ilKLrLY~~Q~gkcmYsgqei~I~rL~dk~~~eIDHi~P~Sr~~DDS~~NkVLv~s~~Nq~KgnqtP~e~~~~k~s~~Wq  640 (1088)
  T COG3513         561 ILKLRLYYLQNGKCMYSGQEIDIHRLSDKGYYEIDHIVPQSRTWDDSIDNKVLVLSSENQEKGNQTPYEEFVGKNSAKWQ  640 (1088)
  T ss_pred             hhHHHHHHHhcCccccCccccchhhccccceeeceecccccccccccceeEEecccccccCCCCHHHHhhhhHHHHH

  Q ss_pred             HHHHhccCCchh------------------HHHHHHHHH-------HHHHHHHHhCCC---------eEEeechhhhhhh
  Q Thu_Feb_20_16:  255 RILRKAKAPLKD------------------AAAVNSTRW-------ALFNALKATGLP---------LTTGTGGQTKFNR  300 (425)
  Q Consensus       255 ~i~~~~k~~lkd------------------a~~vn~tR-------~l~~~Lk~~~~~---------V~~~~G~~Tk~~R  300 (425)
                        ++.+..-.+=..                  ...+|.|||       .|-++|.-.+..         |.+..|.+|...|
  T Consensus       641 k~~~~vL~s~~~~tK~eR~ll~~~Dk~~Fi~RnLndTRyI~r~v~n~l~~rl~l~~d~n~kl~~kv~V~t~~g~lts~lR  720 (1088)
  T COG3513         641 KLEAGVLASQRPFTKAERILLQEYDKAGFIDRNLNDTRYIARFVANYLKDRLNLSGDENTKLGSKVRVFTLKGQLTSLLR  720 (1088)
  T ss_pred             HHHHHHHhccCchhHHHHhhhhhchhcchhhcchHHHHHHHHHHHHHHHHHHhccccCCccccccceeEEEeccHHHHHHH

  Q ss_pred             hhcCCcc-------cccccceeeccccc
  Q Thu_Feb_20_16:  301 TRLDLPK-------AHWLDAACVGQVDA  321 (425)
  Q Consensus       301 ~~~gL~K-------~H~~DA~cig~~~~  321 (425)
                        ..|||-|       |||+||++||+...
  T Consensus       721 ~~wgl~kvr~~N~~HHA~DAvVVa~a~~  748 (1088)
  T COG3513         721 KTWGLYKVRENNNYHHAHDAVVVACALN  748 (1088)
  T ss_pred             hhcceeeecccccchhhhhhhheecccH

  No 5
>2qgp_A HNH endonuclease; Q39X46, GMR87, X-RAY, NESG, structural genomics, PSI-2, protein structure initiative;
  2.60A {Geobacter metallireducens gs-15} PDB:  4h9d _A
  Probab=99.41  E-value=1.8e-13  Score=119.11  Aligned_cols=60  Identities=28%  Similarity=0.499  Sum_probs=0.0

  Q ss_pred             HHHHHHHhCCcccccCCcCCc--eeEeeeccccCCCCcchhhhhhhhhhhhhhhccCCHHHH
  Q Thu_Feb_20_16:  186 REYLLEKWGRKCTYCGTKDVP--LEVEHIHPRSRGGTDRVSNLTMACHACNQSKGNQDIRDF  245 (425)
  Q Consensus       186 rey~l~k~~~~C~yCg~~~~~--Lev~HI~prs~gGsd~~sNl~l~C~~Cn~~K~n~~~~~f  245 (425)
                        +.+...+++..|+|||....+  ++||||+|+|.||+|.++||+++|.+||+.|++.++.||
  T Consensus        27 ~~~~~~~~~g~C~yCg~~~~~~~~~iDHiiP~s~gg~~~~~Nlv~~C~~CN~~K~~~~p~e~   88 (112)
  T 2qgp_A           27 QWWKNRIARGICHYCGEIFPPEELTMDHLVPVVRGGKSTRGNVVPACKECNNRKKYLLPVEW   88 (112)
  T ss_dssp             HHHHHHHHHTBCTTTCCBCCGGGEEEEESSCTTTTCCCSTTTEEEEEHHHHHHHHTCCSHHH
  T ss_pred             HHHHHhHhCCCcCCCcCCCCCCcccccceeeecCCCCCChhHhHHHhhHHHHHHcCCCHHHH

  No 6
>PF13395 HNH_4:  HNH endonuclease
  Probab=99.23  E-value=4.6e-12  Score=96.76  Aligned_cols=46  Identities=50%  Similarity=0.959  Sum_probs=0.0

  Q ss_pred             ccccCCcCCce-------eEeeeccccCCCCcchhhhhhhhhhhhhhhccCCH
  Q Thu_Feb_20_16:  197 CTYCGTKDVPL-------EVEHIHPRSRGGTDRVSNLTMACHACNQSKGNQDI  242 (425)
  Q Consensus       197 C~yCg~~~~~L-------ev~HI~prs~gGsd~~sNl~l~C~~Cn~~K~n~~~  242 (425)
                        |.|||++-..-       +||||+|+|.+++|++.||+++|.+||+.|++++|
  T Consensus         1 C~Y~g~~i~~~~l~~~~~~iDHiiP~s~~~~~~s~~Nlvl~~~~~N~~K~~k~P   53 (54)
  T PF13395_consen    1 CPYCGKPISIENLFKNKYEIDHIIPRSRGGDDSFWNLVLCCKECNRSKGNKTP   53 (54)
  T ss_pred             CCCCCCCCChhhcccCCceeEEEecccccCCCCcchhheECHHHHhcccccCC

  No 7
>PF01844 HNH:  HNH endonuclease PDB:  2QGP _C.
  Probab=99.12  E-value=2.7e-11  Score=88.50  Aligned_cols=44  Identities=41%  Similarity=0.954  Sum_probs=0.0

  Q ss_pred             ccccCCcC---CceeEeeeccccCCCCcchhhhhhhhhhhhhhhccC
  Q Thu_Feb_20_16:  197 CTYCGTKD---VPLEVEHIHPRSRGGTDRVSNLTMACHACNQSKGNQ  240 (425)
  Q Consensus       197 C~yCg~~~---~~Lev~HI~prs~gGsd~~sNl~l~C~~Cn~~K~n~  240 (425)
                        |++||+..   ..++||||+|++.||.+.++||+++|..||+.|+++
  T Consensus         1 C~~C~~~~~~~~~~~~~~v~Hi~~~~~gg~~~~~Nl~~lC~~Ch~~k~~k   47 (47)
  T PF01844_consen    1 CQYCGKPGSDNESLHVHHIIPRSKGGKNDLENLILLCPSCHRKKHDK   47 (47)
  T ss_dssp             -TTT--B--GG-GEEEEESS-TTTT---STTTEEEEEHHHHHHHH--
  T ss_pred             CCCCCcCcccCcceEeECcCchhcCCCCCHHHHHHHhHHHHHHHhcCC
```

```
  No 8
>COG1403 McrA Restriction endonuclease [Defense mechanisms]
  Probab=98.76  E-value=5.4e-09  Score=91.49  Aligned_cols=61  Identities=36%  Similarity=0.641  Sum_probs=0.0

  Q ss_pred              hHHHHHHHHhCCcccccCCc-CCceeEeeeccccCCCCcchhhhhhhhhhhhhhhccCCHHHH
  Q Thu_Feb_20_16:  184 EVREYLLEKWGRKCTYCGTK-DVPLEVEHIHPRSRGGTDRVSNLTMACHACNQSKGNQDIRDF  245 (425)
  Q Consensus       184 ~vrey~l~k~~~~C~yCg~~-~~~Lev~HI~prs~gGsd~~sNl~l~C~~Cn~~K~n~~~~~f  245 (425)
                        .+|..++.+|+++|+||+.. . -++++||+|++.||.+..+||.++|..||..|++....+.
  T Consensus        59 ~~r~~~~~~d~~~c~~c~~~~~-~~~~dHiip~~~g~~~~~~Nl~~lc~~c~~~k~~~~~~~~  120 (146)
  T COG1403         59 KTRRAVLLRDNGLCQYCGSVGT-DLEVDHIVPLSRGGASAWENLETLCERCHNKKGSRLPGKA  120 (146)
  T ss_pred              HHHHHHHHhccccccccCCcCC-CceeeEEEeeccCCccHHHHHHHHHHHhcccccccccccc


  No 9
>PF14279 HNH_5:  HNH endonuclease
  Probab=97.05  E-value=0.00038  Score=56.63  Aligned_cols=39  Identities=38%  Similarity=0.837  Sum_probs=0.0

  Q ss_pred              ccccCCcCCcee--EeeeccccCCCCcchhhhhhhhhhhhhhhc
  Q Thu_Feb_20_16:  197 CTYCGTKDVPLE--VEHIHPRSRGGTDRVSNLTMACHACNQSKG  238 (425)
  Q Consensus       197 C~yCg~~~~~Le--v~HI~prs~gGsd~~sNl~l~C~~Cn~~K~  238 (425)
                        |.||++......  .|||+|-|-||.....|   +|.+||..-|
  T Consensus         1 Ci~C~~~~~~~~~~s~EHIIP~sLGG~~~~~~~---vC~~CN~~~g   41 (71)
  T PF14279_consen   1 CIYCNKEKSESNFSEEHIIPESLGGKLKINN---VCDKCNNKFG   41 (71)
  T ss_pred              CccCCCCCCccCCCccccCchhcCCcccccc---hhHHHhHHHh


  No 10
>PF13391 HNH_2:  HNH endonuclease
  Probab=96.99  E-value=0.0006  Score=53.06  Aligned_cols=44  Identities=27%  Similarity=0.269  Sum_probs=0.0

  Q ss_pred              ccccCCcC----CceeEeeeccccCC--CCc-----------chhhhhhhhhhhhhhhccC
  Q Thu_Feb_20_16:  197 CTYCGTKD----VPLEVEHIHPRSRG--GTD-----------RVSNLTMACHACNQSKGNQ  240 (425)
  Q Consensus       197 C~yCg~~~----~~Lev~HI~prs~g--Gsd-----------~~sNl~l~C~~Cn~~K~n~  240 (425)
                        |+.||..+    ..+|+-||+|.|.+  +.+           +++|.+++|.++|+.-.+.
  T Consensus         1 C~itg~~~~~~~~~~~eaaHI~P~s~~~~~~~~~~~~~~~~~~~~~~~~Ngl~L~~~lH~~fd~~   61 (66)
  T PF13391_consen   1 CVITGIRDPSPWELLEAAHIVPFSLGSWWMNNWFGEYANDWISPSNGLLLRPDLHKLFDRG   61 (66)
  T ss_pred              CcCCcCCcCCCCCcEEEEEcccCccCCCChhhhhhhhhccCCCccEEEEcCHhHHHHHCCC


  No 13
>1e7l_A GP49, recombination endonuclease VII; resolvase, holliday junction, DNAse; HET: SO4; 1.32A {Bacteriophage
  T4} SCOP: a.140.4.1 d.4.1.5 PDB:  2qnc _A*  1en7 _A  1e7d _A  2qnf _A*
  Probab=91.18  E-value=0.29  Score=45.98  Aligned_cols=53  Identities=19%  Similarity=0.421  Sum_probs=0.0

  Q ss_pred              hHHHHHHHHhCCcccccCCc----CCceeEeeeccccCCCCcchhhhhhhhhhhhhhhc
  Q Thu_Feb_20_16:  184 EVREYLLEKWGRKCTYCGTK----DVPLEVEHIHPRSRGGTDRVSNLTMACHACNQSKG  238 (425)
  Q Consensus       184 ~vrey~l~k~~~~C~yCg~~----~~~Lev~HI~prs~gGsd~~sNl~l~C~~Cn~~K~  238 (425)
                        +.++-+++..+..|+.|+..    ...+.|||..+-.  |+.+=.=--++|..||...|
  T Consensus        10 ~~~~~l~~~Q~g~C~IC~~~l~~~~~~~~~vDHdH~~~--g~~tG~VRGlLC~~CN~~lG   66 (157)
  T 1e7l_A          10 EEKQKFYDAQNGKCLICQRELNPDVQANHLDHDHELN--GPKAGKVRGLLCNLCDAAEG   66 (157)
  T ss_dssp              HHHHHHHHHTTTBCTTTCCBCCSSGGGSEEEECCCCS--STTTTBEEEEECHHHHHHHH
  T ss_pred              HHHHHHHHHhCCcCcccCCcccccccccccCCCCCC--CCCCCcchhhccHhHhchhc


  No 16
>1hjr_A Holliday junction resolvase (RUVC); site-specific recombinase; 2.50A {Escherichia coli} SCOP: c.55.3.6
  Probab=87.67  E-value=1.5  Score=40.99  Aligned_cols=41  Identities=22%  Similarity=0.353  Sum_probs=0.0

  Q ss_pred              EEEEEcCCCCcEEEEEEECC----EEEEEEEEEeCchHHHHHHHH
  Q Thu_Feb_20_16:   56 LELKLDPGSKVTGIAIKQGN----KIIFGAELQHRGQQIKDALLS   96 (425)
  Q Consensus        56 i~lgiDpG~k~~G~av~~~~----~~l~~~el~~r~~~vk~~l~~   96 (425)
                        ..||||||+..+|++|+..+    ..+-.+.+..+.....+||..
  T Consensus         2 ~ILGIDPGl~~tG~gvi~~~g~~~~~v~~G~i~t~~~~~~~Rl~~   46 (158)
  T 1hjr_A           2 IILGIDPGSRVTGYGVIRQVGRQLSYLGSGCIRTKVDDLPSRLKL   46 (158)
  T ss_dssp              EEEEEECCSSEEEEEEEEEETTEEEEEEEEEEECCCSCHHHHHHH
  T ss_pred              EEEEEccCCCCeeEEEEEecCCeEEEEEeeEEECCCCCHHHHHHH


  No 17
>PF02075 RuvC:  Crossover junction endodeoxyribonuclease RuvC; PDB:  1HJR _A.
    Probab=87.43  E-value=1.4  Score=40.65  Aligned_cols=40  Identities=23%  Similarity=0.387  Sum_probs=0.0

  Q ss_pred              EEEEcCCCCcEEEEEEECC----EEEEEEEEEeCch-HHHHHHHH
  Q Thu_Feb_20_16:   57 ELKLDPGSKVTGIAIKQGN----KIIFGAELQHRGQ-QIKDALLS   96 (425)
  Q Consensus        57 ~lgiDpG~k~~G~av~~~~----~~l~~~el~~r~~-~vk~~l~~   96 (425)
                        .||||||...+|+||+..+    ..+-.+.+..+.+ ...+||..
  T Consensus         1 ILGIDPgl~~tG~avi~~~~~~~~~~~i~~G~I~t~~~~~~~~Rl~~   45 (149)
  T PF02075_consen   1 ILGIDPGLSNTGYAVIEEDGGKLRLIDYGTIKTSSKDSLPERLKE   45 (149)
  T ss_dssp              EEEEE--SSEEEEEEEEEETTEEEEEEEEEEE---S--HHHHHHH
  T ss_pred              CEEECCCCCCeeEEEEEeeCCEEEEEEeCeEECCCCCCHHHHHHH


  No 18
>4ep4_A Crossover junction endodeoxyribonuclease RUVC; resolvase, hydrolase; 1.28A {Thermus thermophilus} PDB:
   4ep5 _A
  Probab=86.99  E-value=1.3  Score=41.83  Aligned_cols=41  Identities=17%  Similarity=0.267  Sum_probs=0.0

  Q ss_pred              EEEEEcCCCCcEEEEEEECCEE-------EEEEEEEeCch-HHHHHHHH
  Q Thu_Feb_20_16:   56 LELKLDPGSKVTGIAIKQGNKI-------IFGAELQHRGQ-QIKDALLS   96 (425)
  Q Consensus        56 i~lgiDpG~k~~G~av~~~~~~-------l~~~el~~r~~-~vk~~l~~   96 (425)
                        ..||||||+..+|++|+....-       +-.+.+..+.+ ...+||..
  T Consensus         2 rILGIDPGl~~tG~gvi~~~g~~~~~~~~~~v~~G~I~t~~~~~~~~~RL~~   50 (166)
  T 4ep4_A           2 VVAGIDPGITHLGLGVVAVEGKGALKARLLHGEVVKTSPQEPAKERVGR   50 (166)
  T ss_dssp              EEEEEECCSSEEEEEEEEECSSSSSCEEEEEEEEEECCTTSCHHHHHHH
  T ss_pred              EEEEEccccCceEEEEEEEecCCccceEEEEEeCeEECCCCCCHHHHHHH
```

```
  No 22
>4oge_A HNH endonuclease domain protein; crispr-CAS, CAS9, RUVC, RNA-guided DNA endonuclease, cytoplasmic,
  hydrolase; HET: SPD; 2.20A {Actinomyces naeslundii} PDB:  4ogc _A*
  Probab=80.53  E-value=1.3  Score=52.86  Aligned_cols=74  Identities=19%  Similarity=0.209  Sum_probs=0.0

  Q ss_pred             EEEEEcCCCCcEEEEEEECC------EEEEEEEEEeC--------chHHHHHHHHHHHHHhhhccccCccCCchhhcccC
  Q Thu_Feb_20_16:   56 LELKLDPGSKVTGIAIKQGN------KIIFGAELQHR--------GQQIKDALLSRRQLRRGRRNRKTRYRQPRFLNRNR  121 (425)
  Q Consensus        56 i~lgiDpG~k~~G~av~~~~------~~l~~~el~~r--------~~~vk~~l~~Rr~~RR~RR~rk~ryR~~RF~NR~r  121 (425)
                        .+||||-|..-+|.||+..+      ..+-.+...+.        ......+...|.+.+.+|+.|++++|..|..+--.
  T Consensus        12 YrLGLDlGi~SiGWAvi~~d~~~~p~~Ii~~GvRiF~~g~~~~~~k~~~s~~~~Rr~~R~~RR~~rRrk~R~~~l~~ll~   91 (1101)
  T 4oge_A           12 LRVGIDVGTHSVGLATLRVDDHGTPIELLSALSHIHDSGVGKEGKKDHDTRKKLSGIARRARRLLHHRRTQLQQLDEVLR   91 (1101)
  T ss_dssp             EEEEEEECSSEEEEEEEEECTTCCEEEEEEEEEEECC----------------CHHHHHHHHHHHHHHHHHHHHHHHHHH
  T ss_pred             eEEEEEecccceeeEEEEecCCCCcceeeecceEeccCCCCccccCccccchHHHHHHHHHHHHHHHHHHHHHHHHHHHH

  Q ss_pred             CCCCCChh
  Q Thu_Feb_20_16:  122 PDGWLAPS  129 (425)
  Q Consensus       122 ~~gwL~PS  129 (425)
                        ..||+||+
  T Consensus        92 ~~g~~~~~   99 (1101)
  T 4oge_A           92 DLGFPIPT   99 (1101)
  T ss_dssp             TTTCCCC-
  T ss_pred             HcCCCCCC


  No 37
>4cmp_A CAS9, crispr-associated endonuclease CAS9/CSN1; hydrolase, DNAse, RNA-guided, immunity, crRNA, genome
  editin; 2.62A {Streptococcus pyogenes} PDB:  4cmq _A
  Probab=58.70  E-value=5.3  Score=49.16  Aligned_cols=53  Identities=28%  Similarity=0.396  Sum_probs=0.0

  Q ss_pred             CCEEEEEcCCCCcEEEEEEEC--------------------CEEEEEEEEEeCchHHHHHHHHHHHHHhhhccccCc
  Q Thu_Feb_20_16:   54 QDLELKLDPGSKVTGIAIKQG--------------------NKIIFGAELQHRGQQIKDALLSRRQLRRGRRNRKTR  110 (425)
  Q Consensus        54 qpi~lgiDpG~k~~G~av~~~--------------------~~~l~~~el~~r~~~vk~~l~~Rr~~RR~RR~rk~r  110 (425)
                        ++..||+|.|+--+|-||+.+                    ++.+|.+.|.=-++.    =++||.+|-+||+..+|
  T Consensus         7 k~y~lGLDiGTnSvGWAv~d~~~~~~~~~k~~~~~~~~~~~~~~k~~~G~rlF~~g~t----aaeRr~~R~~RRr~~Rr   79 (1372)
  T 4cmp_A            7 KKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGET----AEATRLKRTARRRYTRR   79 (1372)
  T ss_dssp             -CCEEEEEECSSEEEEEEECTTSCCCEEEEEBCSSSSCCEEEEESEEEEECCCCSC----HHHHHTSCCHHHHHHHH
  T ss_pred             CCceeeeccccCcceeEEEcCcccchhhhhccccCcccCcCCccccccccCcCCcc----hHHHHhhHHHHhhHHHH


  No 44
>COG0675 Transposase and inactivated derivatives [DNA replication, recombination, and repair]
  Probab=50.60  E-value=75  Score=31.60  Aligned_cols=152  Identities=18%  Similarity=0.323  Sum_probs=0.0

  Q ss_pred             EEEEEEeccCCCccCCEEEEEcCCCCcEEEEEEECCEEEEEEEEEeCchHHHHHHHHHHHHHhhhccccCccCCchhhcc
  Q Thu_Feb_20_16:   40 FTLILKEAHPDVPVQDLELKLDPGSKVTGIAIKQGNKIIFGAELQHRGQQIKDALLSRRQLRRGRRNRKTRYRQPRFLNR  119 (425)
  Q Consensus        40 FtI~L~~~~~~~~~qpi~lgiDpG~k~~G~av~~~~~~l~~~el~~r~~~vk~~l~~Rr~~RR~RR~rk~ryR~~RF~NR  119 (425)
                        +.++...++. .....-.+|||.|  ...+++++.+   ........+.....++......-.....+..+.++++..=.
  T Consensus       161 v~~~~~~~~~-~~~~~~~~giDlg--~~~~~~~~~~---~~~~~~~~~~~~~~~~~~~~k~ls~~~~~~~~~~~~~~~~~  234 (364)
  T COG0675       161 VLVEVEEEEP-VPKPGKVVGIDLG--LKNLATVSDN---GTVPIPRDGRKLLKRLKKAQKRLSRKKSRSKRRKKAKLKLA  234 (364)
  T ss_pred             EEEEeccccc-cCCCCCeEEEecC--CceEEEEeCC---ccccCchhHHHHHHHHHHHHHHHHHHhhhhhhhhhHHHHHH

  Q ss_pred             cCCCCCCChhHHHHHHHHHHHHHHHHHhCCcchheeeeeeeehhhhcCCccccceecccccch-hHHHHHHHH--hC--
  Q Thu_Feb_20_16:  120 NRPDGWLAPSLKHRVDTVLTWVNRLRNLAPLGSIAQELVRFDLQKLQNPEISGIEYQQGELQGY-EVREYLLEK--WG--  194 (425)
  Q Consensus       120 ~r~~gwL~PSi~~~~~~~~~~v~~l~~~~Pi~~i~~E~v~fD~qk~~np~i~G~~YQqg~l~g~-~vrey~l~k--~~--  194 (425)
                        +..+  ..-+.+...+..-+++-++...+.+..|++|...=              -+.--..+| ++++.|-+|  ||
  T Consensus       235 ~~~~--~~~~~r~~~~k~~~~~v~~~~~~~~~~~~~~~~~~~--------------~~~i~~~~~~~~~~~l~yka~~~~~  298 (364)
  T COG0675       235 RLRE--RIRNRRKDFHKLAKKLVELADLVGVETLVVEDLVK--------------RRSISDWAFGELRRQLEYKAEWGGI  298 (364)
  T ss_pred             HHHH--HHHHHHHHHHHHHHHHHHHhhheeeeeeeeccccc--------------ccchhhhhHHHHHHHHHHHHHhCCe

  Q ss_pred             -----------CcccccCCcCCceeEeeeccccCCCCcchhhhhhhhhhh
  Q Thu_Feb_20_16:  195 -----------RKCTYCGTKDVPLEVEHIHPRSRGGTDRVSNLTMACHAC  233 (425)
  Q Consensus       195 -----------~~C~yCg~~~~~Lev~HI~prs~gGsd~~sNl~l~C~~C  233 (425)
                                   -+|..||.                    .+.-...|..|
  T Consensus       299 ~v~~~~~~~~tS~~C~~cg~--------------------~~~r~~~C~~c  328 (364)
  T COG0675       299 VVKVVPPYYTSKTCPCCGH--------------------LSGRLFKCPRC  328 (364)
  T ss_pred             eEEEcCCCCCccccccccC--------------------ccCceEECCCC


  No 57
>COG3513 Predicted CRISPR-associated nuclease, contains McrA/HNH-nuclease and RuvC-like nuclease domain
  Probab=36.67  E-value=30  Score=40.66  Aligned_cols=63  Identities=32%  Similarity=0.385  Sum_probs=0.0

  Q ss_pred             CEEEEEcCCCCcEEEEEEECCE-------------EEEEEEEEeCch-HHHHHHHHHHHHHhhhccccCccCCchhh
  Q Thu_Feb_20_16:   55 DLELKLDPGSKVTGIAIKQGNK-------------IIFGAELQHRGQ-QIKDALLSRRQLRRGRRNRKTRYRQPRFL  117 (425)
  Q Consensus        55 pi~lgiDpG~k~~G~av~~~~~~~-------------~l~~~el~~r~~-~vk~~l~~Rr~~RR~RR~rk~ryR~~RF~  117 (425)
                        +..||+|.|---+|-||+..++             +++.+||---|+ .--.+--.|.+.||-||++.+=-+-.||.
  T Consensus         4 ~yilglDIGi~SVGWAvve~de~~~~~~lkd~Gvrif~~ae~pktGeslA~~rRLaRsaRRr~~RRk~Rl~~lkrL~   80 (1088)
  T COG3513         4 AYILGLDIGINSVGWAVVEDDEDEVPIKLKDLGVRIFEGAELPKTGESLALARRLARSARRRIRRRKNRLLYLKRLL   80 (1088)
  T ss_pred             ceEEEeeccccceeeEEeecccccccchHHHHhhhhhhcccccCCcchhhHHHHHHHHHHHHHHHHHHHHHHHHHHHH


  No 75
>PF01385 OrfB_IS605:  Probable transposase;
  Probab=29.04  E-value=39  Score=31.85  Aligned_cols=110  Identities=16%  Similarity=0.156  Sum_probs=0.0

  Q ss_pred             EEEEEEeccC---CCccCCEEEEEcCCCCcEEEEEEECCEEEEEEEEEeCchHHHHHHHHHHHHHhhhccccCccC---C
  Q Thu_Feb_20_16:   40 FTLILKEAHP---DVPVQDLELKLDPGSKVTGIAIKQGNKIIFGAELQHRGQQIKDALLSRRQLRRGRRNRKTRYR---Q  113 (425)
  Q Consensus        40 FtI~L~~~~~---~~~~qpi~lgiDpG~k~~G~av~~~~~~l~~~el~~r~~~vk~~l~~Rr~~RR~RR~rk~ryR---~  113 (425)
                        |++.|.++..   ......-.+|||+|-.+.-.++...+..     ....+..++......+..++...+...+..   .
  T Consensus       106 ~yl~i~~~~~~~~~~~~~~~~~~vgVDlGi~~~a~~~~~~~~~~~~-----~~~~~~~~l~~~~~~~~~~~~~~~~l~~~~~~~~~~~~~~k~  180 (227)
  T PF01385_consen  106 WYLSITVEVEVEPPPPDTEKVVGVDLGIKNLATVSSGDGTR-----FIINGKELKSIKKRLRKLRRRLQRKKKKSKKRKL  180 (227)
  T ss_pred             ccccccccccccccccccceeeeecccccceeeccccccccc-----cccccccccccccccccccchhccccccccccch


  Q ss_pred             chhhcccCCCCCCChhHHHHHHHHHHHHHHHHHhCCcchheee
  Q Thu_Feb_20_16:  114 PRFLNRNRPDGWLAPSLKHRVDTVLTWVNRLRNLAPLGSIAQE  156 (425)
  Q Consensus       114 ~RF~NR~r~~gwL~PSi~~~~~~~~~~v~~l~~~~Pi~~i~~E  156 (425)
                        .|...|-..+  +.=-+.+=....-++|-+.+...=|..|++|
  T Consensus       181 ~~~~~rl~~k--~~~~~~d~~hk~s~~iV~~~~~~~ve~l~~~  221 (227)
  T PF01385_consen  181 RRKLARLHRK--IRNRRKDFLHKVSKKIVKEADEIGVEDLVVE  221 (227)
  T ss_pred             hhHHHHHHHH--HHHHHHHHHHHHHHHHHHhCCeEeeCccEeC
```


A
1
2
3
4
5
6
7
8
9
10
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
45
46
47
48
49
50
52
53
54
58
59
60
61
62
63
64
67
68
70
71
72
74
75
76
77
78
79
80
81
82
84
85
86

B



Supplementary Figure S9. CRISPR repeats and tracrRNA orthologs pairwise identity. (**A**) Sequence identity of each type II CRISPR repeat pair is represented with a color scale from green (lowest) to red (highest identity). The repeat sequences are ordered according to the corresponding Cas9 sequences on the Cas9 phylogenetic tree. The numbers on the left indicate the exact position on the tree. Color bars on the top and left correspond to the colors of Cas9 tree clusters. (see Figure 4 and Supplementary Table S1) Note the high similarity of the repeat sequences belonging to the same, large clusters of the Cas9 tree. (**B**) Sequence identity of every tracrRNA sequence pair is represented with a color scale from green (lowest) to red (highest identity). The tracrRNA sequences are ordered according to the corresponding Cas9 sequences on the Cas9 phylogenetic tree and the numbers on the left indicate the exact position on the tree. Color bars on the top and left correspond to the colors of the Cas9 tree clusters (see Figure 4 and Supplementary Table S1). Note that tracrRNA sequences show higher similarity only in small clusters of closely related loci.

Type II-B

F. novicida 3523

```
GUUUCAGUUACUAGAUUAUCUGAUAUACUA-UCUUAG
GUUUCAGUAGCUAGAUUAUUUGGUACAUACUAGUGUU
UUUUCGUUAAUUAGGUUAUUUGAUAUACUGCUUUCAA
AUUUUAGCGGUUAUAUUAUCUGAUAUACUUUACCCAG
GUUUCGGUAGCUAGGUUAUUUGAUGUACAAUAAAAAC
UUAUAAGUAGCUAGAUUAUUUGAUAUACUGCUGUUAG
GUUUCAGUAGCUAGAUUAUUUGAUAUACUGCUGUUAG
```

F. novicida U112 and Fx1

```
UUUUUAGUAAACUAGACUAUUUAAUAAACUA-UGUUAG
GUUUCAGUUGUUAGAU-UAUUUGGUAUGUACUUGUGUU
UUUUUAUUAUUUGAAU-GCUUUGGUAGUAUGUUGAGAU
GUUUUCGUGAUUAUAU-UAUUUGAUACAAUGUGUCCAA
GUUUCAGUUGCUGAAU-UAUUUGGUAAACUACUGUUAG
```

Gammaproteobacterium HTCC5015

```
AUAACAAGGGAGUAUUUGCUGGAAUAAGCUCUGAAAC
GUUUCAGCUGUUGGUUUGUUGGGAUAAGCUCUGAAAC
```

Distance from 1st repeat: 55 nt instead of ~33 nt spacer

Wolinella succinogenes DSM1740

```
UUUGUUAAAGCUGGAUGGGAUUAUUAUAGAGUGUUGC
GUUUCACAGGCUAAGCGGAUUUGCUAUAAAGUGUUGC
```

Distance from 1st repeat: 37 nt, like other spacers (~33-37 nt)

Legionella pneumophila str. Paris

```
AUUUCAUUAGUUAAAUUUUUAGAUGAGGGAUUAUUGG
GUUUCAGUGGUUGGAUUUUUAGAUGAGGGAUUAUUGG
```

Distance from 1st repeat: 38 nt, like other spacers (~34-36 nt)

Sutterella wadsworthensis 3_1_45B

```
GUUUCAGCGCUUUAGCUUGCAGCGUUAUGAUCUCUAU
GUUUCAGUGCUAUAGCUCGUAGCGUUAUGAUCUUCGC
```

Distance from 1st repeat: 38 nt, like other spacers (~33-39 nt)

Type II-A

Streptococcus pyogenes M1 GAS

GUUUUAGAGCUAUGCUGUUUUGAAUGGUCCCAAAAC
GUUUUAGAGCUAUGCUGUUUUGAAUGGUCUCCAUUC

Streptococcus thermophilus LMD-9

GUUUUUGUACUCUCAAGAUUUAAGUAACUGUACAAC
GUUUUUGUACUCUCAAGAUUUAAGUAACUGUACAGU

Treponema denticola ATCC 35405

GUUUGAGAGUUGUGUAAUUUAAGAUGGAUCUCAAAC
GUUUGAGAGUUGUGUAAUUUAAGAUGACAACCAAGC

Type II-C

Neisseria meningitidis Z2491

GUUGUAGCGCCCAUUCUCAUUUCGCAGUGCUACAAU
GUUGUAGCUCCCUUUCUCAUUUCGCAGUGCUACAAU

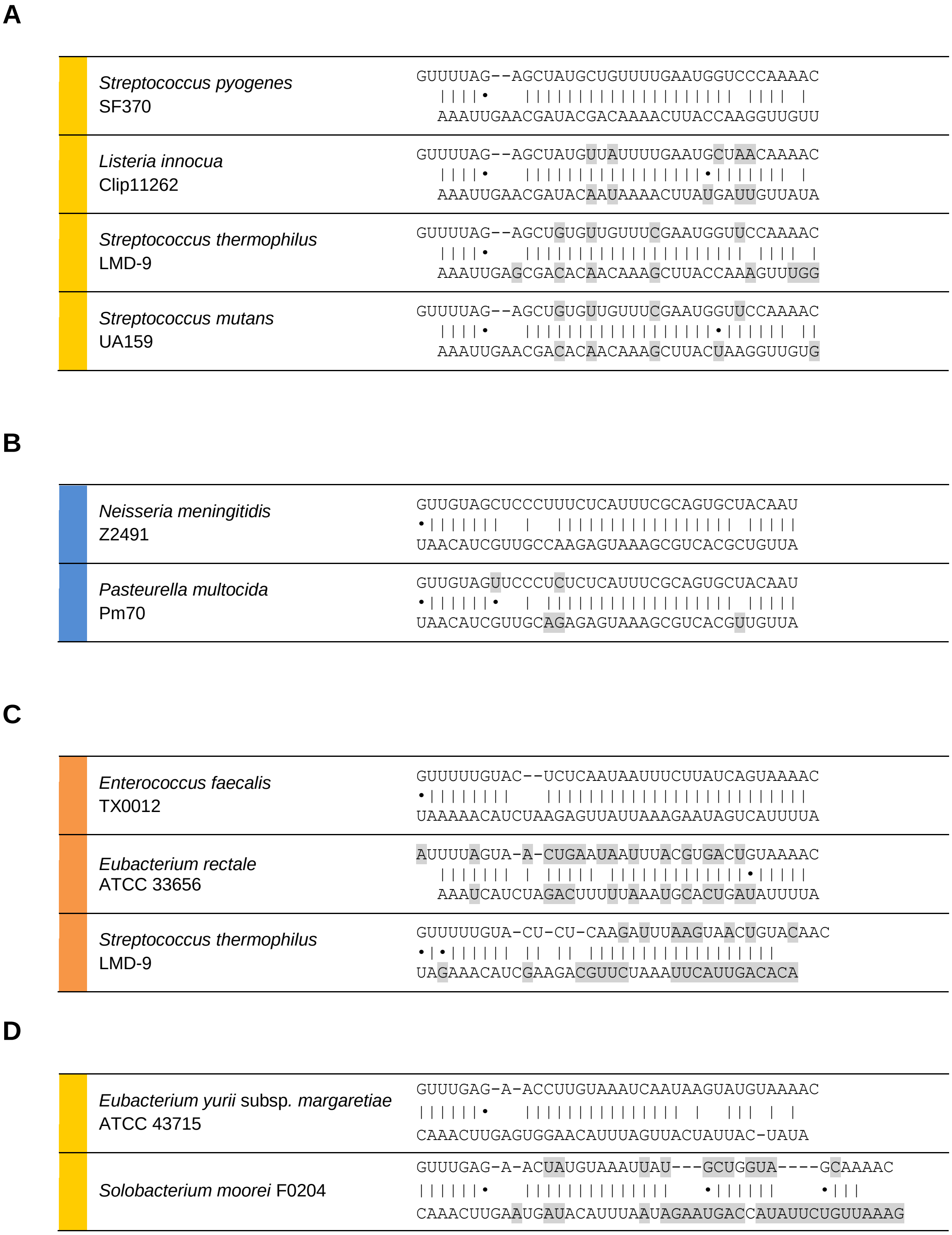
Campylobacter jejuni subsp. *jejuni* NCTC 11168

AUUUUAGUCCCUUUUUAAAUUUCUUUAUGGUAAAAU
GUUUUAGUCCCUUUUUAAAUUUCUUUAUGGUAAAAU

Candidatus Puniceispirillum marinum IMCC1322

UGGGGGGUAGGCUCUCAAUCACCAGAGUGCUAUACU
GUUGCUCUAGGCUCUCAAUCACCAGAGUGCUAUACU

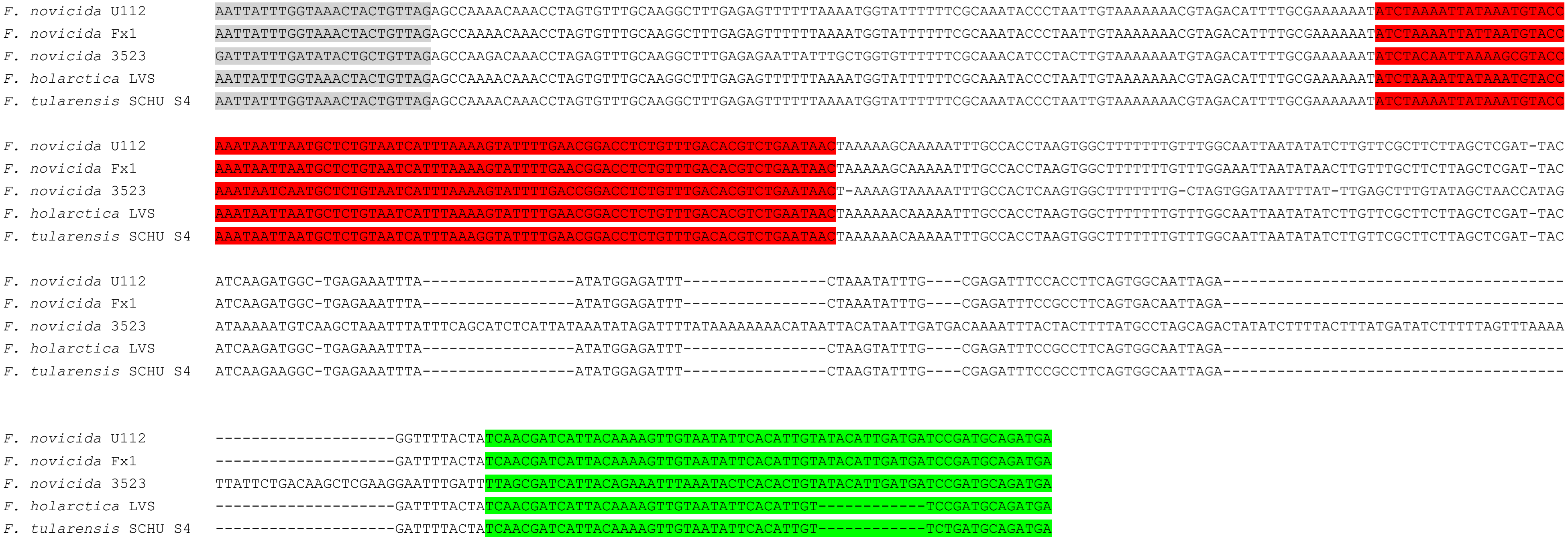
Supplementary Figure S10. Degenerated repeats of type II CRISPR-Cas. Black: consensus repeat sequence; red: degenerated repeat sequence; yellow highlighted: scaRNA repeat. Nucleotides mutated in the degenerated repeats are highlighted in grey. For every strain, the repeats are listed in the direction of 5’ to 3’ end of the CRISPR array. For the type II-B, the degenerated repeat distance from the first, regular repeat in the array is given. Note that unlike type II-B, type II-A and II-C repeats display a lower amount of mutations and their distribution are typically at one of the repeat ends.



Supplementary Figure S11. Compensatory mutations within CRISPR repeat:tracrRNA duplexes suggest coevolution of type II CRISPR-Cas RNA components. Basepairings: upper sequence, repeat (5' to 3' orientation); lower sequence, tracrRNA antirepeat (3' to 5' orientation). Highlighted are nucleotides different from the corresponding ones in top duplexes for every panel. Color bars indicate clusters of Cas9 tree of selected loci (see Figure 4). (**A**, **B**) RNAs associated with very closely related Cas9 orthologs (same BLASTclust group, see Supplementary Materials and Methods and Supplementary Table S1). (**C, D**) RNAs associated with Cas9 orthologs from the same Cas9 phylogenetic tree cluster but from different BLASTCLUST groups. Note the conservation of the cofolding pattern in (**A**, **B**, **C**) and 5' part of (**D**). The conserved structure at the 5' part of the repeat in (**D**) is most likely present in mature crRNA:tracrRNA duplex. Note, that bacterial strains in (**C**) belong to different families (*Bacilli* and *Clostridia*) and in (**D**) to different phyla (*Firmicutes* and *Tenericutes*).

```
F. novicida U112         CAACGTCTAGTTGTTTCCCATGCCATCCATTTAAGTTGATGTTTCACTAAATTCTTAATCCAAAACGAAATATAATTTTGCT-ATAACTATTATAACAAATAATCATTAAACTAACTGTATAAAACTAATAAAAATAACTTTAAAAAAGA
F. novicida Fx1          CAACGTCTAGTTGTTTCCCATGCCATCCATTTAAGTTGATGTTTCACTAAATTCTTAATCCAAAACGAAATATAATTTTGCT-ATAACTATTATAACAAATAATCATTAAACTAACTGTATAAAGCTAATAAAAATAACTTTAAAAAAGC
F. novicida 3523         CAACGTCTGGTTGTTTCCCATGCCATCCATTTAAGTTGATTTTTCACTAAATTCTTAATCCAAAACGAAATATAATTTTGCTAATAGATATTATAACAAATAATCATTAAACTAACTGTATAAAGCTACTAAAAATAACTTTTAAAAGGC
F. holarctica LVS        ------------------------------------------------------------------------------------------------------------------------------------------------------
F. tularensis SCHU S4    ------------------------------------------------------------------------------------------------------------------------------------------------------

F. novicida U112         TATTTTAAAGACCCTTTTAGCTTGGTTAGTTTGAGATTTTTATTTATACTTTATAAATACTGATAATCTCAATATTAGTTATGTTTACATTTTTGTTGTTAGCTATCGCGTTTATAGCAACCCTATATGCTTTATACTACTTTATTTTAG
F. novicida Fx1          TATTTTAAAGACCCTTTTAGCTTGGTTAGTTTGAGATTTTTATTTATATTTTATAAATACTGATAATCTCAATATTAGTTATGATTACACTTTTGTTGTTAGCTGTCGCGTTTATAGCAACCCTATATGCTTTATACTACTTTATTTTAG
F. novicida 3523         TATTTTAGAAACTCTTTTA-TTTTGCTAGTTTGAGCTTTTTATTTATACTCTATAAATACTGATAATCTCAATATTAGTTATGATTACACTTTTGTTGTTAGCTATGGTATTTATGGCAACTTTATATTCTATATACTACTTTATTTTAG
F. holarctica LVS        ------------------------------------------------------------------------------------------------------------------------------------------------------
F. tularensis SCHU S4    ------------------------------------------------------------------------------------------------------------------------------------------------------

F. novicida U112         CTATTGTTAAGCTAATTGTAGCGATATATCATTTTAACTATTTTTTTTCTAAAAATGATTCAAAAAATTAATTAATTCATAAAATAAATTGCTTAAAGTCTATCAAAACTTTGTTTGTCTATTTCTGAAAGTAGTATTTAGATAA-----
F. novicida Fx1          CTATTTTTAAGCTAATTGTAGCGATATATCATTTTAACTATTTTTTTTCTAAAACTGATTCAAAAAATTAATTAATTCATAAAATAAATTGCTTAAAGTCTATTAAAACTTTGTTTGTCTATTTCTGAAAGTAGTATTTAGATAA-----
F. novicida 3523         CTATTATTAAGCTAATTGCAGCCATATATCATTTTAGCTATTATTTTT----AACTGATTCAAAAAA-TAATTAATCCATAAAATAAATTGCTTAAAGTCTCTTAAAACTTTGTTTACATAGCTCTTAGAATTAGGTTTTCATAAGCAAG
F. holarctica LVS        ------------------------------------------------------------------------------------------------------------------------------------------------------
F. tularensis SCHU S4    ------------------------------------------------------------------------------------------------------------------------------------------------------

F. novicida U112         ----AATCTTATGAAAAATTA--TTTTTAAATGAGTTTTTTAGTAAACTAGACTATTTAATAAACTATGTTAGAATTATCAAAGCAAATG-AAAAAGATTAGCTGTAAGTTTCAGTTGTTAGATTATTTGGTATGTACTTGTGTTA----
F. novicida Fx1          ----AAGCTTATGAAAAATTA--TTTTTAAATGAGTTTTTTAGTAAACTAGACTATTTAATAAACTATGTTAGAATTATCAAAGCAAATG-AAAAAGATTAGCTGTAAGTTTCAGTTGTTAGATTATTTGGTATGTACTTGTGTTA----
F. novicida 3523         TAATGATCTTACTAAATATTATTTTTTTTTAATAAG--TTTCAGT-TACTAGATTATCTGATATACTATCTTAGGATTAGTCCAGCAAATGAAAAAAGATTAGCTGTAAGTTTCAGTAGCTAGATTATTTGGTACATACTAGTGTTATAGG
F. holarctica LVS        ------------------------------------------------------------------------------------------------------------------------------------------------------
F. tularensis SCHU S4    ------------------------------------------------------------------------------------------------------------------------------------------------------

F. novicida U112         GTTTAAAGTAGCT----------AGAAAATTCACTTTTAGACCTAC-TTATTTTTATTATTTG----AATGCTTT-GGTAGTATGTTGAGATCAACATTTTGCTGTAAAACTGTTTTCGTGATTATATTATTTGATACAATGTGTCCAAC
F. novicida Fx1          GTTTAAAGTAGCT----------AGAAAATTCACTTTTAGACCTAC-TTATTTTTATTATTTG----AATGCTTT-GGTAGTATGTTGAGATCAACATTTTGCTGTAAAACTGTTTTCGTGATTATATTATTTGATACAATGTGTCCAAC
F. novicida 3523         GTTAGGATTAACCTAACATATAAAGAACAATTAGCTGTAAATTTTCGTTAATTAGGTTATTTGATATACTGCTTTCAATACTGTACTTGAATTGATATTTTGCTATATAACTATTTTAGCGGTTATATTATCTGATATACTTTACCCAGA
F. holarctica LVS        ------------------------------------------------------------------------------------------------------------------------------------------------------
F. tularensis SCHU S4    ------------------------------------------------------------------------------------------------------------------------------------------------------

F. novicida U112         TAGAAACTTTTAGGAATAATTTACAAAATTAATTAA--------------------------------------------------------------------------------------------------------------
F. novicida Fx1          TAGAAACTTTTAGGAATAATTTACAAAATTAATTAA--------------------------------------------------------------------------------------------------------------
F. novicida 3523         TAGAAACTTCTAAATATTATTCGTAAAATTAATTAAGTTTCGGTAGCTAGGTTATTTGATGTACAATAAAAACTTACCTGTTTGATGCTAGGTATTAGTGTTATTTATAAGTAGCTAGATTATTTGATATACTGCTGTTAGGGTATGACA
F. holarctica LVS        ------------------------------------------------------------------------------------------------------------------------------------------------------
F. tularensis SCHU S4    ------------------------------------------------------------------------------------------------------------------------------------------------------

F. novicida U112         --------------------------GTTTCAGTTGCTGAATTATTTGGTAAACTACTGTTAGACAATGGCAAGCTTGATTACTTTTATAGTGTGAT---GTTTCAGTTGCTGAATTATTTGGTAAACTACTGTTAGAAGGATTTAATAC
F. novicida Fx1          --------------------------GTTTCAGTTGCTGAATTATTTGGTAAACTACTGTTAGCTTTAACTCATCAAAATTCACAAATCACCTATTAAAAGTTTCAGTTGCTGAATTATTTGGTAAACTACTGTTAGATCAAATCGACAG
F. novicida 3523         ATGACTATAACCAACACTCGTCCAGAGTTTCAGTAGCTAGATTATTTGATATACTGCTGTTAGTTTTCTTAATCTTAGGGTATCTAAATACAAATTCTCTGTTTCAGTAGCTAGATTATTTGATATACTGCTGTTAGAATAAGCATTTCA
F. holarctica LVS        ------------------------------------------------------------------------------------------------------------------------------------------------------
F. tularensis SCHU S4    ------------------------------------------------------------------------------------------------------------------------------------------------------
```

```
F. novicida U112        CTTAAAATTATCCAAATTTATATGTTTCAGTTGCTGAATTATTTGGTAAACTACTGTTAGTGCCTAATCCTCAGCAGCAACACTTAATTCAAGAATGTTTCAGTTGCTGAATTATTTGGTAAACTACTGTTAGTTACACATTCAGCTGGT
F. novicida Fx1         TAAATACATATATGAGAAGCT--GTTTCAGTTGCTGAATTATTTGGTAAACTACTGTTAGTAACGTTTTGGAATAAGTCTACATTTTCTTTATT--GTTTCAGTTGCTGAATTATTTGGTAAACTACTGTTAGATACAATCTTGTATTAA
F. novicida 3523        AGATATTTTATTTAATTTTTAT-GTTTCAGTAGCTAGATTATTTGATATACTGCTGTTAGTCTAATGTAGAAGCTATATTATCTATTTCACAATT-GTTTCAGTAGCTAGATTATTTGATATACTGCTGTTAGCTGGTGGAGGCGGTCAA
F. holarctica LVS       ------------------------------------------------------------------------------------------------------------------------------------------------------
F. tularensis SCHU S4   ------------------------------------------------------------------------------------------------------------------------------------------------------

F. novicida U112        TCTTTAGCATTATT--AACGTTTCAGTTGCTGAATTATTTGGTAAACTACTGTTAGAGTAATAGCCAATGCTTTTAAACTACTGATATATACGTTTCAGTTGCTGAATTATTTGGTAAACTACTGTTAGGCAATGACAGAAAACAGCACA
F. novicida Fx1         TGAGAGAGACAGATCCAACGTTTCAGTTGCTGAATTATTTGTTAGTTATCGAGCGTAAGACTTGGTTACTACATCCAGA-----------------------------------------------------------------------
F. novicida 3523        ACTGGCGGCGAGGGTGA--GTTTCAGTAGCTAGATTATTTGATATACTGCTGTTAGAGTGCTTGAAACTTCAAGCCTGTTGTTGAAACGG--GTTTCAGTAGCTAGATTATTTGATATACTGCTGTTAGATAATATGCTTAACAATATGT
F. holarctica LVS       ------------------------------------------------------------------------------------------------------------------------------------------------------
F. tularensis SCHU S4   ------------------------------------------------------------------------------------------------------------------------------------------------------

F. novicida U112        GATAATAAAAAACTAGTTTCAGTTGCTGAATTATTTGGTAAACTACTGTTAGCTATAGGGTTACCTATCTTTTGAGTGTTGGCAAATAAGTTTCAGTTGCTGAATTATTTGGTAAACTACTGTTAGGTAATGGGCAGGTTTTGTATGGTC
F. novicida Fx1         ------------------------------------------------------------------------------------------------------------------------------------------------------
F. novicida 3523        TTTCTAGCGATATT-GTTTCAGTAGCTAGATTATTTGATATACTGCTGTTAGTTAAGCCAGATAATGAATATCAAGAGTATGTCGA---GTTTCAGTAGCTAGATTATTTGATATACTGCTGTTAGCAATTATGGAACGATTATGGGGAT
F. holarctica LVS       ------------------------------------------------------------------------------------------------------------------------------------------------------
F. tularensis SCHU S4   ------------------------------------------------------------------------------------------------------------------------------------------------------

F. novicida U112        ATATAGGAGT--GTTTCAGTTGCTGAATTATTTGGTAAACTACTGTTAGACTATTTAAGATAACTCGACCAATATTTGACAAA--GTTTCAGTTGCTGAATTATTTGGTAAACTACTGTTAGATCTGGTGAAGATGTAATCGAAATTGAT
F. novicida Fx1         ------------------------------------------------------------------------------------------------------------------------------------------------------
F. novicida 3523        TGGTTAGATGAAGTTTCAGTAGCTAGATTATTTGATATACTGCTGTTAGCTATATTGGGTTATAAAGTTCAACTTCGAGGCGTTAGTTTCAGTAGCTAGATTATTTGATATACTGCTGTTAGTGCTATGAGCATTGCAGGCAAAACAAAT
F. holarctica LVS       ------------------------------------------------------------------------------------------------------------------------------------------------------
F. tularensis SCHU S4   ------------------------------------------------------------------------------------------------------------------------------------------------------

F. novicida U112        CCGCTTAAGTTTCAGTTGCTGAATTATTTGGTAAACTACTGTTAGCAGTCTTCGCAATGACACAACCACAAGCCGCATAGTTTCAGTTGCTGAATTATTTGGTAAACTACTGTTAGAAGAGATATCTTAACTTTTGGTTTTTGGTTTTAC
F. novicida Fx1         ------------------------------------------------------------------------------------------------------------------------------------------------------
F. novicida 3523        AAAAAAAAGTTTCAGTAGCTAGATTATTTGATATACTGCTGTTAGCTGCTGCCGATACGGGTACAATTCAGGTTATATGGTTTCAGTAGCTAGATTATTTGATATACTGCTGTTAGTGCCATACGATACAGCATATATTTCTTTTTTAAT
F. holarctica LVS       ------------------------------------------------------------------------------------------------------------------------------------------------------
F. tularensis SCHU S4   ------------------------------------------------------------------------------------------------------------------------------------------------------

F. novicida U112        CGTTTCAGTTGCTGAATTATTTGGTAAACTACTGTTAGGCGAGTGATTAGCAGTTGATATATACCAATCTTG------------------------------------------------------------------------------
F. novicida Fx1         ------------------------------------------------------------------------------------------------------------------------------------------------------
F. novicida 3523        -GTTTCAGTAGCTAGATTATTTGATATACTGCTGTTAGACCAAAGCAACATTGCAATTGCAGATAGTGCTGGGTTTCAGTAGCTAGATTATTTGATATACTGCTGTTAGTTGTAGGCTGTAAAGGCTCAAAGGCTGTTATTACGTTTCAG
F. holarctica LVS       ------------------------------------------------------------------------------------------------------------------------------------------------------
F. tularensis SCHU S4   ------------------------------------------------------------------------------------------------------------------------------------------------------

F. novicida U112        -----------------------------------------------------------------------------------------------------------------------------------------GTTTCAGTTGCTG
F. novicida Fx1         -----------------------------------------------------------------------------------------------------------------------------------------GTTTCAGTTGCTG
F. novicida 3523        TAGCTAGATTATTTGATATACTGCTGTTAGCCTCAACTTTAGCCTTAGGCTCAGTCTTTTTGGCGTTTCAGTAGCTAGATTATTTGATATACTGCTGTTAGTGTTACTGGTTTTGCTGAAGGTGGTTATACTGGCAAGTTTCAGTAGCTA
F. holarctica LVS       -----------------------------------------------------------------------------------------------------------------------------------------GTTTCAGTTGCTG
F. tularensis SCHU S4   -----------------------------------------------------------------------------------------------------------------------------------------GTTTCAGTTGCTG
```



```
F. novicida U112       AATTATTTGGTAAACTACTGTTAGAGCCAAAACAAACCTAGTGTTTGCAAGGCTTTGAGAGTTTTTTAAAATGGTATTTTTTCGCAAATACCCTAATTGTAAAAAAACGTAGACATTTTGCGAAAAAATATCTAAAATTATAAATGTACC
F. novicida Fx1        AATTATTTGGTAAACTACTGTTAGAGCCAAAACAAACCTAGTGTTTGCAAGGCTTTGAGAGTTTTTTAAAATGGTATTTTTTCGCAAATACCCTAATTGTAAAAAAACGTAGACATTTTGCGAAAAAATATCTAAAATTATTAATGTACC
F. novicida 3523       GATTATTTGATATACTGCTGTTAGAGCCAAGACAAACCTAGAGTTTGCAAGGCTTTGAGAGAATTATTTGCTGGTGTTTTTTCGCAAACATCCTACTTGTAAAAAAATGTAGACATTTTGCGAAAAAATATCTACAATTAAAAGCGTACC
F. holarctica LVS      AATTATTTGGTAAACTACTGTTAGAGCCAAAACAAACCTAGTGTTTGCAAGGCTTTGAGAGTTTTTTAAAATGGTATTTTTTCGCAAATACCCTAATTGTAAAAAAACGTAGACATTTTGCGAAAAAATATCTAAAATTATAAATGTACC
F. tularensis SCHU S4  AATTATTTGGTAAACTACTGTTAGAGCCAAAACAAACCTAGTGTTTGCAAGGCTTTGAGAGTTTTTTAAAATGGTATTTTTTCGCAAATACCCTAATTGTAAAAAAACGTAGACATTTTGCGAAAAAATATCTAAAATTATAAATGTACC

F. novicida U112       AAATAATTAATGCTCTGTAATCATTTAAAAGTATTTTGAACGGACCTCTGTTTGACACGTCTGAATAACTAAAAAGCAAAAATTTGCCACCTAAGTGGCTTTTTTTGTTTGGCAATTAATATATCTTGTTCGCTTCTTAGCTCGAT-TAC
F. novicida Fx1        AAATAATTAATGCTCTGTAATCATTTAAAAGTATTTTGAACGGACCTCTGTTTGACACGTCTGAATAACTAAAAAGCAAAAATTTGCCACCTAAGTGGCTTTTTTTGTTTGGAAATTAATATAACTTGTTTGCTTCTTAGCTCGAT-TAC
F. novicida 3523       AAATAATCAATGCTCTGTAATCATTTAAAAGTATTTTGACCGGACCTCTGTTTGACACGTCTGAATAACT-AAAAGTAAAAATTTGCCACTCAAGTGGCTTTTTTTG-CTAGTGGATAATTTAT-TTGAGCTTTGTATAGCTAACCATAG
F. holarctica LVS      AAATAATTAATGCTCTGTAATCATTTAAAAGTATTTTGAACGGACCTCTGTTTGACACGTCTGAATAACTAAAAAACAAAAATTTGCCACCTAAGTGGCTTTTTTTGTTTGGCAATTAATATATCTTGTTCGCTTCTTAGCTCGAT-TAC
F. tularensis SCHU S4  AAATAATTAATGCTCTGTAATCATTTAAAGGTATTTTGAACGGACCTCTGTTTGACACGTCTGAATAACTAAAAAACAAAAATTTGCCACCTAAGTGGCTTTTTTTGTTTGGCAATTAATATATCTTGTTCGCTTCTTAGCTCGAT-TAC

F. novicida U112       ATCAAGATGGC-TGAGAAATTTA-----------------ATATGGAGATTT----------------CTAAATATTTG----CGAGATTTCCACCTTCAGTGGCAATTAGA--------------------------------------
F. novicida Fx1        ATCAAGATGGC-TGAGAAATTTA-----------------ATATGGAGATTT----------------CTAAATATTTG----CGAGATTTCCGCCTTCAGTGACAATTAGA--------------------------------------
F. novicida 3523       ATAAAAATGTCAAGCTAAATTTATTTCAGCATCTCATTATAAATATAGATTTTATAAAAAAAACATAATTACATAATTGATGACAAAATTTACTACTTTTATGCCTAGCAGACTATATCTTTTACTTTATGATATCTTTTTAGTTTAAAA
F. holarctica LVS      ATCAAGATGGC-TGAGAAATTTA-----------------ATATGGAGATTT----------------CTAAGTATTTG----CGAGATTTCCGCCTTCAGTGGCAATTAGA--------------------------------------
F. tularensis SCHU S4  ATCAAGAAGGC-TGAGAAATTTA-----------------ATATGGAGATTT----------------CTAAGTATTTG----CGAGATTTCCGCCTTCAGTGGCAATTAGA--------------------------------------


F. novicida U112       --------------------GGTTTTACTATCAACGATCATTACAAAAGTTGTAATATTCACATTGTATACATTGATGATCCGATGCAGATGA
F. novicida Fx1        --------------------GATTTTACTATCAACGATCATTACAAAAGTTGTAATATTCACATTGTATACATTGATGATCCGATGCAGATGA
F. novicida 3523       TTATTCTGACAAGCTCGAAGGAATTTGATTTTAGCGATCATTACAGAAATTTAAATACTCACACTGTATACATTGATGATCCGATGCAGATGA
F. holarctica LVS      --------------------GATTTTACTATCAACGATCATTACAAAAGTTGTAATATTCACATTGT------------TCCGATGCAGATGA
F. tularensis SCHU S4  --------------------GATTTTACTATCAACGATCATTACAAAAGTTGTAATATTCACATTGT------------TCTGATGCAGATGA
```

Supplementary Figure S12. Alignment of *Francisella* sp. CRISPR loci. See Figure 6. *F. novicida* DNA sequence from the 3'-end portion of the radical SAM protein-encoding gene located upstream of the repeat-spacer array to the 5'-end portion of the *cas4* gene was aligned with the *F. holarctica* and *F. tularensis* region between *cas4* and the gene encoding the transposase inserted within the CRISPR locus using MUSCLE with manual corrections. Blue highlighted, 3'-end of the radical SAM protein-encoding gene; yellow highlighted, scaRNA; red letters, degenerated repeats; grey highlighted, repeats; red highlighted, tracrRNA; green highlighted, 5'-end of the *cas4* gene. See as well Supplementary Figure S9 for the degenerated repeat alignments.