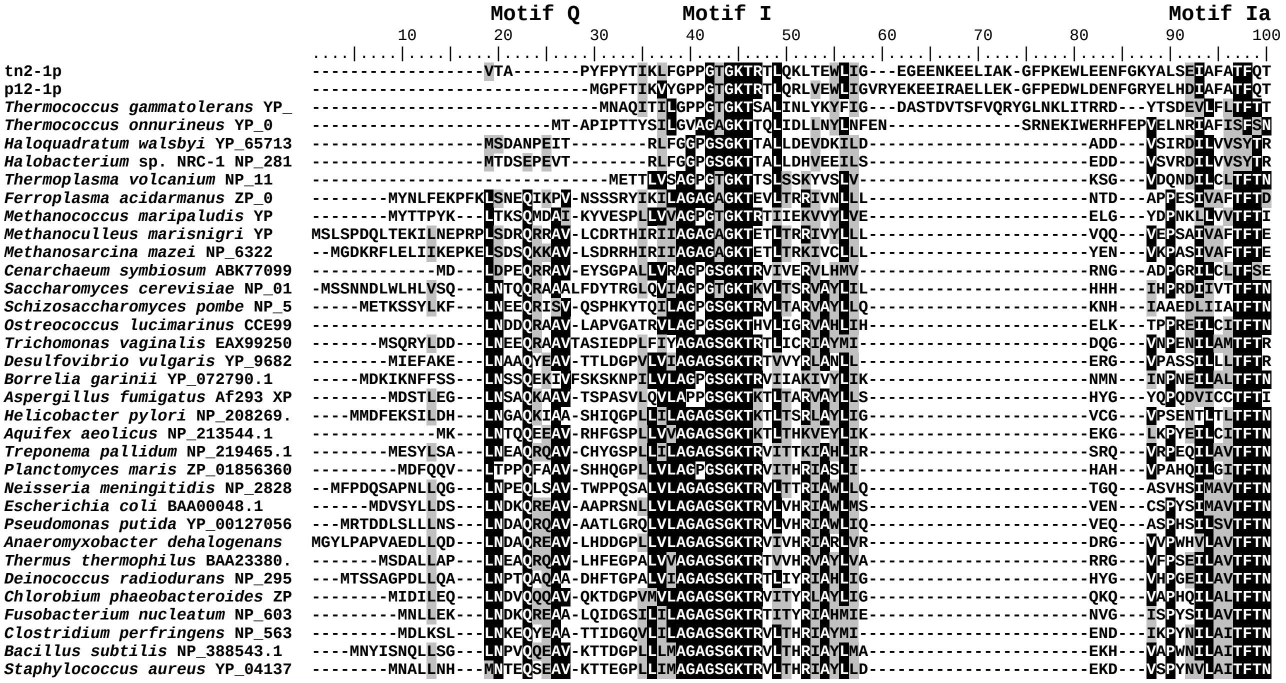
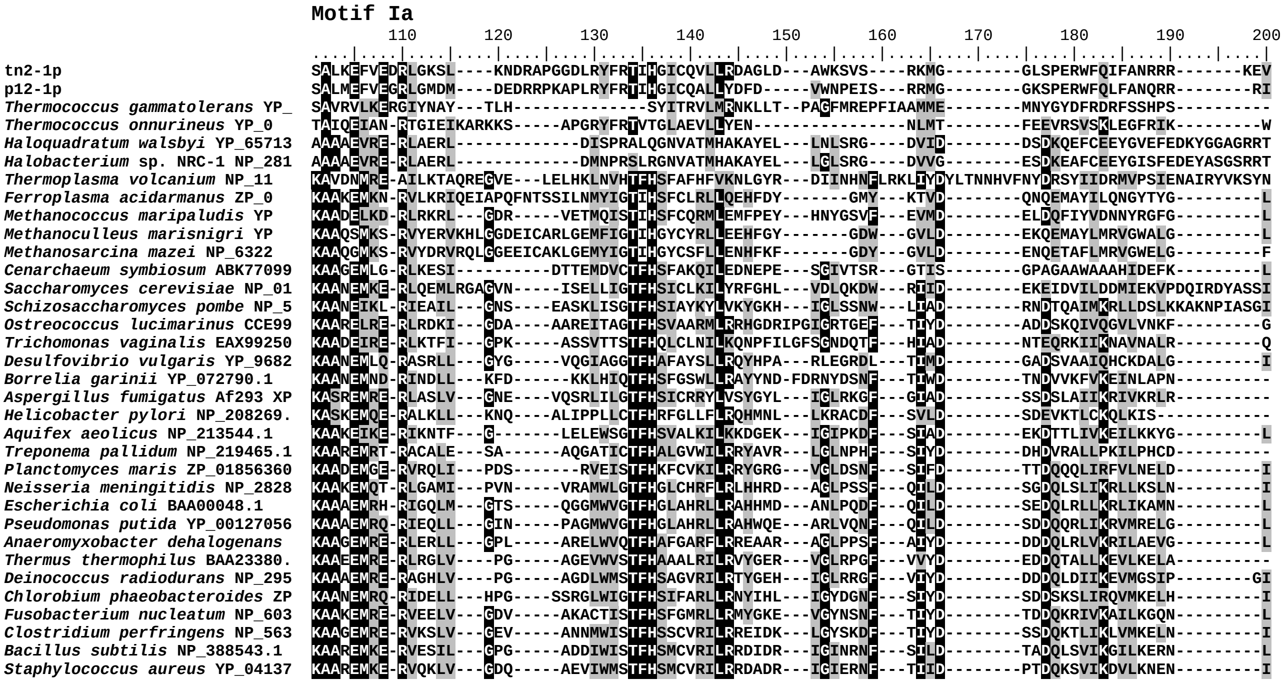
Figure S3



```
                                                  Motif Q             Motif I                                            Motif Ia
                                        10        20        30        40        50        60        70        80        90       100
                                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
tn2-1p                          ------------------VTA-------PYFPYTIKLFGPPGTGKTRTLQKLTEWLIG---EGEENKEELIAK-GFPKEWLEENFGKYALSEIAFATFQT
p12-1p                          -----------------------------MGPFTIKVYGPPGTGKTRTLQRLVEWLIGVRYEKEEIRAELLEK-GFPEDWLDENFGRYELHDIAFATFQT
Thermococcus gammatolerans YP_  ------------------------------MNAQITILGPPGTGKTSALINLYKYFIG---DASTDVTSFVQRYGLNKLITRRD---YTSDEVLFLTFTT
Thermococcus onnurineus YP_0    -------------------------MT-APIPTTYSILGVAGAGKTTQLIDLLNYLNFEN--------------SRNEKIWERHFEPVELNRIAFISFSN
Haloquadratum walsbyi YP_65713  ------------------MSDANPEIT--------RLFGGPGSGKTALLDEVDKILD------------------------ADD---VSIRDILVVSYTR
Halobacterium sp. NRC-1 NP_281  ------------------MTDSEPEVT--------RLFGGPGSGKTALLDHVEEILS------------------------EDD---VSVRDILVVSYTR
Thermoplasma volcanium NP_11    -------------------------------METTLVSAGPGTGKTTSLSSKYVSLV------------------------KSG---VDQNDILCLTFTN
Ferroplasma acidarmanus ZP_0    --------MYNLFEKPFKLSNEQIKPV-NSSSRYIKILAGAGAGKTEVLTRRIVNLLL-----------------------NTD---APPESIVAFTFTD
Methanococcus maripaludis YP    --------MYTTPYK---LTKSQMDAI-KYVESPLLVVAGPGTGKTRTIIEKVVYLVE-----------------------ELG---YDPNKLLVVTFTI
Methanoculleus marisnigri YP    MSLSPDQLTEKILNEPRPLSDRQRRAV-LCDRTHIRIIAGAGAGKTETLTRRIVYLLL-----------------------VQQ---VEPSAIVAFTFTE
Methanosarcina mazei NP_6322    --MGDKRFLELIIKEPKELSDSQKKAV-LSDRRHIRIIAGAGAGKTETLTRKIVCLLL-----------------------YEN---VKPASIVAFTFTE
Cenarchaeum symbiosum ABK77099  -------------MD---LDPEQRRAV-EYSGPALLVRAGPGSGKTRVIVERVLHMV------------------------RNG---ADPGRILCLTFSE
Saccharomyces cerevisiae NP_01  -MSSNNDLWLHLVSQ---LNTQQRAAALFDYTRGLQVIAGPGTGKTKVLTSRVAYLIL-----------------------HHH---IHPRDIIVTTFTN
Schizosaccharomyces pombe NP_5  -----METKSSYLKF---LNEEQRISV-QSPHKYTQILAGPGSGKTRVLTARVAYLLQ-----------------------KNH---IAAEDLIIATFTN
Ostreococcus lucimarinus CCE99  ------------------LNDDQRAAV-LAPVGATRVLAGPGSGKTHVLIGRVAHLIH-----------------------ELK---TPPREILCITFTN
Trichomonas vaginalis EAX99250  -------MSQRYLDD---LNEEQRAAVTASIEDPLFIYAGAGSGKTRTLICRIAYMI------------------------DQG---VNPENILAMTFTR
Desulfovibrio vulgaris YP_9682  --------MIEFAKE---LNAAQYEAV-TTLDGPVLVIAGAGSGKTRTVVYRLANLL------------------------ERG---VPASSILLLTFTR
Borrelia garinii YP_072790.1    -----MDKIKNFFSS---LNSSQEKIVFSKSKNPILVLAGPGSGKTRVIIAKIVYLIK-----------------------NMN---INPNEILALTFTN
Aspergillus fumigatus Af293 XP  --------MDSTLEG---LNSAQKAAV-TSPASVLQVLAPPGSGKTKTLTARVAYLLS-----------------------HYG---YQPQDVICCTFTI
Helicobacter pylori NP_208269.  ----MMDFEKSILDH---LNGAQKIAA-SHIQGPLLILAGAGSGKTKTLTSRLAYLIG-----------------------VCG---VPSENTLTLTFTN
Aquifex aeolicus NP_213544.1    -------------MK---LNTQQEEAV-RHFGSPLLVVAGAGSGKTKTLTHKVEYLIK-----------------------EKG---LKPYEILCITFTN
Treponema pallidum NP_219465.1  --------MESYLSA---LNEAQRQAV-CHYGSPLLILAGAGSGKTRVITTKIAHLIR-----------------------SRQ---VRPEQILAVTFTN
Planctomyces maris ZP_01856360  ---------MDFQQV---LTPPQFAAV-SHHQGPLLVLAGPGSGKTRVITHRIASLI------------------------HAH---VPAHQILGITFTN
Neisseria meningitidis NP_2828  --MFPDQSAPNLLQG---LNPEQLSAV-TWPPQSALVLAGAGSGKTRVLTTRIAWLLQ-----------------------TGQ---ASVHSIMAVTFTN
Escherichia coli BAA00048.1     ------MDVSYLLDS---LNDKQREAV-AAPRSNLLVLAGAGSGKTRVLVHRIAWLMS-----------------------VEN---CSPYSIMAVTFTN
Pseudomonas putida YP_00127056  ---MRTDDLSLLLNS---LNDAQRQAV-AATLGRQLVLAGAGSGKTRVLVHRIAWLIQ-----------------------VEQ---ASPHSILSVTFTN
Anaeromyxobacter dehalogenans   MGYLPAPVAEDLLQD---LNDAQREAV-LHDDGPLLVLAGAGSGKTRVIVHRIARLVR-----------------------DRG---VVPWHVLAVTFTN
Thermus thermophilus BAA23380.  -------MSDALLAP---LNEAQRQAV-LHFEGPALVVAGAGSGKTRTVVHRVAYLVA-----------------------RRG---VFPSEILAVTFTN
Deinococcus radiodurans NP_295  ---MTSSAGPDLLQA---LNPTQAQAA-DHFTGPALVIAGAGSGKTRTLIYRIAHLIG-----------------------HYG---VHPGEILAVTFTN
Chlorobium phaeobacteroides ZP  --------MIDILEQ---LNDVQQQAV-QKTDGPVMVLAGAGSGKTRVITYRLAYLIG-----------------------QKQ---VAPHQILALTFTN
Fusobacterium nucleatum NP_603  ---------MNLLEK---LNDKQREAA-LQIDGSILILAGAGSGKTRTITYRIAHMIE-----------------------NVG---ISPYSILAVTFTN
Clostridium perfringens NP_563  ---------MDLKSL---LNKEQYEAA-TTIDGQVLILAGAGSGKTRVLTHRIAYMI------------------------END---IKPYNILAITFTN
Bacillus subtilis NP_388543.1   ----MNYISNQLLSG---LNPVQQEAV-KTTDGPLLLMAGAGSGKTRVLTHRIAYLMA-----------------------EKH---VAPWNILAITFTN
Staphylococcus aureus YP_04137  --------MNALLNH---MNTEQSEAV-KTTEGPLLIMAGAGSGKTRVLTHRIAYLLD-----------------------EKD---VSPYNVLAITFTN
```



```
                                Motif Ia
                                       110       120       130       140       150       160       170       180       190       200
                                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
tn2-1p                          SALKEFVEDRLGKSL----KNDRAPGGDLRYFRTIHGICQVLLRDAGLD---AWKSVS----RKMG--------GLSPERWFQIFANRRR-------KEV
p12-1p                          SALMEFVEGRLGMDM----DEDRRPKAPLRYFRTIHGICQALLYDFD-----VWNPEIS---RRMG--------GKSPERWFQLFANQRR--------RI
Thermococcus gammatolerans YP_  SAVRVLKERGIYNAY---TLH--------------SYITRVLMRNKLLT--PAGFMREPFIAAMME--------MNYGYDFRDRFSSHPS----------
Thermococcus onnurineus YP_0    TAIQEIAN-RTGIEIKARKKS-----APGRYFRTVTGLAEVLLYEN----------------NLMT--------FEEVRSVSKLEGFRIK---------W
Haloquadratum walsbyi YP_65713  AAAAEVRE-RLAERL-------------DISPRALQGNVATMHAKAYEL---LNLSRG----DVID--------DSDKQEFCEEYGVEFEDKYGGAGRRT
Halobacterium sp. NRC-1 NP_281  AAAAEVRE-RLAERL-------------DMNPRSLRGNVATMHAKAYEL---LGLSRG----DVVG--------ESDKEAFCEEYGISFEDEYASGSRRT
Thermoplasma volcanium NP_11    KAVDNMRE-AILKTAQREGVE---LELHKLNVHTFHSFAFHFVKNLGYR---DIINHNFLRKLIYDYLTNNHVFNYDRSYIIDRMVPSIENAIRYVKSYN
Ferroplasma acidarmanus ZP_0    KAAKEMKN-RVLKRIQEIAPQFNTSSILNMYIGTIHSFCLRLLQEHFDY-------GMY---KTVD--------QNQEMAYILQNGYTYG---------L
Methanococcus maripaludis YP    KAADELKD-RLRKRL---GDR-----VETMQISTIHSFCQRMLEMFPEY---HNYGSVF---EVMD--------ELDQFIYVDNNYRGFG---------L
Methanoculleus marisnigri YP    KAAQSMKS-RVYERVKHLGGDEICARLGEMFIGTIHGYCYRLLEEHFGY-------GDW---GVLD--------EKQEMAYLMRVGWALG---------L
Methanosarcina mazei NP_6322    KAAQGMKS-RVYDRVRQLGGEEICAKLGEMYIGTIHGYCSFLLENHFKF-------GDY---GVLD--------ENQETAFLMRVGWELG---------F
Cenarchaeum symbiosum ABK77099  KAAGEMLG-RLKESI----------DTTEMDVCTFHSFAKQILEDNEPE---SGIVTSR---GTIS--------GPAGAAWAAAHIDEFK---------L
Saccharomyces cerevisiae NP_01  KAANEMKE-RLQEMLRGAGVN-----ISELLIGTFHSICLKILYRFGHL---VDLQKDW---RIID--------EKEIDVILDDMIEKVPDQIRDYASSI
Schizosaccharomyces pombe NP_5  KAANEIKL-RIEAIL---GNS----EASKLISGTFHSIAYKYLVKYGKH---IGLSSNW---LIAD--------RNDTQAIMKRLLDSLKKAKNPIASGI
Ostreococcus lucimarinus CCE99  KAARELRE-RLRDKI---GDA----AAREITAGTFHSVAARMLRRHGDRIPGIGRTGEF---TIYD--------ADDSKQIVQGVLVNKF---------G
Trichomonas vaginalis EAX99250  KAADEIRE-RLKTFI---GPK-----ASSVTTSTFHQLCLNILKQNPFILGFSGNDQTF---HIAD--------NTEQRKIIKNAVNALR---------Q
Desulfovibrio vulgaris YP_9682  KAANEMLQ-RASRLL---GYG-----VQGIAGGTFHAFAYSLLRQYHPA---RLEGRDL---TIMD--------GADSVAAIQHCKDALG---------I
Borrelia garinii YP_072790.1    KAANEMND-RINDLL---KFD------KKLHIQTFHSFGSWLLRAYYND-FDRNYDSNF---TIWD--------TNDVVKFVKEINLAPN----------
Aspergillus fumigatus Af293 XP  KASREMRE-RLASLV---GNE----VQSRLILGTFHSICRRYLVSYGYL---IGLRKGF---GIAD--------SSDSLAIIKRIVKRLR----------
Helicobacter pylori NP_208269.  KASKEMQE-RALKLL---KNQ----ALIPPLLCTFHRFGLLFLRQHMNL---LKRACDF---SVLD--------SDEVKTLCKQLKIS------------
Aquifex aeolicus NP_213544.1    KAAKEIKE-RIKNTF---G-------LELEWSGTFHSVALKILKKDGEK---IGIPKDF---SIAD--------EKDTTLIVKEILKKYG---------L
Treponema pallidum NP_219465.1  KAAREMRT-RACALE---SA------AQGATICTFHALGVWILRRYAVR---LGLNPHF---SIYD--------DHDVRALLPKILPHCD----------
Planctomyces maris ZP_01856360  KAADEMGE-RVRQLI---PDS-------RVEISTFHKFCVKILRRYGRG---VGLDSNF---SIFD--------TTDQQQLIRFVLNELD---------I
Neisseria meningitidis NP_2828  KAAKEMQT-RLGAMI---PVN-----VRAMWLGTFHGLCHRFLRLHHRD---AGLPSSF---QILD--------SGDQLSLIKRLLKSLN---------I
Escherichia coli BAA00048.1     KAAAEMRH-RIGQLM---GTS-----QGGMWVGTFHGLAHRLLRAHHMD---ANLPQDF---QILD--------SEDQLRLLKRLIKAMN---------L
Pseudomonas putida YP_00127056  KAAAEMRQ-RIEQLL---GIN-----PAGMWVGTFHGLAHRLLRAHWQE---ARLVQNF---QILD--------SDDQQRLIKRVMRELG---------L
Anaeromyxobacter dehalogenans   KAAGEMRE-RLERLL---GPL-----ARELWVQTFHAFGARFLRREAAR---AGLPPSF---AIYD--------DDDQLRLVKRILAEVG---------L
Thermus thermophilus BAA23380.  KAAEEMRE-RLRGLV----PG-----AGEVWVSTFHAAALRILRVYGER---VGLRPGF---VVYD--------EDDQTALLKEVLKELA----------
Deinococcus radiodurans NP_295  KAAAEMRE-RAGHLV----PG-----AGDLWMSTFHSAGVRILRTYGEH---IGLRRGF---VIYD--------DDDQLDIIKEVMGSIP--------GI
Chlorobium phaeobacteroides ZP  KAANEMRQ-RIDELL---HPG----SSRGLWIGTFHSIFARLLRNYIHL---IGYDGNF---SIYD--------SDDSKSLIRQVMKELH---------I
Fusobacterium nucleatum NP_603  KAAKEMRE-RVEELV---GDV-----AKACTISTFHSFGMRLLRMYGKE---VGYNSNF---TIYD--------TDDQKRIVKAILKGQN---------L
Clostridium perfringens NP_563  KAAGEMRE-RVKSLV---GEV-----ANNMWISTFHSSCVRILRREIDK---LGYSKDF---TIYD--------SSDQKTLIKLVMKELN---------I
Bacillus subtilis NP_388543.1   KAAREMKE-RVESIL---GPG-----ADDIWISTFHSMCVRILRRDIDR---IGINRNF---SILD--------TADQLSVIKGILKERN---------L
Staphylococcus aureus YP_04137  KAAREMKE-RVQKLV---GDQ-----AEVIWMSTFHSMCVRILRRDADR---IGIERNF---TIID--------PTDQKSVIKDVLKNEN---------I
```

....

```
                                                              Motif II                     Motif III
                                        310       320       330       340       350       360       370       380       390       400
                                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
tn2-1p                          IKDYDDMLIDAYDALKAGVISLPT-----RVLIVDEFQDLSPLQFEIFKMLAR-------DKELVVIAGDDWQTIFTWMGAD-----PRFLVNYPA----
p12-1p                          IVDYNDMLMDGYDYLKSGVIHLPT-----KVLIVDEFQDLSPLQFEIFKLLAK-------DKELVVIAGDDWQTIFTWMGAD-----PRFLIKYPA----
Thermococcus gammatolerans YP_  IFDYVSVLRYFLTT-TPATLAPEKEGSEPRVMIIDEAQDFSPLQWAVVKRLRS-----IGLLDYVVAAGDPNQSIYSFQGADPSEF-ENFMAS-------
Thermococcus onnurineus YP_0    IVDYEDILMRAYDY---RNDIVVDL----EYMIIDEAQDNSLLDYATLLPIAK------NNATELVLAGDDAQLIYDFRGANYKLF-HKLIERS------
Haloquadratum walsbyi YP_65713  LVGFADMLERV-----KQRSLVPQV----DYLVIDEFQDITTLQYDVYEKWRS-------SVDRVLIAGDDDQVVYAWQGADPNLL-LDANR--------
Halobacterium sp. NRC-1 NP_281  LVGFADMLERV-----EQRSLLPNV----EYLVIDEFQDITSLQYDVYEEWRP-------HMDRVLIAGDDDQVVYAWQGADPRLL-LDTTV--------
Thermoplasma volcanium NP_11    VIDYNDMLIKFINLSEK---------PHYKYVLIDELQDVSDLEAKIAGSVAE----------NIYAVGDKKQSIFGFQGGGF----SNFISNFPN----
Ferroplasma acidarmanus ZP_0    IISFSQLIYFAVNKIKDNKEKISNI----KYLLVDEYQDINKVQYDLIKQIGE--------NASIFAVGDPRQSIYEWRGSNYKYF-EEFSNDFSN----
Methanococcus maripaludis YP    KLDFAMLQKQFYHMLLNNEDALNSVREMFDYIFVDEYQDTNPIQDAIIRLIAE-------PKFNITVVGDEDQSIYGFRGASVENF-RQFLERYSGPEKI
Methanoculleus marisnigri YP    RLTFNRMVQLAVENLQHRPEVAEHV----KYLIVDEYQDINHAQDELIRLIGL--------GGSIFVVGDPRQTIYQWRGSDEQYF-EEFTTRYPE----
Methanosarcina mazei NP_6322    RLTFNRMINLALTNLQEKQNVLSDV----KYLIVDEYQDINRAQEKLIQLIGL--------NSEIFIVGDPRQTIYQWRGSDEKCF-EDFAVKHID----
Cenarchaeum symbiosum ABK77099  IIDFDDMVSEAVRLFRARDDVRSEYGGRYSHILVDELQDANYAQLELVRLIAG----------NITAVGDENQSIYGFQGAYTSGF-ENFDDITGG----
Saccharomyces cerevisiae NP_01  TLDFDDLLMYTFRLLTRVRVLSNI-----KHVLVDEFQDTNGIQLDLMFLFAK---GNHHLSRGMTIVGDPDQSIYAFRNALAHNF-LEMGRKCPI----
Schizosaccharomyces pombe NP_5  LADFDDLLLNFILLLQKQPDCVRNI----KHILIDEFQDTSKIQYFLVKLLAL-------QNSDITIVGDPDQSIYGFRSAEIRNL-NQMSEDFEG----
Ostreococcus lucimarinus CCE99  AFDFDDLLSATVAMLEMCPDVRSYYEHRWSQVLVDEFQDTSTTQYELIRKLSQ-------PQGSVFVVGDADQAIYGWRGAEVANIRTQFDEDFKD----
Trichomonas vaginalis EAX99250  LIDFTDFMTYTQELLLKHQRVAYKYRKLYQYILIDEFQDTSDINFAILKLILQQSMTRQPPNRRVTIVGDPHQSIYSFRGANPDNI-KQFLEMFPD----
Desulfovibrio vulgaris YP_9682  LLDYDDLLFEVERLLRERPDLLAHLRDRYRYIMVDEYQDTNLVQARLVQLLAG-------EGGNVMAVGDDAQSIYAFRGANVQNI-LSFPQFFPG----
Borrelia garinii YP_072790.1    AFDFADLIIKPILMLRHSKSLKESVQSKFKVIFVDEYQDTNYSQFLFLKELYS-------DGMYFMVVGDEDQSIYSFRGARIENI-LEFEKTFDN----
Aspergillus fumigatus Af293 XP  LLDYDDLLLRCADLLRRHPQCVSNV----QAVLVDEFQDTNHIQYELMNLFAV-------RNRRITVVGDPDQSIYGFRSAEIKNL-GRMQQHYKD----
Helicobacter pylori NP_208269.  LVDFDDLLFLSLKILQDNETIAKETSERYHYIMVDEYQDTNALQLEFLKKLSF-------THHNLCVVGDDDQSIYGFRGADISNI-LNFSKHFKG----
Aquifex aeolicus NP_213544.1    LLDFSDLMRELYNLLL-VDEVREKYRNTFKYIMVDEYQDTNNIQYEILKLLA---------NKNICAIGDPNQCIYEWRDARPDNI-LRFIEDF-N----
Treponema pallidum NP_219465.1  TVDFGDLIMLPVQLLREHQDVAEQLHARWQVVMVDEYQDSNVAQFHFLQVLTG-------AHTYLCVVGDDDQSIYRFRGAEVKNI-LTFPEFFQN----
Planctomyces maris ZP_01856360  AVDFDDLLLHVASLLKGSPELRATLDERYQYILVDEYQDTNLAQYQIIMGLSL-------NTRNLCVTGDPDQSIYGWRGAKIENI-LQFERDFPD----
Neisseria meningitidis NP_2828  VVDFAELMLRSYEMLQSNEILRRHYQNRFNHILVDEFQDTNKLQYAWLKLMAG-------GNTAVFAVGDDDQSIYRFRGANVGNM-TALMEEFHI----
Escherichia coli BAA00048.1     LVDFAELLLRAHELWLNKPHILQHYRERFTNILVDEFQDTNNIQYAWIRLLAG-------DTGKVMIVGDDDQSIYGWRGAQVENI-QRFLNDFPG----
Pseudomonas putida YP_00127056  VIDFSELLLRALDLWRDHPGLLEHYQRRFRHVLVDEFQDTNAVQYAWLRLLAR-------GGDSLMAVGDDDQSIYGWRGAKIENI-HQYTADFPD----
Anaeromyxobacter dehalogenans   AVDFGDLLLRPVRLLEEDAALRARWAGRFRYLLVDEFQDTNPVQYRLLQLLTG-------PRRNVCVVGDDDQAIYRWRGADVSNI-LGFDGDFPG----
Thermus thermophilus BAA23380.  ALDFGDILLYALEAFRGGRGGPQARAQRARFIHVDEYQDTSPVQYRFTRLLAG-------EEANLMAVGDPDQGIYSFRAADIKNI-LDFTRDYPE----
Deinococcus radiodurans NP_295  AIDFGDLITETVRLFKEVPGVLDKVQNKAKFIHVDEYQDTNRAQYELTRLLAS-------RDRNLLVVGDPDQSIYKFRGADIQNI-LDFQKDYPD----
Chlorobium phaeobacteroides ZP  ALDFDDLLIKPLELFQEQPAVLEQLQDYFRYIMIDEYQDTNRAQYLVAKSLAA-------VHRNIFVVGDDAQSIYSWRGADITNI-LNFQDDYPD----
Fusobacterium nucleatum NP_603  AMDFSDILLNTYKLLQK-PEILEKVQNKYKYIMIDEYQDTNNLQYKIIDLIAR-------KSSNLCVVGDENQSIYGFRGANILNI-LNFENNYNN----
Clostridium perfringens NP_563  ALDFDDLIVKTVELFRKHNDVLEFYQNKFKYIMIDEYQDTNKAQYEFAKLLAA-------KHKNICVVGDDDQCIYAWRGADIRNI-LDFEKDYPE----
Bacillus subtilis NP_388543.1   SLDFDDLIMTTIKLFDRVPEVLEFYQRKFQYIHVDEYQDTNRAQYMLVKQLAE-------RFQNLCVVGDSDQSIYRWRGADITNI-LSFEKDYPN----
Staphylococcus aureus YP_04137  ALDFDDLIMTTINLFERVPEVLEYYQNKFQYIHVDEYQDTNKAQYTLVKLLAS-------KFKNLCVVGDSDQSIYGWRGADIQNI-LSFEKDYPE----


                                  Motif IV
                                        410       420       430       440       450       460       470       480       490       500
                                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
tn2-1p                          -DLQIILRKTWRLPEKILREALTYAKANL---RSGVFKEMVSTGKRGAIGIVSNFSTKNDNDLDLLA-RWVLNELKR-GH-----SVFVLART-NRQALH
p12-1p                          -DLEIVLRKTYRLPEKILREALSYAKANL---RRSVFKDMVSTGKKGAFAHIRYFSTRREAHMDQLA-KWVLAELKK-GH-----SVMILTRT-NSQALH
Thermococcus gammatolerans YP_  -GTVVVLKRSYRLPKQVYNKAEMLTRWLR---THF---EYEPREELGKAITEYLKKSNLLKM-NRTTGEKVIQILESVMD--EDKTVFVLTRT-NKQARE
Thermococcus onnurineus YP_0    -EIILNLTETRRFGSEIANLATAIIDDMN---YIQ-KREVLSAATHSTKVAHIDLF-------QMMS---ILQNMAT-----TDLTVYILART-NAVLNY
Haloquadratum walsbyi YP_65713  -DEDVVLPNSYRLPSKVLSVVNQEIRHID---KRQ-EKDLHPRKQGGTV--------EGIES-PSML-DLARNVRHTVERSDGEETLMILFRA-RYQMFQ
Halobacterium sp. NRC-1 NP_281  -TDDQVLPNSYRLPSQVLEIVQQEISHID---TRQ-EKDLKPRKEGGSV--------EAIDS-PSML-DLVRNVRRT-VEATDDDTVMILFRA-RYQLFD
Thermoplasma volcanium NP_11    -HQLKYLTEDHRNGKAILKYAKEFFKKST-------KGGLDEDLEKLTSVRDHDGYVEVIEG-SY---SDILSVLDR----FPRGSVGIIART-NAQVIN
Ferroplasma acidarmanus ZP_0    -AEEFTLNENRRSTKSIVNISNTIVKQFP---DNPGYPDMVPVREKSGDVIATEYENASREA-EGIA-NKIKCEIKK-DK--KYSKFAVLLRSVNTSGDL
Methanococcus maripaludis YP    KAHMIKLEENFRSPEEIVDAFDKFMGPHR----TF-EKKIISKNGHYSDPVLIVSESPEEES-IKIA-EWIEKLITKHNV--DYGDIAVLFKSVRNHANE
Methanoculleus marisnigri YP    -AVTIHITENRRSTVAVIETANGFSDTFE---DQH-YDHMDPVRREQGGAYLAEMQSDITEV-DWIA-DQIQALVRE-EK-CSYSDIALLFRSVSTSAPP
Methanosarcina mazei NP_6322    -AETISINENRRSGRQIIEVSNQFSDYFE---DE--YDHLDPVRKEEGAVYIAKLSDNIHEA-KWIA-DQIEKHVDA-GK-CSYGDIGILLRSVNTSAPP
Cenarchaeum symbiosum ABK77099  -AETISLNNNYRCTKTIAGLAAQCIANTP---S---GSQAGAVHPEGSPVEIVECADEDSEA-ACIA-DRIVQIRSP-GGESAGYSIAVLTRK-IADGKK
Saccharomyces cerevisiae NP_01  EYSTIILVENYRSSQKILNTSEILITQQN-KGRQN-RAPLRAQFDLDFPPVYMNFPAYFLEA-PSLV-RELLYLKAL-PNLFTFNDFAILVRQ-RRQIKR
Schizosaccharomyces pombe NP_5  -TQVLHLERNYRSAKPILELALSIISQDK---SRP-KKGLKSNHISSLKPHYRLFETNNKES-YWIA-REIKRIVGSCPELIFYNDIAILVRS-SSLTRS
Ostreococcus lucimarinus CCE99  -VQTHMLTTNYRSTATIVEAAQAVIKESA---FPS-PLDLVANTPGGRDVAIVEANDDREEA-EFIA-LEARKSVKE-DPDLRYSEIAVLYRT-NSQARV
Trichomonas vaginalis EAX99250  -AGHVTLSKNYRSCESIVKASQSIIKFDG---N---ETELISMIPGGNPIKFISASDAYAEA-DSIC-TEIERMVYP-GSSIQYRDICIMFRM-KKVASE
Desulfovibrio vulgaris YP_9682  -TRIIKLEENYRSTQPVLDLTNAILAEAP---QAY-RKHLYTSREDGVLPQVVRPLSDLTQA-SIVV-SRVSELLRT----YPPHEIAVLFRA-GFHSYH
Borrelia garinii YP_072790.1    -VIKFYLVQNYRSNSNIVSIANGVISKNK---NRY-EKQITTQNGSGKRMKFLVFQSTSDEA-EYFS-----NLLVA-ND----IETAILYRF-NSQSFH
Aspergillus fumigatus Af293 XP  -TSVILLEDNYRSSGSILRSAQDVIEQDS---SRP-SRKLQPTHCPGTMPVLRKLPTAEAEA-QWIV-LEIKRSMAMTGKLLNYSDFAILLRS-AALSRH
Helicobacter pylori NP_208269.  -AKIVKLETNYRSSAEILACANSLISHNQ---HRH-IKTLQSFKGSHKSVVCKEYLTQKEES-LDVA-YQIKALLKK-GE--NLENIAILYRL-NGLSRS
Aquifex aeolicus NP_213544.1    -PKIIKLELNYRSREPILRVANAVLEASTLEWKDL-IPKLRGVRGEGQKPYVRRFQDEEEEA-LWIS-RKIKEL--A-GE-YELKDIAVLVRV-GYITDV
Treponema pallidum NP_219465.1  -TQIIRLEYNYRSTDAILRVADSVVKKNQ---DRL-GKALIAQRTGGTKPRLFLLNNQDEEA-ALCV-HLIQEARAR-GI--PYADWAILYRV-NAQSLS
Planctomyces maris ZP_01856360  -CKTIRLEQNFRSTKEILRVADDLISHNQ---RRK-VKSLFTENEEGSPVELLCFQDERHEA-DDIA-RRIRNAVDQ-GK-QEYTDYAIFYRV-NSLSRE
Neisseria meningitidis NP_2828  -DAPVKLEQNYRSVGNILASANAVIENND---ERL-GKNLRTDAEAGDKIRYYSAFTDLEEA-QFIV-DETKALERE-GW--DLDEIAVLYRS-NAQSRV
Escherichia coli BAA00048.1     -AETIRLEQNYRSTSNILSAANALIENNN---GRL-GKKLWTDGADGEPISLYCAFNELDEA-RFVV-NRIKTWQDN-GG--ALAECAILYRS-NAQSRV
Pseudomonas putida YP_00127056  -AEMIRLEQNYRSTGGILKAANALIANNS---GRL-GKELWTDMGEGEPLTLYAAYNEHDEA-RYVV-ETIESLVKQ-GN--ARNEIAILYRS-NAQSRV
Anaeromyxobacter dehalogenans   -AKVVKLEQNYRSTRTILDAAHAVISRAR---RRR-EKRLWTESGAGAPLALLVGEDEGAEA-ERIA-RAVAAERGR-GT--SGDEIAVLYRT-NAQSRP
Thermus thermophilus BAA23380.  -ARVYRLEENYRSTEAILRFANAVIVKNA---LRL-EKALRPVKRGGEPVRLYRAEDAREEA-RFVA-EEIARL----GP--PWDRYAVLYRT-NAQSRL
Deinococcus radiodurans NP_295  -AKVYMLEHNYRSSARVLEAANKLIENNT---ERL-DKTLKPVKEAGQPVTFHRATDHRAEG-DYVA-DWLTRLHGE-GR--AWSEMAILYRT-NAQSRV
Chlorobium phaeobacteroides ZP  -VSAFKLVENYRSTQTILQAANGIIKRNS---RQI-EKELVSHQSTGEPLTLLESFNEKREA-EKVG-ENIQSLKLVKGY--DYRDFAVFYRT-NAQSRV
Fusobacterium nucleatum NP_603  -AKIVKLEENYRSTTIILDAANELIKNNK---SSK-DKKLWTQNGKGDLIKVLACDNGRDEV-SRII-EFIRENHQN-GV--AYRDMTILYRT-NAQSRI
Clostridium perfringens NP_563  -AKVVKLEQNYRSKGNILEAANKVIKNNA---ERK-FKILRTEQVEGEKISIHRAYSDKMEA-AFVV-SEIEKLKNDNGY--KYKDFAVLYRT-NAQSRN
Bacillus subtilis NP_388543.1   -ASVILLEQNYRSTKRILRAANEVIKNNS---NRK-PKNLWTENDEGIKISYYRGDNEFGEG-QFVA-GKIHQLHST-GK-RKLSDIAILYRT-NAQSRV
Staphylococcus aureus YP_04137  -ANTIFLEQNYRSTKTILNAANEVIKNNS---ERK-PKGLWTANTNGEKIHYYEAMTERDEA-EFVI-REIMKHQRN-GK--KYQDMAILYRT-NAQSRV


.... .... ....
```

```
                                                                                                                        Motif V
                                        710       720       730       740       750       760       770       780       790       800
                                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
tn2-1p                          ARRFLEKILK---------------------------------------EGITPIAGIPE----------------------DHELYVDTIHSSKGLEAD
p12-1p                          ARRFIEKILR---------------------------------------EGLRPMDGIPE----------------------EHELYIDTIHASKGREAD
Thermococcus gammatolerans YP_  DVLSLARFIE-------NHYLDVLDPEEK--------------------RVLAGLRKKRN----------------------GRAIYVDTMHAAKGEEAD
Thermococcus onnurineus YP_0    DVLGI----------LLLAYEQTTAK-----------------------RILTTAKNTNF------------------------KIKLSTIHSAKGSEAD
Haloquadratum walsbyi YP_65713  -TAAAGDMLR-----KVTSFQRK--------------------------SVEAYFDG--------------DYGGMD-----PSRIRVGTIHSAKGREAD
Halobacterium sp. NRC-1 NP_281  -PSAAADMLR-----KVTRYQRR--------------------------SIDAYFEG--------------DYRGMD-----PDRVRVGTIHSAKGREAD
Thermoplasma volcanium NP_11    IYMNLQEYLY-----MYGINDRQ--------------------------DLLNYLEVCDQ----------ELTESSN-----TERISVLTVHKAKGLEFD
Ferroplasma acidarmanus ZP_0    ILANIGRFTK-----VLGDYEQSVRYGGGNKNIQKEMK-----------DLCWFINTYAN-------HSYQEAKIEDLSK--DNVVTVSTIHQSKGLEWQ
Methanococcus maripaludis YP    QLKNLAKFTT-----ILEKFKKNT-------------------------GSNSFKSFIYLLTKIPDNKLEDAASPEG-----QDVVKIMTIHQAKGLEFP
Methanoculleus marisnigri YP    IMANLGRFNT-----LLTDYETAIMLGGRRRNWETDLK-----------GLCWYLNTYAS-------SKYEEQTGDDIGG--VDAVQLVTIHQSKGLEWP
Methanosarcina mazei NP_6322    IMANLGRFSS-----LLTDFESANMLGGRQRNWERDLK-----------NLCWYMNSYAS-------GSYEEQPGDDIRG--VDAVQLMTVHQAKGLEWP
Cenarchaeum symbiosum ABK77099  GAEHLREACR-----LATEYDLLDPRGGLD-------------------GLVGYLWAANS-------DKNLRSDADA-----GGAVEVTTIHKSKGREFD
Saccharomyces cerevisiae NP_01  RHKNIELLKNYFLALLSKSESSDKEKNEAIKAATDEAEPIENKVITPKEYLRNFFNSLSLHSDAAEEEESESNKDAKIKREKNGFVTISTIHGAKGLEWP
Schizosaccharomyces pombe NP_5  KWENVMELVQ-----QSDNISCIFYELDYKISTIV--------------LLQNFLTQIAL--------VNEEQKEGE-----SQKVTISTLHAAKGLEWP
Ostreococcus lucimarinus CCE99  RWAFVQELIN-----LAKEPPQDIDETIQDRVGLE--------------ALGSFLEGISL------LTSAESKEEEG-----GDTVKLMTMHASKGLEFN
Trichomonas vaginalis EAX99250  FSEAVQDMIN-----DRNETFQLLLEESKRFHIRLLAANEGTPLSSQK-CLQKFIDSITL----------ETTSSNM-----KNAVTLSTIHQMKGLESP
Desulfovibrio vulgaris YP_9682  RQQGLEQLVQ-----IASAYR----------------------------DLDLFISDLSL--------EDPGTEEED-----RDSVVLSTIHSAKGLEWS
Borrelia garinii YP_072790.1    KLRNIDELIN-----SGIEYSGTFE------------------------GLAIFLENSSL---------SPLISGDF-----KSNIFLSSIHGVKGLEFD
Aspergillus fumigatus Af293 XP  RWANVEELLA-----QAGDVAAPAAEKGQDDSLPEIQGLAQQQAHQGEEALSQFLANVAL-----STEVLSEDGDQP-----QEKVTISTIHAAKGLEWP
Helicobacter pylori NP_208269.  REGFVKELLT-----LVKEYFKTNPTH----------------------SLLDFLNESVL----------DAHNTEN-----AQKVSCMSVHMSKGLEFK
Aquifex aeolicus NP_213544.1    RIENVKEFLS-----SLKDFYAKAY------------------------TLEDLLAEITL----------TSEEEEE-----ENAVKILTIHSAKGLEFP
Treponema pallidum NP_219465.1  CAVNVQELMN-----AASLYACSHE------------------------GLVSFLEHIQL-------DQNMADEGGA-----ADAVHLITIHNTKGLEFR
Planctomyces maris ZP_01856360  RIANVNELLS-----TAKKYDDIYGEET---------------------TLEGFLEVSTL-------ASATDQLDTD-----AGRVTLMTLHAAKGLEYP
Neisseria meningitidis NP_2828  RLDNLDELVN-----AAIEFKPEDSNFETLPDNISDDPAF---------PILAFLSNAAL-------ESGENQAGAG-----EKAVQLMTVHAAKGLEFN
Escherichia coli BAA00048.1     RIENLEELVT-----ATRQFSYNEEDEDLM-------------------PLQAFLSHAAL-------EAGEGQADTW-----QDAVQLMTLHSAKGLEFP
Pseudomonas putida YP_00127056  RVENLEELVS-----AARNFESTDEDADLS-------------------PLSAFLGHASL-------EAGDTQADEH-----EDSIQLMTLHSAKGLEFP
Anaeromyxobacter dehalogenans   RAENLAELVA-----AAREFDESIAGEPPPGDPDEVRPP----------PLARFLEQIAL--------LGEADAATP-----EGRVALMTLHAAKGLEFE
Thermus thermophilus BAA23380.  RLENVEELLR-----AAKEAE----------------------------DLQDFLDRVAL-------TAKAEEPAEA-----EGRVALMTLHNAKGLEFP
Deinococcus radiodurans NP_295  RLENLEELVS-----AAEEWSQDEANVGG--------------------SIADFLDDAALLSSVDDMRTKAENKGAP-----EDAVTLMTLHNAKGLEFP
Chlorobium phaeobacteroides ZP  RHENLQELLS-----MARDFSDHNPEAN---------------------SLDEFLNGISL-------ATDYNEVQES-----DNYVSLMTVHAAKGLEFP
Fusobacterium nucleatum NP_603  RIENIDEFKN-----SILELENVVGEL----------------------RLNEYLENVSL-------ISATDNLEEK-----SDYVKLMTIHNSKGLEFP
Clostridium perfringens NP_563  RIENLEELVS-----DAVEFEKSNEEDK---------------------SLSAYLEKVAL-------VQDMDDLEAE-----DNYVMLMTVHSSKGLEFP
Bacillus subtilis NP_388543.1   RLENIDEFLS-----VTKNFEQKSEDK----------------------TLVAFLTDLALIADIDQLDQKEEESGG------KDAITLMTLHAAKGLEFP
Staphylococcus aureus YP_04137  RLENIDEFMS-----VPKDYEENTPLEEQ--------------------SLINFLTDLSL--------VADIDEADT-----ENGVTLMTMHSAKGLEFP

                               Motif V                                                            Motif VI
                                        810       820       830       840       850       860       870       880       890       900
                                ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
tn2-1p                          TVFLINEMPRRNW-------------------RRYFKRPE--D---------------------LEAE-ARVWFVGMTRARKGLLIINT-----------
p12-1p                          -VVFLINEMPRRN-------------------WRRFF-RSEEE---------------------LEAE-ARVWFVGMTRARKALGIVWT-----------
Thermococcus gammatolerans YP_  -VVIVIDFVNRKI-------------------EKEIQ-RDENA---------------------LKEE-VRVLYVAMTRARERLYVLTS-----------
Thermococcus onnurineus YP_0    -VVFLINSV---------------PHK-----TKMKILEN------------------------YEGE-KRVLYVAVTRARKFLFIVDQ-----------
Haloquadratum walsbyi YP_65713  HVFVATDLTEKVV-----------EQM-----AASVDESAVPGVDEFTARTDPVPIL-------TDNE-RRVFYVGMSRARERLVLLEN-----------
Halobacterium sp. NRC-1 NP_281  HVFLATDLTEKVV-----------EQM-----AATADPETVPDDVEFTSKTSPVPIL-------TDNE-RRVFYVGMSRARERLVLLQD-----------
Thermoplasma volcanium NP_11    HVIYLPKTTRDNF-----------SVVDLCAYELIRE-KTKRDISTE-----------------LEDEDTRVDFVAITRPVHTLIIITD-----------
Ferroplasma acidarmanus ZP_0    -TVFLPSIINKRF-----------PSLTHFKSNLMID-QSLYEYSRYETS--------------LDSE-YRLFYVAVTRAKNYCLMSNF-----------
Methanococcus maripaludis YP    -VVVLGGVTERRYGARADDRLVNIPKQ-----FMMNKHDV--D---------------------FDEEVRRVFYVGMSRAQKLLTISAI-----------
Methanoculleus marisnigri YP    -LVFVPCMVRQRF-----------PSSMVGKGRTWLISPELFDVAKYEGD--------------LESE-RKLMYVALTRAKDVAVVSYF-----------
Methanosarcina mazei NP_6322    -LVFVPAMNARRF-----------PSSMAGREQTWMIPRNMFDAEKYEGD--------------IESE-KKLFYVALTRAKDVLVVSHF-----------
Cenarchaeum symbiosum ABK77099  -AVFIADAVEDRLPLKYKGRKYQVPTTLLHGRRPQGK-EKDRH---------------------MQEE-RRLLYVAMTRAKKNLCITYS-----------
Saccharomyces cerevisiae NP_01  -VVFIPGCEEGII-----------PCVFNDDKKDESEEDEEEDQENSKKDASPKKTRVLSVEDSIDEE-RRMFFVAQTRAKYLLYLSNTVTVEDVDRPRI
Schizosaccharomyces pombe NP_5  -VVFLPCLCENII-----------PHS-----RSDD----------------------------LDEE-RRLLYVGATRAQALLYLSSF-----------
Ostreococcus lucimarinus CCE99  -SVFISGVEEGLI-----------PFV-----RNGDS-ED--D---------------------QDEE-VRLFYVGITRAKRKLMLCHA-----------
Trichomonas vaginalis EAX99250  -VCFIMRFNQSIL-----------PMA-----DKDSDDSGLGDQSVS-----------------LNEE-RRILYVAMTRARQKLFLTCA-----------
Desulfovibrio vulgaris YP_9682  -AVIILDLVEDRF-----------PSR-----HAVTR-AD--D---------------------FEEE-RRLMYVACTRARERLMLFVP-----------
Borrelia garinii YP_072790.1    -RVVISGLEKGLL-----------PAE-----IEALS-ED--R---------------------LEEE-RRLFYVAITRAKSELIIALN-----------
Aspergillus fumigatus Af293 XP  -VVFVPAVYNGII-----------PHS-----RAEDS----------------------------DEE-RRLLYVAMTRAQALLYLSYPLRQSRNEEAAT
Helicobacter pylori NP_208269.  -HVFVIGLEEGFF-----------PH------RGFNQ-ES--D---------------------LEEE-RRLAYVAITRAKEELQLSYV-----------
Aquifex aeolicus NP_213544.1    -VVFLPRLEEGIL-----------PHH-----RSQES-ER--E---------------------LEEE-RRLFYVAITRAKDLLFMSYT-----------
Treponema pallidum NP_219465.1  -RVILTGLENGVF-----------PRD---------D-EA--D---------------------IQEE-RRLMYVACTRAMDSLYLTAC-----------
Planctomyces maris ZP_01856360  -VVFIVGVEQNLI-----------PHE-----RVLRE-EFRDG---------------------LEEE-RRLFFVGITRAEQNLYLTQT-----------
Neisseria meningitidis NP_2828  -AVFLTGMEEGRF-----------PSE-----MSLAE-RG--G---------------------LEEE-RRLMYVAITRARKRLYITMA-----------
Escherichia coli BAA00048.1     -QVFIVGMEEGMF-----------PSQ-----MSLDE-GG--R---------------------LEEE-RRLAYVGVTRAMQKLTLTYA-----------
Pseudomonas putida YP_00127056  -YVFLVGMEEGLF-----------PHK-----MSLEE-PG--R---------------------LEEE-RRLAYVGITRAMRQLVMTYA-----------
Anaeromyxobacter dehalogenans   -AVFLAGMEDGTL-----------PYE-----RPWADQTASERDES------------------LDEE-RRLCYVGMTRAKRRLTLSLA-----------
Thermus thermophilus BAA23380.  -VVFLVGVEEGLL-----------PHR-----NSVST-LE--G---------------------LEEE-RRLFYVGITRAQERLYLSHA-----------
Deinococcus radiodurans NP_295  -VVFIVGVEQGLL-----------PSK-----GAIAEGPS--G---------------------IEEE-RRLFYVGITRAMERLLMTAA-----------
Chlorobium phaeobacteroides ZP  -VVFITGLEERLF-----------P-------LNYYE-TD--E---------------------LEEE-RRLFYVAVTRAQQKVFLSWS-----------
Fusobacterium nucleatum NP_603  -IVFLVGFENEIF-----------PGT-----RAMFD-EK--E---------------------MEEE-RRLCYVALTRAEKKLYLSHA-----------
Clostridium perfringens NP_563  -TVFLVGMENGIF-----------PNA-----ACFEH-DN--Q---------------------MEEA-RRLCYVGITRAEEKLYMTSA-----------
Bacillus subtilis NP_388543.1   -VVFLMGLEEGVF-----------PHS-----RSLME-EA--E---------------------MEEE-RRLAYVGITRAEQELYLTNA-----------
Staphylococcus aureus YP_04137  -IVFIMGMEESLF-----------PHI-----RAIKS-EGDHE---------------------MQEE-RRICYVAITRAEEVLYITHA-----------
```

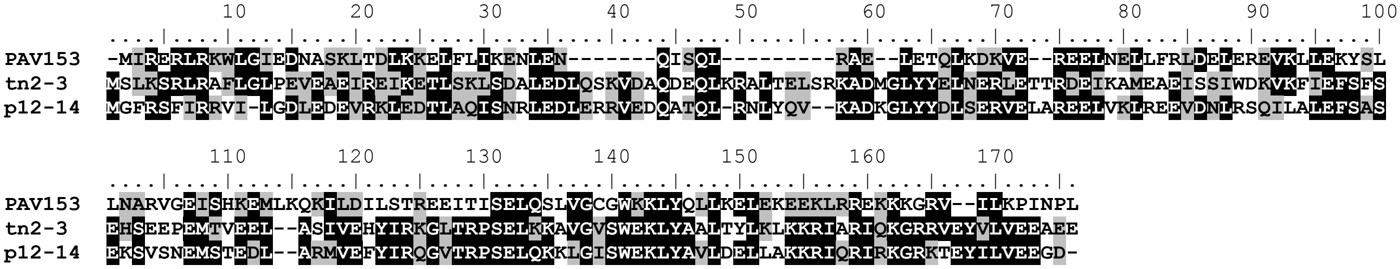
Figure S5

```
Rep-pXZ1          75 ANT--FFADLDFKEK-KDCKPGIYKDDP-------ETFELELCYEEGEKTVYVHRPPLQELLRKL-EEVK--LTPVTFIIDSGNGYQILWKNKYPVEAEEWK
Rep-pTIK4        142 AKR--LFTDLDFKEKIKDPDGGLKQKALNRECIELEDHALDCYYLDGETVVHVRRPQLKEILQKIENELG--IKI-SGVNDSGNGYHLFLELENEIDIKEWE
p12-17p          121 KYTLTLVVDIDV-----------------------EEIHKSKVVESEEKAFELAKRAWDELKPKL-EGIG--VKP-RYVFFTGGGVQL-WFVAPGLEPIEVI
tn2-12p          120 GFTLALVVDIDN-----------------------TKIHDTRIIENEEEAFEASKREWEALKPKL-QELG--FLP-RWILYTGGGLQL-WFVSDKLEPISVI
MvolDRAFT_1375    92 DNISIIPIDIDVLKDVLNY----------------GDPKKDEKFLEKDIT-KITDLEAHKLVEDYYKNLNLPIKP-TFSFFTGHGIQLYYLLENSITEEELL
MvolDRAFT_0800   147 AWS-YIPIDFDI-----------------------DEWK-DKEPTKEEMNQKINNI-FNKLPDDY-------TKP-HLIVFTGGGLRFIYYIDRPISRLELP
MJ_ECL28         161 AWS-YIAFDFDV-----------------------EEWKTNKMPTEEEIMEKLIKY-LSKFVEKD-------ILP-HKVAFTGGGLRFILYPERPILEEELI


Rep-pXZ1         164 KREEELIR--TLQKAGLPAD------------PQVKDISRVLRLPGSINHRNGR----------------EAKIIYVGEIV-----AVLPEK-YERIINA
Rep-pTIK4        239 KLENDFIK--ALQAIGLEVD------------DAVKNPSRVLRLPGSVNQRNGR----------------RASVIYYN-------------------FNI
p12-17p          195 DRASRVIP--PVLNAMLPEG---------YSVDNIFDRARIVRVPLTINYKYKT--------PDERPLEIRGRLIEFNDVR-----TPLGE------VLD
tn2-12p          194 DRASEIIP-----NIMNGVNGVKGLLSEGFKADNIFDPARIVRAPLTFNHKYRTIIKDEDGTERVVPTQVKGRVIEFNDVR-----ISLTE------FLD
MvolDRAFT_1375   176 KF-KKLIEFDNKNNYIVNGD------------SNIYDLARVLRIPFSDNKKYIS---------------KKGRVLEINDVKYDYNAVINQFKNYVNYKDN
MvolDRAFT_0800   215 IF-SKIAE-----EIGHGAD------------SAMYDIARVDRLPGTKNRKEKY------------KSERNCKILYIKNSL-----VPITPEVFIFKFGI
MJ_ECL28         230 LL-RMIAE-----DL--GAD------------IAMYDLARVDRLVGTYNYKEKYGSPRPCVTIAQLTDEEDIEKLLTPIILYEKFGIIENYKEFEVLFNE
```

Figure S6

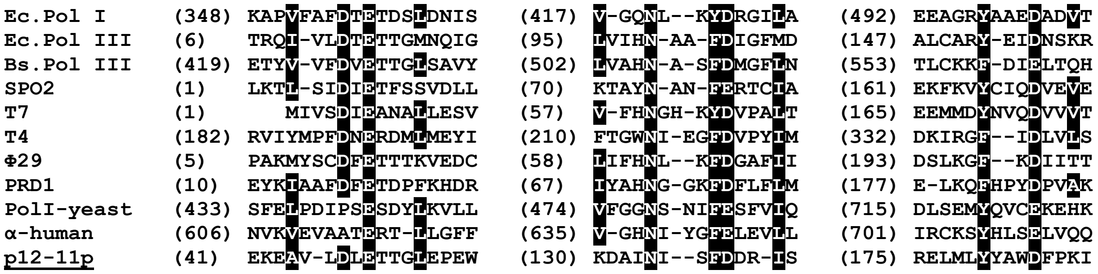
```
                          10        20        30        40        50        60        70        80        90       100
                  ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Mbar_A2246        ------------------------MN--------------------------------------------------------------------------
AMDV3_5           ---------------------------------------MTRDRDF------------------------------------------------------
MvolDRAFT_1394    -----------------MNLKRAFFG--------NGRDVARKNGWE------------------------------------------------------
PAV1_ORF375       ------------------MVAERPFY------LSRGRAALAEFGKR------------------------------------------------------
tn2-8p            ------------------------VS------SVKGMQIHARKGDE------------------------------------------------------
p12-10p           ------------------------MG------TLRGKAMLARKGEE------------------------------------------------------
HborDRAFT_3717    ------------------MSTQNPTD---TTETKRKEEILAEFPDPHTTSKQYTGDQRCHDDFWDIKRAFGIPKMEWDATYTEEEAEGKAYAEAGERKLR
pNG4027           MPRGNYRYQPATAAHPLMLIGEDPDGDGYTVRTQQGQPIVIGRPAP----SQYPSDAH------EIRQRFNIPAFDPEAFYRRTGA-----ALSG-----

                         110       120       130       140       150       160       170       180       190       200
                  ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Mbar_A2246        --SWEEKLFKEFILSPGQHG---------LENGNTGSGKTTLMFWLVDGFLQARNKL----PASQWETLVWFDRGKSSEILQLSRM-APVRLLIPESCNI
AMDV3_5           --VLPEYVTLDRDGENGHNV---------ILLGAPRQGKTYLLKYLTYRACIGRDLYDDDKSNIEPQACIWRAR-THDRYLEFFKL-GIGILAIPSGGDF
MvolDRAFT_1394    --LTEKSLERSYA-DGGKEC---------LITGPPGTGKSTLMLRFAMQ-LLD-NEF-----------VIWRGR-DTESYHYLDDWEEKVLIHHHINDEV
PAV1_ORF375       --LWKELVTRDPQ-TGGKEA---------LIIGRPGAGKTTLLLNLMIVNMVENEEV-----------VIWRGR-DLEQIHKLPHWEDISVFYHYYKDDM
tn2-8p            --LLEKLLSQDPA-YGGRQA---------AIYGAVGSGKTTLLLWLAQV-LQARGEW-----------IIWRGR-DTDGWHMLHSWKQRVKLLVHEKDEV
p12-10p           --LLNKLISQDPT-FGGRLA---------LITAPQGAGKTTLLLKLAEA-LMG-NEY-----------VIWRGR-DLAQFHRLPDWRKKVKIFAHKHDEL
HborDRAFT_3717    KKLLPNIIFHDWDPDDHQLSVGGTDW---LAVGPPGSGKSNLALWLAVRLMEANNEK-----------VIWRASTSRSEWLPFAPW---ARVCLPKGVDI
pNG4027           --VYPDMLIHIEDAADGPDTVTSPKASLTLIEAPKGSGKSTAIDQFATYQMEHNGER-----------IVRNGRQQSSDWRRLRDW---TTLWLPEGIDV

                         210       220       230       240       250       260       270       280       290       300
                  ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Mbar_A2246        EI----NFFEDENGEEVQHYEIEKIY-----FQDPKEVWNLLD---RDHINVLCLQRFI-LHPSLVAPTTAKLFESLILKAMNYDLKPAGYTYSQLK---
AMDV3_5           TL---IRINNDDSKVEITLDDLEEQGMEYFVYDSPQDIVDHLV---PGKVICILSPG------NLMEE---------------------AEFYADLGEKI
MvolDRAFT_1394    DI---FKITEDSRVKSDLNKKS---------YKNIKELVSNFE---KGVLNVVYTPTTYKVSPKLIETIEST--NTLTFSDKEKNTKNNNLFWFEFFDYY
PAV1_ORF375       TV---YKVH--GTSYEKVEVTTEE-------YKNPKDLLKKLK---PNKVNVVYAPYDYPFSVELFEEMRRN--EYFKVKKEELMRFNPNYFWYELFWQV
tn2-8p            HV---YKVPYGGTTATEITKSLNIEK-----YSTPEDLLTKLE---QDKINIVYEPTFYQITQELATYILAR--TNIRIPSWQMKEAKGIYFWFELFEKL
p12-10p           TF---YKLEHGSEYAEEIQIKVHK-------YETPEDILKQLE---KDKLNVIFEPSFYSISEELALEVLRR--SGLRINKKQLSEMKSAYFWFEFLYRL
HborDRAFT_3717    EA----RFVPKDPTQEGFEVDLEEIVREVVRYRDPVHLNHDLLK--PGQFHVVYPDPKMRRLQEIYENDQEKQYDGLEFSEDD----PLNHWWFGWMHAR
pNG4027           NGRWMEAVDRDAPTPEELVRAVRY-------YTDVFDLIEQLQDHPKGTFNVVYPDPLFRGCERAMAIPDTTTEQPKFTPQGEDKPTPVVHWWFGFLAAR

                         310       320       330       340       350       360       370       380       390       400
                  ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Mbar_A2246        ----------GKNPRVAIFLDELNNIAPSRGQGFSG-DQ--QAGAILQHNAEQLRSQNVRLVASSHGWRKLRPGIRSSFQFLISLRGANYPSSE-QPKLY
AMDV3_5           NLR-------QSRQWVHMVIDEAGDLFPP-FNQHT---M--KTQEKFINAASDFSKHLVNCIMASHEMSGLDWRLGQKFPWTIYKKGATRRKRAGAAPQI
MvolDRAFT_1394    LTL-------DYAKWTSIFIDEADDLFPATFKGIA---Y--RILERFTTNFKDYRKSWISLYCGVHNSFDLDYHIVGKFQYYIYLKAASVPNS----RDI
PAV1_ORF375       TKR-------KDRRWLGFYIDEFDDIFPSVVTGIEY--W---FVRFFKNKFKDFRKARVSLYAATHSYSDVHWMILPKFQTYIYLQDATVPPK----RKV
tn2-8p            ITR-------KDRRWITIIVDEADDLIPANPQGLQ---W--RLQEWVKNSVKDFRKNLITLILAVHNAEDCDHRVIRKLPIRIYLQGARVDKD----SLM
p12-10p           LLR-------PDRRWISILIDEIDDTFPETPTGLQ---W--KLQAWVKDIAKDLRKAFITLIGTTHTNTHVDWKIRSKMQARIYLSGAQLEED----SLI
HborDRAFT_3717    IAE-------GPHHWSSIMFDEIGDIAPEAARNDEYAHY--QKVELLKDCWVDARKYGHSIFAFGHSEADIHSMIRRKIRWRVTMNGTANPTN------S
pNG4027           TFRGTRVDSDGDRNWMTVYIDEFGMLAPESTSGGEAGHWTYECVEIMSDTAKEMRSAGVSLVGAAHHEEDVHNHWLKEFDYWVELSNRERPNRT---TKA

                         410       420       430       440       450       460       470       480       490       500
                  ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
Mbar_A2246        RFNRLFEKLEPGQAIIS----------------FPS---KTFTDKI-RIPWYGK------------------------------GEDLGYIRYIGHLEP-
AMDV3_5           LRQDYINRLEKKEAIITF-GAFYDK------FTFPMIP---------------------EDKRMDYKLITM-----------------------KRTFV-
MvolDRAFT_1394    -DKSVVKRLRKGEYLIE--NGNFGF------QSFKKLKVRDYLIRA-SIRTI----KHPEEEQEESKIPISSNHGFKDKVIAKYYQDKDSTQCLAYLQQ-
PAV1_ORF375       -EPTLPLGLEPGQMIID--SGRFGI------YEFSPLPQPNYDLQV-EIRRKKGV-NYDEELKALESIVEFGKSGLKYR-VLEIAKRDGKEAALAHIKS-
tn2-8p            TDKTITQRLNPGEAKIEWRGFGYGG------FKFPPLERRDYAIIV-EKKWTGELPKFEEDKRTAGSIIQE---------LREIAEEEGIEEALERAWE-
p12-10p           NDKEFPLRLAIGTALIEWYGLGFGL------FKFSKLPEKNYDLLV-QKKWLGPIPTLTEDKRREGSIAEM---------VRRVAMKEGIETALKLIDD-
HborDRAFT_3717    GDVVGFRRVPMEDDIMSSRDVGMFMPFSESEFQYPPVKAAHIPSPI------------------------------------------------------
pNG4027           ETPKPFRGLKQDQNLFASRPPGYGLCYTESRFSEFQFTDLGYPHETPEFQFRLSVPESAAEDESDDAVQVATDEAGQPTIVEEYTAVDGAVHELRVLAPG

                         510       520       530       540       550       560
                  ....|....|....|....|....|....|....|....|....|....|....|....|..
Mbar_A2246        ----EHKPRTQKSSIGLEDLVTALKAVVSQ--------------------------------
AMDV3_5           --IESEGNVT--FSEKVGR-------------------------------------------
MvolDRAFT_1394    --IYANNEITKRYYYKLIREFFEFLKKEYDIDTSLKELHTCYENKLITTKYMLNLKTKILVN
PAV1_ORF375       --LKEQGKVSESYYYKLRKLVLEGA-------------------------------------
tn2-8p            --LVQEGVITKRYYYQIRKALREES-------------------------------------
p12-10p           --LKAQGEITPRWAWELKRRIKSEVPSKTP--------------------------------
HborDRAFT_3717    -------DMTLKVSYA----------------------------------------------
pNG4027           SGLIDIGGDEPEIVESPSSPYPDGSFPDSPLRETAEGYELLFERAAGEIIVAARVSHH----
```

Figure S7



```
                 10        20        30        40        50        60        70        80        90       100
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
PAV153  -MIRERLRKWLGIEDNASKLTDLKKELFLIKENLEN-------QISQL---------RAE--LETQLKDKVE--REELNELLFRLDELEREVKLLEKYSL
tn2-3   MSLKSRLRAFLGLPEVEAEIREIKETLSKLSDALEDLQSKVDAQDEQLKRALTELSRKADMGLYYELNERLETTRDEIKAMEAEISSIWDKVKFIEFSFS
p12-14  MGFRSFIRRVI-LGDLEDEVRKLEDTLAQISNRLEDLERRVEDQATQL-RNLYQV--KADKGLYYDLSERVELAREELVKLREEVDNLRSQILALEFSAS

                110       120       130       140       150       160       170
        ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|.
PAV153  LNARVGEISHKEMLKQKILDILSTREEITISELQSLVGCGWKKLYQLLKELEKEEKLRREKKKGRV--ILKPINPL
tn2-3   EHSEEPEMTVEEL--ASIVEHYIRKGLTRPSELKKAVGVSWEKLYAALTYLKLKKRIARIQKGRRVEYVLVEEAEE
p12-14  EKSVSNEMSTEDL--ARMVEFYIRQGVTRPSELQKKLGISWEKLYAVLDELLAKKRIQRIRKGRKTEYILVEEGD-
```

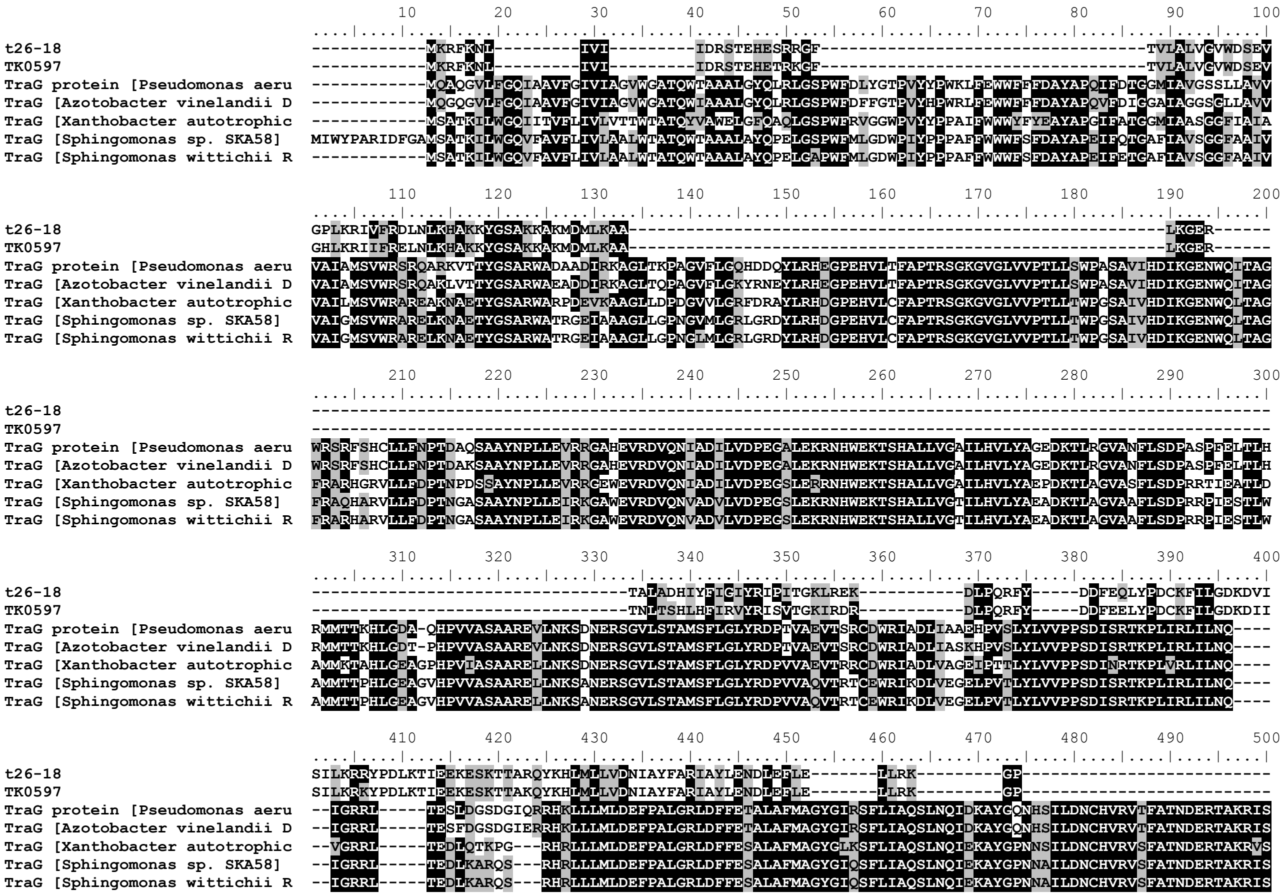
Figure S8



```
Ec.Pol I      (348) KAPVFAFDTETDSLDNIS   (417) V-GQNL--KYDRGILA   (492) EEAGRYAAEDADVT
Ec.Pol III    (6)   TRQI-VLDTETTGMNQIG   (95)  LVIHN-AA-FDIGFMD   (147) ALCARY-EIDNSKR
Bs.Pol III    (419) ETYV-VFDVETTGLSAVY   (502) LVAHN-A-SFDMGFLN   (553) TLCKKF-DIELTQH
SPO2          (1)   LKTL-SIDIETFSSVDLL   (70)  KTAYN-AN-FERTCIA   (161) EKFKVYCIQDVEVE
T7            (1)      MIVSDIEANALLESV   (57)  V-FHNGH-KYDVPALT   (165) EEMMDYNVQDVVVT
T4            (182) RVIYMPFDNERDMLMEYI   (210) FTGWNI-EGFDVPYIM   (332) DKIRGF--IDLVLS
Φ29           (5)   PAKMYSCDFETTTKVEDC   (58)  LIFHNL--KFDGAFII   (193) DSLKGF--KDIITT
PRD1          (10)  EYKIAAFDFETDPFKHDR   (67)  IYAHNG-GKFDFLFLM   (177) E-LKQFHPYDPVAK
PolI-yeast    (433) SFELPDIPSESDYLKVLL   (474) VFGGNS-NIFESFVIQ   (715) DLSEMYQVCEKEHK
α-human       (606) NVKVEVAATERT-LLGFF   (635) V-GHNI-YGFELEVLL   (701) IRCKSYHLSELVQQ
p12-11p       (41)  EKEAV-LDLETTGLEPEW   (130) KDAINI--SFDDR-IS   (175) RELMLYYAWDFPKI
```

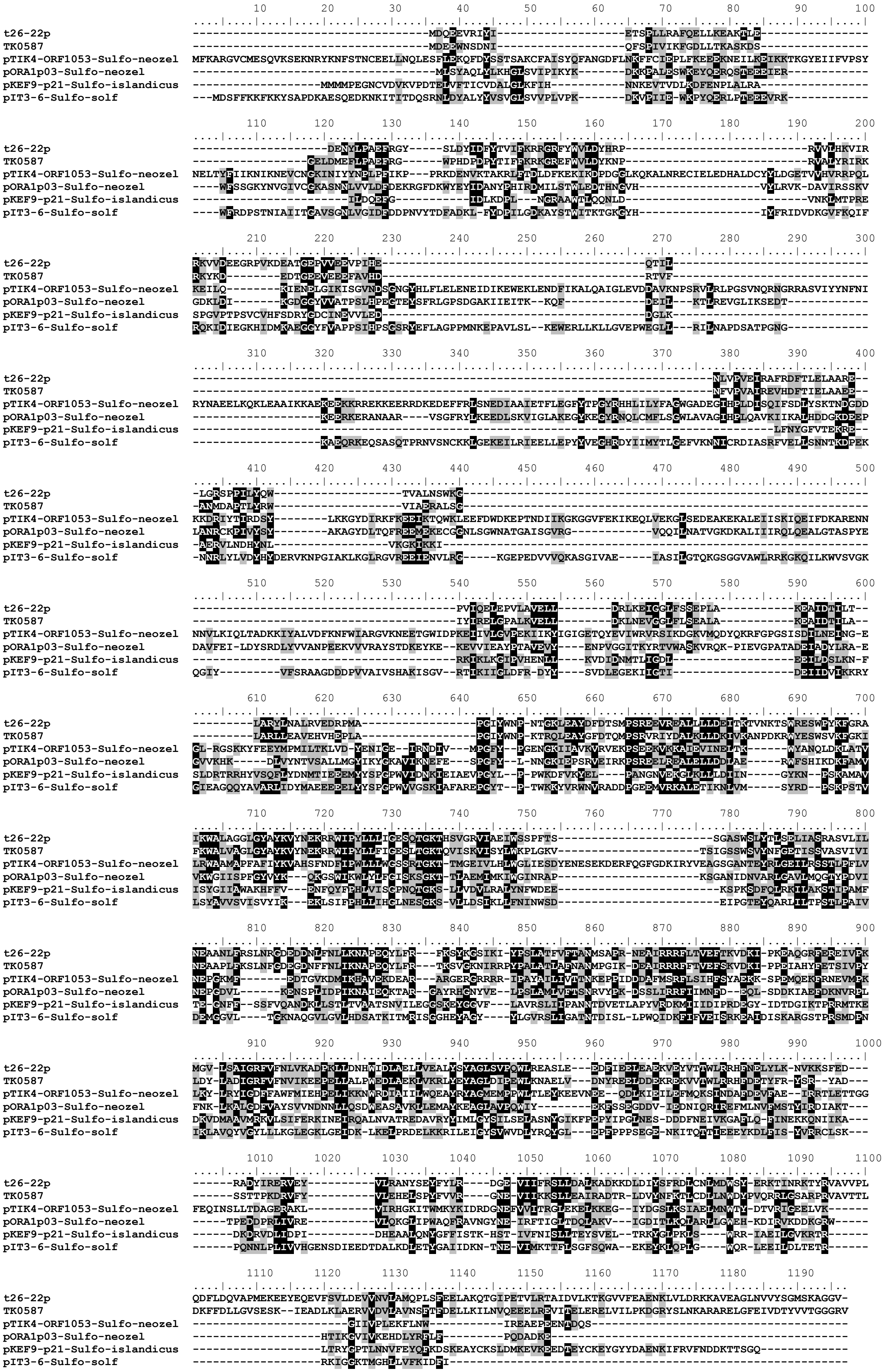
Figure S8 adapted from Bernad *et al*, *Cell*, 59(219-28), 1989 (ref 48). Sequences:
Ec.Pol I: *E. coli* DNA polymerase I. Ec. Pol III: *E. coli* DNA polymerase III ε subunit. Bs.Pol III: *Bacillus subtilis* DNA polymerase III. SPO2: phage SPO2 DNA polymerase. T7 and T4: phages T7 and T4 DNA polymerases. Φ29: phage Φ29 DNA polymerase. PRD1: phage PRD1 DNA polymerase. PolI-yeast: DNA polymerase I from *Saccharomyces cerevisiae*. α-human: Human DNA polymerase α.

Figure S9



```
                                          10        20        30        40        50        60        70        80        90       100
                                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
t26-18                           ------------MKRFKNL---------IVI---------IDRSTEHESRRGF----------------------------------TVLALVGVWDSEV
TK0597                           ------------MKRFKNL---------IVI---------IDRSTEHETRKGF----------------------------------TVLALVGVWDSEV
TraG protein [Pseudomonas aeru   ------------MQAQGVLFGQIAAVFGIVIAGVWGATQWTAAALGYQLRLGSPWFDLYGTPVYYPWKLFEWWFFFDAYAPQIFDTGGMIAVGSSLLAVV
TraG [Azotobacter vinelandii D   ------------MQGQGVLFGQIAAVFGIVIAGVWGATQWIAAALGYQLRLGSPWFDFFGTPVYHPWRLFEWWFFFDAYAPQVFDIGGAIAGGSGLLAVV
TraG [Xanthobacter autotrophic   ------------MSATKILWGQIITVFLIVLVTTWTATQYVAWELGFQAQLGSPWFRVGGWPVYYPPAIFWWWYFYEAYAPGIFATGGMIAASGGFIAIA
TraG [Sphingomonas sp. SKA58]    MIWYPARIDFGAMSATKILWGQVFAVFLIVLAALWTATQWTAAALAYQPELGSPWFMLGDWPIYPPPAFFWWWFSFDAYAPEIFQTGAFIAVSGGFAAIV
TraG [Sphingomonas wittichii R   ------------MSATKILWGQVFAVFLIVLAALWTATQWTAAALAYQPELGAPWFMLGDWPIYPPPAFFWWWFSFDAYAPEIFETGAFIAVSGGFAAIV

                                         110       120       130       140       150       160       170       180       190       200
                                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
t26-18                           GPLKRIVFRDLNLKHAKKYGSAKKAKMDMLKAA--------------------------------------------------------LKGER------
TK0597                           GHLKRIIFRELNLKHAKKYGSAKKAKMDMLKAA--------------------------------------------------------LKGER------
TraG protein [Pseudomonas aeru   VAIAMSVWRSRQARKVTTYGSARWADAADIRKAGLTKPAGVFLGQHDDQYLRHEGPEHVLTFAPTRSGKGVGLVVPTLLSWPASAVIHDIKGENWQITAG
TraG [Azotobacter vinelandii D   VAIAMSVWRSRQAKLVTTYGSARWAEADDIRKAGLTQPAGVFLGKYRNEYLRHEGPEHVLTFAPTRSGKGVGLVVPTLLSWPASAVIHDIKGENWQITAG
TraG [Xanthobacter autotrophic   VAILMSVWRAREAKNAETYGSARWARPDEVKAAGLLDPDGVVLGRFDRAYLRHDGPEHVLCFAPTRSGKGVGLVVPTLLTWPGSAIVHDIKGENWQLTAG
TraG [Sphingomonas sp. SKA58]    VAIGMSVWRARELKNAETYGSARWATRGEIAAAGLLGPNGVMLGRLGRDYLRHDGPEHVLCFAPTRSGKGVGLVVPTLLTWPGSAIVHDIKGENWQLTAG
TraG [Sphingomonas wittichii R   VAIGMSVWRARELKNAETYGSARWATRGEIAAAGLLGPNGLMLGRLGRDYLRHDGPEHVLCFAPTRSGKGVGLVVPTLLTWPGSAIVHDIKGENWQLTAG

                                         210       220       230       240       250       260       270       280       290       300
                                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
t26-18                           ----------------------------------------------------------------------------------------------------
TK0597                           ----------------------------------------------------------------------------------------------------
TraG protein [Pseudomonas aeru   WRSRFSHCLLFNPTDAQSAAYNPLLEVRRGAHEVRDVQNIADILVDPEGALEKRNHWEKTSHALLVGAILHVLYAGEDKTLRGVANFLSDPASPFELTLH
TraG [Azotobacter vinelandii D   WRSRFSHCLLFNPTDAKSAAYNPLLEVRRGAHEVRDVQNIADILVDPEGALEKRNHWEKTSHALLVGAILHVLYAGEDKTLRGVANFLSDPASPFELTLH
TraG [Xanthobacter autotrophic   FRARHGRVLLFDPTNPDSSAYNPLLEVRRGEWEVRDVQNIADILVDPEGSLERRNHWEKTSHALLVGAILHVLYAEPDKTLAGVASFLSDPRRTIEATLD
TraG [Sphingomonas sp. SKA58]    FRAQHARVLLFDPTNGASAAYNPLLEIRKGAWEVRDVQNVADVLVDPEGSLEKRNHWEKTSHALLVGTILHVLYAEADKTLAGVAAFLSDPRRPIESTLW
TraG [Sphingomonas wittichii R   FRARHARVLLFDPTNGASAAYNPLLEIRKGAWEVRDVQNVADVLVDPEGSLEKRNHWEKTSHALLVGTILHVLYAEADKTLAGVAAFLSDPRRPIESTLW

                                         310       320       330       340       350       360       370       380       390       400
                                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
t26-18                           ---------------------------------TALADHIYFIGIYRIPITGKLREK-----------DLPQRFY-----DDFEQLYPDCKFILGDKDVI
TK0597                           ---------------------------------TNLTSHLHFIRVYRISVTGKIRDR-----------DLPQRFY-----DDFEELYPDCKFILGDKDII
TraG protein [Pseudomonas aeru   RMMTTKHLGDA-QHPVVASAAREVLNKSDNERSGVLSTAMSFLGLYRDPTVAEVTSRCDWRIADLIAAEHPVSLYLVVPPSDISRTKPLIRLILNQ----
TraG [Azotobacter vinelandii D   RMMTTKHLGDT-PHPVVASAAREVLNKSDNERSGVLSTAMSFLGLYRDPTVAEVTSRCDWRIADLIASKHPVSLYLVVPPSDISRTKPLIRLILNQ----
TraG [Xanthobacter autotrophic   AMMKTAHLGEAGPHPVIASAARELLNKSDNERSGVLSTAMSFLGLYRDPVVAEVTRRCDWRIADLVAGEIPTTLYLVVPPSDINRTKPLVRLILNQ----
TraG [Sphingomonas sp. SKA58]    AMMTTPHLGEAGVHPVVASAARELLNKSANERSGVLSTAMSFLGLYRDPVVAQVTRTCEWRIKDLVEGELPVTLYLVVPPSDISRTKPLIRLILNQ----
TraG [Sphingomonas wittichii R   AMMTTPHLGEAGVHPVVASAARELLNKSANERSGVLSTAMSFLGLYRDPVVAQVTRTCEWRIKDLVEGELPVTLYLVVPPSDISRTKPLIRLILNQ----

                                         410       420       430       440       450       460       470       480       490       500
                                 ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
t26-18                           SILKRRYPDLKTIEEKESKTTARQYKHLMLLVDNIAYFARIAYLENDLEFLE-------LLRK---------GP--------------------------
TK0597                           SILKRKYPDLKTIEEKESKTTAKQYKHLMLLVDNIAYFARIAYLENDLEFLE-------LLRK---------GP--------------------------
TraG protein [Pseudomonas aeru   --IGRRL-----TESLDGSDGIQRRHKLLLMLDEFPALGRLDFFETALAFMAGYGIRSFLIAQSLNQIDKAYGQNHSILDNCHVRVTFATNDERTAKRIS
TraG [Azotobacter vinelandii D   --IGRRL-----TESFDGSDGIERRHKLLLMLDEFPALGRLDFFETALAFMAGYGIRSFLIAQSLNQIDKAYGQNHSILDNCHVRVTFATNDERTAKRIS
TraG [Xanthobacter autotrophic   --VGRRL-----TEDLQTKPG---RHRLLLMLDEFPALGRLDFFESALAFMAGYGLKSFLIAQSLNQIEKAYGPNNSILDNCHVRVSFATNDERTAKRVS
TraG [Sphingomonas sp. SKA58]    --IGRRL-----TEDLKARQS---RHRLLLMLDEFPALGRLDFFESALAFMAGYGIQSFLIAQSLNQIEKAYGPNNAILDNCHVRVSFATNDERTAKRIS
TraG [Sphingomonas wittichii R   --IGRRL-----TEDLKARQS---RHRLLLMLDEFPALGRLDFFESALAFMAGYGIQSFLIAQSLNQIEKAYGPNNAILDNCHVRVSFATNDERTAKRIS
```

Figure S10



```
                                    10        20        30        40        50        60        70        80        90       100
                            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
t26-22p                     -----------------------------------MDQEEVRIYI-------------------ETSPLLRAFQELLKEAKTLE----------------
TK0587                      -----------------------------------MDEEWNSDNI-------------------QFSPIVIKFGDLLTKASKDS----------------
pTIK4-ORF1053-Sulfo-neozel  MFKARGVCMESQVKSEKNRYKNFSTNCEELLNQLESFLEKQFDYSSTSAKCFAISYQFANGDFLNKFFCIEPLFKEEEKNEILKEIKKTKGYEIIFVPSY
pORA1p03-Sulfo-neozel       ------------------------------------MLSYAQLYLKHGLSVIPIKYK-------DKKPALESWKEYQERQSTEEEIER------------
pKEF9-p21-Sulfo-islandicus  -------------------MMMMPEGNCVDVKVPDTELVFTICVDALGLKFIH-----------NNKEVTVDLKDFENPLALRA----------------
pIT3-6-Sulfo-solf           ---MDSFFKKFKKYSAPDKAESQEDKNKITITDQSRNLDYALYYVSVGLSVVPLVPK-------DKVPIIE-WKPYQERLPTEEEVRK------------

                                   110       120       130       140       150       160       170       180       190       200
                            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
t26-22p                     --------------------DENYLPAEFRGY-----SLDYIDFYTVIFKRRGRFYWVLDYHRP---------------------------RVVLHKVIR
TK0587                      -----------------GELDMEFLPAEFRG------WPHDPDPYTIFFKRKGREFWVLDYKNP---------------------------RVALYRIRK
pTIK4-ORF1053-Sulfo-neozel  NELTYFIIKNIKNEVCNGKINIYYNPLPFIKP---PRKDENVKTAKRLFTDLDFKEKIKDPDGGLKQKALNRECIELEDHALDCYYLDGETVVHVRRPQL
pORA1p03-Sulfo-neozel       ----WFSSGKYNVGIVCGKASNNLVVLDFDEKRGFDKWYEYIDANYPHIRDMILSTWLEDTHNGVH------------------VYLRVK-DAVIRSSKV
pKEF9-p21-Sulfo-islandicus  -----------------------ILDQEFG-----------IDLKDPL---NGRAAWTLQQNLD---------------------------VNKLMTPRE
pIT3-6-Sulfo-solf           ----WFRDPSTNIAIITGAVSGNLVGIDFDDPNVYTDFADKL-FYDPILGDKAYSTWITKTGKGYH------------------IYFRIDVDKGVFKQIF

                                   210       220       230       240       250       260       270       280       290       300
                            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
t26-22p                     RKVVDEEGRPVKDEATGEPVVEEVPIHE---------------------------------------QTIL-----------------------------
TK0587                      RKYKD--------EDTGEEVEEEFAVHD---------------------------------------RTVF-----------------------------
pTIK4-ORF1053-Sulfo-neozel  KEILQ--------KIENELGIKISGVNDSGNGYHLFLELENEIDIKEWEKLENDFIKALQAIGLEVDDAVKNPSRVLRLPGSVNQRNGRRASVIYYNFNI
pORA1p03-Sulfo-neozel       GDKLDI-------KGDGGYVVATPSLHPEGTEYSFRLGPSDGAKIIEITK--KQF------------DEIL---KTLREVGLIKSEDT------------
pKEF9-p21-Sulfo-islandicus  SPGVPTPSVCVHFSDRYGDCINEVVLED---------------------------------------DGLK-----------------------------
pIT3-6-Sulfo-solf           RQKIDIEGKHIDMKAEGGYFVAPPSIHPSGSRYEFLAGPPMNKEPAVLSL--KEWERLLKLLGVEPWEGLL---RILNAPDSATPGNG------------

                                   310       320       330       340       350       360       370       380       390       400
                            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
t26-22p                     -----------------------------------------------------------------------------NLVPVEIRAFRDFTLELAARE--
TK0587                      -----------------------------------------------------------------------------NFVPVAIREVHDFTIELAAEE--
pTIK4-ORF1053-Sulfo-neozel  RYNAEELKQKLEAAIKKAEKEEKKRREKKEERRDKEDEFFRLSNEDIAAIETFLEGFYTPGYRHHLILYFAGWGADEGIHPLDISQIFSDLYSKTNDGDD
pORA1p03-Sulfo-neozel       -------------------KEERKERANAAR----VSGFRYLKEEDLSKVIGLAKEGYKEGYRNQLCMFLSGWLAVAGIHPLQAVKIIKALHDDGKDEEP
pKEF9-p21-Sulfo-islandicus  --------------------------------------------------------------------------------------LFNYGFVTEKRE--
pIT3-6-Sulfo-solf           -------------------KAEQRKEQSASQTPRNVSNCKKLGEKEILRIEELLEPYYVEGHRDYIIMYTLGEFVKNNICRDIASRFVELLSNNTKDPEK

                                   410       420       430       440       450       460       470       480       490       500
                            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
t26-22p                     -LGRSPPILYQW-------------------TVALNSWKG------------------------------------------------------------
TK0587                      -ANMDAPTLYRW-------------------VIAERALSG------------------------------------------------------------
pTIK4-ORF1053-Sulfo-neozel  KKDRIYTIRDSY--------LKKGYDIRKFKEEIKTQWKLEEFDWDKEPTNDIIKGKGGVFEKIKEQLVEKGLSEDEAKEKALEIISKIQEIFDKARENN
pORA1p03-Sulfo-neozel       LANRCKPIVYSY--------AKAGYDLTQFREEMEKECGGNLSGWNATGAISGVRG------------VQQILNATVGKDKALIIIRQLQEALGTASPYE
pKEF9-p21-Sulfo-islandicus  -AERVLNDHYNL-----------------VKGKIKKI---------------------------------------------------------------
pIT3-6-Sulfo-solf           -NNRLYLVDYHYDERVKNPGIAKLKGLRGVREEIENVLRG-----KGEPEDVVVQKASGIVAE-----IASILGTQKGSGGVAWLRRKGKQILKWVSVGK

                                   510       520       530       540       550       560       570       580       590       600
                            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
t26-22p                     ---------------------------------------PVIQELEPVLAVELL--------DRLKEIGGLFSSEPLA-----------KEAIDTILT--
TK0587                      ---------------------------------------IYIRELGPALKVELL--------DKLNEVGGLFLSEALA-----------KEAIDTILA--
pTIK4-ORF1053-Sulfo-neozel  NNVLKIQLTADKKIYALVDFKNFWIARGVKNEETGWIDPKEIIVLGVPEKIIKYIGIGETQYEVIWRVRSIKDGKVMQDYQKRFGPGSISDILNEING-E
pORA1p03-Sulfo-neozel       DAVFEI-LDYSRDLYVVANPEEKVVVRAYSTDKEYKE--KEVVIEAYPTAVEVY----ENPVGGITKYRTVWASKVRQK-PIEVGPATADEIADYLRA-E
pKEF9-p21-Sulfo-islandicus  ---------------------------------------RKIKLKGIPVHENLL----KVDIDNMTLIGDL------------------EEILDSLKN-F
pIT3-6-Sulfo-solf           QGIY---------VFSRAAGDDDPVVAIVSHAKISGV--RTIKIIGLDFR-DYY----SVDLEGEKIIGTI------------------DEIIDVIKKRY

                                   610       620       630       640       650       660       670       680       690       700
                            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
t26-22p                     ---------LARYLNALRVEDRPMA-----------------PGIYWNP-NTGKLEAYDFDTSMPSREEVREALLLLDEITKTVNKTSWRESWPYKFGRA
TK0587                      ---------LARLLEAVEHVHEPLA-----------------PGIYWNP-KTRQLEAYGFDTQMPSRVRIYDALKLLDKIVKANPDKRWYESWSVKFGKI
pTIK4-ORF1053-Sulfo-neozel  GL-RGSKKYFEEYMPMILTKLVD-YENIGE--IRNDIV---MPGFY--PGENGKIIAVKVRVEKPSEEKVKKAIEVINELTK------WYANQLDKLATV
pORA1p03-Sulfo-neozel       GVVKHK-----DLVYNTVSALLMGYIKYGKAVIKNEFE---SPGFY--L-NNGKIEPSRVEIRKPSREELREALELLDDLAE-----RWFSHIKDKFAMV
pKEF9-p21-Sulfo-islandicus  SLDRTRRHYVSQFLYDNMTIEEEMYYSPGPWVIDNKIEIAEVPGYL--P-PWKDFVKYEL----PANGNVEKGLKLLLDIIN-----GYKN--PSKAMAV
pIT3-6-Sulfo-solf           GIEAGQQYAVARLIDYMAEEEEELYYSPGPWVVGSKIAFAREPGYT--P-TWKKYVRWNVRADDPGEEMVRKALETIKNLVM-----SYRD--PSKPSTV

                                   710       720       730       740       750       760       770       780       790       800
                            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
t26-22p                     IKWALAGGLGYAYKVYNEKRRWIPYLLLIGESQTGKTHSVGRVIAEIWSSPFTS-----------------------SGASWSLYTLSELIASRASVLIL
TK0587                      FKWALVAGLGYAYKVYNEKRRWIPYLLFIGESLTGKTQVISKVISYLWKPLGKV---------------------TSIGSSWSVYNFGETISSVASVIVI
pTIK4-ORF1053-Sulfo-neozel  LRWAAMAPFAFIYKVAHSFNDFIPWLLLWGSSRTGKT-TMGEIVLHLWGLIESDYENESEKDERFQGFGDKIRYVEAGSGANTEYRLGEILRSSTLPFLV
pORA1p03-Sulfo-neozel       VKWGIISPFGYVYK---QKGSWIKWLYLFGISKSGKT-TLAEMIMKIWGINRAP---------------------KSGANIDNVARLGAVLMQGTYPDVI
pKEF9-p21-Sulfo-islandicus  ISYGIIAWAKHFFV---ENFQYFPHLVISGPNQTGKS-LLVDVLRALYNFWDEE------------------------KSPKSDFQLRKIIAKSTIPAMF
pIT3-6-Sulfo-solf           LSYAVVSVISVYIK---EKLSIFPHLLIHGLNESGKS-VLLDSIKLLFNINWSD------------------------EIPGTEYQARLILTPSTLPAIV

                                   810       820       830       840       850       860       870       880       890       900
                            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
t26-22p                     NEAANLFRSLNRGDEDDNLFNLLKNAPEQYLFR---FKSYKGSIKI-YPSLATFVFTANMSAPR-NEAIRRRFLTVEFTKVDKI-PKEAQGRFEREIVPK
TK0587                      NEAAPLFKSLNFGDEGDNFFNLIKNAPEQYLFR---TKSVGKNIRRPYPALATLAFNANMPGIK-DEAIRRRFFTVEFSKVDKI-PPEIAHYFETSIVPY
pTIK4-ORF1053-Sulfo-neozel  NEPGKMF-------EDTGVKDMIKHAVEKDEAR---ARGERGRRRR-IPAYAILIVTTNKEPPIDDDAFMSRFLSIHFSYAEKK-SPEMQEKFRNEVMPK
pORA1p03-Sulfo-neozel       NEPGDVL------KENSPLIDPIKNAIEQKTAR--GAYRHGNYVE--LPSLAMIVFTSNRVYPK-DSSLIRRFIIMNFD--EQL-SDDKIAEFDKNVRPL
pKEF9-p21-Sulfo-islandicus  TE-GNFF--SSFVQANDKLLSTLTVAATSNVILEGGSKEYGGVF---LAVRSLIIPANNTDVETLAPYVRDKMIIIDIPRDEGY-IDTDGIKTPRRMTKE
pIT3-6-Sulfo-solf           DEMGGVL----TGKNAQGVLGVLHDSATKITMRISGGHEYAGY----YLGVRSLIGATNTDISL-LPWQIDKFIFVEISRKEAIDISKARGSTPRSMDPN

                                   910       920       930       940       950       960       970       980       990      1000
                            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
t26-22p                     MGV-LSAIGRFVFNLVKADPKLLDNHWIDLAELLVEALYSYAGLSVPQWLREASLE---EDFIEELEAEKVEYVTTWLRRHFNELYLK-NVKKSFED---
TK0587                      LDY-LADIGRFVFNVIKEEPELLALPWEDLAEKLVKRLYEYAGLDIPEWLKNAELV---DNYREELDDEKREKVVTWLRRHFDETYFR-YSR--YAD---
pTIK4-ORF1053-Sulfo-neozel  LKY-LRYIGDFFAWFMIEHPELIKKNWRDIAIILWQEAYRYAGMEMPPWLTLEYKEEVNEE--QDLKIEILEFMQKSINDAFDEVFAE--IRRTLETTGG
pORA1p03-Sulfo-neozel       FNK-LKALGDFVAYSVVNDNNLLQSDWEASAVKLLEMAYKEAGLAVPQWIY--------EKFSSEGDDV-IEDNIQRIREFMLNVFMSTYIRDIAKT---
pKEF9-p21-Sulfo-islandicus  DKVDMAAVMRKVLSIFERKINEIRQALNVATREDAVRYYIMLGYSILSELASNYGIKFPEPYIPGLNES-DDDFNEIVKGAFLQ-FINEKKQNIIKA---
pIT3-6-Sulfo-solf           IKLAVQYVGYLLLKGLEGKLGEIDK-LKELPRDELKKRILEIGYSVWVDLYRQYGL---EPFPPPSEGE-NKITQTTIEEEYKDLFIS-YVRRCLSK---

                                  1010      1020      1030      1040      1050      1060      1070      1080      1090      1100
                            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|
t26-22p                     ------RADYIRERVEY----------VLRANYSEYFYLR----DGE-VIIFRSLLDALKADKKDLDIYSFRDLCNLMDWSY-ERKTINRKTYRVAVVPL
TK0587                      ------SSTTPKDRVFY----------VLEHELSPYFVVR----GNE-VIIKKSLLEAIRADTR-LDVYNFKTLCDLLNWDYPVQRRLGSARPRVAVTTL
pTIK4-ORF1053-Sulfo-neozel  FEQINSLLTDAGERAKL----------VIRHGKITWMKYKIDRDGNEFVVITRGLEKELKKEG--IYDGSLKSIAELMNWTY-DTVRIGEELVK------
pORA1p03-Sulfo-neozel       -----TPEDDPRLIVRE----------VLQKGLIPWAQFRAVNGYNE-IRFTIGLTDQLAKV---IGDITLKQLARLLGWEH-KDIRVKDDKGRW-----
pKEF9-p21-Sulfo-islandicus  ------DKDRVDLIDPI----------DHEAALQNYGFFISTK-HST-IVFNISLLTEYSVEL--TRKYGLPKLS----WRR-IAEILGVKRTR------
pIT3-6-Sulfo-solf           ------PQNNLPLIVVHGENSDIEEDTDALKDLETYGAIIDKN-TNE-VIMKTTFLSGFSQWA--EKEYKLQPLG----WQR-LEEILDLTETR------

                                  1110      1120      1130      1140      1150      1160      1170      1180      1190
                            ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|..
t26-22p                     QDFLDQVAPMEKEEYEQEVFSVLDEVVNVLAMQPLSFEELAKQTGIPETVLRTAIDVLKTKGVVFEAENKLVLDRKKAVEAGLNVVYSGMSKAGGV-
TK0587                      DKFFDLLGVSESK--IEADLKLAERVVDVLAVNSFTFDELLKILNVQEEELREVITELERELVILPKDGRYSLNKARARELGFEIVDTYVVTGGGRV
pTIK4-ORF1053-Sulfo-neozel  -----------------------GIIVPLEKFLNW-----------IREAEPEENTDQS--------------------------------------
pORA1p03-Sulfo-neozel       -------------------HTIKGVIVKEHDLYRFLF---------PQDADKE--------------------------------------------
pKEF9-p21-Sulfo-islandicus  -------------------LTRYGPTLNNVFEYQFKDSKEAYCKSLDMKEVKEEDTEYCKEYGYYDAENKIFRVFNDDKTTSGQ-------------
pIT3-6-Sulfo-solf           -------------------RKIGGKTMGHLLVFKIDFI-----------------------------------------------------------
```

Supplementary figure legends

Fig. S1. Skew analyses
Cumulative GC skews graphics for pTN2, pP12-1 and pT26-2 genomes. (*A*) GC skew graphics starting as the plasmid was sequenced. The three plasmids show a GC frequency inversion exactly were a large intergenic region is found (purple arrows). (*B*) GC skew graphics starting at the large intergenic sequence (purple arrows). All three curves have rounded shapes.

Fig. S2. Origins of replication of the three plasmids.
A. The complete sequences of the three plasmids were divided in windows of 60 bases pairs and the %GC of each window was calculated and visualized in a graphic representation. In green are shown the mean %GC of the genomes. For the three plasmids, the large intergenic region (the purple rectangle) that is found at the GC skew inversion (the purple arrow in the mini GC skew graphics) is localized in an AT rich zone.
B. Furthermore, the three AT rich large intergenic regions (in black and upper cased fonts) contain many sequences that are inverted and highly repeated (colored bases in the sequences and colored arrows). As expected, pTN2 and pP12-1 (close related plasmids) show similar organization with two large clusters of these inverted repeats, whereas pT26-2 has smaller clusters of interspersed inverted repeats. The coding regions next to these sequences (gray and lower cased font) do not contain inverted repeated sequences. These data clearly show that these intergenic regions are very likely the origin of replication in the three plasmids.

Fig. S3. Multiple alignment of Superfamily I helicases including tn2-1p and p12-1p sequences, performed with Muscle software after recovering homologous sequences by BLAST at the NCBI. Motives are indicated according to ref 34. Selected sequences are the following: *Thermococcus gammatolerans* YP_002959996.1, *Thermococcus onnurineus* YP_002307767.1, *Haloquadratum walsbyi* YP_657138.1, *Halobacterium* sp. NRC-1 NP_281177.1, *Thermoplasma volcanium* NP_111258.1, *Ferroplasma acidarmanus* ZP_01709967.1, *Methanococcus maripaludis* YP_001329282.1, *Methanoculleus marisnigri* YP_001047393.1, *Methanosarcina mazei* NP_632214.1, *Cenarchaeum symbiosum* ABK77099.1, *Saccharomyces cerevisiae* NP_012443.1, *Schizosaccharomyces pombe* NP_594341.1, *Ostreococcus lucimarinus* CCE9901 XP_001418014.1, *Trichomonas vaginalis* EAX99250.1, *Desulfovibrio vulgaris* YP_968250.1, *Borrelia garinii* YP_072790.1, *Aspergillus fumigatus* Af293 XP_749523.1, *Helicobacter pylori* NP_208269.1, *Aquifex aeolicus* NP_213544.1, *Treponema pallidum* NP_219465.1, *Planctomyces maris* ZP_01856360.1, *Neisseria meningitidis* NP_282882.1, *Escherichia coli* BAA00048.1, *Pseudomonas putida* YP_001270562.1, *Anaeromyxobacter dehalogenans* YP_463311.1, *Thermus thermophilus* BAA23380.1, *Deinococcus radiodurans* NP_295498.1, *Chlorobium phaeobacteroides* ZP_00531167.1, *Fusobacterium nucleatum* NP_603489.1, *Clostridium perfringens* NP_563176.1, *Bacillus subtilis* NP_388543.1, *Staphylococcus aureus* YP_041370.1

Fig. S4. Unrooted maximum likelihood tree of helicases related to UvrD/PcrA/Rep family (SFI).
All complete archaeal sequences are present. One sequence for each eukaryal species encoding a homologue of theses helicases was selected, and at least one sequence per bacterial phylum. 93 sequences were aligned and 152 homologous positions were selected for

tree calculation by PHYML (Guindon & Gascuel, 2003). The robustness of the tree was estimated by non-parametric bootstrap analysis (100 replicates). Scale bar represents the number of substitutions per site. Phyla are indicated on the right.

Fig. S5 to S10. Multiple alignments were performed with Muscle software after recovering sequences by BLAST or PSI-BLAST at the NCBI.

Fig. S11. CAGs (Clusters of Atypical Genes) analysis of integrated elements in Thermococcales and Methanococcales genomes.
A. Genomic regions surrounding CAGs. Blue columns represent atypical genes (assigned value = 1) whereas blanks between the columns represent non-atypical genes (assigned value = 0). The green horizontal lines above the graphics represent the core genes shared by these elements. Black arrows show the position of integrases and pink columns represent the tRNA genes.
B. CAGs families. pT26-2 (yellow), TKV2 (blue), TKV3 (blue), PHV1 (light blue), MMPV1, (orange) MMC7V1 (pink), MMC7V2 (pink) and MMC6V1 (mauve) share a core gene and their edges are colored in dark gray. These elements may show one or several links with integrated elements detected in other archaea (Cortez et al. Genome Biology 2009). Different thresholds were fixed to draw an edge between two elements, from 2 up to 5 shared genes. mmp is *M. marpaludis*, mmq is *M. maripaludis* C5, mmx is *M. maripaludis* C6, mmz is *M. maripaludis* C7, mja is *M. jannaschii* and mvn is *M. vannielii*. Methanosarcinales: mac is *M. acetivorans* and mma is *M. mazei*. Halobacteriales: nph is *N. pharaonis*, hwa is *H. walsbyi*.

Fig. S12. Unrooted maximum likelihood tree of core genes of the pT26-2 family of integrated elements.
The proteins t26-5p, t26-6p, t26-11p, t26-13p and t26-14p were concatenated and aligned with the concatenation of their homologues among the integrated elements in Thermococcales and Methanococcales using Muscle software. 10 sequences were aligned and 1065 homologous positions were selected for tree calculation by PHYML (Guindon & Gascuel, 2003). The robustness of the tree was estimated by non-parametric bootstrap analysis (1000 replicates). Scale bar represents the number of substitutions per site.

Fig. S13. Modules shuffling of Rep proteins from archaeal plasmids. Homologous modules are in the same color.