Epimerase sequences

```
>B_DELTA gi|218779508|ref|YP_002430826.1| UDP-N-acetylglucosamine 2-epimerase [Desulfatibacillum alkenivorans AK-01]
MKLCLVITTRGNYAKVKSIIQAAENDPDVELQIIVGGGAILAKYGNIADSLTTMGVNVDRRIHFLVEG
ETPVTMAKSAGLAVSEFTTAFENLQPDVVMVIADRFECLSIAMTASYMNIPVAHMEGGEVSGSIDESI
RHAITKLSHVHFPATQEAADRIIRMGEAPETVFPVGATSLDVIAGLDLDNLQPIMEKQASAGVGARV
DLTQPYLTVIQHPVTTEYAENLEHVNQTLNAIHELGMGTIWIWPNMDAGSDGISKGIRIFREKHVDA
PMHFFKSLPIEHYAPLLKNAQAIVGNSSSGIREAAFLGTPCVNIGSRQSGRDRSVNIIDVGYDKAEIKA
AIQKQLDHGRYEENHMWGDGKAGPKILDVLKTFDFTIQKRICY
>B_DELTA_2 gi|148266095|ref|YP_001232801.1| UDP-N-acetylglucosamine 2-epimerase [Geobacter uraniireducens Rf4]
MTRRRICVVTGTRADYGLLYRLMKEIEGDPDLQLQVVATGMHLSPEFGLTYRDIETDGFTIHERVEM
LLSSDTPVGIAKSIGLGVIGFADAFERLRPEIIVLLGDRFEMLAAAQAALVGRIPVAHIAGGDTTEGAF
DEAIRHSITKMSHCHFVTNEAALHRVRQLGENPDAVHLVGSPGIDQIRRLTLLSRHELERDLGLAFRK
RNLLVTFHPATLDEVSASEQMQALFDALDRLGPEVGIILTKPNADTGGRRLSDMIDVYVSERDQAKA
FTSLGQLRYLSLIPQVDAVVGNSSSGLYEAPSFKKPTVNIGDRQKGRLQASSVINCAPVADDILQAVR
AAFGLDCSTAVNPYGDGNASSRIAAILKGLPEPRRLIKKHFFDLNG
>B_DELTA_4 gi|303247652|ref|ZP_07333922.1| UDP-N-acetyl-D-glucosamine 2-epimerase, UDP-hydrolysing [Desulfovibrio fructosovorans JJ]
MDSRLKICIVTGSRAEYGLLYWLLKDIAADPDLKLQIIATGMHLSPEFGLTYRQIEVDGFTIDAKVEM
LLSADTPVAVTKSMGLGVIGFADALDRLAPDIIIVLGDRFEIFAAAQAAMVARIPLAHIHGGETSEGAY
DEGIRHAISKMAQWHFVATEPYRQRVVQLGEAPQRVFNVGAPGLDHLSRTQLLTREELEQVLDMRL
RQPLFVVTYHPVTLGMDSPEKAMKELIAGLRKFEDASIVFTYPNADSGGRALRQIINSFVAEDLQRIR
AYTSLGQQRYLSLIRQADVIVGNSSSGLIEAPALKTATVNIGDRQKGRLKASSVIDASEHRTDIAKAIR
HALSPEFRIQLLQTQSLYGCGDASASILRQLKAPLPTIQKSFFDIKHKY
>B_ALPHA gi|86748660|ref|YP_485156.1| UDP-N-acetylglucosamine 2-epimerase [Rhodopseudomonas palustris HaA2]
MTKRKICVVVGSRANYSSIKSAMRAIQDHPALELQLIVAASAVLDRYGSVVNLIEKDGFRPHARVTM
LIEGETPATMAKSTGLGLIELPTLFEQLGPDVVLTVGDRFETMATTLAAAYMNIPVAHTMGGEVSGTI
DESIRHAVTKFAHIHFPASQGAAERIIKLGELPRHVHMVGCPRIDLVAEILGRSSGGLDAGLFDLGVG
QQFSVDEPFALVSQHPVTTEYGTGEAQITLTLEAVREQGLAAIVLWPNADAGSDDISRGIRKWRERK
LDDRMHFFKNLPIETYVNLMRSAACLVGNSSSGIREGAYIGTPVVNIGTRQHMRDRGDNVIDVGYD
KKQISDAIARQVEHGRYAMDPIYGDGTAGTKIADILVTERVDVQKCITY
>B_ALPHA_2 gi|103487060|ref|YP_616621.1| UDP-N-acetylglucosamine 2-epimerase [Sphingopyxis alaskensis RB2256]
MTSPKRIVYLTGTRADFGLMLPTLRAIDRSPHLELELLVTGMHLSDRFGRTEREVEAAGLKIGRRIPV
PIDDDSGHGMGVSTGLITCAVADYLAETACDVLLLLGDRGEMLAAATAGLFADVPIVHVAGGDRSG
SVDESIRHAISKLAHIHCVSNEDARQRLIRMGEDPDRIFDVGAPGLVGLKKPDRATAAMVRNRYGID
DEESFVLVLFHPVVQQAAEAGAQWRAMFDALAGLPFRYVALMPNADHGTTSIRTEIEAMRDAGQL
TTIDHMPRADYLALLAECRFLIGNSSSGIVEAATFGTPVVNVGDRQFGRLRSANVFDAPPESGAIGRA
IEKAIGFDPQGLRNVYGDPHADIRICEILEKTDFSASALRMKTISY
>B_ALPHA_3 gi|294678483|ref|YP_003579098.1| UDP-N-acetylglucosamine 2-epimerase [Rhodobacter capsulatus SB 1003]
MTREILFVTGTRADFGKIEPLALAARDRGFKVSFLVTGMHMLDRYGLTKIEVHRVQGATVHEFLNQ
REGDPQDTILAKSIIGFSDIFAELRPDLVVFHGDRIEALACALVCATNYIRSAHIEGGEVSGTIDEVFRH
CNTKLAACHFVSSEAAAKRVMTLGEPADRIHVIGSPELDFHARPSGVTLPEVLSRYDIPFDDYGVTLF
HPVTSEAATMGRQAADLFGALEASGRNFVVIAPNNDPGSREIFAVLEQLPRERFRLIPSMRFAHFSEL
MKHAACLVGNSSAGVREAPFLGIPSLDIGTRQTNRAEAPSLFSADAAEREKIAAFLATEWGKRYPPH
TAFGEGRAAERFLEVLADEGFWQGGLQKTFADHG
>B_ALPHA_4 gi|83309178|ref|YP_419442.1| UDP-N-acetylglucosamine 2-epimerase [Magnetospirillum magneticum AMB-1]
```

```
MKRICVVTGTRAEYGPLFWVLKEIDAHADLELQLVVTGMHLSPEYGSTWKTIEADGFAIAAKVEML
LSGDTGVAIAKSMGLGTIGFADAFERLKPDIVVVWGDRFELMAAVQAAVVARIPIAHIGGGDVTEGA
FDDAIRHAISKMAHLHFPIIADSARRLRQLGEDPERIHLTGNASLDHLRRTIFLERAEVETRLNFSLRP
RNVLVTYHPVTLDPEEGRRGWAEMLAALESLGDEVGIVMTAPNADNDSRELMVALEAFVGAHPNA
ILRTSLGSQLYMSTARLCDAVVGNSSSGLLEIPSLGVATVNIGTRQKGRPRADSVIDCPPERAAIRAAM
DRAMAMDCSAVVNPYGDGQSARRIVQVLAGIEDFPALLHKSFVDLAGPGEAGEE
>B_FIRMICUTES gi|302871579|ref|YP_003840215.1| UDP-N-acetyl-D-glucosamine 2-epimerase, UDP-
hydrolysing [Caldicellulosiruptor obsidiansis OB47]
MRKILFITGTRADYGKIKPLMRKVEESKEFELHIFVTGMHMLSKYGSTWKEIEKDGFKNIYKFINQQ
YNFHMDIALSNTIIGLSNFVNELRPDMIVVHGDRLEALAGAIVGAFNNIRVAHIEGGEVSGTIDESIRH
AITKFSHIHFVCNEEAKKRVIQLGEKPESVYVIGSPDIDVMLSDTLPSLREVKEHYEIEFENYAIFIYHP
VTTQIDILHRNIKEVIDALIESGKNYVVIYPNNDEGSNIILKEYERFKMNKRFKIFPSLRFEYFLTLLKH
ADFMIGNSSAGIREAGIYGVRVIDIGIRQKGRYDISKFKNIIHVNENKEEILKAIQKVQLVDKFYISAFG
DGKSAERFINILKSEDVWNIEIQKNFVDIEF
>B_FIRMICUTES_2 gi|253681586|ref|ZP_04862383.1| UDP-N-acetyl-D-glucosamine 2-epimerase, UDP-
hydrolysing [Clostridium botulinum D str. 1873]
MNRKIAVITGTRADYGIYYSVLKAIENHKDLELHLIVCGMHLSPEFGMTINEIEKDGFKIDDKIDTILS
SDSGEAMAKSIGITLMGLTQSLDRIKPDVLLILGDRGEMMAGALAAIHMNIPVAHIHGGEVTGTVDE
SIRHSITKLSHIHFPANEDSRERIIKMGEEKKNVYVVGAPGIDYIKNTEYLSREEVLRRFNLKDDKIFIL
TQHPVTTEKDMVVYQIEETLSAIAELGVQTIISYPNSDNGGREIIKVIEKYREKYDFLKVFKNLSQVE
YLSLLNTADIMIGNSSSGIIEAPSFKLPVINIGTRQQGRLRACNIIDVSYNRKEILSAIDKVLYNEEFKKE
LKKCENPYGDGHSGERIADILSKVDINHQLIQKRITY
>B_FIRMICUTES_3 gi|118444030|ref|YP_877858.1| UDP-N-acetylglucosamine 2-epimerase [Clostridium
novyi NT]
MKRKIAVITGTRADYGIYYSVLKAIENHKDLELSLIVCGMHLSPEFGMTIEEIQKDGFKIDDKIDTILS
SDSGSAMAKSIGITLMGLTQSLERINPDILLILGDRGEMIAGALAAVHMNIPVAHIHGGEVTGTVDESI
RHSITKLSHIHFPANEDSKRRIIKMGEEEKNVYVVGAPGIDYIKNTKYLSREEVLKRFNLKDDKIFIMT
QHPVTTEKHMVSNQIEETLSAIAKLGVQTIISYPNSDNGGREIIKVIEKYREKYDFLKVFKNLSQVEY
LSLLNTADVMIGNSSSGIIEAPSFKLPVVNIGTRQQGRLRACNIVDISYSKKEILDAIDKVLYDEEFRK
NLENCENPYGDGKAGERIADILSKIQINSDLIQKRITY
>A_EURY gi|91773549|ref|YP_566241.1| UDP-N-acetylglucosamine 2-epimerase [Methanococcoides
burtonii DSM 6242]
MKRKIAVVTGTRADYGIYLPVLKAIQRSSKLDLSLIVTGMHLSETFGHTVDEIEKDGFSIDAKIPLGLL
EDSGASMALDVGICILGLTDALKKIKPDILLVLGDRGEMLATTIAGIYMNIPVAHLHGGEVSGTVDES
IRHAITKLSHIHFPATEESAERIRNLGEDEFRIYVVGAPALDTILSETFVPKEEMGHFFDIDINKPIILVVQ
HPVTTEVGAVERHIRETMDAVVELGEQTVVIYPNADAGGRKIIETIEQYRNYAFIKIFKNIRHVDYLSL
MRTTNVMVGNSSSGIIEAPSFGLPVVNIGTRQTGRQRGQNTIDVDYDKDEIIKAIKVGLYDKDFKRK
ASKCISPYGNGHAGTAIAEILESI
>A_EURY_2 gi|289579837|ref|YP_003478303.1| UDP-N-acetyl-D-glucosamine 2-epimerase, UDP-
hydrolysing [Natrialba magadii ATCC 43099]
MSSTRRKILVLTGTRAEYGLLRSSMEAIQNHDDLTLSIVATGMHLSPQHGMTVEKIREDGFSIDREVL
MQLSGDSETAMAKSLGIGTASLADAFESLDPDVVLLLGDRDEALAGALAASHMNIPVAHVHGGDS
AHGAMIDESIRHAITKFSHIHFPASERSAERIKKLGEESWRITIAGAPGLDDILAGEYEDPESVLQKYD
LDPDRRLLMVLQHPVTTQPDAAGEQMAATLDAVESTEAQAVIIYPNSDAGSNQMIDEIESRSFGADV
RTFRNMPRKEYLGLMAATDVMVGNSSSGIIEAPSFDLPVVDIGPRQRGRERTQNTFSAEHESEQIREK
IYQCLNEGVLDENQDNPYDYGGAGSRICERIRTINIDENILRKKLTFGRWF
>A_EURY_3 gi|150400855|ref|YP_001324621.1| UDP-N-acetylglucosamine 2-epimerase [Methanococcus
aeolicus Nankai-3]
MKNRKIKKVAVVTGTRAEYGILKPLIEKINDDNDLELQLMVTGMHLLKKFGYSIKEIENDGFPIASKI
RMYDEDSLGELSYHGVSLGRAVSEFTREFVYLNPDIVLVIGDRLEALAPVLSASTLNIPIGHIHAGDST
DSGHIDEQIRFAISRFSHLLFAPTEKCVERLTKMGEEPWRAYNVGALGLDSILSYKPLTKEELFKKLNL
NGDNPVAIIIFHPVIHEYSTIDHQIESIMKAVIETKINTVVIYPNNDLGSKKIIEIIKKYSNFENIKLFENLE
HNIYISLMYHANLMIGNSSSGIIEAPSLGLPVINVGSRNTGREHGDNVLFVKPVSKEIVEAINKALYDI
DFIEKVKKKNNPWGDGKTSERLVNILKNTEINEHFMRKRILY
```

```
>B_BACTERIODETES tr|Q9RGJ0|Q9RGJ0_BACFR UDP-N-acetylglucosamine 2-epimerase
OS=Bacteroides fragilis GN=wcgL PE=3 SV=1
MEIKLDYSDIKFRYDGKLRLLIIVGTRPEIIRLAAVINKCRRYFDCILAHTGQNYDYNLNGVFFHDLGL
QAPDVYMDAVGDDLGSTMGNILNASYKLMSHLRPDAVLVLGDTNSCLSVISAKRLHIPIFHMEAGN
RCFDECLPEETNRRIVDIISDMNLCYSEHARRYLNASGVAKERTYVTGSPMAEVLSENLSAIESSDIH
ARLGLRKGQYILLSAHREENIDTDKNFASLFEGINAMAEKYDMPVLYSCHPRSRNRLESSGFKLDSR
VIRHAPLGFHDYNCLQMHAYAVVSDSGTLPEESSFFTSVGHSFPAVCIRTSTERPEALDKGCFILAGID
KASLLQAVDTAVEMNRNGDNGVPVPDYMDRNVSTKVVKLIQSYTGIVNKIVWRKS
>B_BACTERIODETES_2 tr|Q9F752|Q9F752_BACFR UDP-N-acetylglucosamine 2-epimerase
OS=Bacteroides fragilis GN=wcgT PE=3 SV=1
MLKVMTIVGTRPEIIKLSRVMAELDKYTEHIMVHTGQNFDYELNEIFFQELRIRKPDYFLDAAGKNA
AETIANVIRKSDELMDQVKPDALLLYGDTNSCISVISAKRRKIPIFHMEAGNRCFDQRVPEEINRKIVD
HLSDINMPLSEHARKYLLAEGLRPETVIKIGSPMTEVLIYHKAEIEENDVLEKEGLKKGDYFIVSTHR
EENVDSEKNFSDLLSSLNAIVDKYHKKVIVSTHPRTRKKLESIGFINSNPMIEFMKPFGFMEYIKLQQ
NAFCVISDSGTITEESSILHFPAITIRQAHERPEGMDEGTLIMTGLNSDRILESIEIVTSQYAEGADVIHSI
PDYASDNVSKKVVRIILSYTDYINRTVWHKEI
>B_BACTERIODETES_3 tr|Q9RGJ8|Q9RGJ8_BACFR UDP-N-acetylglucosamine 2-epimerase
OS=Bacteroides fragilis GN=mnaA PE=3 SV=1
MKKIMLVFGTRPEAIKMAPLVKEFQKYPKEFQTIVCVTGQHREMLDQVLHLFHIRPNYDLNIMKQG
QDLYDVTARVLIGMRDVLAEVRPDLLLVHGDTTTSTAAALSAFYQQIPVGHIEAGLRTHNVYSPWPE
EMNRQVTGRIATYHFSPTLLSRQNLLNEGVKGDFIIVTGNTVIDSIYMVVDRIRHDKLLEVQLRNVLS
TSGYDVKRLSNDKKLVLITGHRRENFGDGFISMCQAIKALSEKYPNVDFVYPMHLNPNVRKPIHEVF
GDELSDLNNLFFIEPLEYLSFVYLMEKSIIVLTDSGGIQEEAPGLGKPVLVMRDTTERPEALEAGTVK
LVGTNYDKIVNEVSALLDDSDYYNTMSKAVNPYGDGKACSKIVNFLLA
>B_BACTERIODETES_4 tr|Q64V44|Q64V44_BACFR UDP-N-acetylglucosamine 2-epimerase
OS=Bacteroides fragilis GN=BF1886 PE=3 SV=1
MCAKKQIIPQESVNLFLENCKNTKMKITIVAGARPNFMKIAPITRAIDAAKAQGKRISYRLVYTGVEN
DNSLDASLFADLHMKAPDAYLGVNGNNPTELTAGIMIAFERELTENPTHVVLVVDDLTATMSCAIVA
KKQNIKVAHLVAGTRSFDMSMPKEINRMITDGLSDYLFTAGMVANRNLNQTGTENETVYYVGNILI
DTIRYNRNRLIKPVWFSVLGLKEHEYILLTLNRHVLLNNKENLQELMETLLKKANGMPIVAPLHTYV
RDAIKALGITAPNLHIMPTQSYLSFGYLMNQAKAIVTDSGNVAEEATFLGIPCITLNTFAEHPETWRTG
TNELVGEDPAALGACMDKLMNGEWKQGTLPERWDGRTAERIVQILLGE
>B_BACTERIODETES_5 tr|Q64MP1|Q64MP1_BACFR UDP-N-acetylglucosamine 2-epimerase
OS=Bacteroides fragilis GN=BF4509 PE=3 SV=1
MKKILLVFGTRPEAIKMAPLVKALQRDTEHFETKVCVTAQHRQMLDQVLEVFDIIPDYDLNIMAPN
QDLYDITTKVLLGLRDVLKDFCPDTVLVHGDTTTSMAASLAAFYRQVAVGHVEAGLRTYDMLSPW
PEEMNRQVTDRICTYYFAPTGKSKQNLLQENIDAKKIFVTGNTVIDALLMAVDIISKKPGIKEKLHQE
LRDKGYEVGQREYILVTGHRRENFGEGFLHICKAIRELAALHPEMDIVYPVHLNPNVQKPVYELLSG
VDNVYLISPLDYLPFIYAMQHSTLLLTDSGGVQEEAPSLGKPVLVMRNTTERPEAVEAGTVKLVGTD
AEAIVSNVTELLRNKELYRRMSETHNPYGDGHACERILSALTR
>B_BACTERIODETES_6 tr|A6GZ45|A6GZ45_FLAPJ UDP-N-acetylglucosamine 2-epimerase
OS=Flavobacterium psychrophilum (strain JIP02/86 / ATCC 49511) GN=fnlC PE=3 SV=1
MKKLKVMTVVGTRPEIIRLSRVLMALDESEAIEHIIVHTGQNYDYELNQIFFEDLGLRKPDYFLEAA
GKTATETVGNILIKIDPLLEQLQPDAFLVLGDTNSCLCAIPAKKRQIPIFHMEAGNRCFDQRVPEETNR
KIVDHTSDVNLTYSDIAREYLLREGLPADRIIKTGSPMFEVLHHYLPQIKKSDVLSRLNLEEGKYFVV
SSHREENINSEKNFKGLIESLNLIAEKYNYPIIVSTHPRTRNMIDKMKVQVRPEVQFLKPLGFHDYNA
LQMRSYAVLSDSGTISEESSTLNFRALNIRDAHERPEAMEEASVMMVGLSPERIMQGLVQLQSQEIGE
KRNFRSVADYSMPNVSQKMVRIILSYTDYINRVVWSKK
>B_BACTERIODETES_7 tr|Q64XB8|Q64XB8_BACFR UDP-N-acetylglucosamine 2-epimerase
OS=Bacteroides fragilis GN=BF1108 PE=3 SV=1
```

```
MEIKLDYSDIKFRYDGKLRLLIIVGTRPEIIRLAAVINKCRRYFDCILAHTGQNYDYNLNGVFFHDLGL
QAPDVYMDAVGDDLGSTMGNILNASYKLMSHLRPDAVLVLGDTNSCLSVISAKRLHIPIFHMEAGN
RCFDECLPEETNRRIVDIISDMNLCYSEHARRYLNASGVAKERTYVTGSPMAEVLSENLSAIESSDIH
ARLGLRKGQYILLSAHREENIDTDKNFASLFEGINAMAEKYDMPVLYSCHPRSRNRLESSGFKLDSR
VIRHAPLGFHDYNCLQMHAYAVVSDSGTLPEESSFFTSVGHSFPAVCIRTSTERPEALDKGCFILAGID
KVSLLQAVDTAVEMNRNGDNGVPVPDYMDRNVSTKVVKLIQSYTGIVNKIVWRKS
>B_BACTERIODETES_8 tr|Q64Q32|Q64Q32_BACFR UDP-N-acetylglucosamine 2-epimerase
OS=Bacteroides fragilis GN=BF3656 PE=3 SV=1
MKKKILLTFGTRPEAIKMAPLVKELQEKKDCFEILVCVTGQHREMLDQVLHIFDISPDYDLNIMRQG
QDLYDITSRVLLGMRTILIETKPDIVLVHGDTTTSTATALAAFYQQIPVGHIEAGLRTHDIYSPWPEEM
NRQITARIATYHFAPTILSKQNLINEGVKEDFIFVTGNTVIDALYWVVNKIKTNEGLSKAIKDALKSSG
YNIDNRGTEKKIVLITGHRRENFGEGFIHMCHAIKTLAEKYPDVDFIYPMHLNPNVRKPINDVFGEQ
KFINIFFIEPLEYLSFVYLMDKSFLVLTDSGGIQEEAPGLGKPVLVMRDTTERPEALEAGTVKLVGTD
YYKIVKEVSKLLESKTYYESMSKAVNPYGDGCACSRIVTLLQEV
>B_DICTYOGLOMALES tr|B5YCQ1|B5YCQ1_DICT6 UDP-N-acetylglucosamine 2-epimerase
OS=Dictyoglomus thermophilum (strain ATCC 35947 / DSM 3960 / H-6-12) GN=DICTH_0427 PE=3
SV=1
MKVVSIVGARPQFIKLAPFSAELRKNGIKEVILHTGQHYDENMSELFFKELEIPEPDYNLGIGSGSHG
EQTGRMLIGIEEVLVKEKPDVVIVYGDTNSTLAGALASAKIHIPLAHVEAGLRSFNKKMPEEINRIVA
DHLSDILFCPTETAVENLKREGIEKGVYLVGDVMFDALMHFSKLAKEKSKILEKLSLNPKDYYLITV
HRAENTDNPERLKNIFSAISELDKEVIFPIHPRTKNRLKELGLEGYLRGRVRIIDPVGYLDMIELEKNA
LAILTDSGGVQKEAFWLRVPCITLREETEWVETLKYGWNILVGSNKERILEAIKNIKNGKEISFENDY
ASPKMREVLIKELERRREDDR
>B_BACTERIODETES_9 tr|A6GZ15|A6GZ15_FLAPJ UDP-N-acetylglucosamine 2-epimerase
OS=Flavobacterium psychrophilum (strain JIP02/86 / ATCC 49511) GN=neuC PE=4 SV=1
MNKKILFLTGTRADFGKIKSLIQILEAHPNFEPYIFVTGMHLQEEYGYTFLEVQRCGFRNIHTFQNHT
HETTMDLTLAKTITGLSAYVKDCQPDMIVIHGDRVEALAGAIVGSLNNILVSHIEGGEVSGTIDELIRH
ATSKMSHIHFVSNDQAKKRLLQMGELHESIFIIGSPDVDIMFSKSLPDLQTAKKYYEITFNDYAVVMF
HPVTTEAKYIAQYANDFVHVLLADTHNYVVVFPNNDLGSHAIIQAYEKLKNNKRFRVFPSLRFEYFL
TLLKNAQFIIGNSSAGIREAPYYGLPIINIGTRQQNRAAHTHIINVDYNKLHIQEALSSIDSHKVQPSKN
DFGQGNSAKLFLASLEKNDIWQINYQKQFNDI
>B_DICTYOGLOMALES_2 tr|YF19|B5YF19_DICT6 UDP-N-acetylglucosamine 2-epimerase
OS=Dictyoglomus thermophilum (strain ATCC 35947 / DSM 3960 / H-6-12) GN=DICTH_1307 PE=3
SV=1
MKKKISLVFGTRPEAIKMAPVAKIIKESEFFDLQIILTAQHRELLDQVIEIFGLKSDYDLNIMQEKQTLT
HITVKVLQGLDIIWQKDPPDMVLVHGDTTTTFAASLAAFYKKLPIGHVEAGLRTYNKYQPYPEEMN
RHLTGVLADLHFAPTQRAKDNLINERVPKENIFITGNTVIDALLFVHRNMNQLKPQDLVKNLPEKFIL
VTAHRRENWGEPLKNIVLALDEILKEFEDFYVVFPVHPNPLVREQVYSVLKDNKRAILIPPVDYVTM
VYLLDKCYLVLTDSGGLQEEAPSLGKPVLVLREVTERPEAVEAGTVKIVGTSKESIVREVRRLILNKE
EYVKMSKAINPYGDGKASERIRDILLYYFGFIDTPPSEFNP
>B_CYANOBACTERIA tr|Q05U86|Q05U86_9SYNE Putative UDP-N-acetylglucosamine-2-epimerase
OS=Synechococcus sp. RS9916 GN=RS9916_38627 PE=4 SV=1
MALIMRKICVVTGTRADYGLLRWVIDGISKSSMLQLQIVATGMHLSHEFGLTFKEIENDGYNIDLKV
EMLLSSDSSVGITKSIGLGIIGFADAFKELNPDIVLLLGDRFEIFAAATAALISKIPVAHCHGGETTEGAF
DESIRHSITKMSHLHFVAAREYQERVLQLGEDPENVHLVGGLGIDNIKKLSLLDKSSLESELEFKFQD
TNYLVTFHPVTLDTYSGTEQITELLRALSNLVNTGIIFTMPNSDTGSRPLMNLINNFCSDHEYASSYTS
LGQIKYLSTLKYVDAVIGNSSSGLLEAPTFKTATINIGDRQKGRLMASSVINCDPNCISIQNAINHLNS
HAFKQSLNTTVNPNGSGGSADKIIKVLESYKYPTKLQKKFHSYP
>B_CYANOBACTERIA_2 tr|Q05WR5|Q05WR5_9SYNE UDP-N-acetylglucosamine 2-epimerase
OS=Synechococcus sp. RS9916 GN=RS9916_33397 PE=3 SV=1
```

```
MVVLLKPMAGKPRVTIVLGTRPEAIKLAPVIQEFQACQLLETRVVLTGQHREMVTQVMDLFSLKAD
QDLNLMAPRQTLTHVTCAALQGLRDDFQAYPPSLVLVQGDTTTAFAAALAAFYEQIPVGHVEAGLR
TDNILDPFPEEANRRLISQVAQLHFAPTQQSESNLLASSVVGKVMVTGNTVIDALLRMAERAPQLDD
VAIDWSRQRVILATVHRRENWGERLQSIAEGMLQVLNDHPDTALLLPLHRNPTVREPLQALLGNHP
RVVLTEPLDYDRLVAAMKGCTLLLTDSGGLQEEAPALGKPVLVLRRTTERPEAVEAGTARLVGTDAG
TIAEEAHRLLSDPQAYDAMARAVNPFGDGQASARILEAARALLEV
>B_CYANOBACTERIA_3 tr|A0YRX0|A0YRX0_LYNSP UDP-N-acetylglucosamine 2-epimerase OS=Lyngbya sp. (strain PCC 8106) GN=L8106_19768 PE=3 SV=1
MINSRIRVAVVLGTRPEAIKLAPVIQQLQLSSDFETHVILTGQHREMVKQVMQLFDLDADYNLEIMQ
RSQTLTDITWRSLQGLEGLFQQLRPDIIIVQGDTTTAFAAALAAFYQKIPVGHVEAGLRTNDLLNPYP
EEANRRLISQLTQFHFAPTTQAVDNLRNSGVLGEVHHTGNTVIDALLSVVKCQPELKVAEGSWSQYR
LILATVHRRENWGIPLTQIAESFLKILDDFPDTALLLPLHRNPTVREPLQKILGNHPRIFLTEPLDYAEL
VAAIDRCYFVLTDSGGLQEEAPSLGKPVLVLRKTTERPEAITAGTAKLVGTETAEIVEAATELLNNINT
YQKMATAINPFGDGHAAEYIVQIIRSHFEGKDRS
>B_CYANOBACTERIA_4 tr|A2C1S1|A2C1S1_PROM1 UDP-N-acetylglucosamine 2-epimerase OS=Prochlorococcus marinus (strain NATL1A) GN=NATL1_08731 PE=4 SV=1
MDKKVLLFVTGTRADFGKMEPLAREAFNNGFKVIFFVTGMHMMREYGLTKEEVHKNKDIQIFEFS
NQKYGDKLDTILSNTVRGFSNYVKEINPDIVIIHGDRIEAIACSLVCSTNNIISAHIEGGEVSGTIDEVF
RHCNTKLCTFHLVSSNEAKKRVRQMGEPEKNIFVIGSPELDIHGRKSGVDLLQVKERYKIDFKEYGIC
IFHPVTTEENQIKIQAENLFKSLSISNRNFVIILPNNDPGSIYICNEIDKLNSNNFRIIPSMRFNYFSELMK
NSSLIIGNSSLGVREAPFLGIMSINIGTRQNKRALTQSIYNCSGQSIPEIVDAIGKFWNKKTTSHKGFGS
GNSRKKFLKFINSDKIWNQSTQKSFEEL
>B_CYANOBACTERIA_5 tr|A2C1Q6|A2C1Q6_PROM1 UDP-N-acetylglucosamine 2-epimerase OS=Prochlorococcus marinus (strain NATL1A) GN=wecB PE=3 SV=1
MKRISVIIGTRPEAIKFGPLILAFLKTKEIDLRIISTGQHYELVDQVNELFKIVPNKNLKIMVPGQSLTKI
TNEVLIGLKEDFNEYPPDLVLVQGDTTSAFSAALAAFYEKIPIGHIEAGLRTNQIMLPYPEEANRRIISQ
IASIHFAPTKIAFENLKKESVLGEVYLTGNTVVDSLLFISEKAQIPKIKNVDFIKQKIILATVHRRENWG
ANLKQIAKGLKKILDEHLDYILILPMHPNKSLREPLEEILGVHERAILTESLSYNSLVGTLKHTKLLLT
DSGGLQEEAPTFGVPVLVLRDSTERPEAIKAGTAKIVGSNPNKIFKEANNLLTNQKEYQKMSKAINPF
GDGKASERIVKYCIEFLERNKK
>B_THERMOTOGAE tr|A5IJG4|A5IJG4_THEP1 UDP-N-acetylglucosamine 2-epimerase OS=Thermotoga petrophila (strain RKU-1 / ATCC BAA-488 / DSM 13995) GN=Tpet_0308 PE=3 SV=1
MRVLSLVGARPQIIKEAMLHREFKEKGIEEILVHSGQHYDYNMSDVFFEVLEIRQPHYNLNVGSGTH
GEMTGKIMIEFEKVLLREKPDLVLVYGDTNTTLAGALVAAKLKIPVAHVEAGLRQHPKDMPEEINRI
VTDRVSQILFCPSELAVENLRKEGITEGVYFTCDVMYDLFLKMKPLFRYDLFRTLGLKENGYIVCTIH
RDFNTDVPERLREILEQLRRLSKRYEVVFPVHPRTAKRIREFGLAELLNEILVIEPADYLNMMGLIEKS
RFVITDSGGLQKEAYWCGKRAIVVMPDTGWRELVEAGWNVLSEPDEITEKSEYIDNHVSLPENVYG
EGNASEKLAEVILQL
>B_THERMOTOGAE_2 tr|A5IJF6|A5IJF6_THEP1 UDP-N-acetylglucosamine 2-epimerase OS=Thermotoga petrophila (strain RKU-1 / ATCC BAA-488 / DSM 13995) GN=Tpet_0300 PE=3 SV=1
MKVLSLVGARPQIIKEAVLHRRFKEKGIKEVLVHSGQHYDYNMSDVFFEILQIRKPDYNLNVGSGTH
GEMTGKIMIEFEKVLLKEKPDLVLVYGDTNTTLAGALVAAKLKIPVAHVEAGLRQHPKDMPEEINRI
VTDRVSQILFCPSKLAVKNLEREGITEGVYFVGDVMYDLFLKMEDRFRYDVFEKLGLKENEFILVTL
HRDFNVDDPVKLRKILEQLKRISREKKVVFPIHPRTKNRVKEFGLENLLEGMVVIDPVDYLNLMGLV
KRCWKVVTDSGGLQKEAYFAGKRAIVIMPDTGWRELVEAGWNKLASEENLFDVTMEEDCSEYPSG
LYGDGNAAGKIVEVIEEWFRSLRGIH
>B_THERMOTOGAE_3 tr|A5INE8|A5INE8_THEP1 UDP-N-acetylglucosamine 2-epimerase OS=Thermotoga petrophila (strain RKU-1 / ATCC BAA-488 / DSM 13995) GN=Tpet_1716 PE=3 SV=1
MIRVLSVFGTRPEAIKMAPLVKKLEEEERNVESLVCVTAQHRQMLDQVLEVFDIKPDFDLNIMKERQ
SLADITVNALSGLYDLIEELKPDIVLVQGDTTTTFAGALAAFYHRIPVGHVEAGLRTNDRYSPFPEEIN
RRLTGVLSTLHFAPTKRNRENLLRENVMGKIYVTGNTVIDALRYTVKESHVFENPVLRNMDFSDGR
YILLTSHRRENIGKPLENICKAVRRIVEEFEDVRVIYPVHMNPAVREIVFPMLENMERVFLIDPVNVID
MHNLMARSYLIMTDSGGIQEEAPALGRPVIVLRRETERPEAIEAGVAVLGGVEEERIFELAKKLLLDR
EEYEKMAKAVNPFGDGRASERIVKAILHEFGLSDPPEEFG
```

>B_ACIDOBACTERIA tr|Q1ILE3|Q1ILE3_ACIBL UDP-N-acetylglucosamine 2-epimerase OS=Acidobacteria bacterium (strain Ellin345) GN=Acid345_3306 PE=3 SV=1
MKLVTVVGARPQFIKSGPVSLAIEKHNRENGVTIEEILVHTGQHYDSEMSQVFFEEMNLRTPKYNLE
VGSGNHGEQTAQILARCEKVLMDEKATALMVYGDTNSTLAAALAAVKLHIPVFHVEAGLRSFVRE
MPEEVNRVLTDHISDLLFAPTDTAVENLKAEGITKGVELLGDVMYDAIQQHLQTAKMSSSILNTLALT
PGGYALMTMHRASNTDDPALLGQILSAISEIAKNIRVVWPVHPRARKRMEDFGIHAEGITLISPASYL
DMMMLVSNASLVLTDSGGLQKEACWMRVPCVTLRDETEWVETVASGWNTLAGADRDQILMAAR
KALTSKPSETPGAQHAGASERIAASIRDYMTKRG
>B_ACIDOBACTERIA_2 tr|Q1ISX2|Q1ISX2_ACIBL UDP-N-acetylglucosamine 2-epimerase OS=Acidobacteria bacterium (strain Ellin345) GN=Acid345_1025 PE=3 SV=1
MHFLHVVGARPNFMKAAPLIRALEQRGSRQTLVHSGQHYDRNMSTVFFDQLGIRKPDVNLQVGSG
SHAQQTAAIMSRVEPVLLNQRPDAVIVYGDINSTVAVALVCAKLGIKLIHVEAGLRSFDRSMPEEINR
LVTDQLADVLFTPSLDGDENLHREGIPDNKVHFVGNIMIDTLVRLLPLAELRFADLAAKFNLIKFGLV
TLHRPSNVDDISHLAPLLFALDRIAEDLPLLFPVHPRTLQHMQEFSINLHHLQILEPLPYIDFLSLQQRA
ALVITDSGGIQEETTYLGIPCLTVRENTERPVTVTLGTNLLVGSDFHRMESEARKVIAGNKKCGSIPPL
WDGHTSDRIASILINCGSFPGDPNDVKNSHSLLVDASVSA
>B_ACTINOBACTERIA tr|A7VMX0|A7VMX0_STRKA UDP-N-acetylglucosamine 2-epimerase OS=Streptomyces kasugaensis GN=kasQ PE=3 SV=1
MALRVGIVYGTRPEAIKLAPLVLALDADPGFEPVIITTGQHRDMLDEINELFGLRPRHNLDIMRPGQR
LSAMASRIVGELGDPLLDELVDVAVVQGDTSTAFAAAYAAACERIPVAHLEAGLRTGDRFEPFPEEIN
RRLITQLADLHFAPTADAAGNLLAEGVRSDDVYVTGNTVIDAMHLVLDRPGDSANRELDAFTEGRQ
TVLLTMHRRESWGIPMGRVAAAVAELCRSRPTLRFVIPLHPNPEVRRVFRSHLSSLSQVLLCEPLRYSE
FIRLMHRAVLVLTDSGGVQEEAPTLGKPVLVLRDRTERPEGIAAGCARLVGTDPALIVKEVGRLLDDP
EAYEAMRRPGIFCYGEGDAAARCLEALRERWPSSPDASAELRTYR
>B_ACTINOBACTERIA_2 tr|Q3V7C2|Q3V7C2_STRKA UDP-N-acetylglucosamine 2-epimerase OS=Streptomyces kasugaensis GN=kasQ PE=3 SV=1
MALRVGIVYGTRPEAIKLAPLVLALDADPGFEPVIITTGQHRDMLDEINELFGLRPRHNLDIMRPGQR
LSAMASRIVGELGDPLLDELVDVAVVQGDTSTAFAAAYAAACERIPVAHLEAGLRTGDRFEPFPEEIN
RRLITQLADLHFAPTADAAGNLLAEGVRSDDVYVTGNTVIDAMHLVLDRPGDSANRELDAFTEGRQ
TVLLTMHRRESWGIPMGRVAAAVAELCRSRPTLRFVIPLHPNPEVRRVFRSHLSSLTQVLLCEPLRYSE
FIRLMHRAVLVLTDSGGVQEEAPTLGKPVLVLRDRTERPEGIARAAPGWWAPTRHSSSKRSADCSTT
PRRTRRCGVRASSATARGTPRHGAWRHCASAGCLPPTRRPDSGRTGDRGRPGPAPAGRCRTDRHRP
GCRGTPGGR
>B_ACTINOBACTERIA_3 tr|A8L039|A8L039_FRASN UDP-N-acetylglucosamine 2-epimerase OS=Frankia sp. (strain EAN1pec) GN=Franean1_2163 PE=3 SV=1
MPDQGGLPAAGLHAVVIAGARPNFIKVKPVLDALAADGARSSFVHTGQHYDEAMSDVFFADLGLR
QPDYHLEAGSGSHAVQTAAVMTAFEPLLARLAPDVVVVVGDVNSTLACALVAAKANVPVAHVEAG
LRSGDRTMPEEINRIVTDRVSDLLFAPSPEGVVHLLSEGARPESVHLAGNVMVDTLLACRERARSRP
VLADLGLTPGEYGLVTLHRPSNVDDPAVLDGLLTALGEIARRCPLVFPVHPRTASRLAGRLPDGVRAL
GPAGYLDCVALQMGARLVLTDSGGVQEESTVLGVPCLTLRESTERPITVTEGTNRVVGCSPAAIMAG
AFEVLDRPPPPRCPDLWDGHAGRRISAVLGRTLERRARRGGGSRPRGRAEAVPDPLLPTSPADRGA
>B_ACTINOBACTERIA_4 tr|A3PW76|A3PW76_MYCSJ UDP-N-acetylglucosamine 2-epimerase OS=Mycobacterium sp. (strain JLS) GN=Mjls_1351 PE=3 SV=1
MGGVTEVHLVAGTRPEAIKLAPLVPALRAQGMTPVFVASGQHPTMVHQALDAFGLEPDVTLSIDRG
SGSQAELMAALTMKLEKHWQQRTPAAVVVQGDTTTVLAAAMVAFWAKLPIAHLEAGLRSHDLAA
PFPEEGNRKLVGQISRLHLAPTARARANLEREGTPAADIVVTGNTVIDAVLGIAARGGPVTDSRVAAF
VERARAGASRLVLVTAHRRESWGEPLDRVLNAVALLLEKYSDVEVVLPAHPNPAVAEQVRAVLGAH
PRVLVTEPLAYPVLVGALAASTLVLSDSGGIQEEAPSFGVPVIVLREVTERMEAVDAGCAILVGTDRD
AVLGNACRLLDDPDERTAMVSKGNPFGDGRAAERSAAAIAWMLGRTDGRPAEFRF
>B_ACTINOBACTERIA_5 tr|A8L0M7|A8L0M7_FRASN UDP-N-acetylglucosamine 2-epimerase OS=Frankia sp. (strain EAN1pec) GN=Franean1_5202 PE=3 SV=1

MGVSHRIPAARPRPQTGHTATSPNTTGLDDGGAGGQTVPGQGARGQSIPGQGAGGPGVPGQGAGGP
GVGAQGVGGQNVGGQGAEEQNVADQGAAGESTAGWGPAERGGMTAMTGTAGTTARTGPSVPTAP
STPIAPTGPTVPTVPNRRLPPDRQGAAGRMTMPDRVPRTEIMGVPAPVPRGRPGPAEHRARVMILVG
TRPEIVKLSRIIAALERAVDVCLVHSGQHYDYELNQVFFDELGIRKPDHFLDAVGASAAETIGRVIARS
DAVFVDESPDALLLYGDTNTTLAVIAARRRHIPVFHLEAGNRCFDDRVPEEINRRLVDHLSDINLPLTE
HARRHLLAEGLPAQRIFVTGSPMKEVLDHYAPLVDASPVLTNLGVTAGHFLVVSAHREENVDAPELL
IGLLETLNALAARYRVPIIVSTHPRTRDRLDALEASGRAPATDGLVRFCRPFGFADYIALQRAAQCVIS
DSGSLTEEASLLGFPAVMIREAHERPEGVDHGVAVSCLPRPDRVLAAVDLVVDAAQGDRAPRIVPDY
DVDDVSRRVVRIIVSHIDYVRRTVWFERPPVGTSEPTPGGTSLTLP

>B_ACTINOBACTERIA_6 tr|A8LFA2|A8LFA2_FRASN UDP-N-acetylglucosamine 2-epimerase OS=Frankia sp. (strain EAN1pec) GN=Franean1_0381 PE=3 SV=1

MTDSSDGRTVLLVYGTRPEAIKMAPVVHELGRTGPLRPVVAVTGQHRSMLDQVNDLFGIVPDYDLG
ILRDRQSLAGVTTRSLGGLESVMAEVRPDVVVVQGDTTTAFTGALAAFYQGLPVVHMEAGLRTED
PASPFPEEINRRLTSQLADLHLAPTPSARANLLAEGIRPESVLVTGNTVIDALLHVTAAPPPPDRLLAE
VRQRSGDGRRQVLLVTAHRRESWGEPLARVGAALARLAEQRPELLIVLPVHRNPVVRETVLPPVEG
FPNILVADPVDYAAFAHLMKAATVVLTDSGGIQEEAPSLGKPVLVLRDNTERPEGVQAGTACLVGTE
PNRIVAAVDRLLDDPVAYAAMAEAINPYGDGQAARRTVAAIAHRFSDAPAPDSFVPRPRTITIPLQPAA
PGQELTGSAAGRREESGT

>B_AQUIFICAE tr|A8V4S1|A8V4S1_9AQUI UDP-N-acetylglucosamine 2-epimerase (Fragment) OS=Hydrogenivirga sp. 128-5-R1-1 GN=HG1285_04773 PE=4 SV=1

FTGTRAEYGLLKPLMYEIKKDKDLELQIIASGMHLSPEFGLTYKEIEKDGFTIDEKVEMLLSSDTPFGI
SKSIGLGVIGYTDALERLKPDITVVLGDRFEALAFVIASFTLKIPIAHLYGGETTLGVLDEGYRHAITK
LSYLHFTSTEEYKKRVIQLGEEPERVFNVGALGIDNIKKLKLLSKEEVEKRLGKKLKKRNLLITFHPE
TLKKGSSENQFKELLKALDELEDTLLIFTKANADTEGRKINYLIDEYASKNKEKTVVFTNMGQLLYL
STMQFVDAVVGNSSSGIIEAPSFKIGTINIGDRQKGRIRAESIIDCKPDYESIKKSINKLFSQNFRQKLK
NVNNPYENKNTARKIKNILKSFELDNLEKKFFDIPVRY

>B_ALPHA_5 tr|B9JEP0|B9JEP0_AGRRK UDP-N-acetylglucosamine 2-epimerase OS=Agrobacterium radiobacter (strain K84 / ATCC BAA-868) GN=Arad_4847 PE=3 SV=1

MRHVMVIVGTRPEAVKMAPVIKALYARPEAFRCTVVSTGQHREMLRQTLTSFGLSVDVDLDIMQPD
QTLASLTGAAIGACEKIFQAARPDVVLVQGDTTTVLSASLAAHYAQIPVGHVEAGLRTYERYNPFPE
EMNRKLVTSLATLHFAPTERSARQLRKEGVPPSKVFVTGNTVVDALEELRGRVTTGDVSPAVRERVA
RSQGRFVLVTCHRRESFAHDLNVIVEAIAMLADRFPDRTFFFPVHLNPNVRALVMPRLNGIENVVLA
DPVPYADILFCLSSAELVLTDSGGLQEEAPSFGVPVIVLRRTTERPEGVRAGFSRLVPIEQENIVSLASS
WLRSHRKARLAGRPNPYGDGNAAARIVDILSKEPLR

>B_ALPHA_6 tr|Q11C10|Q11C10_MESSB UDP-N-acetylglucosamine 2-epimerase OS=Mesorhizobium sp. (strain BNC1) GN=Meso_3697 PE=3 SV=1

MTKKVLCIVGTRPEAIKMAPVILALKQQPWAVVHVLATAQHRQMLDQVLNFFGIVPDSDLNVMRP
NQDLTELTARLLVGAGEVLDRERPDVVLVQGDTTTVMAVAIAAFYRRIALGHVEAGLRTGDIRNPFP
EEANRVIAGRLARWHFAPTEGAKQNLLREGVSEDRIFVTGNTVIDALLNTAERRIDSGLNLPEGRRIV
LITAHRRENFGEPIRRICAALKELALRNPDVHLVYPVHPNPNVKEVAETMLAGLQNVQLCAPLDYAP
FVSVMRRSHLIISDSGGVQEEAPALGKPVLVLRDETERPEAVAEGVVKLVGTDHDYIVSQVQLLLDD
ADAYKNMSRGVSPYGDGMASQRVIEILACDTAGRG

>B_ALPHA_7 tr|A8GRP0|A8GRP0_RICRS UDP-N-acetylglucosamine 2-epimerase OS=Rickettsia rickettsii (strain Sheila Smith) GN=A1G_02600 PE=3 SV=1

MLKVMTIVGTRPELIKLCCVISEFDKYTHHILVHTGQNYAYELHQVFFDDMGIRKPDYFLEVAADNT
AKSIGLIIEKVDAVLEKEKPDAVLFYGDTNSCLSAIAAKRRKIPIFHMEAGNRCFDQRVPEEINRKIIDH
ISDVNITLTEHARRYLIFEGLPPELIFKSGSHMPEVLDRFMSKILKSDILDKLSLTTKQYFLISSHREENV
DVKNNLKELLSSLQTLIKEYNFPVIFSTHPRTKKRLEDLEEFKELGDKIRFLPAFGFTDYVKLQMNAF
CILSDSGTITEEASILNLPALNIREAHERPEGMDAGTLIMSGFKAERVLQSVKAITEEHENNKRLQDIV
PDYADAGLVSKKILRIVLSYVDYVNHTVWFKK

>B_ALPHA_8 tr|Q2WB69|Q2WB69_MAGSA UDP-N-acetylglucosamine 2-epimerase OS=Magnetospirillum magneticum (strain AMB-1 / ATCC 700264) GN=amb0102 PE=3 SV=1

MSAMAAPKVLCVVGARPNFMKIGPVIRALKGAGMAAPLVHTGQHYDEAMNDRFFADLGLPRPDVNLEVGSGSHAQQTAEIMRRFEPVLDDLAPQALLVVGDVNSTIACALVAAKKGVTVVHVEAGLRSYDRAMPEEINRVLTDQISDLLFTTERDALANLTREGIDPARVHFVGNVMIDTLVMNRPRAVPATATLEAVGADFRTGYALVTLHRPSNVDDPVVLARLLGGLAEIARQTPLVFPIHPRTRARIEAAGLMDRLDNARTVLLPPAGYLEMLGLMAGARLVITDSGGVQEETTALGVPCITVRENTERPITVSEGTNTVVGTDPALLLAAWRDVLASGGKTGRIPEFWDGKASERIASILYQHLR

>B_DELTA_6 tr|Q6MMF3|Q6MMF3_BDEBA Putative UDP-N-acetylglucosamine-2-epimerase NeuC OS=Bdellovibrio bacteriovorus GN=neuC PE=4 SV=1

MKRRICVLSGTRADYGLLYWLMKELQTYSSVDLQIIATGMHLSPEFGMTYKFIEADGFLINEKVEMLLSGDSVTALCKSMGLGVMGISEALARLSPDLLVVLGDRFEALAGAQAAMIHRIPIAHLHGGEATEGLIDEAIRHSITKMSHLHFVAAEAYRNKVIQLGESPDRVFNFGATGIDNIKRLPLFNREEVEERIPVLKGAKGPVFLVTYHPVTLLNSSPEQAMADLLSAIEEFSDACVILTKSNADTAGRVINEMIDQFCLKHPERAMSSVSLGQKLYLSVMSIANVVVGNSSSGIIEAPAMNIPTVNIGPRQQGRAKAPSVLDCNETKSEIVSGIQTVLENSFMASRKDKHVYGDGNTAVKVAEVLASYDLDGIIFKKFYSIDGVK

>B_DELTA_7 tr|Q39VK1|Q39VK1_GEOMG UDP-N-acetylglucosamine 2-epimerase OS=Geobacter metallireducens (strain GS-15 / ATCC 53774 / DSM 7210) GN=Gmet_1489 PE=3 SV=1

MDKLNVMFVFGTRPEAIKMAPLVKELQKQAGQFKTIVCVSAQHRQMLDQVLQLFDIRPEYDLDIMKPGQDLFDITSNVLLGLKPVLTKEQPDILLVHGDTTTTMSAALAAYYCRIPVGHVEAGLRTYNKFSPFPEEMNRRTAGALTDLHFAPTETARRNLLSEGVADTSIFVTGNTVIDALFSVIGEINANRELKESLSTQFQFLDPHKRLILVTGHRRENFGAGFESICRAIARIAERYPDVEILYPVHLNPNVREPVRRLLGNGKLSNVHLIEPVDYLPFVYLMAQSYLIITDSGGVQEEAPSLGKPVLVMRDTTERPEAIGAGTVKLVGTREESIYREAARLLDAADAYQVMSQALNPYGDGRAAERIAKCLADFPNTCMGGQLA

>B_DELTA_8 tr|Q6MME0|Q6MME0_BDEBA UDP-N-acetylglucosamine 2-epimerase OS=Bdellovibrio bacteriovorus GN=capG PE=3 SV=1

MKKMKVMTVLGTRPEIIRLSRVLAKLDQHCEHKIVHTGQNYDFELNEIFFQDLGVRKPDFFLNAAGESATETIGNIIKTVDKVMGEFLPEAVLVLGDTNSCLSVIAAKRRKVPIFHMEAGNRCFDQRVPEETNRKIVDHTADINLTYSSIAREYLLREGLPPDRVIKTGSPMFEVLTHYRKQIDSSDVLSRLGLKSGQYFVVSAHREENIESDVNFDKLVESLNLVAETFDIPVVVSTHPRTQNRIDKRGSKFHSNVRLLKPLGFSDYNHLQLHARAVLSDSGTITEESSIMNFPALNIREAHERPEGFEEASVMMVGLDKIRIMQALRILETQPRADDRLLRQVADYSMPNVSDKVVRIIQSYTDYVNRVVWRK

>B_DELTA_9 tr|Q39VI6|Q39VI6_GEOMG UDP-N-acetylglucosamine 2-epimerase OS=Geobacter metallireducens (strain GS-15 / ATCC 53774 / DSM 7210) GN=Gmet_1504 PE=3 SV=1

MNVFLVAGARPNFMKIAPLYRESLTREGVTCRIVHTGQHYDYEMSQTFFDDLQLPQPDYFLNAGSGSHAEQTARVMVTFEELCAKELPDLVIVVGDVNSTLACSIVAKKAGIRVAHVEAGLRSFDLSMPEEINRMVTDSIADLYFVTETSAVRNLFNEGKPPEQIFLVGHVMVDNLLHQMVQVHLNGAVAPELRELKQKAGEYVFMTLHRPSNVDDEETFRGIAFAVNEIAARIPIIFPVHPRTRKMMESFGIQLHKGVTCLPPLSFSESLYFWKDARVVMTDSGGLQEETTALGVPCVTIRENTERPITVEMGTNVLAGTDPDRILTRVFDAAEGRGKRGKVPPFWDGKAAERIWDVIGRKYGKVPDLSEACHELMECVGS

>B_DELTA_10 tr|A7HA71|A7HA71_ANADF UDP-N-acetylglucosamine 2-epimerase OS=Anaeromyxobacter sp. (strain Fw109-5) GN=Anae109_1410 PE=3 SV=1

MRILIVFGTRPEAVKLAPLIHDLRQRPGADVEVCLTGQHREMVTQVVSFFGVDVDHDLEIMRPNQTLSDVAARTLTGVDRILESRRPDWVIVQGDTSTCLATALAAFHRKVRVAHVEAGLRSGDPHAPFPEEMNRVLTTPIASLHLAPTSRAKANLRAERVPEDRIRVVGNTGIDALLLAVQTLRDRGLDATYGARFPFLRPSRPLVLVTGHRRESFGQPFEELCEAIRDVATGDDVDVVYPVHLNPNVREPVFRILSGLSNVHLVEPVEYPALVWLAQRSRFILTDSGGIQEEASALGKPVLVMRDVTERQESVEAGVSRLVGTSREVIREACNSLLRDEATYARMARRVDLYGDGKASARIGDALGLPPSGASREWAAA

>B_DELTA_11 tr|Q30W97|Q30W97_DESDG UDP-N-acetylglucosamine 2-epimerase OS=Desulfovibrio desulfuricans (strain G20) GN=Dde_3255 PE=4 SV=1

MAKKKILFITGTRADFGKLCSLIDKVSEHTEFEYCIFVTGMHTLSRYGYTVDEVMKKYSSFRLEGGFRNVHVFMNQVHGESMDMVLGNTIFGLSRYVSEYQPDMIVVHGDRVEALAGSIVGSLRNILVAHVEGGELSGTIDELIRHAVSKMAHLHFVANGDSRRRLMQMGEAEETIYPIGSPDVDLMFSPNLPSKESVLAYYNIPFTEYAITLLHPVTTDLEQTKKMAGAVVDAMLDSDDNYIVIYPNNDHGSDIILDEYDRLSGEGRIAVYPSLRVESFLVLLRAAKYLLGNSSAGIRETPCYGVPSINIGSRQDGRFCCSSIINVPGSYDAIVKALADVRKMVPHPPRFEFGRGNSAEQFIKILMDEKTWLTKPQKQFVDHHFALCR

```
>B_ZETA tr|Q0EZA1|Q0EZA1_9PROT UDP-N-acetylglucosamine 2-epimerase OS=Mariprofundus ferrooxydans PV-1 GN=SPV1_07506 PE=3 SV=1
MSDIKRMKVVTVVGTRPEIIRLSRVIACLDRSVDHKLVHTGQNYDYELNEIFFQEMGIRKPDYFLNA
VGASPVQTIGRILETIDPVLEQEAPDAMLLLGDTNSCLAVIAAKKRKIPIFHMEAGNRCFDQRVPEES
NRKLVDHLSDINMPYSSHAREYLLREGLPADRIIKTGSPMHEVLHHYSGNIENSSVLSQLQLSPGDYF
VVSCHREENVDSDSNIRRFVHVLNSLAEKYDRKIIVSTHPRTRARLESLGVELHSHVALLKPLGFFDY
VHLQMHAYAVLSDSGTITEESSILNFPALNIREAHERPEGMEEGSVMMVGLDLAQILQALSILETQQR
GEQRNLRQVADYAPLNVSEKVVRIIVSYTDYINRNVWKKHL
>B_EPSILON tr|Q7X522|Q7X522_CAMCO PtmD OS=Campylobacter coli GN=ptmD PE=4 SV=1
MSKRKICIVSATRAEWYLLRNLCLEIQNDKDLELQIIATGAHLSPEFGLTYKEIEKEFKIAKKIPILLAS
DDKISLCKAMALANIGFSEAFESLKPDIVVILGDRYEMLSVASVCLMMNLPIAHLCGGELTLGAIDDS
IRHSISKMAHLHFVSTQTYKNRLLQLGEEKNRVFNVGSLASTIMKNINFLDKNELAKVLEFDLGQNI
YLITYHPVTLNLKNSKKEIDILLKTLDELNNASLIFTKANADENGLLINEILKDYCEKNPKKAKLFDN
LGSQKYLSIMKIAKAMIGNSSSGISESPFFKTPCINIGSRQKGRLRTQNIIDCEFENLDQAFEKLESKDF
KQALHHFKNPYENDKNPNIIIKTCLKNANLDTILQKNFIDLRQNLGYK
>B_EPSILON_2 tr|A6DDY8|A6DDY8_9PROT UDP-N-acetylglucosamine 2-epimerase OS=Caminibacter mediatlanticus TB-2 GN=CMTB2_04387 PE=4 SV=1
MKNICVVTSTRAEYGLLYWLMKEIENDKKLQLQLIVTGSHLSPEFGMTIKEIEKEFKIDKKIEMLLSS
DSKIGISKSMGLLQISIAEAFEELKPDFLVVLGDRYELIPIVSTAVVFNIPVVHLHGGEITQGAIDEYFR
HAVTKMSYLHFTSTEEYKKRVIQMGEEPNRVFNVGAFGVENIFRLKLLNKKEFEDSINFKLNKKNLL
ITYHPETLNTQNTKKDFKEILDVLDELKDTNFIFTKANSDTNGRIINKMIDEYVSKNGNKAISFDSLG
QIRYLSALKHVDAVIGNSSSGIIEAPSFKTATINIGNRQKGRIKAKSIIDCEPKKESIKKAFEKLYSKEFQ
ELLKSVKNPYQKEKYPSKKVVDVLKNFEVKSSKNFYDMECKGIK
>B_GAMMA sp|P52642|RFBC_SALBO UDP-N-acetylglucosamine 2-epimerase OS=Salmonella borreze GN=rfbC PE=3 SV=1
MSKVLFVFGTRPEAIKMAPLVIEFKNNPAIEVKVCVTGQHREMLDQVLDFFEIEPDYDLNIMKQKQS
LGSITCSILTRLDEILASFMPAHIFVHGDTTTTFAASLAAFYQNIKVWHIEAGLRTWNMNSPFPEEGNR
QLTSKLAFFHAAPTLQAKDNLLRESVKEKNIIVTGNTVIDALLIGIKKITGSTGDVREIISLKNKLNLD
KKIILVTLHRRENQGELLRTICDDIKQLALEHDDIEIVFPVHMSPRIREVVNEKLSGVVNIKLVEPLAYP
GFIWLMNNAHFILSDSGGVQEEAPSLQKPVLVARDTTERPEVIENGAAMLVDPRIPNNIYSSCKKLLS
DERLYEKMSQAGNPFGDGKASKKILDYFVSLEDIK
>B_GAMMA_2 tr|A6A3Z4|A6A3Z4_VIBCH UDP-N-acetylglucosamine 2-epimerase OS=Vibrio cholerae MZO-2 GN=wecB PE=3 SV=1
MKKVLIVFGTRPEAIKMAPLVQQLCQDNRFVAKVCVTGQHREMLDQVLELFSITPDFDLNIMEPGQ
TLNGVTSKILLGMQQVLSSEQPDVVLVHGDTATTFAASLAAYYQQIPVGHVEAGLRTGNIYSPWPEE
GNRKLTAALTQYHFAPTDTSRANLLQENYNAENIFVTGNTVIDALLAVREKIHTDMDLQATLESQFP
MLDASKKLILVTGHRRESFGGGFERICQALITTAEQHPECQILYPVHLNPNVREPVNKLLKGVSNIVLI
EPQQYLPFVYLMDRAHIILTDSGGIQEEAPSLGKPVLVMRETTERPEAVAAGTVKLVGTNQQQICDA
LSLLLTDPQAYQAMSQAHNPYGDGKACQRIADILAK
>B_GAMMA_3 tr|Q9L9K6|Q9L9K6_PASMU UDP-N-acetylglucosamine 2-epimerase OS=Pasteurella multocida GN=bcbA PE=3 SV=1
MFRCNNRFLVVFGTRPEAIKMAPLVEGLKKNNLNFRICVTAQHREMLDQVLELFNIIPDYDLDIMNQ
TQTLSTVTSSILEKIQPVIDDYKPNVIFVHGDTATTLAASLAAYYNQIDIAHIEAGLRTNNIYSPWPEEG
NRKLTAALAKYHFTPTQATKENLLRENIDPSKIYITGNTVIDALFLVNKKIENNLTLLQKYKRNFSDFL
GKKVILITGHRRENFGDGFENICAAISSLAELHPDVQFVYPVHLNPNVREPVNRLLKNKNNIHLIQPL
DYFSFIFLMKNAYLILTDSGGIQEEAPSLGKPVLVMRKTTERLEAVQAETVKLVGTTQELIIDSVNELL
TDIEAYNKMSKAHNPYGDGRAVERILNVFKI
>B_GAMMA_4 tr|A6A5R6|A6A5R6_VIBCH UDP-N-acetylglucosamine 2-epimerase OS=Vibrio cholerae MZO-2 GN=A5A_0266 PE=3 SV=1
MKVLTIFGTRPEAIKMAPLVKVLATDISIESKICVTAQHREMLDQVLALFEIVPDYDLNLMRPGQDLF
DVTSGVLLGLRDVLNDFKPDLVLVHGDTATCLGATLAAFYMQIPVGHVEAGLRTGNIYSPWPEEAN
RVLVSKLATWHFAPTQRNKESLVQEGINPERIFVTGNTVIDALQWVVKKIQHSEEIKKQVHESLTASG
LTSLNILDERFVLITGHRRENFGSGFENICSALKKLAQSNPETHFIYPVHLNPNVQTPVNRLLGKLNN
VHLIKPLGYEAFVYLMQHSYLVLTDSGGIQEEAPGLGKPVLVMRDTTERPEAVEAGTVKLVGTSVDS
IVRELQALLNNESKYHNMALAHNPYGDGNACDRVLAVIRGQAK
```

```
>B_GAMMA_5 tr|A6A4N6|A6A4N6_VIBCH UDP-N-acetylglucosamine 2-epimerase OS=Vibrio cholerae MZO-2 GN=A5A_0249 PE=3 SV=1
MIKVLSVFGTRPEAIKMAPVIETLKNDPRFDSRVCVTGQHRQMLDQVLELFEIVPNYDLNIMKPGQD
LTDVTTGILQGLRDVFSQFKPDYVLVHGDTATTLSTTIAAYYHQVKVGHVEAGLRTNNIYSPWPEEG
NRKVTGSLANLHFAPTSTSQNNLLAENIPADTIVVTGNTVIDALFMVRDKLNNDPELQSCFVKQFDF
LKTGRRVVLITGHRRESFGGGFERICQAVSELATKFIDVDFVYPVHLNPNVREPVNRFLSGQPNIFLIE
PQDYLPFVYLMDRSDIILTDSGGIQEEAPSLGKPVLVMRDTTERPEAVEAGTVKLVGTDVSRIINEVSI
LLTDKQAYKKMSVAHNPYGDGAASQRILDAIARKLN
>B_GAMMA_6 tr|Q9AHM7|Q9AHM7_PASMU UDP-N-acetylglucosamine 2-epimerase OS=Pasteurella multocida GN=ecbA PE=3 SV=1
MFRCNNRFLVVFGTRPEAIKMAPLVEGLKKNNLNFRICVTAQHREMLDQVLELFDIIPDYDLDIMNQ
TQTLSTVTSSILEKIQPVIDDYKPNVIFVHGDTATTLAASLAAYYNQIDIAHIEAGLRTNNIYSPWPEEG
NRKLTAALAKYHFTPTQATKENLLKENIDPSKIYITGNTVIDALFLANRKIENNLTLLQKYERNFSAFL
GKKIVLITGHRRENFGDGFENICAAISSLAELHPDVQFVYPVHLNPNVREPVNRLLKHKKNIHLIQPL
DYFSFIFLMKNAYLILTDSGGIQEEAPSLGKPVLVMRKTTERPEAVQAGTVKLVGTTQELIIDSVNELL
TDIEAYNKMSKAHNPYGDGRAVERTLNVFKI
>B_GAMMA_7 tr|A1S831|A1S831_SHEAM UDP-N-acetylglucosamine 2-epimerase OS=Shewanella amazonensis (strain ATCC BAA-1098 / SB2B) GN=Sama_2332 PE=4 SV=1
MRKISVVTATRAEYGLLRGLLEDINAASELELQLIVTGTHLSPEFGLTLRQIEEDGFCINKKVEILLSSD
TAVGVSKSMGLALISFAEVFDELKPDILVVLGDRYELIPIVSAANIARIPVAHLSGGELTEGAIDELIRH
SVTKMSQLHFTAMEEYSRRVVQMGELPSRVFTVGEIGLDNLKRIQLSSKVEFENSIHRKLMPKNLLIT
YHPETTQEVEKVRNDFELILKALDQLDNTLLIFTNANSDVGGRSVNMMIDEYVDSHPEKSIAFTSLG
VKRYLSALQYVDAVVGNSSSGIVEAPSFRIATINIGERQKGRVRASSVVDVDVDSEQILKALEAIYSD
DFKAHLKETVNPYGKGDSVQKVIAVLKSVDLDSLKRKPFYDLTIG
>B_GAMMA_8 tr|Q9CM35|Q9CM35_PASMU UDP-N-acetylglucosamine 2-epimerase OS=Pasteurella multocida GN=wbjD PE=3 SV=1
MKKLKVMTVVGTRPEIIRLSRVIVACDKHFQHILVHTGQNYDYELNEVFFHDLGIRKPDYFLNAAGK
TSSATIGNVIIAVDQLLEETQPEALLVLGDTNSCLAVLPAKRRKIPTFHMEAGNRCFDMRVPEEINRRI
VDHTADINLTYSTIARDYLLAEGLSPDLIIKTGSPMFEVLHFYHEQIKHSKILDKLGLTPFQYFVVSAH
REENINSEPHFLDLIHTLNAVAKKYQYPVIVSTHPRTRKRIEELNITFHPLIQLLKPLGFLDYNKLQLM
AKATLSDSGTINEESSILNFPALNLRQAHERPEGMEEAAVMLVGLQAERVLQGLAILDEQRRGQERD
LRLVADYSMPNVSAKVIRILLSYTDYVNRVIWKKY
>B_GAMMA_9 tr|A5IGF4|A5IGF4_LEGPC N-acylglucosamine 2-epimerase OS=Legionella pneumophila (strain Corby) GN=neuC PE=4 SV=1
MIRKIIYVTGTRADYGLMREVLKRLHQSEDIDLSICVTGMHLDALYGNTVNEIKADQFSICGIIPVDL
ANAQHSSMAKAIGHELLGFTEVFESETPDVVLLLGDRGEMLAAAIAAIHLNIPVVHLHGGERSGTV
DEMVRHAISKLSHYHFVATEASKQRLIRMGEKEETIFQVGAPGLDEIMQYKTSTRDVFNQRYGFHPD
KKICLLIYHPVVQEVDSIKIQFQSVIQAALATNLQIICLEPNSDTGGHLIREVIQEYIDHPDVRIIKHLHR
PEFIDCLANSDVMLGNSSSGIIEAASFNLNVVNVGSRQNLRERSDNVIDVDVTYDAILTGLREALNKP
KIKYFNCYGDGKTSERCYQLLKTIPLHSQILNKCNAY
>B_GAMMA_10 tr|Q48ML9|Q48ML9_PSE14 UDP-N-acetylglucosamine 2-epimerase OS=Pseudomonas syringae pv. phaseolicola (strain 1448A / Race 6) GN=PSPPH_1083 PE=3 SV=1
MAGQQPRMKVLSIFGTRPEAIKMAPLVRALAAEPGIDSRICITGQHQSMLQQVLDMFELKADYSLD
VMRPDQTLNSLTAALYAAIDPILDEMKPDKVLVHGDTTSAMVAAMSAFHRRIPIGHVEAGLRTGDIR
QPWPEEMNRRCIDLISDHLFAPTAESRRNVLGERLQGISFVTGNTVIDALHLTAQRIDSNRQLRHALD
RQFSFLVPERRVLVVTGHRRENFGDGFLNICKALGELARRDDIQIVYPVHLNPNVLGPVTEHLGDLP
NVHLIKPLDYLSFVRLMQRAHVILTDSGGVQEEAPSLGKPVLVMRDVTERPEAVAAGTVRLVGTET
DAIIRGVNALFDDDALWQRASHAANPYGDGKASARIVDALMGRPVDEFVAELPRYRTDPVDVQLD
TLIQPQPQHQHNQMRSMAG
>B_GAMMA_11 tr|A1S7W0|A1S7W0_SHEAM UDP-N-acetylglucosamine 2-epimerase OS=Shewanella amazonensis (strain ATCC BAA-1098 / SB2B) GN=Sama_2261 PE=3 SV=1
```

```
MAKKKVLTVFGTRPEAIKMAPLVNALASDDRFEAKCCVTAQHREMLDQVLELFEITPDYDLDLMK
AGQTLNEVTAGIIQKLKPVLQEFKPDVVLVHGDTATTFAASLAAYYEQIAVGHVEAGLRTGNIYSPW
PEEANRRLTGVLTKYHFAPTETSKQNLLKENFEPSNIIVTGNTVIDALLMVKEKIDSDTDLNHTIAEQF
PFIDENKKLILVTGHRRESFGGGFERICEALATVARMHSDVQILYPVHLNPNVREPVNRILGDIDNIFLI
EPQQYLPFIYLMDKAYIILTDSGGIQEEAPSLGKPVLVMRDATERPEAVEAGTVKLVGTDVKRIVSSL
NELLCDDLAYQHMSVAHNPYGDGKACHRILKALI
>B_BETA tr|Q79DD9|Q79DD9_NEIME GlcNAc-6-P epimerase OS=Neisseria meningitidis GN=synX
PE=4 SV=1
MKRILCITGTRADFGKLKPLLAYIENHPDLELHLIVTGMHMMKTYGRTYKEVTRENYQHTYLFSNQI
QGEPMGAVLGNTITFISRLSDEIEPDMVMIHGDRLEALAGAAVGALSSRLVCHIEGGELSGTVDDSIR
HSISKLSHIHLVANEQAVTRLVQMGEKRKHIHIIGSPDLDVMASSTLPSLEEVKEYYGLPYENYGISMF
HPVTTEAHLMPQYAAQYFKALELSGQNIISIYPNNDTGTESILQELLKYQSDKFIAFPSIRFEYFLVLL
KHAKFMVGNSSAGIREAPLYGVPSIDVGTRQSNRHMGKSIIHTDYETKNIFDAIQQACSLGKFEADD
TFNGGDTRTSTERFAEVINNPETWNVSAQKRFIDLNL
>B_BETA_2 tr|Q79DN7|Q79DN7_NEIME SiaA OS=Neisseria meningitidis GN=siaA PE=4 SV=1
MKRILCITGTRADFGKLKPLLAYIENHPDLELHLIVTGMHMMKTYGRTYKEVTRENYQHTYLFSNQI
QGEPMGAVLGNTITFISRLSDEIEPDMVMIHGDRLEALAGAAVGALSSRLVCHIEGGELSGTVDDSIR
HSISKLSHIHLVANEQAVTRLVQMGEKRKHIHIIGSPDLDVMASSTLPSLEEVKEYYGLPYENYGISMF
HPVTTEAHLMPQYAAQYFKALELSGQNIISIYPNNDTGTESILQELLKYQSDKFIAFPSIRFEYFLVLL
KHAKFMVGNSSAGIREAPLYGVPSIDVGTRQNNRHMGKSIIHTDYETKNIFDAIQQACSLGKFEADD
TFNGGDTRTSTERFAEVINNPETWNVSAQKRFIDLNL
>B_BETA_3 tr|A8V8Z2|A8V8Z2_NEIME SynX OS=Neisseria meningitidis GN=synX PE=4 SV=1
MKRILCITGTRADFGKLKPLLAYIENHPDLELHLIVTGMHMMKTYGRTYKEVTRENYQHTYLFSNQI
QGEPMGAVLGNTITLISRLSDEIEPDMVMIHGDRLEALAGATVGALSSRLVCHIEGGELSGTVDDSIR
HSISKLSHIHLVANEQAVTRLVQMGEKRKHIHIIGSPDLDVMASSTLPSLEEVKEYYGLPYENYGISMF
HPVTTEAHLMPQYAAQYFKALELSGQNIISIYPNNDTGTESILQELLKYQSDKFIAFPSIRFEYFLVLL
KHAKFMVGNSSAGIREAPLYGVPSINVGTRQSNRHMGKSIIHTDYETKNIFDAIQQACSLGKFEADD
TFNGGDTRTSTERFAEVINNPETWNVSAQKRFIDLNL
>B_BETA_4 tr|Q82SQ8|Q82SQ8_NITEU UDP-N-acetylglucosamine 2-epimerase OS=Nitrosomonas
europaea GN=wbpI PE=3 SV=1
MQKKVYLIAGARPNFMKIAPIVRALQQQDRLTYKIVHTGQHYDREMNDVFFEELGIPQPDIFMAAG
GGSHSQQTAKIMVAFEEYCQTEPPDAVLVVGDVNSTLACSIVAKKLHIPVAHVEAGLRSGDMCMPEE
INRLVTDSITDWFFVTEPSGQQHLLQEGKPASAIHYVGHVMVDNLLYQVEQLARADTSNFETSQLKT
SLGNQSYGVVTLHRPSNVDSPDALERISLTLKQIAKRLPLVFPAHPRTQNNLKKFNIDLGPNILLMGP
QAYMPFLHLWKDATLVLTDSGGLQEETTALGVPCITIRENTERPITVNEGTNILAGTRPERILAAVDDI
LGGRGKQGRRPHLWDGNAARRIVEILTRELHP
>B_BETA_5 tr|Q82X14|Q82X14_NITEU UDP-N-acetylglucosamine 2-epimerase OS=Nitrosomonas
europaea GN=rffE,wecB,nfrC PE=3 SV=1
MKILIVVGTRPEAIKMAPVILALKKEPWANVRVLATAQHRNMLDQVNEFFGIDPDIDLNIMRPNQAL
TTLTARLLPELDDVLQAEKPDAVLVQGDTTTVMTVALACFYHHIPVGHVEAGLRTWDMQNPFPEEA
NRVITGKFARWHFAPTEGSRQNLLKEGVADSKIIVTGNTVIDALLMSASKDLQLGIELDSNKRLVLVT
SHRRENFGEPFRNICRALQTLAEKNPDVQFLYPVHPNPNVKDVAHEFLAGLHNFTLCEPLDYASFIAA
MKRAYIILTDSGGIQEEAPALGKPVLVLRDETERPEAVEQGVVKLVGPNYDAIVQETQCLLDDEFAYR
AMARGISPYGDGKAAERIVQVLRKYFT
>B_BETA_6 tr|Q63IQ3|Q63IQ3_BURPS UDP-N-acetylglucosamine 2-epimerase OS=Burkholderia
pseudomallei GN=wecB PE=3 SV=1
MKKILLAFGTRPEAIKMAPLVRMLQTRADVDARVCVTAQHRQMLDQVLALFDIAPAYDLNVMRQS
QTLADVTTGILQTIGAVFDDFDPDIVLVHGDTTTTLAVSLAAFYRYLPIGHVEAGLRSGDIWSPWPEE
LNRRVTDAVSSWHFAPTERAQHNLFSEGVPTQGVVLTGNTVIDALQDVKRMLDADAPLAREIAAQF
PFLGHDERVVLITGHRRESFGEPFAHFCDALRTLARRYPGVRFVYPLHLNPNVQGPAHALLDGLPNV
HLIAPQEYLSFVFLMSRAHFIITDSGGIQEEGPALGKPVLVTRDTTERPEAIQAGTARLVGTDTERIVG
EASRLLDDDDAYDEMSRATNPYGDGHASERIAHALLNMPHAADTARFPLAAIEAPCNALALGLRAL
RSA
>B_SPIROCHAETES tr|Q8VTY7|Q8VTY7_LEPIN UDP-N-acetylglucosamine 2-epimerase
OS=Leptospira interrogans PE=3 SV=1
```

```
MKKLKVSTIIGTRPEIIRLSRVLAKLDQYCDHIMIHTGQNYDYELNEIFFNDLEIRKPDYFLNAAGTSG
AATIGNVIIKVDELLAQVQPDAVLVLGDTNSCMAVIPAKRRKIPIFHMEAGNRCFDQRVPEEINRRIVD
HTADINLTYSNIAREYLLREGLPSDMVIKTGSPMFEVLNYYLEGIKKSDILEKLKISEGKYFVVSAHR
EENIDSDKNFAKLIDIINTIAEVFKFPVIISTHPRTQKKINVSNVSFNPLVQLLKPLGFKDYNKLQLSAK
AVLSDSGTITEESSILNFPALNIREAHERPEGMEEASVMMVGLEKERVLQVLQILEKQPKGEGRLLRN
VSDYSMPNVSEKVVRIIHSYTDYVNRVIWKKY
>B_SPIROCHAETES_2 tr|Q9AEF1|Q9AEF1_LEPIN UDP-N-acetylglucosamine 2-epimerase
OS=Leptospira interrogans PE=3 SV=1
MKKLKVSTIIGTRPEIIRLSRVLAKLDQYCDHIMIHTGQNYDYELNEIFFNDLEIRKPDYFLNAAGTSG
AATIGNVIIKVDELLAQVQPDAVLVLGDTNSCMAVIPAKRRKIPIFHMEAGNRCFDQRVPEEINRRIVD
HTADINLTYSNIAREYLLREGLPSDMVIKTGSPMFEVLNYYLEGIKKSDILEKLKISEGKYFVVSAHR
EENIDSDKNFAKLIDIINTIAEVFKFPVIISTHPRTQKKINVSNASFNPLVQLLKPLGFKDYNKLQLSAK
AVLSDSGTITEESSILNFPALNIREAHERPEGMEEASVMMVGLEKERVLQVLQILEKQPKGEGRLLRN
VSDYSMPNVSEKVVRIIHSYTDYVNRVIWKKY
>B_SPIROCHAETES_3 tr|Q9S4G9|Q9S4G9_LEPIN UDP-N-acetylglucosamine 2-epimerase
OS=Leptospira interrogans PE=3 SV=1
MFKVMTIVGTRPELIKMSRVIAELDRNFKHILVHSGQNYDYELNQVFFEDLEIRKPDHFLNAAGESA
AATIAQVLLKADEVFEKEKPDALLLYGDTNTCLAVISAKRRKIPIFHMEAGNRCFDQRVPEELNRKV
VDHLSDINLVLTEHARRYLLAEGIKPETIIKTGSHMDEVLKYYKAKIEKSDILEKIKLEKNKFFIVSSH
REENVNTVENLKKLLESLNAICEEYGLPVIVSTHPRTRKRLEDFKEVEMNPLITFLKPFGFFDYVQLQ
MSAFCILSDSGTITEEASLLDLPAVTIRNTHERPEGMDVGTLIMSGLSKERILESVRIVVDQYRAGIRH
YRVVDDYSAGQVSKKIVSIVQSYTDYINRTVWHKS
>B_CHLOROBI tr|B3EI03|B3EI03_CHLL2 UDP-N-acetylglucosamine 2-epimerase OS=Chlorobium
limicola (strain DSM 245 / NBRC 103803) GN=Clim_2390 PE=3 SV=1
MKKLLLAAGGKPGALLLAPLYNALKKNGIFKPVAVFAAAQSAEPLSRELATCFGIGEAGHTITLGED
SPGQQLAAVITGMEPIIAREQPVLVMVCGSDNAALGAALAATKLGVPVAVADAGLRCHDRSDADEIN
RILIDSIADLHFISEHSGEYNLINEGVADEKVFFSGNLSIDTLAVLMEQAERDTIAAELGLQPKKYALV
LLNPAGAIAAREPLEMTLRLLKKMSGLTTVLVPLAAGLDELMKQHKLDKAFGGVEGLVMIAHPGH
AGLLTLLRDSVLLITDSEELQAESTVMNVPCLTMMDNTARPATIEIGTNVLVGVDEEDIMSRIHDILHS
GEHAHSSKRNKIPEKWDGAAASRIVEVLGRVL
>B_SPIROCHAETES_4 tr|Q8F5M5|Q8F5M5_LEPIN UDP-N-acetylglucosamine 2-epimerase
OS=Leptospira interrogans GN=wecB PE=3 SV=1
MKKLKVSTIIGTRPEIIRLSRVLAKLDQYCDHIMIHTGQNYDYELNEIFFNDLEIRKPDYFLNAAGTSG
AATIGNVIIKVDELLAQVQPDAVLVLGDTNSCMAVIPAKRRKIPIFHMEAGNRCFDQRVPEEINRRIVD
HTADINLTYSNIAREYLLREGLPSDMVIKTGSPMFEVLNYYLEGIKKSDILEKLKISEGKYFVVSAHR
EENIDSDKNFAKLIDIINTIAEVFKFPVIISTHPRTQKKINVSNASFNPLVQLLKPLGFKDYNKLQLSAK
AVLSDSGTITEESSILNFPALNIREAHERPEGMEEASVMMVGLEKERVLQVLQILEKQPKDEGRLLRN
VSDYSMPNVSEKVVRIIHSYTDYVNRVIWKKY
>B_SPIROCHAETES_5 tr|Q8F5Q7|Q8F5Q7_LEPIN UDP-N-acetylglucosamine 2-epimerase
OS=Leptospira interrogans GN=LA1610 PE=4 SV=1
MKRKICVITGTRADYGLLRWLILEISKSSKLDLQIIATGMHLSPEFGLTYKEIENDGFLIHKKIEILLSSD
STVGVSKSIGLGLIGFSEAFADLNPDIILVLGDRFEILAATVAALISRIPVAHLHGGETTEGAFDEGIRHS
ITKMSHLHFVAANEYRNRIIQLGEDPSNVFLVGGMGIDGIKKSNLLKKEELESSLNIKFKKNNLLITFH
PVTLEESTAKVQMIELLLSLETLSPEIGLIFTMPNADTDGRIIFELVKEFTSSHSNAWYFTSLGQTRYLS
CLQFVDAVVGNSSSGIIEAPSFKIGTINIGDRQKGRLRAKSIIDCEPKKIEIIDAFKRLYSSDFQKNLSTT
VNPYGEGGASEKIVRILEKIDIQGILKKKFFNLQF
>A_CRENARCHAEOTA tr|A1S0Y0|A1S0Y0_THEPD UDP-N-acetylglucosamine 2-epimerase
OS=Thermofilum pendens (strain Hrk 5) GN=Tpen_1715 PE=4 SV=1
MTKVVSIVGARPNFVKLAAVAETFDREFEHTVIHTGQHYDYEMSKVFFEQLRLRDPDIHLGVGSGS
QGYQVGEIVKKAEEHLKSINPDVVVVYGDTNSTLAGALAAAKAGYPVAHVEAGLRSHDMKMQEEI
NRRVVDHVSALLFAPTPSARDNLLKENVPGKVFLTGDVHVDVLQKWIESIQRSNILGRLGLLEGEYV
MATIHRAENVDDPAKLAEVIKCLKELAALTKVVFPIHPRTRSRAEEAGLEKELYSNPNLVATKPLGYV
DFLRLVTGSKLVVTDSGGVQREAYLLGKPAIVLRDRTEWIELVQAEWVRLADANPARLVQEYKKLL
EKPPAPQPGLLGDGKAAERITKTISELMRR
```

```
>A_KORARCHAEOTA tr|B1L6H3|B1L6H3_KORCO UDP-N-acetylglucosamine 2-epimerase OS=Korarchaeum cryptofilum (strain OPF8) GN=Kcr_1306 PE=4 SV=1
MRLIISVGTRPEIIKMAPLYEALSKLNEIELLLVHTGQHYDWEMSGIFFKELGVEEPDINLNIGSDDQV
SQTSAVMKEIGKIIESYEPDGVIAVGDTNSVLGTAIAASKMEVPFIHVESGLRSYDFSMPEEINRRISDH
LASLNFAPTSRAFSNLMEEGISPKTTFLSGNTIVDSVIKVLRKVSAKRTLEKFDLDGAQPLVTLTLHRK
ENTEYEHKLIGVLKALEELDDITFVWPIHPRTQKALKTFDLWNKLKSLKNVRILEPLGYLDFLGLLIS
SDLVMTDSGGVQEEAATLKRPCIILRENTERPEIIEMGFGEIAGTNPTAIISLVRKYLYQANLIERLKRT
PNPFGDGSSSKIIASVIQRIWDLKSLRRPPTFKGGSPHYLAFRVDESMRGYTVASFREACGYDVVSIY
DASGRAIHFDEGTPLIPNYIVRVRGDPVNYGRFKKLVKI
>A_EURYARCHAEOTA tr|Q46GF0|Q46GF0_METBF UDP-N-acetylglucosamine 2-epimerase OS=Methanosarcina barkeri (strain Fusaro / DSM 804) GN=Mbar_A0052 PE=4 SV=1
MRIASIVGVRPQFVKASVVSRELRKKNEEILIHTGQHYDYQMNKLFFEELNIPEPEYHLDIGSGSHGY
QTGEMLKKIEEVLIKEKPDLVLTYGDTNSTLAGALAASKLHIKTAHVESGLRSFDKAMPEEINRIVTD
HCSDILFCPTENAVENLKNEGITKNVYLTGDVMVDSLLYNRKIAEDKSTILNDLGLQSKGYLVATIHR
ASNTDNEENLKNIVNAFSELKETIVFPLHPRTDKFLKKYGLYDRLKSSVILIEPLGFFEFIKLMNNAK
MILTDSGGVQKEAYVLKVPCITLRENTEWIETVNDEWNVLVGTNKEKIVKMVKEFRPFLEKHRDRF
GNGDASKNILLIIDCLKKQEVNL
>A_EURYARCHAEOTA_2 tr|Q9V179|Q9V179_PYRAB WlbD UDP-N-acetylglucosamine 2-epimerase OS=Pyrococcus abyssi GN=wlbD PE=4 SV=1
MRPAIIFGTRPEIIKLSPVIRAFIKNDIQPILIHTGQHYDYEMSSIFLEELELPEIDYHLEVGSGTQAEQT
GIAMIKIEKVLMNEVPDVSIVQGDTNTVLAGALASVKLLIPVAHVEAGLRSFDRTMPEEINRILADHS
SEVLFAPTLEAKKNLEREGIRENVFVVGNTIVDAVLQNSKIAEKKSKILSELGLEPKSYILVTAHRKEN
VDNKDRLKKLVDILTSLPMQVVYPVHPRAEKRLKEFGLWDRLKSSGVMLLKPLGYLDFLKLEKNA
FIVMTDSGGIQEEAIILNVPCLTLRYNTERPETVKAGGNILVGVEKDLALRYVKKLIEDEAFYKRMAS
AKNPFGDGRSGERIVSILMTLWESGRLKVKTSNFITPQKA
>A_EURYARCHAEOTA_3 tr|Q9HM99|Q9HM99_HALSA UDP-N-acetylglucosamine 2-epimerase OS=Halobacterium salinarium GN=uae PE=4 SV=1
MDAAPTIYEDRLAAQMDRGDFVLAVVTATKPDFYKQAPVVAAAEANDVPCFVIHTGQHYDDVLGH
GLEEYGIEPHIAADLGIRGDLTQKTAEMMLAVKTLAERLEAWPDTTVLPMVHGDTHAAGVFPQAW
MFATNQQVAHNEAGLRGMAPAYETTDDPAAVIREQWRGDWHVERTEPFPEQYDTFVGSAASIYQFA
PVALNREHLEREGYPRSVDGHERMPVVGNSVVDAIEMKRDHDGESVFDVYPVLEARDDWLRVDIH
RRANLLPERFTAIVEGVIELVERGYNVNFVELNATKRALENYGYRERLQRLADERENFLFTGLWKK
HAHVYEFLESGQCFAEFTDSGSMQEELNEIEAALCLTARFSTDRPETVFDANTNLLVPPTSGAFVAD
MITHVAETDDVRERMRSGQALYGADVGEEIVEFLQARADDGVFDWAHERLGFDAGSEQAFDYL
>A_EURYARCHAEOTA_4 tr|Q46AM3|Q46AM3_METBF UDP-N-acetylglucosamine 2-epimerase OS=Methanosarcina barkeri (strain Fusaro / DSM 804) GN=Mbar_A2140 PE=4 SV=1
MKIVSIVGARPQFIKCAPLSRLIREKHEEILIHTGQHYDTGMSDIFFDELKIPKPNYNLGVGSNSHGVQ
TGKMLIEIEKILLRESPDLVLVYGDTNSTLAGDLAASKLHIKTAHIEAGLRSFDRSMPEEINRVLTDHT
SDLLFCPTETAVLNLKKEGITTGVYNVGDVMLDSLKYNIRIAEQKATILGKLNLNSKEYIVATVHRAS
NTDSFENLSSITNAFCHTGVSIVFTVHPRTEKYLKQYGLWNKLCEKVKVIPPLGYLEMLKLMVHAK
KILTDSGGVQKEAYMLGVPCITMRENTEWVETIEDGGNVLVGTDYEKIMDAILNFKGVPVKGNFFG
NGNACAEICKTLNKL
>A_EURYARCHAEOTA_6 tr|A6UU99|A6UU99_META3 UDP-N-acetylglucosamine 2-epimerase OS=Methanococcus aeolicus (strain Nankai-3 / ATCC BAA-1280) GN=Maeo_0485 PE=4 SV=1
MKIAIILGTRPEIIKLSSIIRELQNFSKENFYSCKTCEANLKHNQSNTHKDNKNIEYFIIHTNQHYSKNM
DEIFFKELNLPTPKYNLNVGSGTHGEQTAKMIDGIEKILISENPDVVIVQGDTNTVLAGALSASKLHI
KVAHVEAGLRSYDRNMPEETNRVLTDHISNYLFAPTEIAKHNLLKEGINKNVFVVGNTIVDATIQNIE
IAENIYKYNNMSIKTNINEDYFLLTLHRAENTDNKERLTNIVNAIIKATEQYNKKIIFPMHPRTEKKLK
EYNLFENLQKNSKIEIIEPVGYLEFLLLEKNAKLILTDSGGVQEEACILGAPCITLRDNTERPETIYIGS
NILVNADINKILDGIGVMANKKINGNNPFGDGNSGKAIVKILLEN
>Saccoglossus2 XR_086474.2 PREDICTED: Saccoglossus kowalevskii UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase-like (LOC100367162), miscRNA
```

QRCRDGIVGISIPDFGRSTPQFMDGKKDPHDVFMYQNGFQNGHRRTPSSSSLGSESLDRPMRVCVAT
CNRADYSKLAPIMLGLKEDKDFQLQVIVMGCHLIDDYGSTYRMIEQDKIKIDAKLHTIVRGEDEAA
MVESVGLALVKLPDIFLRLKPDILIVHGDRFDALSTATAAALMNIRIIHIEGGEISGTIDDCIRHSITKLA
HYHVCCTERARGRLQAMCEDNERILLAGCTSYDMLLKSNTKNCQLIIDRWVEKGVKPKEYIIALQH
PVTTDLHNSLKMFDLTLEALLEFGKKTVLLFPNIDAGSKDMTRILRRKKIENHPNFSTAKHIPFEEFIV
LVANCACIIGNSSAGVREAGAFGTPVINLGSRQTGRETGENVMHCRDADTKEKILHALELQYQKQY
PPSYIYGDGHAVSRIIKFMKELDVIDSIQKQFIFPPMPDGVSQDIDHILEIQSAIAVDLGGTNLRVSIVSK
AGEIMHMMSKPTPSDNTARMESLLTLLVDASQIATKLNCRVLGVGVSTGGRVNPHEGLVMHSTKSL
EGWSAIDLRTPISSALHLPVWVDNDGNCAALGERKFGHGKGKEDFLAIATGTGIGGGLVLNGRLVH
GSNFCAAELGHIAVSIDGPACMCGSSGCVEAYSSGLALMREAKKLHQAGELIVDGMKMT?
NEIISAKHLIQAAKLGNKKAEKIIHLGGEALGTAVISILHLLNPTLVVLCGVLAPTYVDIVRETIENKAL
PSARGVKVQVSLLKEPALMGAASLVLEYATRRIY
>Strongylocentrotus1 gi|115637356|ref|XP_795233.2| PREDICTED: similar to UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase [Strongylocentrotus purpuratus]
MVEIAFNRFLNQQHQRKLSGTMTSLNGSGVKANCGGDMDTKRKFRVCVATTNRADYSKLGNIMQ
AIKDDEMLELSTIVLGCHLIDDYGSTYRLIEKDGFTIDSRLHTIVRGEDEAAMVESVGLAMVKLPDI
MLRLKPDVVIVHGDRFDVLSLAACAALMNIRIVHVEGGEVSGTIDDSIRHTISKLAHYHVCCTERAH
KRLLAMCEDNDRILLAGCPSYDKLLSTDVVDCQHILNRWLKGEGKSKEYIVALQHPVTTNIADSLK
MYSLMVDALMEFNKRVIMLFPNIDAGSKDIVRIMRQKGVETNPMFYPVKHIPFEEFIILVANAGCMI
GNSSAGVREAGAFGTPVVNLGSRQTGRETGENVLHCRDADTTQKIHHALEIQYQRQFPPSYIYGDG
HAVPRIIKFLKSIKQEDTIQKQFIFPPMPESPSIDIDHILETQSALAVDMGGTQLRVALIAADGEIKLRRS
RPTPHTTNEDRMKALLELLLEATKESYALNCRVLGVGISTGGRVNSKEGLVMHSTKAIAGWDEIDLR
TPIETTLHLPVWVDNDGNCAALGERKFGHGRGVKDFITIATGTGIGGGIVLDKRLIHGTNFCAAELG
HIKVSLDGPACQCGSHGCVEAYSSGMALKREAMKLHEAGELLVDGMIVEKGEEVTAKHLVQAAKT
GNQKAKKILDTGAQALGTAITSLLHVLNPKLVILCGVLSNVYLEPVREAVKTHSLPSAVSEVEIVSSQ
LVEPALLGAASLVLEYATRRTY
>Symsagittifera JF826132
MKIAVATCNRADYSKLAPVLRGLKQDTTFQVSIVVMGSHLIDDYGNTYRFIERDGNQIDSMLHTLVR
GETEGAMVESMGLAMIKLPDILNRLQPDIVMVHGDRFDAMSVAISAAVMNVRVLHLEGGEVSGTID
DVIRHSITKLAHYHICCTKGARLRIESMCEDSSRVLLAGCPAYDELIHTDVSKYHEAFERWLPNHVGT
KNFIICVYHPVTTAIAESIQHFGLLLDALVHFDVRTVILFPNVDAGSKELVRCIRLRNLENHEKISCYK
HVPFSEFVYLMGNCGLMIGNSSAGIRESNVFGTAVINIGTRQRGRQSGANVVHVKNPTSAEDLLWHI
RSQFGKVYPRDYIYGDGRAVERIVKFIKEIDCDSPVEKSFNFDRNSSGLRAPSKASFGAEFPLSRQAH
SLQLALAVDLGGSFIRVALVDDRGNIERMRRTEMVDCPDERIAIIVRMANDLLENKRDLVVGVGVST
GGRVNSETGEILFATKVLSGWGGVALKTRLQEQLGLPCYVENDGNCAALAEVHFGSHKMEDMVVL
HFGTGIGGGIIQDGCLLNGSSYSAGEFGHIVVCFDDGPDCMCGNSGCVEAYAGGWALNKLAKEMEL
SSNWRKEGADDQKPVTPGQPNKPTYLTDLASDGVDYAVQHINRAVRAVASALLTIYSSYNPPVAILA
GPLAPVYFDGVKNKLEERSAILGRNFTLLQSDMTEMSLKGAATLVLNNPSRAVPQNAIV*
>Takifugu tr|Q4W392|Q4W392_TAKRU Putative UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase (Fragment) OS=Takifugu rubripes GN=glcne PE=2 SV=1
MFGLKSHPDEFELEVVVLGSHLIDDYGNTFRMIEQDDFDIGSKLHTIVRGEDEAAMVESVGLALVKL
PDVIQRLHPDILVVHGDRFDALALATAAALMNIRILHVEGGEVSGTIDDSIRHAISKLAHYHACCTRM
AEQHLIAMCEDHSRILLAGCPSYDKLLLSHQKEDYMDIIKSWLGDNVKEHDYIVALQHPVTTDIKNS
IKIYGLMLDALLSFNKTTLILFPNIDAGSKEMVRVMRKKGIEQHPNFRAVKHIPFEQFIQLVNHAGCM
IGNSSCGVREAGAFGTPVINLGTRQTGRETGENVLHVRDADTHNKIYHALELQFGKRYPCSKIYGDG
NAVPRILKFLSSIDLDEPLQKTFCFPPVKDPISQDIDHILETQSALAVDLGGTNLRVAIICMRGNIVRKY
TQANPKTFEARMQLILKMCSDAMRDAVFLNCRILGVGVSTGGRVNPQEGVVLHSTKLIQEWSAVDL
RTPISDALHLPVWVDNDGNCAALAEKKFGHGKGVENFVTVITGTGIGGGIIHQNELVHGSTFCAAEL
GHIMVSFDGPECSCGSRGCIESIASGMALQREAKRLHDEDLLKVDGLDMKISDPITAAHLINAARLG
NSKANIVLNKASTALGMGIINILHIMNPSLVILSGVLGSYYQAPVQRTIMERALFSAQSVKVVISDLEE
PALLGAASMVLDYATRRIY
>Tetraodon tr|Q4RJ92|Q4RJ92_TETNG Chromosome 18 SCAF15038, whole genome shotgun sequence OS=Tetraodon nigroviridis GN=GSTENG00033530001 PE=4 SV=1

MTRVGSAAPAAPRAAGQSQSPAPPPVQSSPSPCRIAMTTPPSASFSPFFSNATGFCKLRTLMNAGIKTE
LYSWRMQRGRVKMDRRRMEDQNQCRKKLRVCVATCNRADYSKLAPIMFGLKSHPDEFELEVVVL
GSHLIDDYGNTFRMIEQDDFDIGSKLHTIVRGEDEAAMVESVGLALVKLPDVLQRLQPDILLVHGDR
FDALALATAAALMNIRILHLEGGEVRSGTIDDSIRHAISKLAHYHACCTRMAEQHLIAMCEDHTRILL
AGCPSYDKLLLTHHKEDYMDIIKSWLGDKVKDHDYIVALQHPVTTDIKNSIKIYGLMLDALLSFNKT
TLILFPNIDAGSKEMVRVMRKKGIEQHPNFRAVKHIPFEQFIQLVCHAGCMIGNSSCGVREAGAYGTP
VINLGTRQTGRETGENVLHVRDADTHNKIYHALELQFGKRYPCSKIYGDGNAVPRILKFLRSIDLEEP
LQKTFCFPPVKDPISQDIDHILETQSALAVDLGGTNLRVAIICMRGKIVKKYTQANPKTFEARMQLILK
MCSDAVRDAVFLNCRVLGVGVSTGGRVNPQEGVVLHSTKLIQEWSSVDLRTPISDALLLPVWVDND
GNCAALAEKKFGHGKGVENFVTVITGTGIGGGIIHQNELVHGSTFCAAELGHIMVSFDGPECSCGSR
GCIESYASGMALQREAKRLHDEDLLKVDGMDLKISEPITAAHLINAARLGNSKANIVLNKASTALGM
GIINILHTVNPSLVILSGVLGSYYQAPVQRIISERALFSAHSIKVVTSDLEEPALLGAASMVLDYATRRT
Y

>Xenopus gi|118404390|ref|NP_001072728.1| glucosamine (UDP-N-acetyl)-2-epimerase/N-acetylmannosamine kinase [Xenopus (Silurana) tropicalis]

MEKNINQRKLRVCVATCNRADYSKLAPIMFGIKAEPDRFVLSVVVIGSHLIDDYGNTYRMIEQDDFD
IQARLHTIVRGEDEASMVESVGLALVKLPDVLNRLNPDIIVVHGDRFDALALATSAALMNIRILHIEG
GEVSGTIDDSIRHSITKLAHYHACCTRSAEQHLIAMCEDHDRILLAGCPSYDKLLSVNNKDYMSVIK
MWLGDDAKSGEYIVALQHPVTTDIKHSIKMFEFTLDALLSFNKKTLILFPNIDAGSKEMVRVMRKK
GVEHHPNFRAVKHVPFEQFIQLVAHAGCMIGNSSCGVREAGAFGTPVINLGTRQTGRETGENVLHV
RDADTQNKIIHALQLQFGKRYPCSKIYGDGNAVPRIVKFLKSISLDEPLQKKFCFPPVKESISQDIDHIL
ETLSALAVDLGGTNLRIGIVSMTGEIIKKYVQPNPKTYEDRIELILKMCVEAASEAVKLNCRILGVGIS
TGGRVNPREGVVLHSTKLIQEWSSVDLRTPLSDTLHLPVWVDNDGNCAALAERKFGQGKGIEDFVT
VITGTGIGGGVIHNHELVHGSSFCAGELGHIMVSFDGPDCMCGSRGCVEAYASGIALQREAKKLHDE
DMLLVEGMSVKNDESVSAVHLIQAAKFGNTKASNILKTAGTALGIGVINILHTINPSLVILSGVLANQ
YVNVVKDVIRQRGLASIQNVDVVVSSLSDPALLGAASMVLDYTTRRTY

>Gallus gi|71895297|ref|NP_001026603.1| bifunctional UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase [Gallus gallus]

MEKNGNNRKLRVCVATCNRADYSKLAPIMFGIKAEPQFFELDVVVLGSHLIDDYGNTYRMIEQDDF
DIHTRLHTIVRGEDEAAMVESVGLALVKLPDVLNRLKPDIMIVHGDRFDALALATSAALMNIRILHIE
GGEVSGTIDDSIRHAITKLAHYHVCCTRSAEQHLIAMCEDHDRILLVGCPSYDKLLSAKNKDYMSVI
RMWLGEDVKPRDYIVALQHPVTTDIKHSIKMFELTLDALISFNKRTLVLFPNVDAGSKEMVRVMRK
KGIEHHPNFRAVKHVPFDQFIQLVAHAGCMIGNSSCGVREVGAFGTPVINLGTRQTGRETGENVLHV
RDADTQDKILHALQLQFGKQYPCSKIYGDGNAVPRILKFLKSIDLKEPLQKKFCFPPVKDNISQDIDHI
LETQSALAVDLGGTNLRVAIVSMKGEIVKKYTQLNPKTYEDRLGLILKMCVEAASEAVNLNCRILGV
GISTGGRVNPREGIVLHSTKLIQEWSSVDLRTPISDALHLPVWVDNDGNCAALAERKFGHGKGIENF
VTLITGTGIGGGIIHQHELIHGSSFCAAELGHIVVSLDGPECPCGSQGCIEAYASGIALQREAKKLHDE
DLLLVEGMSMKNEEVVSAAHLIQAAKLGNAKAESILRTAGTALGLGVVNILHTMNPSLVILSGVLAS
HYVNAVKDVIHRQALSSVKTVDVVVSNLADPALLGAASLVLDYTTRRIY

>Homo gi|45360235|gb|AAS59258.1| UDP-N-acetylglucosamine-2-epimerase / N-acetylmannosamine kinase [Homo sapiens]

MEKNGNNRKLRVCVATCNRADYSKLAPIMFGIKTEPEFFELDVVVLGSHLIDDYGNTYRMIEQDDF
DINTRLHTIVRGEDEAAMVESVGLALVKLPDVLNRLKPDIMIVHGDRFDALALATSAALMNIRILHIE
GGEVSGTIDDSIRHAITKLAHYHVCCTRSAEQHLISMCEDHDRILLAGCPSYDKLLSAKNKDYMSIIL
MWLGDDVKSKDYIVALQHPVTTDIKHSIKMFELTLDALISFNKRTLVLFPNIDAGSKEMVRVMRKKG
IEHHPNFRAVKHVPFDQFIQLVAHAGCMIGNSSCGVREVGAFGTPVINLGTRQIGRETGENVLHVRD
ADTQDKILQALHLQFGKQYPCSKIYGDGNAVPRILKFLKSIDLQEPLQKKFCFPPVKENISQDIDHILE
TLSALAVDLGGTNLRVAIVSMKGEIVKKYTQFNPKTYEERINLILQMCVEAAAEAVKLNCRILGVGIS
TGGRVNPREGIVLHSTKLIQEWNSVDLRTPLSDTLHLPVWVDNDGNCAALAERKFGQGKGLENFVT
LITGTGIGGGIIHQHELIHGSSFCAAELGHLVVSLDGPDCSCGSHGCIEAYASGMALQREAKKLHDED
LLLVEGMSVPKDEAVGALHLIQAAKLGNAKAQSILRTAGTALGLGVVNILHTMNPSLVILSGVLASH
YIHIVKDVIRQQALSSVQDVDVVVSDLVDPALLGAASMVLDYTTRRIY

>Mus gi|4456673|emb|CAB36908.1| UDP-N-acetylglucosamine 2-epimerase/N-acylmannosamine kinase [Mus musculus]

```
MEKNGNNRKLRVCVATCNRADYSKLAPIMFGIKTEPAFFELDVVVLGSHLIDDYGNTYRMIEQDDF
DINTRLHTIVRGEDEAAMVESVGLALVKLPDVLNRLKPDIMIVHGDRFDALALATSAALMNIRILHIE
GGEVSGTIDDSIRHAITKLAHYHVCCTRSAEQHLISMCEDHDRILLAGCPSYDKLLSAKNKDYMSIIR
MWLGDDVKCKDYIVALQHPVTTDIKHSIKMFELTLDALISFNKRTLVLFPNIDAGSKEMVRVMRKK
GIEHHPNFRAVKHVPFDQFIQLVAHAGCMIGNSSCGVREVGAFGTPVINLGTRQIGRETGENVLHVR
DADTQDKILQALHLQFGKQYPCSKIYGDGNAVPRILKFLKSIDLQEPLQKKFCFPPVKENISQDIDHIL
ETLSALAVDLGGTNLRVAIVSMKGEIVKKYTQFNPKTYEERISLILQMCVEAAAEAVKLNCRILGVGI
STGGRVNPQEGVVLHSTKLIQEWNSVDLRTPLSDTLHLPVWVDNDGNCAAMAERKFGQGKGQENF
VTLITGTGIGGGIIHQHELIHGSSFCAADVGHLVVSLDGPDCSCGSHGCIEAYASGMALQREAKKLHD
EDLLLVEGMSVPKDEAVVALHLIQAAKLGNVKAQSILRTAGTALGLGVVNILHTMNPSLVILSGVLA
SHYIHIVKDVIRQQALSSVQDVDVVVSDLVDPALLGAASMVLDYTTRRIH
>Paracentrotus1 GLEAN3_05241
MYSLMVDALMEFNKRVIMLFPNIDAGSKDIVRIMRQKGVETNPMFYPVKHIPFEEFIILVANAGCMI
GNSSAGVREAGAFGTPVVNLGSRQTGRETGENVLHCRDADTTQKIHHALEIQYQRQFPPSYIYGDG
HAVPRIIKFLKSIKQEDTIQKQFIFPPMPESPSIDIDHILETQSALAVDMGGTQLRVALIAADGEIKLRRS
RPTPHTNNEERMKALLELLLEATKESYALNCRVLGVGISTGGRVNSKEGLVMHSTKAIAGWDEIDLR
TPIETTLHLPVWVDNDGNCAALGERKFGHGRGVKDFITIATGTGIGGGIVLDKRLIHGTNFCAAELG
HIKVSLDGPACQCGSHGCVEAYSSGMALKREAMKLHEAGELLVDGMIVEKGEEVTAKHLVQAAKT
GNQKAKKILDTGAQALGTAITSLLHVLNPKLVILCGVLSSVYLEPVREAVKAHSLPSAVSEVEIVSSQ
LVEPALLGAASLVLEYATRRTY
>EU_Cow [cowbropro] CAOG_03577T0 | CAOG_03577 | Capsaspora owczarzaki ATCC 30864 UDP-N-
acetylgluc
MPSFRLPPFVRKPTVLALVFMLELVVALALLSSSTGLNGWKLGNPAPDNVDVQLSEQQPGASGDAV
SPLPAALAAQLHQLAQHRQNQARRRPLALVFVGTRPEVIKMAPLVHEMQENSVNSHVFDTIVVSTG
QHREMLAQTLRGFDLKIDLGLSVMQHNQNLGALFAAVVAQSTAVIASLRPEVVLVQGDTSTALACA
TAAYYERVPVAHVEAGLRSFDFNHPFPEEMNRRAIDSFATFHFAPTEYAAEAIRRENTCARNIVVTGN
TGVDAVLARLNADPPAAARRLIAQLQTCTGGACDVNDAPPVIILVTSHRRENLGEPLESICRAVKRIV
EDNPVVRVVFPVHLNPNVQKVVFAHLKNVDRITLTDPLDYDVLTYLLQEVTIVLSDSGGLQEEAVSI
GKPVVLLRQTTERPEGVYAGTVRLAGTNEDRVVSDTQNVIDNLVEYHQRIKGQNVFGDGAASGRIA
NVLRQYLAADSAQRRKYLAPCTQEKQAELANVVYGHS
>EU_Cow_2 [cowbropro] CAOG_05274T0 | CAOG_05274 | Capsaspora owczarzaki ATCC 30864 UDP-N-
acetylgluc
MQANVLTPADSSSTTPLSASPVPSGGASLVGVAARIAVRPTATRVLSIFGTRPEVIKFYPVLKEMDRRS
ESLLSITCVTGQHRQMIDPLLSLFDIATDIDLNLMTQGQTLSNLSARLFTALAPVFEAVRPHLVLVQGD
TTTAMIAAMCAFYYRVPVGHIEAGLRTNDRYNPFPEEVNRRIVSVVGNLHFAPTNYAAEALRAERID
MQTVFVTGNPVIDAILLIREKAPSAAARQLLDRVEAVDAATNATLPAGQKARHILLTAHRRENHGEGI
AAICRAVKRIIAQHKDVNVWYPVHLNPAVCVPVRQELGKVERVHLLDPLEYDVFVHFLDRMYMVM
TDSGGIQEEVTALAKPTIVLRETTERPEGVDAGVTKLVGIHEDTIVKEANVLLTDEQTFAKMSCKCFP
FGDGTAAKQIVDIMLMSVSEVKLSIGA
>EU_Cow_3 [cowbropro] CAOG_03163T0 | CAOG_03163 | Capsaspora owczarzaki ATCC 30864 UDP-N-
acetylgluc
MSKQLLNIMNSSLLTPVPSIPTTPLSPSPSAAMLACGASYPNVAARIAERPTATRVLSIFGTRPEVIKFY
PVLKEMDKRSESILSITCVTGQHRQMIDPLLSLFDIATDVDLNLMTQGQTLSNLSARLFLALAPVFEA
VRPHLVLVQGDTTTAMIAAMCAFYYRVPVGHIEAGLRTNDRYNPFPEEVNRRIVSVVGNLHFAPSSY
AADALRAERIDMNTVFVTGNPVIDAIMLIREKVPSAAARQLLDRVEAADAATNAVLPAGQKARHILL
TAHRRENHGEGIAAICRAVKRIIAQHKDVHIWYPVHLNPAVCVPVRQELGKMERVHLLDPLEYDVF
VHFLDRMYMVMTDSGGIQEEVTALAKPTIVLRETTERPEGVDAGVTKLVGINDDTIVKEATLLLTDD
QTFAAMSRKCFPFGDGTAAQQIVDILLTSLSTNKLSIGA
>EU_Enterocytozoon tr|B7XKK1|B7XKK1_ENTBH UDP-2-acetamido-2,6-dideoxy-beta-L-talose 2-
epimerase (Fragment) OS=Enterocytozoon bieneusi (strain H348) GN=EBI_23619 PE=4 SV=1
MKKLKVVTVVGTRPEIIRLSRVMAELDKHCDHILVHTGQNYDYELNEIFFRDLGIRTPDYFLNAAGV
SGAQTIGNVIIAVDRVLEDVKPEALLVLGDTNSCMAVIPAKRRRIPTFHMEAGNRCFDMRVPEEINRR
IVDHTADINLTYSSIARDYLLREGLPPDMVIKTGSPMFEVLNYYRDGIEASDVLKRLGWRR
>EU_Micromonas tr|C1EJI1|C1EJI1_9CHLO Udp-n-acetylglucosamine 2-epimerase/CMP-N-
acetylneuraminic acid synthetase OS=Micromonas sp. RCC299 GN=MICPUN_64942 PE=4 SV=1
```

MPACPNTEADKASLYAPDKCTDMKTTLRRKICVVTSNRSDWSKLKLVAINLRKLCTSQDNQEASDIQ
VDIICLGSHLLHELGATKNIVKEDFPNAYELHTLVAGDSVESMTDSVGFGIVKLTSLLCALKPNIVLVH
GDRFDAFCAAIAANMLNLTIAHVEGGELSGTVDGTLRHAITKLSHLHFTCTPEAARRIRGMGENPAS
IFVTGCPSYESLFAVSATCWEDEKMDQFFNGTPFKLKPNKFILVIMHPVTNDLEESNTLYGSLLSCLFS
RKTPTVMFYPNVDPGNKSMIQTLHKHQKADPASTSWLRLVTHMPHAKFTALMRHASAMVGNSSA
GIRESCVFGIPTLNLGSRQEGRRVPANVTTLVKPSIRSIDCWFDNELGKRYAQSTMYGFPDSAKRIAH
HLSRIDTSAGQLKQFWEPRYALLPPLQPRQYVSRTQAILADSTSSPTSSPIGRCKILGLITARGGSKGIP
GKNIIDLNGKPLIQYTIEAALSSKQLDRVILSTDSDEIAEVAQNCGCEVPFRRPSELAADDSSHLACIV
HALNILRETECFVPDFVVILQPTSPFRKSIDIDSCINIMLTSSCDMVLSVCESSLNLSKNFYFAADGTLS
PFAESTAEIDYTPRQKLLKTYAENGAVYVLRTQSLLYPPDNAPNVGSFRSADTKGYEMPVERSLDIDN
PFDLHVARLLMAKPF
>EU_Naegleria gi|290995697|ref|XP_002680419.1| predicted protein [Naegleria gruberi]
MSKIFKLIIACTTLLVFSFVFFRLLTKSPINSDNNIFDRNSNTKQETPTIPHKNNLPRVVIVLGTRPEAIK
CAPLISHLKSNYSSELELIVLSTGQHSEILKQTLAVFDQHVDIDLELMEPNQSISSFFTLAYQRISLEFER
QGPISIVLVQGDTTTSLAAALAASYLKIPVGHVEAGLRSYDFNSPFPEELNCKMIDSFARLLFAPTDFS
KAALVREGICPQHVHVTGNTGIDSFYSLMSVPPKIIPEFLKNIKSFKGSNDHAAETLAIESHHRIVILVT
MHRRENIPHFKEMCNAIKTISNTFGKNVMIILPVHPNPNAKSVVNEVLSALDNVQLVDPIAYDIFPHV
ISEADIIVTDSGGIQEESASIGKPVILMRDTTERPEGIYIGTIKKIGVNYFHIVKAMTDAIKDSQNSKQIL
SKHIFGDGKASMRISKIVRDFVTSKLPDSIECSTQFIQDSIAESVYSQNFTIRK
>EU_Naegleria_2 gi|290982189|ref|XP_002673813.1| predicted protein [Naegleria gruberi]
MPRVVIVLGTRPEAIKCAPLISHLKSNHHSELEVIVLSTGQHSEILKQTLGVFHQHVDIDLKLMTPNQ
SISSFFTLAFQKISLEFEKQGPISMVLVQGDTTTSLVAALAASYLKIPVGHVEAGLRSYDFNSPFPEELN
RKVIDSFASLLFAPTEFSKAALVREGICPQHVHVTGNTGIDSFYSLKSLPPKIIPELLQNIRSFKTKHVD
ATTSSHRTVILVTMHRRENIPHFKEMCNAIKTISNTFGKNVMVILPVHPNPNAKSVVHEVLSSLDNVQ
LVDPIAYDIFPHVISEADIVVTDSGGIQEESASIGKPVILMRDTTERPEGIYIGTIKKIGVNYHHIVKAM
TDAIKDSQNSKLILSKHIFGDGKASMRISKIVRDFVTSKLPESSECSTQFIQDSIAESVYSQNFLPFESR
KKELLERSKPKDPLTMQELFALPSAYTPATKTSDEFSVTVVVGLYKRTGLVKRWMQALLSQTHQPK
VVWMVFFASPKAEEIENEIQEAKLLATSNNISLFINKGEMQLKYFGRFQLALQSPTKYIAVFDDDCIP
EKRFLEAAIHTINTNNYRGILGTKGTPTAEDYFYGPVSGSEQIIEVDVVGGSWFMEREWVKLLFREK
MFSWATGEDWHLCSNARKYANIRCFVMPVSRTLETNSYSGDYMQISNNGDTTGRVQGTSEARSLIIQ
KETQRGNRLMNSYRNNDLRTAFVFLSNQNEGRTILKMLSSLELPFTIQYSLGVADQSIAQIDSDELSKI
TKFKSFNDFMLSREFDTAQTNLSSAAETMYHFDMSIQQQQATAVILVGSSPTPSTLAVATACQINKIPV
INFVLTEKSNKLVESVSVLTVRSTPIDGNLLDETKRILTETLTKIFT
>EU_Naegleria_3 gi|290992983|ref|XP_002679113.1| predicted protein [Naegleria gruberi]
MSKIFKLSVVCTTLLVFSFVFFHILNINKSSADSDNMLVRNSNGKREGIIIPHKNNLPRVLIVLGTRPEA
IKCAPLISHLKSNYSSELEVIVLSTGQHSEILKQTLGVFHQHVDIDLKLMTPNQSISSFFTLSFQKISLEF
ERQGPISMVLVQGDTTTSLVAALAASYLKIPVGHVEAGLRSYDFNSPFPEELNRKMIDSFASLLFAPT
DFSKAALVREGICPQHVYVTGNTGIDSFYSLKSTTPKIIPELLQNIKSFKSNVLDTTSSTTTSSSAVESH
RIVILVTMHRRENIPHFKEMCNAIKRISNTFGKNVMVILPVHPNPNAKSVVNEVLSSLDNVQLVDPIA
YDIFPHVISEADIVVTDSGGIQEESASIGKPVILMRDTTERPEGIYIGTIKKIGVNYLDIVKAMTDAIKD
SQNSKQVLSKHIFGDGQASMRISRIVRDFITSKLPDSRECSTQFIQDSIAESVYTQNFLPFEIRKKDLLE
RSKPKEQLTMQELFSLPSTYVPTRNTSDEFGVTAVVGLYKRTGLVKRWMQALLSQTHQPKVIWMVF
FASPKADEIEKEIQEAKLLASSNNISLFVNKGEMQLKYFGRFQLAIQSPTRYLAVFDDDCIPERRFFEA
AMHTINTNNYRGILGTKGTPSAEDYFFGPVSGSEHIIEVDVVGGSWFMEREWVKLMFREKMFSWAT
GEDWHLCSNARKYANIRCFVMPVSPTLETNSYSGDYMNISNNGDTTGRVQGTSEGRSLIIQKETQRG
NRLMYSYRGNDQRTAFVFLSNQEEGKTILTMLSSLELPFAIQYSLGVADQSIAQINLDELSKITKIKSF
NDFMLSREFDTAQANLSSVAETIYHFDMSIQQQQATAVILVGSSPTPSTLAIATACQINKIPLINFLLSEN
SNKLIESVSLLTVHIPIEGSNGNLLDETQRSLTETLIKIFN
>EU_Ostreococcusl tr|A4SAA4|A4SAA4_OSTLU Predicted protein OS=Ostreococcus lucimarinus (strain CCE9901) GN=OSTLU_25793 PE=4 SV=1

```
MKDIESERFDRTAYSAEIKIAIVFGTRPEAVKMAPVIQAVARSSTLSAILISTGQHKQMLEQVLRQFSL
QDKIQHELALMKPNQQLAELTSSAVRAVDGVLRSSKPDAVLVQGDTTTAFITSLAAFYLKIPVGHIEA
GLRTRDIYSPFPEEVNRQCISVMATYHFAPTEHAAKNLYDEGRRTNVFTTGNTVTEPSDRVIELSKVV
KTVSTLRDVRLLLLTAHRRENLGEPILNIFTSIEKLLQEYPDVVVIYPIHLNPMSDHAPPTTTHLRRLLI
VPPLDHADLLFMMKESFFVMTDSGGIQEEAVTLGKPVLVLRDTTERPEGVLAGAAKLVGHGAESIY
TEAASLLKDPDSYRSMSGSKKTYGDGNAAGNIVAKEKQKERRGTVPKK
>EU_Ostreococcusl_2 jgi|Ost9901_3|25765|estExt_fgenesh1_pg.C_Chr_180008
MGRLKSILVLDRRRARTLISLVSLVSLVSLCFAALLSVNKHAGWPRKGDPPTPSLDSMKDIESERFDR
TAYSAEIKIAIVFGTRPEAVKMAPVIQAVARSSTLSAILISTGQHKQMLEQVLRQFSLQDKIQHELALM
KPNQQLAELTSSAVRAVDGVLRSSKPDAVLVQGDTTTAFITSLAAFYLKIPVGHIEAGLRTRDIYSPFP
EEVNRQCISVMATYHFAPTEHAAKNLYDEGRRTNVFTTGNTVVDALYAILKTEPSDRVIELSKVVKT
VSTLRDVRLLLLTAHRRENLGEPILNIFTSIEKLLQEYPDVVVIYPIHLNPMSDHAPPTTTHLRRLLIVP
PLDHADLLFMMKESFFVMTDSGGIQEEAVTLGKPVLVLRDTTERPEGVLAGAAKLVGHGAESIYTE
AASLLKDPDSYRSMSGSKKTYGDGNAAGNIVRILETMPRFARPLLYFTMRKTFG
>EU_Paulinella tr|B1X5F8|B1X5F8_PAUCH UDP-N-acetylglucosamine 2-epimerase OS=Paulinella
chromatophora GN=PCC_0762 PE=4 SV=1
MTTPPLVSIVLGTRPEAIKLAPVIMSFQKCGALRTRTILTGQHREMVNQVMKLFQLSFDKDLALMEP
NQTLTHLTATILEGLRKEFLVHRPSLVLVQGDTTTALASAMAAFYEQIPVGHIEAGLRTNDIDDPFPEE
GNRCLISQITKLHFSPTIQSAANLQASGVIGEIHITGNTVIDALLSMSQKKSTVCQINGLNWKDKRVIL
ATVHRRENWGTRLNGIAYGFFKLLERHSDVALLLPLHKNPAVRKPLNDILKDHPRAFLTEPLAYDQL
VAAMRSCSLVLTDSGGIQEEAATFGKPVLILRRTTERSEAVEGGTARLIGTDTTDILNEASLLLKDDSA
YRAMSQFNNPFGDGKASDRILTAARTFLNC
>EU_Physcomitrella tr|A9U8C8|A9U8C8_PHYPA Predicted protein (Fragment) OS=Physcomitrella patens
subsp. patens GN=PHYPADRAFT_104339 PE=4 SV=1
MDILVVPSMVSEGFGMTALEGMLCGKPVVAFASGGLGELMVSTGNGHLAVPAGDVQGLADKVAGL
LDDPAALANTGFHNEATARAMYGIETYQRRLDEIMNQWRTLYPHWLMPGNRAPLILAEPEVSAQPPI
EPPDEPAPAPRKARRSRGRTSFRIRIIRKRRVKKAAKRSRKRKPGKAVRKLKLMKPEIIRLSLILPKLDR
YAERHVLVHTGQNFAERLSGIFFSQLGLRKPDYVLQDSAASLGTQLSAMFSQMEDILEREKPDTILLL
GDTNSALCAILAERLGYPVVHMEAGNRCFDLN
>EU_Ricinus tr|B9TNA0|B9TNA0_RICCO Polysialic acid biosynthesis protein P7, putative (Fragment)
OS=Ricinus communis GN=RCOM_2068830 PE=4 SV=1
MPDGGRTEGERMSMIRKICVFTGTRAEYGLLRPLMQAIATRPGATLQTLVTGAHLAESTGATWREIA
NDGLPIDERVEVLLDGGGDESICTSIGLGVMRYAEALKRLAPDMLVILGDRYEAIAAAVAATVCKVPI
AHIHGGELTLGAMDDAFRHSITKMSYLHFTSTEAYRKRVIQLGEQPARAHNVGALGVENIRTLNLYD
RAEVEKRLDLPAGQPYLLVTFHPATLESQSPELQLKALLNALDAFPDHVCVFTGANADPGGAGLNR
LLADH
>EU_Ricinus_2 tr|B9TD82|B9TD82_RICCO UDP-n-acetylglucosamine-2-epimerase, putative (Fragment)
OS=Ricinus communis GN=RCOM_1782450 PE=4 SV=1
MEAGNRCFDQRVPEELNRKVLDHLSDINMVLTEHARRYLLAEGIPADRIIKTGSHMQEVLEYYMPNI
EASNVLERLNLSSQQYFLLSTHREENVDSPANLQDLLDTLQALVKKYDMPVVVSTHPRTRQRLEKL
GVGSLDNRIHFLKPFGFFDYIKLQKEAFCILSDSGTITEEASLLNLSAVTIRNAHERPEGMDEGTLIMC
GLKTERVLDAVNIVTHQHKKTDTKLNIVPDYLGGLVSRK
>EU_Salpingoeca [pspscfgens] scf7180000054676|GENSCAN_predicted_peptide_30|366_aa
MGNFTIVAGARPNFMKIAPIIHEIQKLQKEKSGVSFRLIHTGQHYDKKMSGDFFEQLDIPQPHANLGA
GGGTQAEQTAAIMVAFEKELMENRPDLVLVVGDVTSTLSCSIAAKKLQIDVAHVEGGIRSGDLGMPE
EINRMVTDSITDHFFTTSEIANQNLRNSGFSEDKIHFVGNTMIDTLLAQMPKFQKPEGAIFDQLEAGN
YFVMTMHRPANVDQEHKLKAMIDAILDGTKGLPIIFPVHPRTAKNLQAIGIEAPNLNMNPLGYLEFN
YLVKNAKGVITDSGGITEEASVMNVPCITLRDNTERPETIDLGTNELVGTNPEKLKPYLDKIMSGDW
KKYKGIPLWDGKTAERIVKILEDKYPSVK
>EU_Spyzello [spunprot] SPPG_08155T0 | SPPG_08155 | Spizellomyces punctatus DAOM BR117 UDP-N-
acetylglu
```

```
MGTRPETIKMFPIIRALNTHGINPIICVTGQHQEMVEPLLRLFDIKPHVNLNVMKSGQSLNSLTERIVG
RMTKAVTHYRPDWLLVQGDTTSAFVASVVAFHEKIAVGHVEAGLRTYKRYSPFPEETSRRLIAGLAS
YHFVPTQHAANNLKAEGIPSENVYITGNTVIDALQWVASLEPSPEVASSLLEIEEAAKSHTSQYRLLL
VTMHRRENLGSPLISVCKAIKRIVDGFPDVHIVLPVHMNPTVQETVTQFLGNDKRITLLGPLSYEFFA
QLLKRATLLLTDSGGVQEEGTAFSKPILVLRNDTERPEGVTAGVAKLIGTDEDNVFRHVQELLTDKQ
AYAAMAIKTFPYGDGTAGSKIVDIVLKGSAQIAVNRPTIASDKLVEAESAKTTLTNVSESTSAASIAPR
PRSFVRAPTKPRQLNLTALLELPSSYPENGDGKGLNGITAVISGYNRVEVIPRLLGSLFNQTVPPLEIWI
TVFASKSVAAIKQAVDDFANSHKAQKIPIKFLQGDVQLKYFGRFEVGLQIPTKYAVFFDDDCLPGAN
VFRNFLHVQNIVGGEFRGLYGAKGHIVPVQNSNNYKETYGLGQVIHPEVITQVDLVGGIWFMEKDW
IKLMFRENAVTWETGEDFQITYTLSKYANLPTFVFPVAASDPSSHLVTPDYRAISASGDTTHGAISVN
GKEMDIHRLRDYITFKHFTRGHTRVLMSEMWRPNEMRILYLVDTIQDARLFKPLHEYIWATTLQKPV
PNIRLRPFPVLIGRYPQSQLLSELGISWTLESFHTGVFDLGVASEFSRKSRAVDIATEVMMSFQQVIET
VRPDLVIVPNSPDDPAIVAAAITARSFNGFNVVAWNKKNAKDVLLPDESNSNERLTPLKDLVDGIIET
NEMVARILEEIAVASWLARRRDVRTDFRN
```

ROK sequences

```
>A_CRENARC_2 tr|A1RT73|A1RT73_PYRIL Glucokinase OS=Pyrobaculum islandicum (strain DSM 4184 / JCM 9189) GN=Pisl_0979 PE=4 SV=1
MGKFFYSALKQVYVRLFLGIDIGATWTRALLIDEHGDIINRVKVKTSVNPIADVVDIVKRWQFHAVG
VGSIGPLDLKSGWVVNSPNSPTRRFPLVEPLKELGKPIVVANDCVAAVWGEYVFKYNVENMVYITLS
TGVGVGAIVNGNLLLGKDGNAHELGHAVIDFKSARRCGCGGRGHFEAYVGGAHIPRVYQEATGDA
PASPEEIFRRYRHGDEKARKFINLWLDALAAGIATVVAAYDPELLIVGGSIALNNWDIISRELPARLRD
YLSLREPEILKASFGDDEVAVGAAALAYKTPDTLKKFGYPITASSQS
>A_EURY_2 tr|Q2FS47|Q2FS47_METHJ ROK OS=Methanospirillum hungatei (strain JF-1 / DSM 864) GN=Mhun_2804 PE=4 SV=1
MQHGEGRIAGAADIGGTNTRVGLIREDGKIVRIEKFPTPVSGNAEDIPLAIARALMDIAGDIPLAGLGI
AAAGPLNIREGILDHPPNIPFDFVPIVAPLKEATNLPVIFQNDCRAAVLGEVCAGGARGYETVVYITIS
TGIGGGICTNGKVITGRGGNAGEIGHFPVDTTYNLTCTCGLSGHWEGYASGRGIPFFFREWCTAHDL
PCIYSTTPEILRFSATDPRYAGFRDALAQVNGRGLSSVIVAYDPDCIILDGTVIQRNPDLLDQALVYTD
RYLDLPPCIFSPLNGDAPLIGAAMAVFHPKMI
>A_THAUMARCHAEOTA tr|A9A5Z4|A9A5Z4_NITMS ROK family protein OS=Nitrosopumilus maritimus (strain SCM1) GN=Nmar_1576 PE=4 SV=1
MLYKLGVDLGGTKTEAILLDDSLNVLERKRVPTPKNNYSEILDTISNLVLELSSNTLDYSLGICTPGAI
SKKTGLIKNSNTQCLIGKSLKEDLEKKLKKTIVMENDANCFVMAESKMGAAKNFDLVFGVIMGTG
VGGGITVNGKLHSGRTNIAGEWGHHTLHRNGNPCYCGKTGCVETYISGPALEQKWELLSGESKSVP
EILSNLDNDIGKTWKSEFLENFGYSLANVIDILDPDAIVLGGGLSNIDFLYTEGKKSVYEKVFSDLVD
TPILKNELGDSAGVYGAALLN
>B2_Actino gi|221195876|ref|ZP_03568929.1| fructokinase [Atopobium rimae ATCC 49626]
MVDGPRIGALEAGGTKMVLAIGTREGKIFEREVMPTTEPAGIVSKMIAWFAEREIDALGVGAFGPTC
VNPELPDYGKILSTPKQGWVNYDFLGALRAGLGVPIGYDTDVNAACLGEALFGSARGLKNVVYLT
VGTGIGAGVLLGGKLLHGMLHPEAGHIPIEREKDDPLAMSVCQYHVSCLEGLASGPSIEKRWGLPAS
ELDEKQEVWELEATYLAKALAVYVLCYSPQRIILGGGVMKQTQLFPLIRQKLLENLNGYINTAELSA
IDSYVVSDGCSGNQGILGALALGLQSLDS
>B2_ALPHA gi|302382790|ref|YP_003818613.1| ROK family protein [Brevundimonas subvibrioides ATCC 15264]
MSVFAGVELGGTKVMVGFGSGPDDLSDPIRIPTTTPDQTLAEVERLIAGVAAHTTLQGIGVATFGPAR
LDRAAPDWGRILPTPKPGWTGAVIAPRLAEAFGVPVAFDTDVAGAAMGEGHWGAAQGLRDHAYV
TVGTGVGVGLIVNGMPLHCALHPEAGHIKVRRDPARDPFEGVCPFHGDCLEGLVSGPALAKRTGQR
GETLTADDPVWDLVADYLAQAMATLCFVAAPRRIVIGGGVGSHPTVLAATRLRLRDELGGYLPHLA
SAGAIETFLVPPALGDRSGVLGAIALARALHDTSQTTKAPDGPASQGTLEEDPHDRRQPDG
```

```
>B2_Bacteriodetes gi|268315688|ref|YP_003289407.1| ROK family protein [Rhodothermus marinus DSM
4252]
MAERPLLGGLEAGGTKFVCAVGTGPDDIRALERFPTTTPEETLGRVIDFFRRQPEPIAALGIGSFGPVD
PDPGSPTYGYITTTPKPGWAHTDVAGTLRRALNVPVAFDTDVNAAALGEQRWGAGRGLHTFVYLTI
GTGIGGGVIVNGRRHHGRQHPEIGHLFIPRLPGDDRPGHCPFHGDCLEGLASGPAIAARWGRPAPELP
PNHPAWDEVAQYLAFGLANLILTLSPQRLILGGGVMHQTHLFPRIRRHVAACINGYVALPDLDTFIVP
PALGDRAGVLGALALAEEVMSGLKHGIL
>B2_CHLORO gi|163846165|ref|YP_001634209.1| ROK family protein [Chloroflexus aurantiacus J-10-fl]
MTHAPLYGGIEAGGTKWVCAIGTGPDDIRAEVRFPTTTPAETLANAIAFFRAHEPEHLAAIGVGSFGP
VDLNPASPSYGSITTTPKPGWAHTDVVRTLHQALGRPIGFDTDVNVALLGERQWGAARDCDVAVYI
TVGTGIGGGAVVGGKLVHGLIHPEMGHMRLVRDPARDPFPGICPYHSDCLEGLACGPALKARWQTP
AEELPADHPAWELEADYLGQALANLLCILSPERIIIGGGVMSQPQMFPLVRAATQRWLNGYLQHPHI
LDHPDRLIVPPALGQRAGVLGAIALAMHTAGGV
>B2_FIRMI gi|160879723|ref|YP_001558691.1| ROK family protein [Clostridium phytofermentans ISDg]
MKIGALEAGGTKMVCAIGDENGNIIERISIPTETPEITMPKIIEFFKAAKVEAIGVGCFGPVDLNRASA
QYGYITSTPKLSWRNFDILGNLKRELNVPIGFDTDVNASALGEATYGITKGLDVSIYITIGTGVGVGV
FINGQLLHGMLHPEAGHILLERHKEDTFGGSCPYHQNCFEGLASGPAIGKRYGKSANELSDCDEVW
KLEAYYIAQALYNYTCMFSPNKIVLGGGVMHQKQLYPLIREEFKKIMNGYIDTKEVRDLENFIIAPSL
NDNQGILGCLELANREMRLLL
>B2_FIRMI_2 gi|310640969|ref|YP_003945727.1| rok family protein [Paenibacillus polymyxa SC2]
MKRDHGEGITTAMTVLGAIEAGGTKFVCGIGNERGEVLERASFPTTTPAETMENVIAFFEGKGIEAL
GVGSFGPIDPIEGSDTYGYITTTPKPHWGNYNLIGKLKEHFDVPMGFDTDVNGAALGESIWGAAKG
LDNCLYITIGTGIGAGALVGGKLVHGLSHPEMGHILVRRHPEDNYEGTCPYHGDCLEGLAAGPALEK
RWKVKGHELSVDHPAWEMEAHYLAQALMSYILILSPQKIIMGGGVMKQDQLFPLIRTKLQQLLNGY
VQHSSLTTEIEQYIVSPGLGDNAGLCGALALAKEKLNS
>B2_FUSOBAC gi|257125414|ref|YP_003163528.1| ROK family protein [Leptotrichia buccalis DSM 1135]
MAIIAAVEAGGTKFICGLGTEDGKIIERVSIPTTTPEETMAQVIEYFKDKEFDVMGVGSFGPIDPVKGS
KTYGYITKTPKPYWSDYNLIGELKKHYDVPMEFDTDVNGAALAESWWGAGENLKNVMYITVGTG
IGAGAVVDGKMLQGLTHPEMGHIFLKRHKDDKFEGRCPFHKDCMEGMAAGPAIEDRWGKKGFELA
DRNEVWDMEAYYLAQAVVNYTLILSPQRIIMGGGVMKQKQLFPLIRKYVLEFLNGYVQKEEILEKIE
DYIVYPGLGDEAGFVGSIALGKIALENNRK
>B2_FUSOBACT gi|269118873|ref|YP_003307050.1| ROK family protein [Sebaldella termitidis ATCC
33386]
MKKIAAIEAGGTKFICGIGNENGEILDKTSIPTKTPEETMKMVIDYFKDKEFEAMGIGCFGPIDPVKGS
ESYGYITKTPKPHWSDYDIVGELKKHFDVPMEFDTDVNSAALGESLWGAGQGLSSVVYITVGTGIG
AGAVINGKMLQGLTHPEMGHIFVKRDKNDPYTGKCPFHKDCLEGLAAGPAIEERWGDKAYNLEER
NEVWEMEAYYLSQALVNYILILSPQKIIMGGGVMKQSHLFPLIRKMVRETLNGYVHKKEILEDIDNY
IVYPGLKENAGLMGSLALGRLALENR
>B2_GAMMA gi|90022284|ref|YP_528111.1| fructokinase [Saccharophagus degradans 2-40]
MKNLFGAIEAGGTKFVCAVGSGPDDLTEVRFPTTTPQETLGKALAFFAPYKGQLKSIGIGSFGPVDLQ
RNSPTYGFVTSTPKPGWADTEFAGVFARLFDLPVGFDTDVNGAALGEQRWGAAQGLDNFIYITMGT
GIGGGVVAGGKLIHGLVHPELGHMFVPKQPDDTFAGQCPYHGDKCFEGVAAGPAIGARWGTPAHEL
PADHPAWDLQARYVATALASLVCSFSPQRIILGGGVMGQAHMLNSVREKTVQVLNGYIQSPAILENI
NEYLVLPGLGDKAGILGAMVLAEQAFDNR
>B2_GAMMA_2 gi|254787264|ref|YP_003074693.1| fructokinase [Teredinibacter turnerae T7901]
MNEPLYAGIELGGTKTICLIGTGVDSIRDQLQIPTTNPADTLGQIHTFLASQGELAGIGIGAFGPVNIDP
GSANYGCIESTPKPGWSHTSVVPFFRERFSCPINLDTDVNAAAIAEHQHGNGKGLRNFIYITVGTGIG
GGALIEGAPVRGNSHPEMGHIALPRHMADETFLSACPYHQNCAEGLASGSALRKRWGMPLNEFPPE
HPAWDMQASYLAEFFHSLTLLFSPQRIIVGGGVSSEQLLARVRTALYKKLNGYVDALKLEASLESYL
CLPELAGNAGPLGSLMLAYPEYRRH
>B2_PLANCTO gi|283781228|ref|YP_003371983.1| ROK family protein [Pirellula staleyi DSM 6068]
```

MTHSHQPLWGAIEAGGTKFVCVVGYGPGERLLARQQFATGANPAALMQQVVTWFLARQAEHGAL
AGLGVASFGPVDLHPASRTYGQITTTPKPGWQNADILGPLRSALAGIPIALDTDVNGAALGEHRWGA
AQGLDDFVYITAGTGIGGGGMARGRLLHGMVHPEMGHLGLPRIAGDTFEGACPFHGRCWEGLCSG
PAIAQRAGRPAETIPPDDPAWDLTIRYMAHALANITYVLSPRKIILGGSVRKAGLLGEEALFQQLRLRL
REVLAGYIASPALTQDGIETFVVPPTLGDDAGICGAIALAAQQIEPPAPSSSN
>B2_SPIRO gi|307719480|ref|YP_003875012.1| hypothetical protein STHERM_c18040 [Spirochaeta thermophila DSM 6192]
MLGGIEAGGTKWVCAVASSPDCILEEVRFPTTTPEETIGRALSFFREMEARHGRLAALGIGCFGPVDV
GEGSPSWGFITTTPKPGWRDTGVAGVFQEELGVPVGFDTDVNAAALGECSYGAAQGLANVVYLTV
GTGIGAGVVVDGRPVHGLVHPEAGHVLVRRHPDDTYGGRCPFHGDCLEGMASGPALAERWGMRG
EEIPPEHEAWRMEAFYLAQGVMNLVLAVSPERVILGGGVMQQGHLFPLVRSELLRLLAGYVGHPAIT
EQVEDYLVPPGLGQRAGIVGALVLAARTAP
>B2_Thermobaculum gi|269925285|ref|YP_003321908.1| ROK family protein [Thermobaculum terrenum ATCC BAA-798]
MSSNNQVYGGIEAGGTKFVCAVGTGPDDIRSSITFPTTYPEETLEKAVSFFRQATKGLTLVAIGIASFGP
VDPNPLSPTYGYITSTPKPGWANTNFAGYIRESLGVPVGFDTDTNVAALGEHRWGAAVGLETFIYLT
VGTGIGGGGLVNGKLIHGMLHPEMGHLAIPHDKEEDPFEGWCKFHKDCLEGLASGPAIESRWGVRG
ENLPADHPAWQLEAKYLAFGLVNYIYTLSPQRIIMGGGVMQQSHLFPLIRQRVQHLLNGYIQADDVL
QRIDQYIVPPKLGTRAGVLGAIALAQDTVGD
>B_Acidobacteria_4 tr|C1F498|C1F498_ACIC5 ROK family protein OS=Acidobacterium capsulatum (strain ATCC 51196 / DSM 11244 / JCM 7670) GN=ACP_1127 PE=4 SV=1
MRRIDLTRAELASSETARRINRDILLEIIRARQPISRADLARVSGLQRSTVSQIIEQLIEELWIREGAVAR
LPRGRRPTMLVLNEDVVVLTADLHPRKAAIAVIDLNGRVLSHAELSLSHDPAKVATAILECMRRLRQ
AHPEKSFRGVGVSLPGRVDSQTQRLIFAPNLHWPDFDLRQALAKGLRMPVEMENAANACLTSELWF
GRMQGVQNAVLITVSEGIGGGILANGQLVTGQNGMAGEFGHISLDPRGPRCGCGQRGCWETFASCK
AALRHYQESSGASRRIAYQDLLALGAEKNPHAVGALTEQARQIGRGLRLVIASLSPELVLIAGEVTSA
WNLVAPALRKEMEAQWLGGTMPRIEPTFDSDAARLRGAAAMLLQRRASAIDRNGKVKTKPKPRT
>B_Actinobacter_2 tr|Q9RFM6|Q9RFM6_STRLI Putative xylose operon regulator OS=Streptomyces lividans GN=xylR PE=4 SV=1
MSAPPHEAQPARPGRALPDTQQGMRRRNLSRVMHTVSAEGPLSRAAVASRIGLTRAAVSTLVDELIR
SGLLEELGPERPGRVGRPGSALAVSGQGPAGIGAEVGVDHLAVCAVDLRGRVRARAVRYGSNRGRS
PEPVLEQLTGLVRQVVSQAETEGLWPAGLAVAVPGLVARDGRTVVRAPNLDWHDADLGALLPADLP
PTVDNEANFGALAELWLGDGTPRDFLHVSAEIGIGAAVVFDGRLCAVPAGSPVSWGMCRSIPTGPGA
PAVGAGAWNSTPVRRRSCARRAWSRGEDRVGLLAGRAAEGDEDVRRALREAGTALGIALTGAVNL
LDPEAVVLGGALADSRPGCCRRCGTNWPGARRVPPARSRCPSWVRRAAARSRALRGPAVLDDPGA
VAERA
>B_Actinobacter_4 tr|A1A0I4|A1A0I4_BIFAA Transcriptional regulator OS=Bifidobacterium adolescentis (strain ATCC 15703 / DSM 20083) GN=BAD_0436 PE=4 SV=1
MTAKRRILAFEIGTFFTRYVVFEDGRMGIPGTVATPVDSVESFYQALAHIVNGQRAPLDGIAMSVPGF
IDVSKQVAVTAGALGMLYKHEIGKELQEYLDKPVPTWMENDANCAAMAEKLSGNAVKLDDFALITI
DTGIGGALFLGGGIRRGKDWRAGELGMMIPNYETGGFNTMQNYLSTIVLAEDYAKEFDVPTGSIVP
ATLFRRLDEPRVRKIVDKWIDYLAIAIFNTAAATDPECILLGGGICREQQLLPMVNAALDRIPQWGDF
RTSVKRCRHTNNAGLIGAYYAFETEVGGLTDVPIR
>B_Alpha gi|13476027|ref|NP_107597.1| glucose kinase [Mesorhizobium loti MAFF303099]
MSIEAALAIDLGGTELRAALVDRDGKILAFAAVPTQAQAGPDVVIGQIEALAATVHAEAPGLAIVGV
GVGAPGPLDPLAGIAVGPPTLAGWQDVPLADILERRLGLPVRLENDANAAALGEWRFGAGHGARS
LVFVTVSTGIGGGVVADGRILHGRRGLAAEIGHMTITNEGERCVCGVVGCFEAIASGTALGRRANAA
TSAFDGSTLRRLSANAEVTGRHVVEAARLQDDLALALLEEEARWLGVGFTNLLHLYSPDVLVVGG
GIANGLDLMHPVIEATIRQRAMRAYRDVPVVQAQLGRHAGLVGAASLVLFDDGSLAARMPVGPSTF
PEARRDFNG
>B_ALPHA_13 tr|B9JLU4|B9JLU4_AGRRK Transcriptional regulator protein OS=Agrobacterium radiobacter (strain K84 / ATCC BAA-868) GN=Arad_7079 PE=4 SV=1

MTEPAVATSSPRGIRQRNELAALRALYQFGPLSRAELARRLGLNRSSSGHIIAGLTLDGLVREVSDGD
QARGGHAHAGRPGIMIELVPGAVYFLGIEIGVEHISAVEIDLGGNIVSTTVEPFDGASAGVAATVERAI
EMVLGAIPSSRLERCEGIGVAVPAQMDKHGFVRLAPLLRWENVQLAELVRESLPVSVPVVAENDGN
AFAIGASYGRNDKHSGVTLFLVMESGVGGGIIANGSLFRGANGLAGEIGHLRISSASEPSRSLEEVLG
LEHIMTEYRKVPAVAAPTFGHFLADVRDRVPGAVSIAEEWARALAFGLIQACRVIDADRIILGGSVAA
LYPLMAARVAHHIQLAQEASFPLPSIGVNEEETVGPAFGAACILHQRFLSLESQRFAEEAG
>B_ALPHA_14 tr|B9JL48|B9JL48_AGRRK Transcriptional regulator protein OS=Agrobacterium radiobacter (strain K84 / ATCC BAA-868) GN=Arad_9637 PE=4 SV=1
MKKVALAFDLGGTELRAALVDEEGNLLSFSAVPTQAAEGPDAVVRQIEVLAATVLAKTPDLLPIGIGI
GAPGPLDPEAGVVIAAPTLTGWNEVPLADILSSRFQLPVRLENDANAAAVGEWRYGAGRGARSIVF
VTVSTGIGGGVIADSRILHGRRGLAAEIGHMTITNEGERCFCGAVGCFEAVASGTALGRQATARTRRS
DGSMLRKLSANADVTGRHVVDAARNGDALALELLNAEARWLGIGFTNLLHLYSPDVLVMGGGISH
GFDLLHDTIMATICDRAMPAYRDVPIVAAQLGRHAGLIGAASLILGSEAQASSEPSSLVGSGNPDADIS
AGTKEASNG
>B_ALPHA_15 tr|Q982N1|Q982N1_RHILO Transcriptional regulator OS=Rhizobium loti GN=mll8569 PE=4 SV=1
MALRGTNQEFGRPYNRRIVLESIRLHGPIARGEIASRVGLTVQTVSTIVRELEEQGYILSVREEPRGRG
LPPATLRINPEGGYAVGIHITPLGINAALINLSGDVIESTYREAPNATPDHAFEMIGAMVIELTGLRAGG
RVLGVGLALPGPFGVESMSFVGPTTMTGWQDVALRERLAASTGLPAFFETDMAAAAMGERLYGLG
AQFSEYYYLYFGVGLGGVMVHEGSVQRGAWGNAGEVGHIPVVPGGEACPCGNRGCLERYLSLEAL
RRWNGTEADWVAEVAPIFHNAVAVIENLFDPETVILGGLASTDLLERLAGSTGGLHNSVSARKDRIAP
RLMVARGGQHSVLRGAAALAVSGVLSPRFGQMFAAERERGRDLLTAKEIAA
>B_ALPHA_19 tr|Q2RR18|Q2RR18_RHORT N-acetylglucosamine kinase OS=Rhodospirillum rubrum (strain ATCC 11170 / NCIB 8255) GN=Rru_A2630 PE=4 SV=1
MTLRLGLDLGGTKTEIIALDDEGRILLRRRRPSPRAAYGATLDCLAALVTEAEAELGRQGSVGVAMP
GAISPASGLVKNANSHWLNGQRLDHDLAERLGRPVRVANDADCFALSEATDGAAAGASSVFGVILG
TGVGAGIVVNGRLLAGPNAIAGEWGHMPLPWPGDDERPGPDCYCGLKGCVETFCSGPGLAADHQ
KSTGHAIEGPALLALAQAGDAQAQASLDRHADRLARALAVVINILDPQVIVLGGGLGQMPHLYQAL
PRLWTPWVFSDRVDTRLVAPRHGDSSGVRGAAWLWPPSSTLETDPHA
>B_ALPHA_2 tr|B8GZM5|B8GZM5_CAUCN Glucokinase OS=Caulobacter crescentus (strain NA1000 / CB15N) GN=CCNA_02549 PE=4 SV=1
MIRFGVDFGGTKIEVAALNAAGDFVARVRKPNPGNYDEALEVVAALVADAESMAGGSCARLGLGIP
GSISPRTGLIRNANSTYLNGRPFGENLETRLARPVRIANDANCLALSEAADGAGAGASVVFAAIVGT
GCGGGVVVDGKIINGHNGIGGEWGHAPLPWPKPEEYPGPDCWCGRKGCLETWIAGPAFARDAGFA
NGQAAMQAIGEGDVKASAALDRYVDRLARSLAVVCDLIDPDIIVLGGGMSNVDALYERLPAAIAPH
VFSDIFETPVRKAVHGDSSGVRGAVWLWPPEV
>B_Alpha_2 gi|222082872|ref|YP_002542237.1| transcriptional regulator protein [Agrobacterium radiobacter K84]
MKKVALAFDLGGTELRAALVDEEGNLLSFSAVPTQAAEGPDAVVRQIEVLAATVLAKTPDLLPIGIGI
GAPGPLDPEAGVVIAAPTLTGWNEVPLADILSSRFQLPVRLENDANAAAVGEWRYGAGRGARSIVF
VTVSTGIGGGVIADSRILHGRRGLAAEIGHMTITNEGERCFCGAVGCFEAVASGTALGRQATARTRRS
DGSMLRKLSANADVTGRHVVDAARNGDALALELLNAEARWLGIGFTNLLHLYSPDVLVMGGGISH
GFDLLHDTIMATICDRAMPAYRDVPIVAAQLGRHAGLIGAASLILGSEAQASSEPSSLVGSGNPDADIS
AGTKEASNG
>B_ALPHA_21 tr|B9J8H0|B9J8H0_AGRRK Sugar transcriptional regulator protein OS=Agrobacterium radiobacter (strain K84 / ATCC BAA-868) GN=frcR PE=4 SV=1
MSPPQDPEETPTSPLILDPSGGANQVRVRAHNERLVLSLVRRHGALSKADIARRSGLSAQTVSVIMR
ALEKDGLLSRGEPVRGRVGQPSIPMRLNPDAVLSFGVKIGRRSADLVLMDFVGQIRMQLHQIYPYPL
PQDILSFVTSGIQELESRLNTEERGRLAGLGIAAPFELWNWAEEVGAPDGAMEVWRGVDLQAEIAA
RVPYPVYLQNDATSACGAELVFGVGPHYPDFVYFFIGSFLGGGIVLNSSIFVGRTGTAGALGPLPVRG
RNGETLQLLEIASIFVLENLLRDHGFDPQPLWYSADNWIDFGEPLEIWIQETAKALAQAIVAAASVID
FSAAVIDGGFPNWVRERVVRATIKEAAELDLQGVVMPEIIEGMVGPQARAIGGASLPIFARYLIDQNI
LFKEIEHAEGH
>B_ALPHA_27 tr|B8H3N8|B8H3N8_CAUCN Fructokinase OS=Caulobacter crescentus (strain NA1000 / CB15N) GN=CCNA_01192 PE=4 SV=1

```
MAGRPMSRIAAIELGGTKVMVAFGSGPDDLSPPLRIPTTTPAETLARIEDALAAEQGRFDAIGVASFG
PIRLDPAAPDWGHILKTPKPGWSHADVAARLVRRFDRPLALDTDVNGAAVAEGLWGAAKGLGDYA
YVTVGTGVGVGLVVNGAPTHGLLHPEAGHILVRRDAALDPFTGSCPFHGDCLEGLISGPALAARTG
APGESLSKDDPVWALVADYLAQLVANLALIASPRRVIIGGGVGGNPQLLEQTRTRLQTHLAGYLAPL
EQRSDIDAFVAAPGLGANSGLLGAVALGLRHDAILRQDLMP
>B_Alpha_3 gi|116248710|ref|YP_764551.1| putative glucokinase [Rhizobium leguminosarum bv. viciae
3841]
MQQVAIGIDLGGTQVRAALVDEQGRILARAAEPTDALAGPDRVLAQICGLTDGLLAASNPASVVGV
GVSAPGPLDTVAGVASNIPTLSGFVDFPLKAELQKRFPFPVDLENDAIAAAIGEWQFGAGKGLDNLV
YVTVSTGIGGGVVSDGRVVRGRKGMAAHVGHMSVVPNGELCPCGNRGCFEAYGSGTAFARRAQM
RAMETSGTTIGSDGGAIDSRGVFAAARDGDRLANQLVDEEAEILGRGFTSLIHIFSPDIIVMGGGLSHE
FDRLQPGIQAYITQWAMPAFRDVKVVLAALDQNSGLVGAAALAFLTGKVPAIDQL
>B_ALPHA_3 tr|B9JJT4|B9JJT4_AGRRK Transcriptional regulator protein OS=Agrobacterium radiobacter
(strain K84 / ATCC BAA-868) GN=Arad_9050 PE=4 SV=1
MKLKGDQSTSRAMNRRLILNLLRQEGPKSRAEIATITGLSPAAVTFVIADLIEEGHVTEGKAVAGFSG
RRPIPVEINYTNGLAIGFKLMVGSVECVVTDLATSPLASMRLTLADHDPENVAKTLAAAVPLLVQYA
ARPNAQLAGIGISMPGVIDNDQATCVRSHRFNWNNVPLASILAQKVKVPVWLEDDTNAYAIAQQLF
GVGRQHRNMAVLAIGVGISCALIIEGKLYRGANGAAGKFGHTLHEENGRLCECGKRGCLMAYHSQ
TSMLRTWRETTNRSNSDGLPEMLNAVEEGEAAVGDILREAGIGIGKALANLVNVTDPEVIVVGGEAV
SFGEAFFEPLRSTLAAHTFRASPPLLPDWEDNSWARGAAALVTQKFFDFETSGGTTGSTEALGTTSA
A
>B_ALPHA_6 tr|A4WP29|A4WP29_RHOS5 ROK family protein OS=Rhodobacter sphaeroides (strain
ATCC 17025 / ATH 2.4.3) GN=Rsph17025_0233 PE=4 SV=1
MTAGRPAVPGREMTVRQGQETAERPAAAEPGLRGSNQSGMRARNERLVLSLVRRAGALAKSDIAR
MTGLSAQTVSVIMRELERDGLLCRGEPIRGRIGQPSVPMSLAAEGAYFLGLKIGRRSAELVLIDFLGQ
PVAARRRIYSYPTPDAVVDFTAEALPLLTGSLPPEARARVGGMGIAMPFQLWNWVQYVGAPQSEMD
AWRERDIQTELAALTGLPVHLRNDATAACGAELVFGTGERPKDFLYFYFAYFIGGGLVLNGHLYAGR
TGNAAGVGPMPVPDRSGRMRRLLNVASMSVLAEAMEQAGEDPAALWENPDHWEVSAPVLSDWL
DGAAAGLASAILSASALVEMETVMIDGWMPPEIRAEMVRRTHAALHRLDLSGIEPPQIREGTVGAQ
ARALGAAAIPLSQRYLVEAWGAAADG
>B_ALPHA_7 tr|A4WWD6|A4WWD6_RHOS5 N-acetylglucosamine kinase OS=Rhodobacter sphaeroides
(strain ATCC 17025 / ATH 2.4.3) GN=Rsph17025_2813 PE=4 SV=1
MLIAFDIGGSRIRAARVFAPDHLEPLGELPMPPSFPGFIRALGQLVPRASRTVAISIAGVIDPASGRLTSA
NLPAVNGRRLAGDLAAALGRPVWIGNDADCFVLTEARRGVARGHRNVFGIILGSGVGGGLVLDGAL
VAGAGGITGEWGHGPVLDQRPLGHDLPRLRCGCGQTGCLDTVGGARGIERLHLHLSGRGLDSREIL
SAWRAGEMAAAETVSVWLELVSGPLAVVLNVVGSSVVPVGGGLANDRDLVAALDRAVRQRLLRPT
TEPVLRPAIHPEPGLVGAALAGLETFG
>B_Aquificae tr|O67468|O67468_AQUAE Transcriptional regulator (NagC/XylR family) OS=Aquifex
aeolicus GN=xylR PE=4 SV=1
MKKGIDIGGTFIKVLWEDGKKEKHYIKDLKKDREALMEKIREVIKAGNPSGVGIAVAGFTSLNGIVY
KSPNIPALDKTDFRKVLEDINIPFVVGNDVNLGAFGEWYYDYRDSKILLLVAVGTGLGTGLVYKGEV
FFGVCGSALEMGHHTVKKGGELCNCGRKGCWEAYCSSYAIEREYKKLTGRTLKDYEIAEKAKEGEE
EALKVVESFKDYLITGLMNGVHIFNPDRIVLAGGVIETFKEFLRDVPEKVKEISEELPALCLKISFSRA
GEFMGARGALAYALKYL
>B_Aquificae_2 tr|C0QRM3|C0QRM3_PERMH Xylose repressor OS=Persephonella marina (strain DSM
14350 / EX-H1) GN=PERMA_1552 PE=4 SV=1
MRRYLGIDIGGTFIKFVYKKGDDIEKGKVYIREIISKNRPDLIVDEIRKIVKKYRPDILGVAVAGLIDKK
TGVLTASPNIKPLENFPFKDELENSLKIPVYIENDASLAAYGEYLYGAGKGSEILICLTLGTGLGGGAVI
NGKLLTGVSGSAMEIGHTTIEMDGLPCHCGRKGCLESYVSSYGLERIYYLYTDQKISSSQIITLANEG
DLTAMRSMERFSEYLSVGLMNIVHIFNPDRIVLAGGITENYPAVVDMAVSNLKNIAFHLPFRDLTVKR
AELKEFSGAYGALGYAENESR
>B_Bacteriodetes tr|A7V3D0|A7V3D0_BACUN Putative uncharacterized protein OS=Bacteroides
uniformis ATCC 8492 GN=BACUNI_02075 PE=4 SV=1
```

```
MNQALLKEIEMGSKNALLKKRIITHYIYNGSSTITDLSKELDLSVPTVTKFISEMCDDGYINDYGKLE
TSGGRHPSLYGLNPESGYFIGVDIKKFAINIGLINFKGDMVELKMNIPYKFENTQEALEKLCKLISNFT
KKAGINPEKVLNICINISGRVNPESGYSFSMFNFSELPLAEVLAKKIGYSVCIDNDTRAMTYGEYMQG
CVKGEKNIIFVNVSWGLGIGIIIDGKVYTGKSGFSGEFGHINVFDNEILCHCGKKGCLETEASGSALH
RILLERINNGECSILSSRIATKEDPLTLDEIIAAVNQEDLLCIEIVEEIGQKLGKQIAGLINLFNPELVIIGG
TLSLTGDYITQPIKTAVRKYSLNLVNKDSVIITSKLKDKAGIVGACMLARSRMFES
>B_Bacteriodetes_6 tr|A7V6R2|A7V6R2_BACUN Putative uncharacterized protein OS=Bacteroides
uniformis ATCC 8492 GN=BACUNI_03275 PE=4 SV=1
MLNTLMLYTMKLSIDLGGTNIRIAQVEKGNCLNKVSVPCLAQQDASTVLNQLSQLIRNMMNEQVD
GIGIGVPSIVDPEKGIVYNVANISSWKEIHLKEILENEFKVAVAINNDSNCFTLGESLYGEGKSYTNMV
GVTIGTGIGAGVVIGRRLYGGQYMGAGEIGSFPYLDSDFEHYCSSFLFKRYGTTGAVVAEKAQQGEQ
AALEIWKEFGRHLGNLIKAILFAYAPQAIVLGGGIVSAFPFFKNAMEQTMQSFPYKIISDNVSVVASH
QKDSSLLGAAALLE
>B_BETA tr|C0YC21|C0YC21_BURPS ROK family protein OS=Burkholderia pseudomallei Pakistan 9
GN=BUH_4484 PE=4 SV=1
MRSPHLGQGSNSANVRRYNERLLLKTLRRAGSASKADLARLANLTGTAVGSIIASLAEAKLIEFAGR
RAEGQRGQPASLIRLDPRGAFGIGVRLDRMRIETALVNFAGDVIGRRSHDTLLPQPAAVLEIVREDVA
AMQQLLCAHERARLTGVGVAQPYNLGSWLREIDLPADAFGAWAEVDFARELDRAIELPVFSENDGN
AAAIAELFYGCGRQRDDFVYLFLGPAIGGGIAVDGDCLRGVTGNAGDFAMMPVPPSRLPSVPKPSG
AWDLLITRASLNGLARHLRHRGETLESRADLEACIARGGQAVDEWIDDCVDALAPALRAVLAVLESP
VVVLDADTDAGLIDTLIARLHTALAALAPEARGTPVLVRGTFGPDAGAIGAATLPMFFNFSPRAGILR
GAGVDSQEARYAAF
>B_BETA_2 tr|Q3SGS8|Q3SGS8_THIDA ROK family protein OS=Thiobacillus denitrificans (strain ATCC
25259) GN=Tbd_2216 PE=4 SV=1
MAPSLRFGIDLGGSKIELIALDRDGREILRRRVPTPQGDYPATVAAIAALVGHAESGLGQQGSVGVGT
PGAVSPASGRMKNCNSTCLNGQPLREDLERALGREVRVANDANCLALSEATDGSAAGAESVFGVIL
GTGVGGGVVVHGRLLQGANAIAGEWGHSPLPYFQFAGAQADRAMTGHHPATGEAIVHPWPQPREL
DAAPACYCGKKGCIETWLSGPGLAADHVRYGGEDLPAHEIVQLANAGYGPCSATLARYEERLARAL
AGVINLVDPDVIVLGGGLSNIARLYDTVPRLWPRYVFSDRVDTKFVPPKYGDSSGVRGAAWLWE
>B_Chlamydiae tr|Q6MCP0|Q6MCP0_PARUW Putative glucokinase OS=Protochlamydia amoebophila
(strain UWE25) GN=glk PE=4 SV=1
MLKETDCVIGIDLGGTKIGIGVLNVSGTLIDSVRLKTDFKHGPASVEKQIMQAIQDLKNRTKVEIKGI
GIGVAGQIDEETGVVRFAPNLPGWHQVTLRKNLEREAEIPVKVVNDVRAITWGEWLYGAGKHYQD
LICLFVGTGIGSGIVCQGKMQKGDNNTFGEVGHMTIDFHGPRCTCGNNGCFEAFAGGWGIARQAKE
LILADNQSGQSILEKAGGYLENVSAKAVIEAYHSGDPLALLILEKVKQALIAGCINLVNAFNPACLILG
GGVLDGIPEILSFIDKGIRETALKTATDKLQIKTALLGKNVGIIGSGAVILDVLKNNVG
>B_CHLOROBI tr|B3EGU5|B3EGU5_CHLL2 ROK family protein OS=Chlorobium limicola (strain DSM
245 / NBRC 103803) GN=Clim_2184 PE=4 SV=1
MSQWAIGIDLGGTAVKAAIVSRKKGILKNRTVPTDTASGPEGIVSQLAVMIASLYTEASAELSRQDFS
GIGFGAPGAVDIEAGTLSYPPNLPGWTTFPLRSELERALLAKLPKSVPVVIENDANAAAYGEAVYGA
GRNFRDFLMVTLGTGVGGGIVLNRKLYRGPNGTAGEIGFMIVDFQSPAVHAGIHGTIEGMIGKERIVE
YACGLIRDNPEAGSLLASLCGQDFSSLSPRHIEQAAKMGDQLSLAVWNHVGAILGTGFACVTSLMDI
RKFVIGGGISAAGTLIFEPAYRQLLRSTLPSMHDGLELVPAELGNSAGIYGAAALCFS
>B_Cyanobacteria tr|Q065Y2|Q065Y2_9SYNE ROK family sugar kinase OS=Synechococcus sp. BL107
GN=BL107_14345 PE=4 SV=1
MSSDQVIGIDLGGTAIKLARFNRSGDVLAELQIPTPQPPVPGAVTMALCEAIDQLDPDHLADLVGVG
LPGPMDANARVARVCINLSGWDDVPLADWLEPRLQRQVTLANDGNCAVVGEAWLGAARGFEDVV
LLTLGTGVGGGVILGGALFTGHNGAAAEPGLIGIHPDGPPCNSGNQGSLEQFASIAALRRLCDVDPRE
LSDRADAGDLQAKAIWQQYGTTLGVGIASLVYVFTPQRVLLGGGLAAASAHFLPDVRREVEARVQ
AVSREGLSIEACALGNGAGRLGAARLALQRLSG
>B_Cyanobacteria gi|119490672|ref|ZP_01623077.1| ROK [Lyngbya sp. PCC 8106]
```

MNNKILALDFGGTKLAAATLISGQQKWQNTRRQLSGNSPNATTDLEIMRSLIRELLGNEKPLAIGISF
GGPVDAKTGIVRLSHHVSGWENVPLREILEAEFNVPVRVDNDANVAALGEYKFGAGQGCESLLYIT
VSTGVGGGWILNNRLWQGTQGMAGEIGHTVVQPDGPLCLCGKRGCVERLASGPYIAQQVRQSLNK
QPQQGEILRHLVNHQLDKITAEIVSQAANEGDELCQNAIKLAGWAIGVGIGNAANLINPQRFILGGGV
TKAGDLFWHQVRQISRQTALPEVDFEIVSAQLGDEAPLWGAVALAETGLEN
>B_Cyanobacteria_5 tr|Q3MAV3|Q3MAV3_ANAVT ROK OS=Anabaena variabilis (strain ATCC 29413 / PCC 7937) GN=Ava_2265 PE=4 SV=1
MISAVVVWCEELNVGNSQVIGIDLGGTAIKLGRFSEDGTCSQSLTVETPQPATPEAVFLVMVDAIAQI
DPDNETIAIGVGTPGPADAQGRIAQIAINLPQWENVPLADWLETKTNKPTVIENDANCAGIAEAWLG
AGRHYQNFIMLTLGTGVGGAIFLDGKLFVGHRGAAGELGLITLQPDGPMCKSGNPGSLEQYTSIKAI
RRLTGKEPAELGMLAQAGDIEALQFWQEYGQNLGIGLTSLIYVLTPQAIVLGGGVSASHEFFLPTLKA
EIEQRVMSTSRAGLQILPAELGNSAGMVGAARLAWKKFGNG
>B_DELTA_2 tr|A0LFI7|A0LFI7_SYNFM N-acetylglucosamine kinase OS=Syntrophobacter fumaroxidans (strain DSM 10017 / MPOB) GN=Sfum_0489 PE=4 SV=1
MLRLGVDLGGTKIEIIALDDGGNELARFREATPRGDYRSTLEAVARLVSKVEAELGGERGSVGIGTPG
AVSRATGLLRNSNSVCLNGKPIARDLEAVLGRPVRIANDANCFALSEYRDGAARGARVVFGAIVGTG
TGAGIVVGGEVLEGVNAIAGEWGHNPLPWPKDDERPGPRCYCGRYGCIETFLSGPGMAALHALSTG
DRLPPEEIVARADEGDPSCEHTLLRYEDRMARSLAHVINILDPDVIVLGGGMSNIRRLYRNVPRLWG
AYVFSDRVDTRLVPPRFGDSSGVRGAAWLWG
>B_DELTA_4 tr|A0LJN6|A0LJN6_SYNFM ROK family protein OS=Syntrophobacter fumaroxidans (strain DSM 10017 / MPOB) GN=Sfum_1953 PE=4 SV=1
MSSKRYAVGVDLGGTKIAVALVDDRGEVLKHARYLTFVREGPEAVRDQIIGAVKEIRKGTKTRPAGIG
IGVAGQIARDDGMVRFAPNLGWRNIPLGEQLRAITRLRVVVVNDVRAAAAGEWAFGAGKDCGDLI
CMFVGTGIGGGIVAQGRMLHGCGNSAGEIGHVVVDMNGPLCHCGRRGCMEALAGGWAIAQKARD
AILLDPALGTPLLRLAKGQINNVTTELVATAFRMGDPLARQLIDRAAEALSVGAVSLVNAFNPCRLIL
GGGVVNGLPELIERVREGIRHHALVTAAESVTVVPASLGDDAGVIGAAVLAMQSPA
>B_DELTA_5 tr|C4XQQ9|C4XQQ9_DESMR Fructokinase OS=Desulfovibrio magneticus (strain ATCC 700980 / DSM 13731 / RS-1) GN=scrK PE=4 SV=1
MGEKAFRGAVEAGGTKFVCAVGAGLDDIRRPENRAVFPTGDDPAATLAAVAGWLKSREDRRGAPL
AALGVASFGPVCLDKALPDYGRVMTTPKPGWSGFDLLGALGRAFPGRPIGFDTDVNGAALGEREW
GAARGLDDFLYVTIGTGIGVGGLAGGRLLHGLTHPEMGHIGLRRLAGDVFPGVCPFHGDCWEGLCS
GPAMAARTGTAAEDLSADHPAWEHEAAYVAEALATVTYALSPRRIILGGSVPKGGRLGEEGFFAKVR
ERFVATLGGYLPDERLTARVAEYIVPPGLGELAGVAGAYCLACAAPGGREGETRPKP
>B_DELTA_9 tr|Q1CYA3|Q1CYA3_MYXXD Glucokinase OS=Myxococcus xanthus (strain DK 1622) GN=glcK PE=4 SV=1
MPTLGIDLGGTFARAAVVDEVGKLIASSKVALVERSPSGVVETIAQAASDAVMAAGVPLGACGVAA
AGQIHKDSGVLSVAPNLGWRNVPLGALLTDRLGQPVRVVNDLAAAAWGELHAGAGRGSQDMLVV
FVGSGVGSAIIAGGRLVDGGGGVAGELGHIKVVPGGRRCGCGELGCLEAYAGGHNLIAQTRELLAS
GGAPEVARLTGGDPARITPVTLEQAAEAGDVAAGEVYARAAQFLALAVANQVTMLNPARLVLGGGV
LRHCPGLRRRVEEGVRAWSSTTSREGLLIADAELGDDSGLIGAALLVK
>B_Dictyoglomi tr|B5YAN7|B5YAN7_DICT6 Transcriptional regulator, XylR-related OS=Dictyoglomus thermophilum (strain ATCC 35947 / DSM 3960 / H-6-12) GN=DICTH_1696 PE=4 SV=1
MVFIRTGRGGRASQQREVNKTIILNAFLRCRNISRTEIAKRFNLSKSTVTRLVNTLIDEGMIIEIEPSSE
KKIGRKAIVLGLNPEYKNALAIKVGVTHTYLSKIDFAMNIKEIKSFLTPKNPQDFLDKLEEYTKMLFP
EGLNKTHAVGIGIPGIVDNTFKNVVVAPNLNWKNLPLGDMIADRIKKVFSVEIPVKMDNEANMAVV
AEGMLGTKIEYNDLNIVYVYVGEGIGTGLILDGKLYRGRYNTAGEFGHMTVMKDGLKCKCGNLGC
WERYASLGSEIALKAGFDSEKETVEVSDEKVLKKYINELAIGLINIINGLNPDVIILGGPLIKKDTKEV
WEFIRQEIKKIVEEKSITSDAGKVRIELTSFLDYPAELVGAGIWAFWDIFEGPVLSTV
>B_EPSILON tr|A6Q9R4|A6Q9R4_SULNB Putative uncharacterized protein OS=Sulfurovum sp. (strain NBC37-1) GN=SUN_1270 PE=4 SV=1
MSKLYIDIGGTYLRSELLKNGKTFKEKVSSRGISLSEYLEQKLGAYPDIAEIGISFAGQVDHGKIVSSP
NIAVKEYDIKKYIEKKYPVSLKIDNDLNCAMLAEKEDIKRKNMALLYIGTGMGSAVLEQGEIVRGER
NLAYEIGHVPFKKAPFRCGCGKDNCLELFSSGSGLKKWYTYYGLPQMTLEELRKSKDKYAKKIYQN
FQEGLFRAAATLVTLANPKVLVLGGGVVSANHFLKEKVEKKIGKYALASNLEGLEIKLSQLKNASIE
GAKQL

```
>B_FIRMIC tr|A5LLF6|A5LLF6_STRPN ROK family protein OS=Streptococcus pneumoniae SP6-BS73
GN=CGSSp6BS73_09359 PE=4 SV=1
MTYYVAIDIGGTNIKYGLVDQEGQLLESHEMPTEAHKGGPHILQKTKDIVASYLEKGPVAGVAISSAG
MVDPDKGEIFYAGPQIPNYAGTQFKKEIEESFTIPCEIENDVNCAGLAEAVSGSGKGASVTLCLTIGTG
IGGCLIMDRKVFHGFSNSACEVGYMHMQDGAFQDLASTTALVKYVAEAHGEDVDQWNGRRIFKEA
TEGNKICMEGIDRMVDYLGKGLANICYVANPEVVILGGGIMGQEAILKPKIRTALKEALVPSLAEKT
RLEFAHHQNTAGMLGAYYHFKTKQS
>B_FIRMIC_14 tr|B2TPN7|B2TPN7_CLOBB ROK family protein OS=Clostridium botulinum (strain
Eklund 17B / Type B) GN=CLL_A3008 PE=4 SV=1
MANSGTIKNININSIRNILDSNKSMTKSDIARYTGLSFPTVSNTIEYLLEKGEVIDCGLKDSSGGRCAK
NYSLNPMYLVSLSLYLEGFEIYWFITDHCGNKIQDGRKNCKNKILKCIEDTIISMKLNYSQLASIMIGI
ASNINSGKITSHMEYEELYDIDVIKYLKDKYEIPINVENDMKVAARGYWARHENKDIQAVASIYMGK
NGMGSNMVIDGKIWSGTSNFAGEIHYLPINGDSKKYPMYEFNDIDTNIVQYYGKIIQSYIALINPNLIV
LYSNSYIIDKLEEIKFYCKCRIPENAMPKIIISDEFIEDYEYGLSKMANELMD
>B_FIRMIC_15 tr|B2TQR4|B2TQR4_CLOBB ROK family protein OS=Clostridium botulinum (strain
Eklund 17B / Type B) GN=CLL_A3335 PE=4 SV=1
MIKKQVNNMEVKKNNRNRIFRYLCKCDKTSNSEIAYELKMSVPTVAQNTKELMERGLIEETGEFQS
TGGRRAKAFSVISDSRLAVGLDITKNHFGLLLTNLKGEILKYDRFDYQYSNDKLYYCEIDKKIEEFLG
KNLRNRETILGIGISFPGIVNLEKEIVSYSHMLGLQTLLFTEVSQFFSYPCCFLNDANAGAYAEGINKE
FQKRFFYLSLSNTVGGAIFNCDELIQGENFRCGEVGHITVIPDGVPCYCGKLGCLDVYCSAKNLSDV
TDGKLSLFFTALERKEKEILKIWDKYTTYLSVAINNIHMVLDCDIILGGYVGSYLENHINDIRQKVLQ
RNTFSEDGMFVKTCNYKVGAAAFGAALKVIEAFIKQV
>B_FIRMIC_17 tr|B3WBA9|B3WBA9_LACCB Transcriptional regulator OS=Lactobacillus casei (strain
BL23) GN=LCABL_29280 PE=4 SV=1
MQYLAIDVGGTTTKYGLVSATGALSQQGEQPTMRASLADFMASLTGLIRLHKAHVAGIGLALPGVI
DSQQGLVKASATLPFLEGLVLGTQLTTAADLTVPILIENDGNAAALAEHWRGNLAGTMNSAMVVLG
TGVGASLFLNGQLYHGSHHVAGEPSFMVTNGLTPIMREQTAAGLSAVATINAMADTLGVHEEPIGQR
VFQALTDNTSEAAVILRTFTRGVAAMIYNMQTVLDLEKVIIGGGISAQPRVIKEIRDDIEAYQQVTSLS
ARTIRLPVVEPAKYRNAANLIGAVAPLVVRG
>B_FIRMIC_2 tr|A5LN49|A5LN49_STRPN Fructokinase OS=Streptococcus pneumoniae SP6-BS73
GN=CGSSp6BS73_07779 PE=4 SV=1
MTKLYGSLEAGGTKFVCAVGDENFNVVEKTQFPTTTPIETIDKTIEFFSKFDNLAGLAVGSFGPIDIDK
NSKTYGFITTTPKPNWANVDLLGALRRALNVPMYFTTDVNSSAYGEMVARNNAGGRIENLVYYTIG
TGIGAGVIQRGEFIGGVGHPEMGHYYVARHPMDIEKEFKGVCPFHKGCLEGYAAGPSLEARTGVRG
ETIELNNPVWDVQAYYIAQAAVNATVTFRPDVIVFGGGVMAQQHMLDRVREKFTSLLNGYLPVPD
VRDYIVTPAVAGNGSATLGNFVLAKEVSK
>B_FIRMIC_4 tr|Q4MXR3|Q4MXR3_BACCE Transcription regulator protein OS=Bacillus cereus G9241
GN=BCE_G9241_3299 PE=4 SV=1
MLVKVVSDLSEQFVTQKSIKETILRGIRKVLLERGSATKVELSNTLEISFPTISKFIEKMKQDGEVTLA
GLDDSSGGRRAKRYEYNPEYMLGLAIFLEKNETNYTIFNCLGEVKEQGSTSSMLIDTGVNLLSKHIE
GLIATFPKISSISIGVPGSVDNGRIFYIPGYEKFQNFNLKSHLEEQFSMPVVIENDMNAAVLGYYKNTG
NYDNSSLVYLYSGQNGPGAGIMVNGDVVRGSTFFSGEISFVPQYDNKNFLQALRSEDSNNPEEYNID
AITRLIATCIAIINPHGFIFCDDEVNQFVIDQIVKSCPQYIPAEHIPKITVSNWKEDYLYGLKSLGLDLMI
TRTNKEN
>B_FIRMIC_6 tr|Q4MQF6|Q4MQF6_BACCE ROK family protein OS=Bacillus cereus G9241 GN=xylR-2
PE=4 SV=1
MKEYIAFDIGGTQIKYGIVSETGTVLKHKTVPTEIHLGGEQIIQKLILLSKKLMGEHTILGIGISTAGIV
DVNKGIVTGGADHIPGYSTIPIINRLQEVLKVPVSIDNDVNCAAFGEKWNGSGREKGNFIMLTLGTGI
GGAIFIDGELYRGHSFSAGEWGNMLIEGKAFEEVASISGLIHLVRKYKGEGDWNGKTIFELYDKGDR
EVTQAVEVFFKHLAIGISNLAYIFNPEMIVIGGGITDRGNQFLKEVKEEVEKYLQKEIYNNCEIELAQN
GNCAGMIGSIYHFLHHHK
>B_Firmicute gi|239624996|ref|ZP_04668027.1| ROK domain containing protein [Clostridiales bacterium
1_7_47_FAA]
```

```
MRFDKYVMAQMNKKMILNMIIQKGPINRAEIARLSGLSVPTVMKITDEFSQSRLIRTIGKRESTGGR
QPELIEINKEEYFCIGLDVGHNRLKVVVMNLAGEIKLRRAIPTGDTLPPETAIYRMISLVREILGGMGE
QKQRLLGLGIGMPGLLDPDTGMVNFSPDFGWEHVDLLGRFQNEFGFPVLIENANRVMALGERWFG
AGRSAEDFLCVNLGHGIGSALVFDGEIYHGNSGSSGEIGHITLEKDGPLCECGNHGCLEALASGRAIA
KHGLELAASGKGKRILELAGGRKEEIEAETIFRAAMEGDGEAEDILDNAIEYIGIAIAGVVNLFDPELI
IFEGGLMKSSSYLLPRLKETIRRHQMHLAGRNVQLLKGNLGEDITAVGAATLLLQDLLNHGGEIKKG
R
>B_GAMMA sp|P44554|NAGK_HAEIN N-acetyl-D-glucosamine kinase OS=Haemophilus influenzae
GN=nagK PE=3 SV=2
MYYGLDIGGTKIELAVFNEKLEKLYSERVPTPKTDYEEWLNTIVDLVNRADEKFGEVGTVGLGVPGF
VNQQTGLAEIANIRVADNKPILCDLSTRLGREVRAENDANCFALSEAWDTENQQYSTVLGLILGTGF
GGGFVLNGKVHSGQVGMAGELGHLQLNYHALKLLGWDNAPIYQCGCGNKACLDNYLSGRGFEM
LYQDLKGETLSARKIINLFYQSNESAVDFVNLFVELAAISIGNIITAFDPHMIVLGGGLSNFDYLYEALP
KALPPHLMRKAKVPPIKKAKHGDSGGVRGAAALFLTK
>B_GAMMA_10 tr|C2I369|C2I369_VIBCH ROK family protein OS=Vibrio cholerae TM 11079-80
GN=VIF_001087 PE=4 SV=1
MYYGFDVGGTKIEFGAFNAQLERVATERVATPTDDYAKLVETIAGLVHKYDAQFGVEGTVGLGIPG
MEDADNGCVLTVNVPAAKGKPLRADLEAKLGRAVKVENDANCFALSEAWDDELKEAASVMGLIL
GTGFGGGLVYEGKVFSGRNHVAGEIGHMRLPIDAWFHLGEKAPLLGCGCGNKGCMDNYLSGRGFE
LLYEHYYGEKKKAIEIITAQKEGEAKAVEHVERFMELLAICFANIFTANDPHVVVLGGGLSNYDLIYE
EMPKRVPKHLLSVAKCPKIVKAKHGDSGGVRGAAFLNIK
>B_GAMMA_14 tr|B5NZT5|B5NZT5_SALET Transcriptional regulator Mic OS=Salmonella enterica
subsp. enterica serovar Heidelberg str. SL486 GN=mic PE=4 SV=1
MVADSQPGHIDQIKQTNAGAVYRLIDQLGPVSRIDLSRLAQLAPASITKIVREMLEAHLVQELEIKEA
GSRGRPAVGLMVETEAWHYLSIRISRGEIFLALRDLSSKLVVEECLPLPLTEATPLLERIITHVDRFFTR
HQQKLERLTSIAITLPGIIDTENGVVHRMPYYEDVKEMPLGDALERHTGVPVYIQHDISAWTMAEAL
FGASRGARDVIQVVIDHNVGAGVITDGHLLHAGSSSLVEIGHTQVDPYGKRCYCGNHGCLETIASV
DSVLELTQLRLNQSMSSMLHGQPLTVDSLCQAAMQGDLLAKDIISGVGTHVGRILAIMVNLFNPQKI
LIGSPLSKAADILFPAIADSIRQQALPAYSRNTVVESTQFTNQGTMAGAALVKDAMYNGSLLIRLLQG
>B_GAMMA_17 tr|Q48D96|Q48D96_PSE14 ROK family protein OS=Pseudomonas syringae pv.
phaseolicola (strain 1448A / Race 6) GN=PSPPH_4529 PE=4 SV=1
MPLDLDTPLHAHQPLSMNERKLLDILRRRGVITRATVSTEMDLAQQSVHRLIEELISRGLLRSGERVK
NGRGQPSPRIELVNEAVYVIGVSINTDSAVVCIADLGCNVLEQVTLRTPPLSRNSTLDSLDQTIERMLQ
RNGIEAQRVIGMGFAIAGFFLENRQINAPEPLRDWSLIDLQPILEERFGMPVWLENNATTAAIGESLVG
VGAWASNFIYLSFNFGFGAGVVINGKPYFGSHGNAGEITLYNDEESINRPALRYLLDELHQNGVQVD
SIEDLRLRFDPDWPGVDTWLARVKPTLDRLVNALAGLFDPQAVVFGGQLPPELGRRLIAATAFWGA
HRYNAPPPRPQLLLSETNGDAAAIGAALVPLKERFFV
>B_GAMMA_19 tr|A9KUF8|A9KUF8_SHEB9 ROK family protein OS=Shewanella baltica (strain OS195)
GN=Sbal195_1433 PE=4 SV=1
MSVLCLDLGGTKLMLAQVEGKTLLDTWRYPVPADGNFEQLFDFLVTCIHSHLTPETYGISIGIPGMV
DMQSGTLLEVLNIPALTATQLAQQLKNTFEMDVVVNNDANLFALGEAVLNRNQDMLGITLGTGVG
AGVIFNGQLYSGKHCAAGEIGSLSYRDGIIEHYCSGQYFTTHHHMSGEHLYQKACEGDSQALQAFA
HFGEHLAHMIAQTLLVYDPKDIILGGSVSQSFPFFIEALNQKLQSLVYGPQLADLTISASQHHNAALIG
AAQWFLQQKDSVK
>B_GAMMA_20 tr|A9KVD9|A9KVD9_SHEB9 ROK family protein OS=Shewanella baltica (strain
OS195) GN=Sbal195_1506 PE=4 SV=1
MMRMGVDLGGTKIELVALSDEGNELFRKRVTTPRDYQGTLAAVVNLVKEAEATLGEQGTVGVGIPG
VVSPYSGLVKNANSTWINGHPLDVDLGELLQREVRVANDANCFAVSESVDGAAAGASVVFGVIIGT
GCGAGVAINGKVHGGGNGIGGEWGHNPLPWMTKEEFNTTRCFCGNPDCIETFISGTGFVRDYNAAL
TAAGTVRAAAKSGADIMLLVDEGDAIAVAAFDRYMDRLARSLAHVINMLDPDAIVLGGGMSNVAAI
YPRLPALLAHYVVGRECHTPVVQNLYGCSSGVRGAAWLWEK
>B_GAMMA_23 tr|Q3BZN1|Q3BZN1_XANC5 Putative transcriptional regulator, ROK family
OS=Xanthomonas campestris pv. vesicatoria (strain 85-10) GN=XCV0051 PE=4 SV=1
```

MGAGRGMQQAHGTISTSVAPTQGRGRIKLSCLIKAIMTTTNVSPCMPSLFANAAIRSHYRRRSDGQA
AGSSERLLLDLIRRAGQFERADLPRASGLSVPGIKGIVDPLVAHGLLQLRPSLRRGRGQPNAQVSLVP
GYAYSVGVSVMVDGFAVVLIDFAGQVRGMRQLTAFPLTLAVVRAQLPALVEALLTAAGVDRQHVFG
VGLSMTGPRIGDGTRVNPPLSLAAEWMQVELDRFVADCLQLPVWMDNDAHCAALAEAVYGIGRQS
PDLVYLSISDGFAAGVIAAGNVRRGAHGNGGELGRISAITGMARPTLESLRQALAADGHALPDLHT
MLQHYDAAWPQIDAWLDAVEPTVTLAVAAIIALIDPRVIVFGARLPTDLAQRLIARIAFEPAPRRGVAL
PYPTLQVGQVTAHATVLGAAMLPFKETLF
>B_GAMMA_6 tr|B5P1P9|B5P1P9_SALET Putative transcriptional regulator OS=Salmonella enterica subsp. enterica serovar Heidelberg str. SL486 GN=SeHB_A3830 PE=4 SV=1
MQQYIGIDVGGTHVKYGVINSDGEELTHHQFDTPEDASTFTRKWQDVVARCQQDYDIAAIGVSFPG
HINPHNGHAAKAGALAYLDDVNLMELFSGLTDLPLVVENDANCAALGEMWRGAGQHYDNLVCITI
GTGIGGGIIVGRELYRGAHFHAGEFGVMPVGNNGESMHKIASTSGLMASCRQALALPAEEMPPADVI
FERMATDVHLREAVNDWARYLSRGVYSVISMFDPGVMLIGGGISEQEKLYPLLTRHLETFEMWEAL
QVPIQPCQLGNQAGRLGAVWLAQQKLARS
>B_GAMMA_7 tr|B5P7Q5|B5P7Q5_SALET N-acetylglucosamine kinase OS=Salmonella enterica subsp. enterica serovar Heidelberg str. SL486 GN=nagK PE=4 SV=1
MYYGFDIGGTKIALGVFDSTRRLQWEKRVPTPHASYSAFLDAVCELVAEADQRFGVKGSVGIGIPGM
PETEDGTLYAANVPAASGKPLRADLSARLDRDVRLDNDANCFALSEAWDDEFTQYPLVMGLILGTG
VGGGLVLNGKPITGQSYITGEFGHMRLPVDALTLMGFDFPLRRCGCGQMGCIENYLSGRGFAWLYQ
HYYDQSLQAPEIIALWEQGDEQAHAHVERYLDLLAVCLGNILTIVDPDLLVIGGGLSNFTAITTQLAE
RLPRHLLPVARAPRIERARHGDAGGMRGAAFLHLTD
>B_GAMMA_8 tr|A7JDW5|A7JDW5_FRATT ROK family protein OS=Francisella tularensis subsp. tularensis FSC033 GN=FTBG_01633 PE=4 SV=1
MYLAGIEAGGTKFFTTIGDFDGNVIERHRTDTTTSEKTMSEVLKVLKDYQNKYDIKTIGLACFGPIDI
NPNSKTYGYITNTPKIAWQNFDIVNAVKTIFSGPIGFNTDVNAAAICEKLWGCAQDLENLLYLTVGTG
VGGGIICNNKLVQGAMHPEIGHLLIPQNPLDEFKGSCPFHGNCLEGLASGTAINQRWKVAHAGALND
DHIAWQFEAEYLAKAIVNYICSFSPERIILGGGVMHKTILFDMIRKNVTKYLNNYLDYPALKDMTKFI
VPASFGDNTGVKGSLALALETFNNSQAY
>B_GAMMA_9 tr|C2IAQ3|C2IAQ3_VIBCH N-acetylglucosamine-6P-responsive transcriptional repressor NagC ROK family OS=Vibrio cholerae TM 11079-80 GN=VIF_000070 PE=4 SV=1
MIMNGGQIGNVDLVKQLNSAAVYRLIDQQGPISRIQVADVSQLAPASVTKITRQLLERGLIKEVAQQA
STGGRRAISLTTEVKPFHSIAVRIGRDYIQLSLYDLGGNSLVDEHHEFHYNTQDVLMSSLIKQIKQFIQ
QHTALIDQLIAIGVALPGLVNPETGVVEYMPNVAINELPLGATIRDEFHVECFVGNDVRGIALAEHYF
GASQDCQDSILVSVHRGTGAGIIVNGQVFLGYNRNVGEIGHIQIDPLGEQCQCGNFGCLETVATNPAI
TSRVKKLIAQGYESSLSTLDTITIDDVCEHANAGDELAKQALVRVGNQLGKAIAITVNLFNPQKIVIA
GQITAAKEIVFPAIQRNVENQSLKTFHQHLPIVSSQVYKQPTMGAFAMIKRAMLNGVLLQKLLED
>B_SPIRO_2 tr|B9AAM9|B9AAM9_BORBU Xylose operon regulatory protein OS=Borrelia burgdorferi 72a GN=BBU72A_0857 PE=4 SV=1
MKHYLAIDIGGTSTKYSLSDSSGVFFDKNEISTGATSDEQVNILVNIINSYKELSDIAGVAICIPGFVDL
KGNVLRVNAISGFVNYPLKERLESLTGVSTEIENDANCVALAEKFKGNAIDSNNFIAITLGTGIGAGIF
ANGKLLRGNSFMSGEVGFMITGGISNNIPFNCKWESIASVSALRKRVAMRLGKPLKEVSGEFVFDLA
ENGNIHAKNEVDRFFENLSFGIFNLTFILNPEKILIGGGISARPDLIDRIYEKLENLWSLEMAFDNNNNI
KNLVALEPTKFNNESGKIGALFHYFTCKNQNNTSF
>B_Tenericutes tr|A5JIY9|A5JIY9_MYCML Glucokinase OS=Mycoplasma mycoides subsp. mycoides LC GN=glk PE=4 SV=1
MKKILGIDLGGTSAKVGVISQNGDLEHSFSITNPKTKIIENLYFEISKILKTLNVDENDIMLVGITAPGF
VDHNKGIVVMAPNIENGWFNYDLKTEAEFLFKKPVYVINDVNAAALGEYRKGSGLVYKSGLFYWL
GTGIGGAIICDGKLISGSHGFAGEFGHGGSNNHNLKCNCGLNNCIEKVCSATTIPNSLLTILKNKYLEF
YNKHFLNIKDLDMKLLFEIYNNLNKPIELKNSLLEVYDELFNHMSLLIHALDPEVVVIGGGGSLAGN
NLLELFQLGVKNKLTDSYKDIVDFKLALLKNDAGMIGAAFYALEQSLKTN
>B_Thermotogae_2 tr|Q9X2G2|Q9X2G2_THEMA ROK family protein OS=Thermotoga maritima GN=TM_1847 PE=4 SV=1

```
MTEKVSRFPGVFPLIVIGVPGSVDKTHKKLAFAPNLNRWRDIDVEKYFKVFEVYLENDANLAALAE
MMRNKHFGDRKNIVYILVREGIGGGIIIEGKLYKGSFNAAGEIGHMKMYDRGPCFCGRVGCWEANT
SISHCVRQYEKKKPLPGNTMYEKFETLCRIYEEDPLAKEVLDEFTGILIDGIVNLVNILSPEIVIVGGEG
VFLPESVFEVIVSETRRQVHPMDKEVSVEKGSLSNKEVVLEGTSILSSMMISERLV
>B_ZETA tr|Q0EW16|Q0EW16_9PROT ROK family protein OS=Mariprofundus ferrooxydans PV-1 GN=SPV1_12917 PE=4 SV=1
MRIGIDLGGTKTELIALDKTGAARLRFRRPTPAGDYAATINMIADMVAQAEGESGETANIGIGTPGAIS
PRTGRMKNCNSTCLNDQPLQQDLELALRRPVRLSNDANCFALSEAIDGAAARSPVVFGVILGTGVG
GGIVVNGQLLEGVNSIAGEWGHNPLPSANREERPGPPCYCGRKGCIETWLSGPAMRRDHIACGGQD
ITAAEITRMATEGDAACMLTLERYCERLARALAGVINILDPDAIVLGGGLSNISLLYEQVPRLWSSYIF
SDSVSTRLLPPAHGDSSGVRGAAWLWGR
>Branchiostoma1 gi|260836485|ref|XP_002613236.1| hypothetical protein BRAFLDRAFT_210453 [Branchiostoma floridae]
MSSPDTKKEHLNFENRQMKVCVATCNRADYSKLAPIMFCLRDDPLMELRVVVMGTHLIDDYGSTY
RMIQQDGFEVDGYLHTIVRGEDEAAMAESVGLALVKLPDVLVRLKPDLLIVHGDRFDALSLATCAA
LMNIRILHIEGGEVSGTIDDSIRHSISKLAHYHACCTDRARRRLLSMCEDNDRILLSGCPTYDKLLSCN
VRNHDNVLARWAQGAKPKDYIVALQHPVTTNIKDSLKMFELMLDAVLDFGVVTVVLFPNIDAGSK
EMTRTIRQRGLEHNKNFSLCKHIPYDQFIILMANAGCLIGNSSAGIREAGAFGTPVINLGKRQIGREA
GANVLHVKDADTKSKILHAINLQYKKQFPRQGTSFSHIYGDGHACQRIMKFIHTIKNDLDKPIQKAFI
FPPMRDTVSLDIDHILEIQSALAVDLGGTNLRVAIVGQQGEILHKMTEPTPGTNEERMDVLMRLLVES
TSKAVELNCRILGIGISTGGRVNPYDGVVLHSTEILEGWNSIDLRTPISSKLHLPVWVDNDGNCAALG
EKKFGKGLGSEDFITLIVGTGIGGGIFLNNELVHGANFCAAELGHISVCMDGPDCTCGSSGCVESYAS
GLALQREAKKLHDADELLVPGVHLTDGEELTAKHLIQAAQLGNKKAEKVVERACEALGSAICTLLH
TVNPSHIILCGHLAPHYVDGVREVIQRRALPSAANNIQVMVSDLEEPALLGAASIVLEFATRRIY
>Branchiostoma2 gi|260786105|ref|XP_002588099.1| hypothetical protein BRAFLDRAFT_124939 [Branchiostoma floridae]
MDTEGLNVKNRKMRVCVATCNRADYSKLAPVMFALRDDPDMELQVVVMGSHLIDDYGSTYRMIQ
QDEFEVDGYLHTIVRGEDEASMAESVGLALVKLPDVLVRLKPDLLIVHGDRFDALSLATCAALMNV
RILHIEGGEVSGTIDDSIRHSITKLSHYHACCTDRARRRLLSMCEDNDHILLAGCPTYDKLLSCNAKN
YDNVLSKWVGEGVKPKDFIVALQHPVTTDIKGSLKMFDLMLDALLEFNVKTLVLFPNIDAGSKEMT
RMIRLRGLEHHPNFSLAKHIPYDQFIILLANTGCLIGNSSAGIREAGAFGTPVINLGNRQIGREAGENV
LHVRDADTKGKILHAVNLQYQKQFPCSYIYGDGHAVQRIMKFIHKIRNDLEKPIQKSFIFPPMKDTVS
QDIDHILETQSALAVDLGGTNLRVAIISQQGEVLHKMSELTPPTYEQRMDLLVRMLVEATSKAVELSC
RILGIGISTGGRVNPHEGMVLHSTKILEEWSSIDLRTPISSKLHLPVWVDNDGNCAALGEKKFGKGK
GSEDFITLIVGTGIGGGIVLNNELVHGANFCAAELGHISVSMDGPDCMCGSSGCVESYASGIALQREA
KKLHDADELLVPGVHLREGEEVTGKHVIQAAQLGNKKAEKVVETACEALGLAVCTLLHTVNPSHII
LCGHLAPHYVDGVREVIQRRALPSAANSIQIMVSDLEEPALLGAASIVLEYATRRIY
>Danio tr|Q6P3K2|Q6P3K2_DANRE Glucosamine (UDP-N-acetyl)-2-epimerase/N-acetylmannosamine kinase OS=Danio rerio GN=gne PE=2 SV=1
MERLEKPKKERLRVCVATCNRADYSKLAPIMFGIKSHPDIFDLEVVVLGSHLIDDYGNTFRMIEQDD
FDIGSKLHTIVRGEDEAAMVESVGLALVKLPDVLQRLAPDILLVHGDRFDALALATAAALMNIRILH
LEGGEVSGTIDDSIRHAISKLAHYHAVCTLSAERHLISMCEDHSRILLAGCPSYDKLLSAYKRDDYAD
IIKSWIGDDVKEQDYIVALQHPVTTDIKNSIKIYELMLDALISFNKKTLILFPNIDAGSKEMVRVMRRK
GIEQHQNFRAVKHVPFDQFIQLVAHAVCMIGNSSCGVREAGAFGTPVINLGTRQTGRETGENVLHVR
DADTHNKIYHALELQFGKRYPCSKIYGDGNAVQRILKFLQTIDLSEPLQKKFCFPPVKECISQDIDHIL
ETQSALSVDLGGTNLRVGIVSMKGKVVKKYVQLNPKTFEERIELILTMCKQAMADAVHLNCRILGV
GVSTGGRVNPQDGVVLHSTKLIKEWSSVDIRTPLSSALHLPVWVDNDGNCAALAERKFGHGKGVE
NFVTIITGTGIGGGIIQHNELIHGNTFCAAELGHIVVSLEGPECMCGGHGCIEAYSSGLALQREAKRLH
DEDLLLVEGMTLNNKEQVNAIHLINAARLGNSKAETVLHTAGTALGLGIVNILHMINPSLVILSGVLA
VHYETPVRQVIGQRALLTAQGTKVMVSDLEDPALLGAASMVLDYTTRRTY
>EU_Acanthamoeba A new protein sequence entered manually
```

```
MDATPTSELVFAGVEAGGTSFSVGLARGSAESIFARANFPTTTPDETIGRVVAWLREQNAKTPFHALG
IASFGPVDLDRASPTYGYITTTPKPDWGNVDVLSRFSEFNVPTAFETDVNAPAVAHLAQFGQPSCAYI
TVGTGVGVGIAVEGKPVHGLLHPEMGHIFVRMAKGDDFEGTCPFHGACVEGLVASRALAKRFGVE
RTDLSDVADDREEWDHVAYYLAQLCAALVLTVSPHRIVMGGGIMQRPTILPLVHKHVLALLNGYIK
VPAITESGIGNYITLSPFGGNAGLVGACELARLALPHNASPAM
>EU_Blastocystis >gi|300123387|emb|CBK24660.2| unnamed protein product [Blastocystis hominis]
MVLFHTNLRFCTFLFKVSMSDKTKTILQVGGALFATALLGYGVKHCFENSKKKEKCERYFVGIDLG
ATNAKAGVVNDEGELLATASQPLTDYTDKGVVSSLVEVAAKAVTEAGLKWSDISEIGVGSPGTIDFD
NGVVIKASNFPTWDHVPLANLITEATGVAAVLDNDANAACAAECWIGAGKGKSNMVMITLGSGIG
AGVVVNGRVIHGGSGWAGEPGHAIYKVNGTPCGCGQKGCFEKTICRSIPTRKPPCEPSRSSLPKRLW
TLRRPATR
>EU_Cryptosporidium tr|B6AK73|B6AK73_CRYMR Putative uncharacterized protein (Fragment)
OS=Cryptosporidium muris (strain RN66) GN=CMU_004310 PE=4 SV=1
MSGERRSSSNGGASARTSRHLFPFRSDRAPASTGIDAAAASPAPYLTLDIDANHNQRVVFQEIRTHKR
VTRSELANITGLTIPTIVNVSRHLLQAGLIVEAKRPARGRGQPAWELELNPDGAFSAGLNIDRDHLTL
VVLDLSGQIRGRVSAEIGFASPDDARRFFARSWRELVTKAHISPDRVVGLGIALPDELGTLAFPGKPLT
YDAWARTDAADLFRDTFAGPTFVENDAVAAAIGEAQFGHGLEAPSFVYVLISVLLGGGIVIDRTY
>EU_Dictyostellium tr|Q54TJ9|Q54TJ9_DICDI Putative uncharacterized protein OS=Dictyostelium
discoideum GN=DDB_0204621 PE=4 SV=1
MIKNEQKFLGVEAGGTSICLSIAINKPSNVIERTSIPTTSVEETKNKVLEWVKNKGGNNIGAIGIASFG
PIDLDESSETYGYITTTPKPNWGQTNILGWFDDYKCPKGFDTDVNGAAISETFHGLHKRGEITSCAYI
TVGTGVGVGCVVNGSCIHGLVHPEGGHSFCKIQDGDKDFQGTCPFHGNCIEGLVSTGAISKRLGINA
DKLSEINDDHPVWLIIGNYLAQLCANITCLLSPQVIVLGGGVLNRSILYPIIRTETIKILNGYIKSKYLTN
DFIDQYIVQSPFASHAGSIGSLELARRALIKNSKDF
>EU_Ectocarpus gi|298715267|emb|CBJ27916.1| conserved unknown protein [Ectocarpus siliculosus]
MVIYAGVEGGGTTWRVAIADGHPTNITESKSFVTVQDSKEQLKAIKDWLSTRKYDCLGIGTFGPVDP
REGSPTYGYITSTPKPGWNMVDVVGYLSDGSVPCKFDTDVNAPALAEFMWGAKKEGESSCAYITV
GTGVGVGLVVNGQAVHGLMHPEAGHLCLKRMPGDDFPGVDSVFGGASVEGLASTVALAARKNCA
REDLPGVPDSDPVWEATAHSLAGLCASLVLVVSPERIVLSGGVMNRTLLYDKVRKWTRELLNGYID
HPAVTTDAVDDYITPSSFGQNAGMVGSLTLAHIAYEEAGGRGGGAGSPTATKAGGGAACLASPCNV
SVALAAAAVVGLGALAIAKASKK
>EU_Ectocarpus2 >gi|299470076|emb|CBN79253.1| ROK family glucokinase [Ectocarpus siliculosus]
MYASTEHMLFVGLDIGGTNLKAGVIDGTTGGQLLGRAQERLPADRSPEAVVDGLVALCRGLLDEHG
ITWDDILYTGVGCPGQIDREAGVVIGASTFPAWHNVPLANLVQDRTGRPVTVLNDASAAASAEFASR
GSQETIAVLTLGTGIGLGVVCAGRVVTGCRGLVEGGHMIVEPGPNGRLCACGQRGCLEMYASASAV
>EU_Micromonaspusilla tr|C1MJ60|C1MJ60_MICPS Predicted protein OS=Micromonas pusilla
CCMP1545 GN=MICPUCDRAFT_46426 PE=4 SV=1
MSDLDSYSDASHGSPYLTTPPPAAREFHRARPRADPSSPGDDYHSARAANSVNESAASTVADDDAG
VANEWARRVRSLFPTGELERGRKDDVALAENAIAAMRPDEPRARFAGVEGGGTTWVCAVAEGDPD
AIVARAEFPTTTPSETLARVKKWLDARHEEAPFDGLGIATFGPLCLDVASSRYGTITHSPKPGWTRVD
VLKRLSSGFDCPVAIDTDVNAPAASELADLKRDAIEAADGDFTAADAMQNLAYVTVGTGVGVGIIV
GGEPVHGLVHPEAGHARVARFPGDGEKGAPGAFVGACPYHEDCVEGMCSAAAVAKRCGVAVSELS
NVADDDDAWDATAHYVAGLCANLILTTSPQRIVLGGGVLQRRALVSKIRVHVKAQLGGYVKHDMV
NTRAGLRELIVSSRRGNDAGVVGALTLAKGAAEKARGRGGGGGGAAAAAARRGNGGRRKGWRA
HLAHFLVGTVCALVGAKFGVRREERKVDASVGVASAEAAAEEEDASEASESDEV
>EU_Micromonassp tr|C1DZP3|C1DZP3_9CHLO Predicted protein OS=Micromonas sp. RCC299
GN=MICPUN_56468 PE=4 SV=1
MAVPPADDIVPKGKPKRLSYTREKTGWGKAAFKLRFIPFLGGPKERDEDIFEDPDGGSDANGVGSAI
PALQPTSSVKSDLGIVPKVKGKRFAGVEGGGTTWVVAISEGVPENIVERAEFPTTDDPMVTLGGVRA
WLDARAAEGRFDAIGIATFGPVDLDKNSPTYGYITHSPKPGWADVDVLGILADGFDCPAGFDTDVN
APALSELSAMRREMAIDAGAGDGGAEGAEGAEDSIQNLCYVTVGTGVGVGVVCGGQPVHGLSHPE
AGHIRVARLPRDGMPGEPGAFEGGCPYHADCVEGMANASAIARRCGCRVGELSEVPDDHDAWDAA
AHYLAGLCSVLVMTASPQRIVLGGGVLQRKTLVTKVRAQLKRQLGGYVAHDLVSSKRGLVEFIVSS
KLGNDAGIVGALAVAEKAKADWEREREGGWGRAGGKRRYGWREHLLHFGIGVFSSLVTLKVVDG
KITVGKFVF
```

>EU_Monosiga tr|A9V8S7|A9V8S7_MONBE Predicted protein OS=Monosiga brevicollis GN=33963 PE=4 SV=1
MAMLLLPFWLADRLATLASLVTPPKATVYAAVEGGGTSWRAAIAVGRPDNIVERTTFKTEKPEVTLV
AIRKWLDARKYDSLGIGTFGPIQPDRSHPQYGCITSTPKPFWKNAKVVQTLCPSGIPHLFDTDVNAPA
YAEFLALHATNPGLTSLAYITVGTGVGIGLVINGQPVHGLLHPEGGHMLIRRQPGETFAGVCPFHKDC
VEGLVSAPALAARRGVSQEELANLDDDDFIWDAAADALANACANLLLTVSPQAIVISGGVMLRACL
FDKVRARTLELLAGYVDVEQLLQRPEEVIRPSTWGNNAGIMGALHLSKTALERS
>EU_Naegleria_ 69011
MSNELIVGIEGGGTSYKVAIGLNIEHAMENVFITETKTVDETMEPIKQFLIKKESELGGIIKRIGIANFG
PIDVNVGCILPSTPKVSWRSFNIVEYFKREFPQVKHVQFDTDVNGPAMAEYQLVKGSGIKSLAYVTI
GTGIGVGLVINGSTVSGLMHPEAGHIYTPLHPRDMETGFKGFCTFHTEGCLEGMAASPSIVKRRNISI
HDVKSISDDDEIWDIEAHYLAHLCVNLILISSCQVIVMGGGIMNRSILFDKIREKTVQLLNGYHVMIN
NESIKNIIKPSVYGEHAGIKGALHLSQL
>EU_Nematostella jgi|Nemve1|241240|estExt_fgenesh1
MSRYIAGVELGGTSCVAAIAEISNPTTIVHHFEVSTTEYTSTLGALTEYLSAQLKEFNIESYAALGIASF
GPVDLKPDSKTYGYITSTPKPGWKYVEIVGVFKRTLKEGTPIAFDTDVNAPALAEMAAALNTKTIPIP
LPTVIYITVGTGVGVGVTVEGSPVHGLLHPEGGHIIVPRMQGDEYPGCCAYGHECCVEGMIDSQAIA
ERVNVHRHDLHTIDDNNPVWKVVGYYLGALCLNITYLLSPNLIILGGGIMKRRILYTETRHWFKELL
KEYLDSPQFKTDQGLASYIKEPFHGSKAGIVGALELARHKC
>EU_Oryza tr|A2XBC4|A2XBC4_ORYSI Putative uncharacterized protein OS=Oryza sativa subsp. indica GN=OsI_09569 PE=3 SV=1
MYYGFDIGGTKIALGVFDKDLRLQWETRVPTPRESYDEFLTAIAALVAQADERFGVKGSVGIGIPGM
PETDDGTLYAANVPAASGKALRADLSARLERDVRLDNDANCFALSEAWDDEFRQYPLVMGLILGTG
VGGGIVINGKPITGRSYITGEFGHIRLPVDALDIVGREFPLTRCGCGQHGCIENYLSGRGFAWLYEHFY
HQKLEAPQIITLWEQGDAQAREHVERYLDLLAGRLSRFRKNKRRLRERLRQRIFFRDRMMPEAMDK
PRVVVLTGAGISAESGIRTFRAADGLWEEHRVEDVATPEGFARDPDLVQAFYNARRRQLQQPEIAPNP
AHLALAKLEEALGDRFLLVTQNIDNLHERAGNKNIIHMHGELLKVRCAWSGQVLDWKEDVLPEDK
CHCCQFPARLRPHVVWFGEMPLGMDDIYSALAMADVFIAIGTSGHVYPAAGFVHEARLHGAHTVE
LNLEPSQVGSEFEEKHYGLASAVVPEFVDKLLKGL
>EU_Ostreococcus tr|A4S0P4|A4S0P4_OSTLU Predicted protein OS=Ostreococcus lucimarinus (strain CCE9901) GN=OSTLU_32817 PE=4 SV=1
MKLLGIEGGGTTWIARAIEIDVEGGASVSSASSARGEEHFNKRDGGREQRFETTTPEETLRTIREWIEI
NAWDADAIGVATFGPLELNPDKDKYGYITTTPKAGWQDVDVLGSLFGKKDATEEEERAWRGRARL
HTIDQVPLAFETDVNAPAMLEHRALKHELKHVHLVGGESCCYVTVGTGVGVGVVCNGLPVHGML
HPEAGHMFVKMRAGETFAGTCPFHGNCVEGMVGSGALAKRRGVSAAELASLPDDDDIWEHAAHY
LAGMCVNLILTLAPERIVLGGGVMQRECLFSKIRANVRDILQGYLAVDQIMDDAYLRHFIVPPAWGY
QTGLTSALYLAERALQRE
>EU_Ostreococcus_2 tr|Q014K2|Q014K2_OSTTA Chromosome 07 contig 1, DNA sequence OS=Ostreococcus tauri GN=Ot07g04040 PE=4 SV=1
MGAVKLLGIEGGGTTWTARATSIDVDAGDSMSATTVSRVTTDVPRECVTEVFNTTTPTETLGAIRAW
IEVNARDADAIGVATFGPVELNPAKEKYGYITTTPKPGWEDVDVLGALFGPRGEEEGGEPWVGKAR
LKTPNDVPLAFDTDVNAPAALEHRALRRELQNVHRAGGESCCYVTVGTGVGVGVVANGRPVHGM
LHPEAGHMHVRMMDDETFPGTCPFHGNCVEGMCGSNALAKRRGVKPADLVTLPDDDEIWDQCAH
YLAGLCANLILTLAPERIVLGGGVMQRECLFSKIRAKTRELLNGYLSVEQITDDDHLRHFIVPPAWGY
ETGINSALFLAENALKREVA
>EU_Paramecium tr|A0BMN9|A0BMN9_PARTE Chromosome undetermined scaffold_117, whole genome shotgun sequence OS=Paramecium tetraurelia GN=GSPATT00030442001 PE=4 SV=1
MQQSKAALEIGGTFIQVGIGTKTDNKWTLNNKKTFDTRDPEDTLLDLVTYLQQFEFDSIQIASFGPLC
LNKDDPQFGSITSTPKLKWQNFPIATRLSQALKKPFAIDTDVNACAMAEFMLGNHNVRQSLAYITIG
TGVGVGIIVNGQCVHGMLHPEGGHILVAKQQEDKDFKGVCAFHGDCLEGLCTNVAIAKRLNCPITE
LPNISDDHPIWELVGFYLAEACQNILYLLSIEKIVLGGGVMNRKLLYPIIDKHLRRLVNKYVEIPENYI
VEPQVEDVGLIGALLLQ
>EU_Phaeodactylum tr|B7G684|B7G684_PHATR Predicted protein OS=Phaeodactylum tricornutum CCAP 1055/1 GN=PHATRDRAFT_48173 PE=4 SV=1

ASPWPEILARTEIDSSHDHPLVTLRACVDFFEAYKPVGGYHALGIACFGPLGVNVARVEDYGRILATS
PKAAWRGVDLLTPLATCCRGDTHALVTRVDTDVNAPALAEYYLATQRVKHPGAVAYPVDDAVTSAS
SEISSIAYVTVGTGVGVGLVVHGQPVHGRMHPEGGHVAVQPLANDVFPGYSWGQKNCPYEGVHTV
EGLASSVALTERLAQLQQIPTPLSRSVLASLPDDHEVWDHAVNALANLCVTLLLTLSMEKIVMGGGI
LRRALLLPRIQARTVELLNGYLPLPEDMSTLIATSSFGDDIGLIGAMVLAQSSLQLNEHTRESRKRHE
TVMKQTAFKHGLWHGMLVGAVGAALVCKYVWYGPRTRK
>EU_Physcomitrella tr|A9U713|A9U713_PHYPA Predicted protein (Fragment) OS=Physcomitrella patens subsp. patens GN=PHYPADRAFT_103656 PE=4 SV=1
MSLASGGRSVNNRKKSYVAVLLVLAILVSIWAFYPSGSSEVRSVQAMADFAAVTGSGGDDGGYDAY
VKAHSAAKRPDQVIRIEGESFTQTDGPGFEVVHPSGETAVLTPESGSISWNVPIEQAGLYNIRIRYLPV
EGKSSAIERGLTINQQLPFKGADLVTFDRVWGNRDDKIGRDDRGNELRPSQVEKPIWQVESVTDRSG
AVYKGVINVSDIPELIHAPLEAIIRDKHGIEVILENDMNSTVYGFYRKQDYDTEKSIAVATFLEGSLPG
AGLIVDGHIHRGNTRFAGEIAFLPYGISQAEQLRQLHDREAFHSLAAHAVSSLIAVMNPETIALTGSLV
QADDIRLIRQECLKYIPEMHMPQLALLEHPDEDYMYGLTTMTLESLAYSLQLVEKRSKPHPGTVRNI
LFLLFALPGLSFASWVSRTPDVRDTLQVSTAGMGMIIFAIAAGSLFGLLTSGSIVARKGARFVILISSLLI
VVGFVVAGIGTSLASVPVVIIGLIIFGGGYGTAEVALNVEGSAAEKALNRTLLPAFHGFFSVGTLVGA
ALGAAAVAIHLQVMVHFTILALIMAAAVLYSWRYLPEESGKEPA
>EU_Phytophtora sojae |158834|C_scaffold_159000001
MPRFAGVEVGGTTWVAAIAEDHPENILEKFEVDTTTPDETMGAVVAWLKERKFDSIGIASFGPVDLN
KKSPTYGYITSTPKPNWGNTDVVGVFKRAFPDVPIGFDTDVNAPALYEVAYGGHGDISSAVYITVGT
GVGVGVCTNGSAIHGFMHPEGGHIIVPKAPQDVETGFKGVCPFHGDCIEGMVASGSIAARTGVDRR
ELANITDDDPVWDTIAHYLANLCINVTFITSPDVIVIGGGIARREKLFDLIREKFVARVNKYGQQPPVE
>EU_Polysphondilium >gi|281205309|gb|EFA79501.1| hypothetical protein PPL_07552 [Polysphondylium pallidum PN500]
MDSNPHVYAGVEAGGTGFTLAIASGTPDNIVDRVSIPTTTPDETKANVLEWLRGKKFSSIGVASFGPI
DLDTKSSTFGFITTTPKPMWGNTNILGWFDEFQCPKKFDTDVNGAAISETFHGRHERGAISSCAYITV
GTGVGVGVVANEKPIHGLVHPEGGHIFTKLLEDDQFQGTCPFHGNCIEGLVSTGAISKRLGVTADKL
SVIPDDDPVWQIVGHYLAELCATITCIMSPQVIVLGGGVLNRTILYPIIRDELLKILNGYIKSEFLTKEN
VHKYIVQSPFGSNAGIVGALELARRAHLEKN
>EU_Sordaria_macrospora CBI60189.1
MTCDATPWCGTTPAVRRERIEMHSDRIVAGVELGGTKSIAVIGRGREIIDRFRVPTTTPDETLGAVAAK
LAEWRHDHRPAAIGIASFGPISVADGTMLPTPKPHWAGAEIVAPLAQGFDKVAFHTDVTGAALGEGA
FGAAVGLSDFLYVTVGTGVGMGIIAGGRPVTGVMHPEAGHIRVRRRTDDAFAGACPFHGDCLE
>EU_Techamonas AMSG_01753
MAEDGRTSQKVVAAVEAGGTTFVVAVFAAPVTSVWDEPLRRAEMATLRDAQATLDGVVAWISETV
AELGEGNDDGCELVGVGVASFGPVDLNKTSASYGFITTTPKAGWANTDVVGTFAAAFGVPVAFETD
VNAAALAERMAAGAGAEPEDVGDAAFVYVTVGTGVGVGVALDAETAVHGLLHPEAGHMCVGMH
PADAAAGIDNLHVCPFHAWCVEGYVASGGLAARFGVAAADLASVPDDDVGWDAVAWYLAQLVIN
VTLVLSPHTVVLGGGILKRRSLYARVRSHVHTLLNGYLAVPRITDSIEAYLRPPRLAEPGLTGAALLV
>EU_Tetrahymena tr|Q23CY8|Q23CY8_TETTH ROK family protein OS=Tetrahymena thermophila SB210 GN=TTHERM_00052020 PE=4 SV=2
MSTELHSNQKLLAFTLGGFLAGAGSLLLFKKLVCRPRNKPSKCPFTGKTKVVTQRIACIELGGTSIRL
AIGIKETHSDGTQTCKFDQETFKTIETKEPEDNIQQIKEYFENQNIDSVGIASFGPICLDETSEQYGFITT
TPKVSWKNFPLLKRVSEVIPHRKTQRIGFDTDVNAAACAEYNFGNHKAKKSLAYITVGTGVGVGLI
VDGKCVHGLTHPEGGHVLIKPAQGETFQGVCKSHGNCVEGMVTNHALAEKLQTTINELSKIEDSHEI
WNSVAYYLAQLCLNLTLISSPEVIVIGGGIMNRQPLLGLIKQNFIKLLNQYVDHPRLSSNIDDYIVKPF
FTDSGLVGSMVV
>EU_Thalassiosira tr|B8C6P8|B8C6P8_THAPS Fructokinase (Fragment) OS=Thalassiosira pseudonana GN=THAPSDRAFT_263374 PE=4 SV=1
AVGIATFGPAGVNKDTANYGRILEGSPKREWRGVDLITPIRQACGLGTAEDDALVGFDTDVNAPALA
EFRHRYKPMTSLSYVTIGTGVGVGLIVNSQPIHGLLHPEGGHVSIQPLEGDTFKGYSWGSEKSPEDGS
SGGDSNQAQAREVLSTLADDHEVWNHASNAIANLCVSILLLTSCQKIVLGGGIMKRSILFDKIRHRV
WVLLNGYLDSSKLADLIVESSWKELGSGLVGAFALALDA
>EU_Trichomonas tr|A2F1V1|A2F1V1_TRIVA ROK family protein OS=Trichomonas vaginalis GN=TVAG_442070 PE=4 SV=1

MSANDGKKYACGIELGGQTASIAICEKVGEIIYKKKGIKTCEPMTPDEAVANIVAAIKESGYAIDRIGIA
SFGPLDLYKGSIGNTPKPNWGFYPLVKKIQEAFPDCKVSMETDVNAPAYSEYLHLKEQDKSIRSVGY
VTIGTGVGVGVFCDGKPLHGRMHPECGHIMAARVKGDTFEGTCPFHGACFEGLISAQALAKRYGC
QQGELQIIPDSDPVWDIYIEYVAQLTVTMSYVYSLDAFIIGGGIITAKGREWIFDKILARSQQLINNYIH
TPIISKPFHGADAGLVGACAVAINPDVFAVEE

>EU_Trichomonas_2 tr|A2FXV9|A2FXV9_TRIVA ROK family protein OS=Trichomonas vaginalis GN=TVAG_099490 PE=4 SV=1

MLASLSRNFGKKYACGIELGGQTAAFAICENLGSFLYKKKGIKTREPTTPDEAVEAIVEGIKSSGYEV
DRIGIASFGPLDVYKGSIGNTPKPKWGNYPLVASIQKEFPEAQVVLETDVNAPAYSEYLHLNSKDNT
VKSVAYATIGTGVGVGVFCDGKPLHGKMHPEGGHFKPFHLPNDNFKGCCPFHGDCVEGMISAVALS
KRTGLSLQQLPQIATDDPVWDCFTEYAAQLSANCALLYSLDYMVIGGGIVTAKGREYLIEKIQKRTK
ELLNGYIHVPKVIKPFYGGDAGLVGATAVALHPDVFTNN

>EU_Trichomonas2 tr|A2ESX9|A2ESX9_TRIVA ROK family protein OS=Trichomonas vaginalis GN=TVAG_474430 PE=4 SV=1

MKYLCGIELGGTSSSAAIIDEDGKYVLKEKGITAENPATLILTLSNILRNSSYTCETLGIASFGPLNVES
GTIGKSPKKGWYYFHVKAEFRKYFPDIPIAMETDVNAPAYSEFIEFSKKNNSIKSLAYLTIGTGIGLGL
YSDGSIYHGRLHPEFGHTYIKKLQNDTFSGVCHIHGDCAEGLISASAISKRLGISMYEIRDIQNDHPIW
DLYVEYVSQIVANAALAYSLDVFVIGGGVTTDPKRGFLYDRIYSRASELINDYIPMPLVVRPHFDRDA
GLIGATVIARRKFNKINANNDKLFSQIL

>Gallus gi|71895297|ref|NP_001026603.1| bifunctional UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase [Gallus gallus]

MEKNGNNRKLRVCVATCNRADYSKLAPIMFGIKAEPQFFELDVVVLGSHLIDDYGNTYRMIEQDDF
DIHTRLHTIVRGEDEAAMVESVGLALVKLPDVLNRLKPDIMIVHGDRFDALALATSAALMNIRILHIE
GGEVSGTIDDSIRHAITKLAHYHVCCTRSAEQHLIAMCEDHDRILLVGCPSYDKLLSAKNKDYMSVI
RMWLGEDVKPRDYIVALQHPVTTDIKHSIKMFELTLDALISFNKRTLVLFPNVDAGSKEMVRVMRK
KGIEHHPNFRAVKHVPFDQFIQLVAHAGCMIGNSSCGVREVGAFGTPVINLGTRQTGRETGENVLHV
RDADTQDKILHALQLQFGKQYPCSKIYGDGNAVPRILKFLKSIDLKEPLQKKFCFPPVKDNISQDIDHI
LETQSALAVDLGGTNLRVAIVSMKGEIVKKYTQLNPKTYEDRLGLILKMCVEAASEAVNLNCRILGV
GISTGGRVNPREGIVLHSTKLIQEWSSVDLRTPISDALHLPVWVDNDGNCAALAERKFGHGKGIENF
VTLITGTGIGGGIIHQHELIHGSSFCAAELGHIVVSLDGPECPCGSQGCIEAYASGIALQREAKKLHDE
DLLLVEGMSMKNEEVVSAAHLIQAAKLGNAKAESILRTAGTALGLGVVNILHTMNPSLVILSGVLAS
HYVNAVKDVIHRQALSSVKTVDVVVSNLADPALLGAASLVLDYTTRRIY

>Homo gi|45360235|gb|AAS59258.1| UDP-N-acetylglucosamine-2-epimerase / N-acetylmannosamine kinase [Homo sapiens]

MEKNGNNRKLRVCVATCNRADYSKLAPIMFGIKTEPEFFELDVVVLGSHLIDDYGNTYRMIEQDDF
DINTRLHTIVRGEDEAAMVESVGLALVKLPDVLNRLKPDIMIVHGDRFDALALATSAALMNIRILHIE
GGEVSGTIDDSIRHAITKLAHYHVCCTRSAEQHLISMCEDHDRILLAGCPSYDKLLSAKNKDYMSIIL
MWLGDDVKSKDYIVALQHPVTTDIKHSIKMFELTLDALISFNKRTLVLFPNIDAGSKEMVRVMRKKG
IEHHPNFRAVKHVPFDQFIQLVAHAGCMIGNSSCGVREVGAFGTPVINLGTRQIGRETGENVLHVRD
ADTQDKILQALHLQFGKQYPCSKIYGDGNAVPRILKFLKSIDLQEPLQKKFCFPPVKENISQDIDHILE
TLSALAVDLGGTNLRVAIVSMKGEIVKKYTQFNPKTYEERINLILQMCVEAAAEAVKLNCRILGVGIS
TGGRVNPREGIVLHSTKLIQEWNSVDLRTPLSDTLHLPVWVDNDGNCAALAERKFGQGKGLENFVT
LITGTGIGGGIIHQHELIHGSSFCAAELGHLVVSLDGPDCSCGSHGCIEAYASGMALQREAKKLHDED
LLLVEGMSVPKDEAVGALHLIQAAKLGNAKAQSILRTAGTALGLGVVNILHTMNPSLVILSGVLASH
YIHIVKDVIRQQALSSVQDVDVVVSDLVDPALLGAASMVLDYTTRRIY

>Mus gi|4456673|emb|CAB36908.1| UDP-N-acetylglucosamine 2-epimerase/N-acylmannosamine kinase [Mus musculus]

MEKNGNNRKLRVCVATCNRADYSKLAPIMFGIKTEPAFFELDVVVLGSHLIDDYGNTYRMIEQDDF
DINTRLHTIVRGEDEAAMVESVGLALVKLPDVLNRLKPDIMIVHGDRFDALALATSAALMNIRILHIE
GGEVSGTIDDSIRHAITKLAHYHVCCTRSAEQHLISMCEDHDRILLAGCPSYDKLLSAKNKDYMSIIR
MWLGDDVKCKDYIVALQHPVTTDIKHSIKMFELTLDALISFNKRTLVLFPNIDAGSKEMVRVMRKK
GIEHHPNFRAVKHVPFDQFIQLVAHAGCMIGNSSCGVREVGAFGTPVINLGTRQIGRETGENVLHVR
DADTQDKILQALHLQFGKQYPCSKIYGDGNAVPRILKFLKSIDLQEPLQKKFCFPPVKENISQDIDHIL
ETLSALAVDLGGTNLRVAIVSMKGEIVKKYTQFNPKTYEERISLILQMCVEAAAEAVKLNCRILGVGI
STGGRVNPQEGVVLHSTKLIQEWNSVDLRTPLSDTLHLPVWVDNDGNCAAMAERKFGQGKGQENF
VTLITGTGIGGGIIHQHELIHGSSFCAADVGHLVVSLDGPDCSCGSHGCIEAYASGMALQREAKKLHD
EDLLLVEGMSVPKDEAVVALHLIQAAKLGNVKAQSILRTAGTALGLGVVNILHTMNPSLVILSGVLA
SHYIHIVKDVIRQQALSSVQDVDVVVSDLVDPALLGAASMVLDYTTRRIH
>Paracentrotus1 GLEAN3_05241
MYSLMVDALMEFNKRVIMLFPNIDAGSKDIVRIMRQKGVETNPMFYPVKHIPFEEFIILVANAGCMI
GNSSAGVREAGAFGTPVVNLGSRQTGRETGENVLHCRDADTTQKIHHALEIQYQRQFPPSYIYGDG
HAVPRIIKFLKSIKQEDTIQKQFIFPPMPESPSIDIDHILETQSALAVDMGGTQLRVALIAADGEIKLRRS
RPTPHTNNEERMKALLELLLEATKESYALNCRVLGVGISTGGRVNSKEGLVMHSTKAIAGWDEIDLR
TPIETTLHLPVWVDNDGNCAALGERKFGHGRGVKDFITIATGTGIGGGIVLDKRLIHGTNFCAAELG
HIKVSLDGPACQCGSHGCVEAYSSGMALKREAMKLHEAGELLVDGMIVEKGEEVTAKHLVQAAKT
GNQKAKKILDTGAQALGTAITSLLHVLNPKLVILCGVLSSVYLEPVREAVKAHSLPSAVSEVEIVSSQ
LVEPALLGAASLVLEYATRRTY
>Patiria_translation_frame_3 HP128996.1
RLIHGTDFCAAELGHTKVSLEGPPCQCGGTGCVEAYSSGLALQKEAQRLHEEGNLLTEGDHLENNE
KITAKHLVEAARLGNQKAQDIMDRGAKAVASVVVNLLHILNPSRVVLCGVLAPAYIETVRNIVKKEA
LPSAAHAEI
>Saccoglossus2 XR_086474.2 PREDICTED: Saccoglossus kowalevskii UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase-like (LOC100367162), miscRNA
QRCRDGIVGISIPDFGRSTPQFMDGKKDPHDVFMYQNGFQNGHRRTPSSSSLGSESLDRPMRVCVAT
CNRADYSKLAPIMLGLKEDKDFQLQVIVMGCHLIDDYGSTYRMIEQDKIKIDAKLHTIVRGEDEAA
MVESVGLALVKLPDIFLRLKPDILIVHGDRFDALSTATAAALMNIRIIHIEGGEISGTIDDCIRHSITKLA
HYHVCCTERARGRLQAMCEDNERILLAGCTSYDMLLKSNTKNCQLIIDRWVEKGVKPKEYIIALQH
PVTTDLHNSLKMFDLTLEALLEFGKKTVLLFPNIDAGSKDMTRILRRKKIENHPNFSTAKHIPFEEFIV
LVANCACIIGNSSAGVREAGAFGTPVINLGSRQTGRETGENVMHCRDADTKEKILHALELQYQKQY
PPSYIYGDGHAVSRIIKFMKELDVIDSIQKQFIFPPMPDGVSQDIDHILEIQSAIAVDLGGTNLRVSIVSK
AGEIMHMMSKPTPSDNTARMESLLTLLVDASQIATKLNCRVLGVGVSTGGRVNPHEGLVMHSTKSL
EGWSAIDLRTPISSALHLPVWVDNDGNCAALGERKFGHGKGKEDFLAIATGTGIGGGLVLNGRLVH
GSNFCAAELGHIAVSIDGPACMCGSSGCVEAYSSGLALMREAKKLHQAGELIVDGMKMT?
NEIISAKHLIQAAKLGNKKAEKIIHLGGEALGTAVISILHLLNPTLVVLCGVLAPTYVDIVRETIENKAL
PSARGVKVQVSLLKEPALMGAASLVLEYATRRIY
>Strongylocentrotus1 gi|115637356|ref|XP_795233.2| PREDICTED: similar to UDP-N-acetylglucosamine-2-epimerase/N-acetylmannosamine kinase [Strongylocentrotus purpuratus]
MVEIAFNRFLNQQHQRKLSGTMTSLNGSGVKANCGGDMDTKRKFRVCVATTNRADYSKLGNIMQ
AIKDDEMLELSTIVLGCHLIDDYGSTYRLIEKDGFTIDSRLHTIVRGEDEAAMVESVGLAMVKLPDI
MLRLKPDVVIVHGDRFDVLSLAACAALMNIRIVHVEGGEVSGTIDDSIRHTISKLAHYHVCCTERAH
KRLLAMCEDNDRILLAGCPSYDKLLSTDVVDCQHILNRWLKGEGKSKEYIVALQHPVTTNIADSLK
MYSLMVDALMEFNKRVIMLFPNIDAGSKDIVRIMRQKGVETNPMFYPVKHIPFEEFIILVANAGCMI
GNSSAGVREAGAFGTPVVNLGSRQTGRETGENVLHCRDADTTQKIHHALEIQYQRQFPPSYIYGDG
HAVPRIIKFLKSIKQEDTIQKQFIFPPMPESPSIDIDHILETQSALAVDMGGTQLRVALIAADGEIKLRRS
RPTPHTTNEDRMKALLELLLEATKESYALNCRVLGVGISTGGRVNSKEGLVMHSTKAIAGWDEIDLR
TPIETTLHLPVWVDNDGNCAALGERKFGHGRGVKDFITIATGTGIGGGIVLDKRLIHGTNFCAAELG
HIKVSLDGPACQCGSHGCVEAYSSGMALKREAMKLHEAGELLVDGMIVEKGEEVTAKHLVQAAKT
GNQKAKKILDTGAQALGTAITSLLHVLNPKLVILCGVLSNVYLEPVREAVKTHSLPSAVSEVEIVSSQ
LVEPALLGAASLVLEYATRRTY
>Symsagittifera JF826132

MKIAVATCNRADYSKLAPVLRGLKQDTTFQVSIVVMGSHLIDDYGNTYRFIERDGNQIDSMLHTLVR
GETEGAMVESMGLAMIKLPDILNRLQPDIVMVHGDRFDAMSVAISAAVMNVRVLHLEGGEVSGTID
DVIRHSITKLAHYHICCTKGARLRIESMCEDSSRVLLAGCPAYDELIHTDVSKYHEAFERWLPNHVGT
KNFIICVYHPVTTAIAESIQHFGLLLDALVHFDVRTVILFPNVDAGSKELVRCIRLRNLENHEKISCYK
HVPFSEFVYLMGNCGLMIGNSSAGIRESNVFGTAVINIGTRQRGRQSGANVVHVKNPTSAEDLLWHI
RSQFGKVYPRDYIYGDGRAVERIVKFIKEIDCDSPVEKSFNFDRNSSGLRAPSKASFGAEFPLSRQAH
SLQLALAVDLGGSFIRVALVDDRGNIERMRRTEMVDCPDERIAIIVRMANDLLENKRDLVVGVGVST
GGRVNSETGEILFATKVLSGWGGVALKTRLQEQLGLPCYVENDGNCAALAEVHFGSHKMEDMVVL
HFGTGIGGGIIQDGCLLNGSSYSAGEFGHIVVCFDDGPDCMCGNSGCVEAYAGGWALNKLAKEMEL
SSNWRKEGADDQKPVTPGQPNKPTYLTDLASDGVDYAVQHINRAVRAVASALLTIYSSYNPPVAILA
GPLAPVYFDGVKNKLEERSAILGRNFTLLQSDMTEMSLKGAATLVLNNPSRAVPQNAIV*
>Takifugu tr|Q4W392|Q4W392_TAKRU Putative UDP-N-acetylglucosamine 2-epimerase/N-acetylmannosamine kinase (Fragment) OS=Takifugu rubripes GN=glcne PE=2 SV=1
MFGLKSHPDEFELEVVVLGSHLIDDYGNTFRMIEQDDFDIGSKLHTIVRGEDEAAMVESVGLALVKL
PDVIQRLHPDILVVHGDRFDALALATAAALMNIRILHVEGGEVSGTIDDSIRHAISKLAHYHACCTRM
AEQHLIAMCEDHSRILLAGCPSYDKLLLSHQKEDYMDIIKSWLGDNVKEHDYIVALQHPVTTDIKNS
IKIYGLMLDALLSFNKTTLILFPNIDAGSKEMVRVMRKKGIEQHPNFRAVKHIPFEQFIQLVNHAGCM
IGNSSCGVREAGAFGTPVINLGTRQTGRETGENVLHVRDADTHNKIYHALELQFGKRYPCSKIYGDG
NAVPRILKFLSSIDLDEPLQKTFCFPPVKDPISQDIDHILETQSALAVDLGGTNLRVAIICMRGNIVRKY
TQANPKTFEARMQLILKMCSDAMRDAVFLNCRILGVGVSTGGRVNPQEGVVLHSTKLIQEWSAVDL
RTPISDALHLPVWVDNDGNCAALAEKKFGHGKGVENFVTVITGTGIGGGIIHQNELVHGSTFCAAEL
GHIMVSFDGPECSCGSRGCIESIASGMALQREAKRLHDEDLLKVDGLDMKISDPITAAHLINAARLG
NSKANIVLNKASTALGMGIINILHIMNPSLVILSGVLGSYYQAPVQRTIMERALFSAQSVKVVISDLEE
PALLGAASMVLDYATRRIY
>Tetraodon tr|Q4RJ92|Q4RJ92_TETNG Chromosome 18 SCAF15038, whole genome shotgun sequence OS=Tetraodon nigroviridis GN=GSTENG00033530001 PE=4 SV=1
MTRVGSAAPAAPRAAGQSQSPAPPPVQSSPSPCRIAMTTPPSASFSPFFSNATGFCKLRTLMNAGIKTE
LYSWRMQRGRVKMDRRRMEDQNQCRKKLRVCVATCNRADYSKLAPIMFGLKSHPDEFELEVVVL
GSHLIDDYGNTFRMIEQDDFDIGSKLHTIVRGEDEAAMVESVGLALVKLPDVLQRLQPDILLVHGDR
FDALALATAAALMNIRILHLEGGEVRSGTIDDSIRHAISKLAHYHACCTRMAEQHLIAMCEDHTRILL
AGCPSYDKLLLTHHKEDYMDIIKSWLGDKVKDHDYIVALQHPVTTDIKNSIKIYGLMLDALLSFNKT
TLILFPNIDAGSKEMVRVMRKKGIEQHPNFRAVKHIPFEQFIQLVCHAGCMIGNSSCGVREAGAYGTP
VINLGTRQTGRETGENVLHVRDADTHNKIYHALELQFGKRYPCSKIYGDGNAVPRILKFLRSIDLEEP
LQKTFCFPPVKDPISQDIDHILETQSALAVDLGGTNLRVAIICMRGKIVKKYTQANPKTFEARMQLILK
MCSDAVRDAVFLNCRVLGVGVSTGGRVNPQEGVVLHSTKLIQEWSSVDLRTPISDALLLPVWVDND
GNCAALAEKKFGHGKGVENFVTVITGTGIGGGIIHQNELVHGSTFCAAELGHIMVSFDGPECSCGSR
GCIESYASGMALQREAKRLHDEDLLKVDGMDLKISEPITAAHLINAARLGNSKANIVLNKASTALGM
GIINILHTVNPSLVILSGVLGSYYQAPVQRIISERALFSAHSIKVVTSDLEEPALLGAASMVLDYATRRT
Y
>Xenopus gi|118404390|ref|NP_001072728.1| glucosamine (UDP-N-acetyl)-2-epimerase/N-acetylmannosamine kinase [Xenopus (Silurana) tropicalis]
MEKNINQRKLRVCVATCNRADYSKLAPIMFGIKAEPDRFVLSVVVIGSHLIDDYGNTYRMIEQDDFD
IQARLHTIVRGEDEASMVESVGLALVKLPDVLNRLNPDIIVVHGDRFDALALATSAALMNIRILHIEG
GEVSGTIDDSIRHSITKLAHYHACCTRSAEQHLIAMCEDHDRILLAGCPSYDKLLSVNNKDYMSVIK
MWLGDDAKSGEYIVALQHPVTTDIKHSIKMFEFTLDALLSFNKKTLILFPNIDAGSKEMVRVMRKK
GVEHHPNFRAVKHVPFEQFIQLVAHAGCMIGNSSCGVREAGAFGTPVINLGTRQTGRETGENVLHV
RDADTQNKIIHALQLQFGKRYPCSKIYGDGNAVPRIVKFLKSISLDEPLQKKFCFPPVKESISQDIDHIL
ETLSALAVDLGGTNLRIGIVSMTGEIIKKYVQPNPKTYEDRIELILKMCVEAASEAVKLNCRILGVGIS
TGGRVNPREGVVLHSTKLIQEWSSVDLRTPLSDTLHLPVWVDNDGNCAALAERKFGQGKGIEDFVT
VITGTGIGGGVIHNHELVHGSSFCAGELGHIMVSFDGPDCMCGSRGCVEAYASGIALQREAKKLHDE
DMLLVEGMSVKNDESVSAVHLIQAAKFGNTKASNILKTAGTALGIGVINILHTINPSLVILSGVLANQ
YVNVVKDVIRQRGLASIQNVDVVVSSLSDPALLGAASMVLDYTTRRTY
>XENOTURBELLA EC906593.1
LHEQGDLFVDGMKLPNENEEVTSETLVYAARAGNKKAEAVLDRGAKALSAAILTLLHTVNPSLIIIT
GVLGPVYIDKVQDTIKKAALPAAQQVDIKVSKLKEPALLGAASIALDFMTRR

cytidil sequences

>A_ sp|Q58463|Y1063_METJA Uncharacterized protein MJ1063 OS=Methanocaldococcus jannaschii GN=MJ1063 PE=4 SV=1
MKIIGIIQARTGSKRLKNKVLLKLGDRCILEILLERLKKSKKLDDIIVATTIKKEDNAIVELCNSLGVNV
FRGSEKDVLDRFYNASKFYSGDVIVRITGDNPLTSIELIDKQVEYLLKNNFDYVSTKNIILGLSSEVFT
FDALEKAWKNAKEKYQREHVTPYIYENPNLFKVFYLEPPEYLKREGIRLTIDTIKDFKLYLELQKHFD
LINVDIRQIIDFLDKNPQIKNINSNVRQKSYREVEE
>A__2 tr|Q6KZH4|Q6KZH4_PICTO Molybdopterin-guanine dinucleotide biosynthesis protein A OS=Picrophilus torridus (strain ATCC 700027 / DSM 9790 / JCM 10055 / NBRC 100828) GN=PTO1293 PE=4 SV=1
MISIIFAKKSERFPGKHHAMICGEEMIKRISRIINESMLFDKIILYTKDETLYSEYCDIVIDKSKGTLIDSL
MSCLKEYSEFLAVGGDMPLLDYNILKYLMDNYTGRSLAVSSYGFIQPLLSIYNKSIINDLDKYILSGS
KSIYKFIMKSNFKILNINTFKTMSVNTLNDLKEINYYLGC
>A__3 tr|Q8ZW75|Q8ZW75_PYRAE Conserved protein OS=Pyrobaculum aerophilum GN=PAE1934 PE=4 SV=1
MAPLEVKCAGVVLAAGGSARFGSQKLLANFKGRPLVWHAAETLRSAGLETYIVVNSREVASAAGR
VDGIIYNPWWRQGLSTSLKAALIALYQKKCIVWMPGDMPCVKPDTVLKIASACKSGLAVPVYRGAR
GNPVASCRDVYALALGITGDVGLRVLLNAVPTLSLEVEDAGVLADVDFPGDLQRLPC
>A__4 tr|Q465W2|Q465W2_METBF Putative nucleotidyltransferase OS=Methanosarcina barkeri (strain Fusaro / DSM 804) GN=Mbar_A3457 PE=4 SV=1
MDAIVMAGGFGQRLGMGEKPCVELLGKPLIAYVIDTLRASENIDRVFVAVSPVTPRTEIMIQERYKGE
VRVIRTFGGNYVGDMIHAVETAETVGPVMIIMSDLPLINSELIDSVIEKYKEEGKPALSVYVPINVCKG
AGIRPDTVFNKDGKLIVPAGINILDSSQIRKEQEDFNLILDNPKLAINVNTVEDLQHCKEMLQGQD
>A__5 tr|A9A1S8|A9A1S8_NITMS Acylneuraminate cytidylyltransferase OS=Nitrosopumilus maritimus (strain SCM1) GN=Nmar_0149 PE=4 SV=1
MNSAAIVSVRNSSTRLPNKAIMEIKNNFTSIDVVIQRAKKTELPVIIATSTSKEDDIFEDIAKKNDVEIF
RGPLINKIKRWYECFNKFGIQYALLVDGDDLSYSYDIGKRAISELKNKSVDMITHPKDIVTGFFTSAIN
EKGIKKLFSVAPTDEINTDVITRYIEKANLTTDIVPLKDFEKNENVRFTLDYKEDLEFFKKLYENLDIL
SSGKEILDFLNNNKHLIQINFHKQKEFLDNQAKFNESVK
>A__6 tr|A9A1R4|A9A1R4_NITMS Acylneuraminate cytidylyltransferase OS=Nitrosopumilus maritimus (strain SCM1) GN=Nmar_0135 PE=4 SV=1
MKFCNVGIIIQARTGSDRFPKKVLASIEKKPMIWHIVNRCKKVKNIDKIILATTTLKEDYPLISLAKKN
KIEYFRGSKNDVLDRFYQCSTSNNLDIIIRITGDCPLVDPKLIDQFLDFFSHKKYDYVSNTINPTYPDG
LDIEIFSFKALKKAWNMSKKKSNREHVTTFIKHHPEKFKIKNFENNTNLSNYRLTVDHKNDLKLIRKI
YKEFRPNIKFSTKSVISLLNKNPELFKINQNISRNEGYEKSLLNDN
>B_Blast gi|149174391|ref|ZP_01853018.1| CMP-N-acetylneuraminic acid synthetase [Planctomyces maris DSM 8797]
MAGAIGLITARGGSKGVPRKNIKVLAGKPLIAWTIEAALQSQELERVIVSTDDKEIASISRQYGAEVPF
LRPLKLSLDGSSHADVVLHAIDWLIEHEQYESEYITMLQPTSPFRIADDIDGSIRFAREKNAKAVIGM
MEAPSHPVCLRLMTEAGLLVELTPQQEESALRRQLLPEVYAFNGALYVLRTEEFRKTKTFRPQGETY
GYKMPAERSWEIDTEWEFLVASLLMENQVSQTATRTAA
>B_Blast_2 gi|218886935|ref|YP_002436256.1| acylneuraminate cytidylyltransferase [Desulfovibrio vulgaris str. 'Miyazaki F']
MPTPAPHSAETAGPILGLITARGGSKRIPGKNLRIAGGKPLLAWTIEAALQSRLLSRVVLSTDDPQIAA
AGRLYGAEVPFLRPAELATDTSAHILCVLDALDRLRDMDGFAPTAVCLLQPTSPLRQACDIDALLSEA
AATRPPAMVSVNACTEHPYFARSVSATGELMPFVPQNLAYAREQDLEPACFINGAIYYNTVESLRKH
KTFYPEGLRGHLMPRERSLQVDEPFDLHLADLLLSNPMPSTARSTGCENS
>B_Blast_3 gi|295401162|ref|ZP_06811135.1| acylneuraminate cytidylyltransferase [Geobacillus thermoglucosidasius C56-YS93]

MINGKKILAIIPARGGSKGIPKKNIKPLRGKPLIAWTIEEAKKSKLLDKIIVSTDDEEIMNVAKKWGAE
VPFLRPSELARDDTPGIAPVLHALEYFSDYEYVVVLQPTSPLRLAEDIDEAIYLCEKNKSNFCVSVTE
SKIIPDWMFRINNQGMLEPLNSNREIPYQRQKAKKTYVLNGAVYVGRKEALIKTRSFLTPETLPYIMP
NIRSVDIDDMDDFLYCEYILGRNE
>B_Blast_4 gi|88810764|ref|ZP_01126021.1| acylneuraminate cytidylyltransferase [Nitrococcus mobilis Nb-231]
MNDVPRILGIIPARGGSKGVPRKNVRCVAGKPLIGYTIEAAQGASRLSRVVVSTEDAEISAVARACGV
EVLPRPVDLAADDTPGIAPVLHAMEQMSEYDYVVLLQPTSPLRTALDIDAAIETCMSAQAPACVSVS
EAAQSPYWMFRIEAGGKMQPLLSASADEYSRRQNVPPAYALNGAVYVAVTDWLRERQTFLTSETVA
YTMPQERSLDIDSEWDFLIFGLVLEKKP
>B_Blast_5 gi|261855251|ref|YP_003262534.1| N-acylneuraminate cytidylyltransferase [Halothiobacillus neapolitanus c2]
MSDIGSVLALIPARGGSKGLPGKNIRPLKGRPLIGWSIEAARTSRYVSRVVVSSDDEEILAVARDQGAE
TPFRRPASLAGDATPSMDVVLHALDQLAEFEWVVLLQPTSPLRLSADIDAAIEQCLKTNAPACVSVC
EAPASPWWMFEVGAECRMRSFLPAEQRPVRRQDLPDLYALNGAVYVAKTEWLRTSRSFLTEETVAY
VMPPARSVDIDTLFDFQLAECLLGNH
>B_Blast_6 gi|168334400|ref|ZP_02692579.1| CMP-N-acetylneuraminic acid synthetase [Epulopiscium sp. 'N.t. morphotype B']
MEILGIIPARGGSKGLPGKNIKMLGDKPLIAYTIAAAINSKHITRTIVSTDDAAIAEVALKFGAEVPFM
RPECLATDTATSAEVVLHLLKTLEGNGYMPDFICLLQCTSPFRTAADVDSCIEKCLNTGFDACYSVTE
ARSNPYWMKVFEGNQLNSFIDAEMILRRQDLPTVYELNGAIYFAKTEEVIKNKSLHLANATGYVMPI
QKSVDIDTALEFELAKMLMNNI
>B_Blast_7 gi|253997845|ref|YP_003049908.1| acylneuraminate cytidylyltransferase [Methylovorus sp. SIP3-4]
MSRNNEIVCIIPARGGSKGIPGKNLVSLGGKPLLAWSIEAALKSKVLGRVIVSTDSEEIAKVAQQWGA
EVPCIRPNELANDDIHAIHVVLHMLEWLKKEEAYVPAGAMMLLPTSPLRRASDIQGAVKLFLDKKA
SSVVSVVDLGKYMTNLRYLDGDQLVRVAPEENPNAQRQGLKKLHSVNGSIFLARPDILEQSGTFHV
DDALGFVMDNMHSIDINAPEDLNLARKFCEVFEPWKSDKE
>B_Blast_8 gi|281411850|ref|YP_003345929.1| acylneuraminate cytidylyltransferase [Thermotoga naphthophila RKU-10]
MKIVGIIPARRGSKGIRNKNIVNLCGKPLIYYTIKEALKSKVIDKLIVSTDSEKIAKLAKSFGAEVPFIR
PKELATDDAKGIDVILHAMNWFESRGEIFDAVLVLQPTSPLRSHEDIIKAVEVFLEKKANAVVSVCEV
EHHPLWANVLPDDKSMDNFIRKEIRNKNRQELPKYYRLNGAIYLARWNYLKNYKDWFHHKCYALI
MPQERSVDIDSEVDLAVAEYFLKKWGRCYED
>B_ sp|O66914|KDSB_AQUAE 3-deoxy-manno-octulosonate cytidylyltransferase OS=Aquifex aeolicus GN=kdsB PE=3 SV=1
MRRAVIIPARLGSTRLKEKPLKNLLGKPLIRWVVEGLVKTGERVILATDSERVKEVVEDLCEVFLTPS
DLPSGSDRVLYVVRDLDVDLIINYQGDEPFVYEEDIKLIFRELEKGERVVTLARKDKEAYERPEDVK
VVLDREGYALYFSRSPIPYFRKNDTFYPLKHVGIYGFRKETLMEFGAMPPSKLEQIEGLEQLRLLENG
IKIKVLITENYYHGVDTEEDLKIVEEKLKNL
>B__2 sp|Q6G4U6|KDSB_BARHE 3-deoxy-manno-octulosonate cytidylyltransferase OS=Bartonella henselae GN=kdsB PE=3 SV=1
MALEPIILIPARIGSTRLPQKALAEIAGKPMIVHVAEQAKKAAFGRIIVATDHNNIAKVVTAYGHECIIT
CRDHKSGSDRIYEALTHIDPERRYNVILNVQGDLPTITPHEIISALRPLENSLTDIATLGAKIVEENEKT
DPNIVKIIGTPLSHNRFRALYFTRATAPYGDGPLYHHIGIYAYRREALEKFVALKPSPLEQREKLEQLR
ALEHNMRIDVEIVDTIPLGVDTQRDLERVRKILA
>B__3 sp|Q2YPQ5|KDSB_BRUA2 3-deoxy-manno-octulosonate cytidylyltransferase OS=Brucella abortus (strain 2308) GN=kdsB PE=3 SV=1
MLQTMKTLTLIPARLGSTRLPNKPLADICGKPMIVHVADRAAAAKLGRTVIATDSEEIFKVVAAHGH
EAIMTRGDHESGSDRIYEALAKLDPSGEIDAVVNVQGDLPTIDPDTIRRALLPLEDGPADIATLGVEIT
VEEEKTNPNVVKIVGSPLAGNRRLRALYFTRATAPYGEGPLYHHIGLYAYRRSALERFVKLGPSPLEK
REKLEQLRALEAGMRIDVEIVKTVPLGVDTQADLDRARTFCSQAGTI
>B__4 sp|Q3ATT2|KDSB_CHLCH 3-deoxy-manno-octulosonate cytidylyltransferase OS=Chlorobium chlorochromatii (strain CaD3) GN=kdsB PE=3 SV=1

```
MNAIIIIPARLGSTRLPEKMLADIEGEPLIVRTWRQAMQCCRASRVVVATDSVKIAEVLTTYGAEVVM
TSPEARCGSERIAEAARQFACDVVVNLQGDEPLISHETIDLALEPFFSPNPPDCSTLVFPLQPDDWAQL
HDPNQVKVVLNREGYALYFSRSPIPFQRNQLTSTQCYRHVGLYAFKAEVLQCFAALPPTMLEEAESL
EQLRLLEHGYRIRCMVTHDDQPGVNTAEDLELVRTLFKQRHQEA
>B__5 sp|P0CD75|KDSB_CHLTR 3-deoxy-manno-octulosonate cytidylyltransferase OS=Chlamydia trachomatis GN=kdsB PE=3 SV=1
MFAFLTSKKVGILPSRWGSSRFPGKPLAKILGKTLVQRSYENALSSQSLDCVVVATDDQRIFDHVVEF
GGLCVMTSTSCANGTERVEEVVSRHFPQAEIVVNIQGDEPCLSPTVIDGLVSTLENNPAADMVTSVT
ETTDPEAILTDHKVKCVFDKNGKALYFSRSAIPHNFKHPTPIYLHIGVYAFRKAFLSEYVKIPPSSLSL
AEDLEQLRVLEIGRSIYVHVVQNATGPSVDYPEDITKVEQYLLCLSKASF
>B__6 sp|O25016|KDSB_HELPY 3-deoxy-manno-octulosonate cytidylyltransferase OS=Helicobacter pylori GN=kdsB PE=3 SV=1
MIIIPARLKSSRFENKVLEDIFGLPMVVRCAKNANLVDECVVACDDESIMQTCQKFHIKAVLTSKHHN
SGTERCLEAARILGLKNDERVLNLQGDEPFLEKEVILALLEATKNAPFMATCAKVIDEEQAKSPNLV
KVVLDSQNNALYFSRSLIPFLRDFDAKRQTPLLGHIGIYGFHNKEILEELCALKPCVLEEIEKLEQLRA
LYYQKKIAVKIVQSESVGIDTQEDLQNALKIFSPDLLER
>B__7 sp|Q5ZU88|KDSB_LEGPH 3-deoxy-manno-octulosonate cytidylyltransferase OS=Legionella pneumophila subsp. pneumophila (strain Philadelphia 1 / ATCC 33152 / DSM 7513) GN=kdsB PE=3 SV=1
MSHNFHVIIPARYHSSRFPGKLLQEINGITVIERVYRQALLAEPKSVIIATDHDEIADRAIQFGAEVVITS
HTHQTGTDRIAEVIAKGSFAPDDVIVNVQGDEPFIRPKLIQQVACSLTKTKAPVSTLCWPISSLEILNN
PNVVKVVCTRDNHALYFSRSAIPFHRDDKNAYSNTFRHIGLYAYRAAFLLEFVSWPPCTLEQIECLEQ
LRILWSGFSIRVDEACEEPLQDINTKEDLILAQQYFLDTFNV
>B__8 sp|Q8F0C3|KDSB_LEPIN 3-deoxy-manno-octulosonate cytidylyltransferase OS=Leptospira interrogans GN=kdsB PE=3 SV=2
MKKILGVIPARYASSRFPGKPLAKIGDKTMIEWTYRNASRSSVLSELVVATDDVRIHEVVQKFGGRSV
MTSSDHPSGTDRIIEVANQFSEYSIIVNIQGDEPGIEPELIDGVASLKASHPEWAMSTAAVPLLDFSHAI
DFNRVKVIIDRNGKAIYFSRSLIPSQFKTTVPLYRHLGIYGYDRDFLLQYNSLPKSNLEESESLEQLRAI
EAGYGIGIYLSKEAGLSVDTPADLEIVIEDFKKRKWISE
>B__9 sp|Q2VZK3|KDSB_MAGSA 3-deoxy-manno-octulosonate cytidylyltransferase OS=Magnetospirillum magneticum (strain AMB-1 / ATCC 700264) GN=kdsB PE=3 SV=1
MAELTPIIVIPARMQATRLPGKPLADIHGEPMIIHVWRRSVQAGLGPVVVACSEAEVFDAVHAHGGQ
AVMTDPDHPSGSDRVWEAVRKLDPEGRFDAIVNVQGDLPTLDPQIIRAVFAPLAEPGVDVATLVTEIT
NEEERTNPNVVKAVVGLRPGQRVGRALYFSRATVPANAGPHYHHIGLYAYRRDSLERFVSLPQGVLE
SREKLEQLRALENGMRIDCALVDTVPLGVDTPADLERARALLKA
>B__10 sp|Q1DDB3|KDSB_MYXXD 3-deoxy-manno-octulosonate cytidylyltransferase OS=Myxococcus xanthus (strain DK 1622) GN=kdsB PE=3 SV=1
MQSCRTVAVIPARHASTRFPGKPLAIIAGRTMIEHVWRRCQEAQAFDEVWVATDDDRIRAAVEGFGG
KAVMTSPACATGTDRVAEVALGRPDIDIWVNVQGDEPLVDPATLQRLAGLFQDASVRMGTLVRPLEA
DEAASPHVVKAVLALNGDALYFSRSLVPHVREPGTPVQRWGHIGLYGYRREVLLSLAKLAPTPLED
AEKLEQLRALEHGIPIRCAKVTSHTVAVDLPGDVEKVEALMRARGG
>B__11 sp|Q1QR80|KDSB_NITHX 3-deoxy-manno-octulosonate cytidylyltransferase OS=Nitrobacter hamburgensis (strain X14 / DSM 10229) GN=kdsB PE=3 SV=1
MTKTRTLVLIPARMAATRLPGKPLLDIGGLPMVVHVLRRAETAGIGRVAVATDTPEIAAAVTAHGGEA
IMTRADHPSGSDRVFEALDRLDPDGRIETIINLQGDFPTIRPDIIRDVLKPLADPAVDIATLAAEIHSDEE
ATNPNVVKAVGSPVAPRRLRALYFTRATAPHGDGPRYHHIGLYAYRRKALQRFVELPPSPLERQERLE
QLRALEGGMRIDIMIVDDVPRGVDTAADLETARRILASVPIPKLAYPRSNL
>B__12 sp|Q9HZM5|KDSB_PSEAE 3-deoxy-manno-octulosonate cytidylyltransferase OS=Pseudomonas aeruginosa GN=kdsB PE=3 SV=1
MTQAFTVVIPARYASTRLPGKPLQDIAGQPMIQRVWNQARKSAASRVVVATDDERILAACQGFGAE
ALLTRAEHNSGTDRLEEVASRLGLASDAIVVNVQGDEPLIPPALIDQVAANLAAHPEAAIATLAEPIHE
VSALFNPNVVKVATDIDGLALTFSRAPLPWARDAFARDRDSLPEGVPYRRHIGIYAYRVGFLADFVA
WGPCWLENAESLEQLRALWHGVRIHVADARENMLPGVDTPEDLERVRRVLGG
>B__13 sp|B0BXB4|KDSB_RICRO 3-deoxy-manno-octulosonate cytidylyltransferase OS=Rickettsia rickettsii (strain Iowa) GN=kdsB PE=3 SV=1
```

MRHQDVAIIIPSRLSSTRLKQKPLQLIGSITLIERVFKQVNQAGLEHTYVATDSEEIASVITKVGGKVIFTDSAIPTGTDRTYEAFKLIPNNQNINYIVNVQGDMPFIEPSSILKIIEYLKNSKYDIVTPIVKVDRESVKASSNVTVVVDSAGTALYFSRSLIPNGAEEFLYHVGMYGFRKNALEKFVSLKPTFLEKTERLEQLRALENGMTIGTCLVENVPISVDTEEDLKKAVKFYENISKLGL

>B__14 sp|B5F1S0|KDSB_SALA4 3-deoxy-manno-octulosonate cytidylyltransferase OS=Salmonella agona (strain SL483) GN=kdsB PE=3 SV=1

MSFVVIIPARFSSTRLPGKPLVDINGKPMIVHVLERARESGAERIIVATDHEDVARAVEAAGGEVCMTRADHQSGTERLAEVVEKCGFSDDTVIVNVQGDEPMIPAVIIRQVAENLAQRQVGMATLAVPIHSAEEAFNPNAVKVVLDAEGYALYFSRATIPWDRDRFAKSLETVGDTCLRHLGIYGYRAGFIRRYVSWQPSPLEHIEMLEQLRVLWYGEKIHVAVAKAVPGTGVDTADDLERVRAEMR

>B__15 sp|P0A0Z8|NEUA_NEIME N-acylneuraminate cytidylyltransferase OS=Neisseria meningitidis GN=neuA PE=1 SV=1

MEKQNIAVILARQNSKGLPLKNLRKMNGISLLGHTINAAISSKCFDRIIVSTDGGLIAEEAKNFGVEVVLRPAELASDTASSISGVIHALETIGSNSGTVTLLQPTSPLRTGAHIREAFSLFDEKIKGSVVSACPMEHHPLKTLLQINNGEYAPMRHLSDLEQPRQQLPQAFRPNGAIYINDTASLIANNCFFIAPTKLYIMSHQDSIDIDTELDLQQAENILNHKES

>B__16 tr|Q8KHC9|Q8KHC9_PSEAE Similar to Cytidylyltransferase OS=Pseudomonas aeruginosa PE=4 SV=1

MKLAVIPARGGSKRIPRKNIKIFCGQPMIAWSIQAALNSACFDRIVVSTDDAEIAEVARELGAEVPFIRPHVLADDHTGTLPVIRHAIQACAMQGFHATQVCCIYATAPFITTKDILQGLEISERSGITYAFTATTYAFPIQRALKLHPAGGVEMFNPEHFNTRSQDLEEVIHDAGQFYWGKVDSWMQELPFFTVNSCPVLLPRHRVQDIDTVEDWQRAEWLFKAMRSQ

>B__17 tr|A8V931|A8V931_NEIME CMP-NeuNAc synthetase OS=Neisseria meningitidis GN=siaB PE=4 SV=1

MEKQNIAVILARQNSKGLPLKNLRKMNGISLLGHTINAAISSKCFDRIIVSTDGGLIAEEAKNFGVEVVLRPAELASDTASSISGVIHALETIGSNSGTVTLLQPTSPLRTGAHIREAFSLFDEKIKGSVVSACPMEHHPLKTLLQINNGEYAPMRHLSDLEQPRQQLPQAFRPNGAIYINDTASLIANNCFFIAPTKLYIMSHQDSIDIDTELDLQQAENILNHKES

>B__18 tr|A3GPP8|A3GPP8_VIBCH 3-deoxy-manno-octulosonate cytidylyltransferase OS=Vibrio cholerae NCTC 8457 GN=kdsB PE=3 SV=1

MSFTVVIPARYQSTRLPGKPLADIGGKPMIQWVYEQAMQAGADRVIIATDDERVEQAVQAFGGVVCMTSPNHQSGTERLAEVVAKMAIPADHIVVNVQGDEPLIPPAIIRQVADNLAACSAPMATLAVEIEDEAEVFNPNAVKVITDKSGYALYFSRATIPWDRDNFAKADKAIVQPLLRHIGIYAYRAGFINTYLDWQPSQLEKIECLEQLRVLWHGEKIHVAVALEAPPAGVDTPEDLEVVRRIVAERAQ

>B__19 tr|Q7P7V1|Q7P7V1_FUSNV 3-deoxy-manno-octulosonate cytidylyltransferase OS=Fusobacterium nucleatum subsp. vincentii ATCC 49256 GN=FNV1841 PE=3 SV=1

MKFLGIIPARYSSTRLEGKPLKMIEGHTMIEWVYKRAKKSNLDALIVATDDERIYNEVINFGGQAIMTSKNHANGTSRIAEVCEKMTEYDTVINIQGDEPLIEYEMINSLIDTFKENSNLKMATLKHKLIDKEEIENPNNVKVVCDKNDYAIYFSRSVIPYPRKNENISYFKHIGIYGYKRDFVIDYSKMSATPLEETESLEQLRVLENGYKIKVLETTHSLIGVDTQENLEQVINYIKENNIKI

>B__20 tr|Q7X5M6|Q7X5M6_PSEAE Putative uncharacterized protein OS=Pseudomonas aeruginosa PE=4 SV=1

MRDSMKYAVVIPARHASTRLPGKPLLDLCGVPMIVRTYRQCIQAVDAEHVLVATDDERIRAVCEGEGIRTLMTSSRCLTGTDRVAEVAGQVSAEIFINVQGDEPLFNPDDLRKLIDAAQASPEAIINGYCGIADETTFRNPSVPKVVFRPDGRLLYMSRAAIPTTKQGEFSRAWRQVCAYAFPREALRAFAARPTKTSLEEVEDIEILRFLELGWEVKMIEMSDQSISVDNLEDVERVLDAILQRKPRTASVGGDQHAP

>B__21 tr|A8KKK0|A8KKK0_BURPS 3-deoxy-manno-octulosonate cytidylyltransferase OS=Burkholderia pseudomallei Pasteur 52237 GN=kdsB PE=3 SV=1

MTSPLPFVAVVPARLASTRLPNKPLADLGGKPMVVRVAERAREAGAQQVLVASDAQRVLDAVREHGFDAVLTRADHPSGTDRLAEVAAKLGFDDDTIVVNVQGDEPLIDPQLVRDVASHLAAHPSCAIATAAHPIHEAHEVFNPNYVKVVLDAHGVALYFSRAPIPWSRDAYLPHWPNVAAMPAPTCPVYRHIGLYAYRARFLRTYPTLAQAPIEAAEQLEQLRAMWHGERIAVRVTEHAPEAGIDTPADLERVQALFRSRAK

>B__22 tr|B6WU62|B6WU62_9DELT Putative uncharacterized protein OS=Desulfovibrio piger ATCC 29098 GN=DESPIG_01620 PE=4 SV=1

MAAATPCYRISYEQACDPNLVKIVRSDSGAALYFSRARIPFDRDEEGTGAYWGHMGMYAYRPDALK
IFADHEPGMLERTEKLEQLRLLQHGVGIQTVELPPCAPGVDTPEDLERVRAIVKGKEFSTSAPATPATT
SLADIRLVITDVDGVLTDGGLYYGPDGECIKRFHAQDGLGMVMLRKAGIRVAVLSGRDCPALRRRLN
DLGIDMFRLGKVEKRAACESLLQEAGVSAHEALFIGDDLPDLDGFACCGLGVTVANARDEVKRAA
DIVLQTQGGQGAFRELVDMLLPQ
>B__23 tr|Q7P731|Q7P731_FUSNV Acylneuraminate cytidylyltransferase OS=Fusobacterium nucleatum subsp. vincentii ATCC 49256 GN=FNV1768 PE=4 SV=1
MYREKKILCVIPARKGSKRIKWKNIVPLAGSPMLEYTVKCALNSKYIDRVIVSTDSYYIKKLAKKMG
ADTPFIRPKNLATDDAKTIDVLLHAVKYCEEFEKEKYDYLVLLQNTSPLRKSWQVDEAIEKIVSSTLD
SLVSISEVREHPVLMKILSNNKKLIPLLNNLKKNKFRSIYRINGAIFINKIDKNFNSDTILTNNQLPYIM
KRETSIDIDTIEDIKVAEYYLGVEKMKKNQKYILKGEVWRS
>B__24 tr|B1BCE5|B1BCE5_CLOBO Acylneuraminate cytidylyltransferase OS=Clostridium botulinum C str. Eklund GN=CBC_A0051 PE=4 SV=1
MYMNKKILAIIPARGGSKGIPHKNIMKICDKPLISYSIEAAKKSKYIDYILVSTDDVDIKEVSLNYGAK
VPFLRPNEISTDRAKSIDVVLHGINYLKEHNDNFEYVVLLQPTSPLRTSKDIDDAIKNIIESDKDSLISV
CECSENPILMRTIEKEKLKPILEFNGDNLRRQELPTFYVFNGAIYINKVYMLQNKKEFIDENTIPFIMD
SKKSIDIDNMIDAKIAEMILKENKND
>B__25 tr|Q70JW0|Q70JW0_NEIME Cap29eG protein OS=Neisseria meningitidis GN=cap29eG PE=3 SV=1
MNTENLKTVIVVPARYASTRLPGKPLADICGKPMIQHVYERACKVPYIDDVIVAVDDRRVAEVVESF
GGKVIMTSTQHDSGTDRLVEVMGKYAADIYINIQGDEPLIRSEDIALLAQGMKDEQISVGTLCHALP
ASEATNPNTVKVVLSANGNALYFSRSPIPYPRDKEHAYYFKHIGVYAYRKETLAKYSNLKQPDIELSE
KLEQLRLLDAGIDIRVFEVPETGPGVDTPECLERVRAIMSGQLPIDKHSVNLADIRLVITDVDGVLTD
GGIFYNENGECLKRFHVRDGLGIRLLEESGIKVAVLSGRDSPTLRKRIDDLGISYYQLGIKDKHAACIE
LMQEANCLKEQTAYIGDDTIDLPAFSACGLSFAVADTPEYIRNQADITLTLPGGYGAFRELADKILAA
QGKEDVFSSEAGFAQVMHGANQ
>B__26 tr|A4NB24|A4NB24_HAEIN 3-deoxy-manno-octulosonate cytidylyltransferase OS=Haemophilus influenzae 3655 GN=kdsB PE=3 SV=1
MSFTVIIPARFASSRLPGKPLADIAGKPMIQHVFEKALQSGASRVIIATDNENVADVAKNFGAEVCMT
SVNHNSGTERLAEVVEKLAIPDNEIIVNIQGDEPLIPPVIVRQVADNLAKFNVNMASLAVKIHDAEEL
FNPNAVKVLTDKDGYVLYFSRSVIPYDRDQFMNLQDVQKVQLADAYLRHIGIYAYRAGFIKQYVQW
APTQLENLEKLEQLRVLYNGERIHVELAKEVPAVGVDTAEDLEKVRLILAKD
>B__27 tr|C2CKH7|C2CKH7_9FIRM MobA family molybdenum cofactor biosynthesis protein OS=Anaerococcus tetradius ATCC 35098 GN=HMPREF0077_1987 PE=4 SV=1
MKINAILMASGLSKRMGENKLMLEFRGKKIYEYALGLLEEIGFDKVVVASSYEKILKDAEARGFLAL
ENKDNEVGKSSSIKLGVRACDEDAAMMFFVADQPLLSKETCKNLIASFKENPLMTYPRVGMRRGAP
VIFPASYRKKLLSLEADQGGMIFAKDDKTNKVEIGNEDELLDIDTLEAYERLKKDHE
>B__28 tr|C2XJH2|C2XJH2_BACCE Cytidylyltransferase domain protein OS=Bacillus cereus F65185 GN=bcere0025_48950 PE=4 SV=1
MNKQKVVAIIPARGGSKGIPRKNIRLLNGKPLISYAIEVAKKSNLIDKVVVSTDDIEIGNIAKKYGAEVI
MRPDHISSDEVPLDPVIHYTVEKIEEELDESYDIVVTVQPTSPLLSIFTLENVIQKIIKENYDTVLTGLD
DRHLSWKLEQDKFVPNFKERKNRQYLPSEFRESGAVFATKRKCITPNNRMGENITIYVVGSEESIDID
SYTDWWVAEKLLKRKKLIIRVDGYREIGLGHIYRTLTLAHNIFDHEVIFLMDKQYDLGIKLIEKQNFK
IEFFEQDPLPKIREISPDIIINDILDTSTDYMLKLKNMGIKVFNFEDLGPGAEYADGVFNALYPGNVPV
KYFYTGENYYCARPDFINSSTKIIKEHVNKVLITFGGTDPNNLTKKTLDAIVNMPYEFEITVVLGPGY
KYKDAIFKDIDNYSRVINVYTEINNMAEFMLEADVIFSSAGRTMYEIAMIGTPAIIISQNYRELTHLFG
HNYNGFINLGIHHEAREDIIQQSLERLIRDEQLRQMMNNRMLQHDLKRGIERVLSIIFN
>B__29 tr|Q7P734|Q7P734_FUSNV Acylneuraminate cytidylyltransferase OS=Fusobacterium nucleatum subsp. vincentii ATCC 49256 GN=FNV1765 PE=4 SV=1
MGTIALIPVISSKKEGFSDRDMLMLGDSPLAYHTIKVVKASKEFSKIIVFTNSLEYKSIIEKNGVEILFF
VEASYKYNYNIIKKILKENNDIEENFAVFSPLFPFRDEEDIKNAVNLLKKNIDKIDYCISKDVKKDKKI
FDFNRAIFLGKIESYLKNRLSKPKFIVCPLSDESLIKINNKLDFELAIAIYNKRNSHKIIEQKIKKRIEEK
ERLFLEVTDITLIGHSIFDNWDIVEFKGKSVRNLGIGGISTEQYQKFIFDKNKISGKSSIYFVIAGTNDIV
NKNLSYFKISQQINLLIESIYEVSPNAKIFFIETPSVAFRIDRKKEEIFLLNEIIKKNLKKEVTYIPINTFLI
DDFGNLKLEYTYDGLHFSEKGYQKLKEILEKEIEI

```
>B__30 tr|B6WXZ3|B6WXZ3_9DELT Putative uncharacterized protein OS=Desulfovibrio piger ATCC
29098 GN=DESPIG_02973 PE=4 SV=1
MNIIAIIPARMGSSRFPGKPLALIHGVPMVGHVAFRTAMSRCLSATYVATCDTVIEDYCKEASLACVM
TGDHHVRCSTRTAEALLKIEAATGRKADIVVMVQGDEPMVLPGMIDAAVEPMLKDPSINVVNLMA
DMDTLEEFEDPNEVKVVVDRNNDALYFSREPIPSRKKGADKVPMRKQVCIIPFRRDYLIRFNEMEES
PLEICESVDMMRILEHGEKVRMVPTDARTWSVDTPEDLARVTRLMEGDVLMREYAK
>B__31 tr|Q8KIQ0|Q8KIQ0_PSEAE Similar to Cytidylyltransferase OS=Pseudomonas aeruginosa PE=4
SV=1
MIAWSIEAAIESGCFDKVMVSTDDAEIAAVARKYGADVPFMRPVELADDHAGTLPVIRQAIEGYLEK
GVFAEQVCCIYATAPFVRPEDLYQGCTRLEESGAAYAFSVTTFAFPIQRAIRLKEDGRVEMFQPAYMA
SRSQDLEEAYHDAGQFYWGRSEAWLKEIPIFAGNAVPILLPRHRVQDIDTPEDWVRAEWLFKALLSE
ADVSSS
>B__32 tr|Q8KIU4|Q8KIU4_PSEAE Similar to Cytidylyltransferase OS=Pseudomonas aeruginosa PE=4
SV=1
MTTAYSFIFARGNSKGLPGKNIKILGTKPLLAHSIDVARSVREISKIFVSTDCKAIADVAQQYGAEVIW
RPEELATDNSPEWLAWQHAIKTLRDRGDQFDLFVSLPATSPLRSTEDVQNCIYGIDENVDIVVTATPA
ARSPYFNMLVRDERGISTTVCSGDTIYRRQDAPMVYDMTTVAYVTRPEYILGNERLFAGVVRSIVVP
RERAIDIDDIYDFKMAEMLIMEKESNIC
>B__33 tr|A4NAR0|A4NAR0_HAEIN Acylneuraminate cytidylyltransferase OS=Haemophilus influenzae
3655 GN=CGSHi3655_03511 PE=4 SV=1
MTRIAIIPARAGSKGIKDKNLQLVGGVSLVGRTILAAQESGMFDQIVVTSDGENILKEATKYGAKPVA
RPESLAQSDTRTIDAILHCLETLNISQGTAALLQPTSPLRNALDIRNAMEIFLGGKYKSVVSACECEHH
PYKSFTLEGTEVQPIHELTDFESPRQKLPKSYRANGAIYINDIASLFEEKRFFIAPMRFYLMPTYRSIDI
DSTLDLQLAESLISKEF
>B__34 tr|B1BCQ0|B1BCQ0_CLOBO Spore coat polysaccharide biosynthesis protein SpsF
OS=Clostridium botulinum C str. Eklund GN=CBC_A0401 PE=4 SV=1
MNLLIIQAHMGSTRLPGKIMKKIRNKEVLLHVYERCKRAKVVDKIVIATSKNKENNQIEEFCKINNIE
CFRGSENDVLDRYYKCAKVYSPDIIIRVTSDCPLLEPKLIDFWVNNIIKDSVDFIEEEKELFTGFGLDIF
TNNALIKMKRNATLEKQKEHVVGYYYDNKDEFKHKIYPLAEDLKYLYRNYRLTLDTKEDFELIKLL
YEKFYNDNYINLKEVINYIDNNKKILNINENIEQKEY
>B__35 tr|B5H541|B5H541_STRPR Transferase OS=Streptomyces pristinaespiralis ATCC 25486
GN=SSDG_00267 PE=4 SV=2
MRQEPPMTVLAVIPARGGSKGVPAKNLAAVGGVPLVARAVRACLDAPLVTHVAVSTDDPGIAAVARG
AGAEVVLRPAAIAGDTATSESAVLHAMDAHEAVHGAPVDVVLLVQCTSPFISREDIEGVARAVALDG
ADSAVTVAPFHGFVWREADEAAAPAGNSAEQVGGGSVLVDTTTATGSGYGVNHDSSFRPRRQDRP
QDFLETGAAYAMSAAGFREAGHRFFGRTALVHTDPARVLEVDDPHDLARARALAPLLDTAAVPRRE
DIDAVVLDFDGTQTDDRVLIDADGREIVSVHRGDGLGVAHLRKAGLELLILSTEQNPVVAARAHKLK
VPVLHGIDRKDLALEQWCEERGLAPERVLYVGNDVNDLPCFHLVGWPVAVASAHDSVRAAARAVT
ATPGGEGAIREIAAWLLGPTLNTPEPAPMSPKK
>B__36 tr|Q9AGI3|Q9AGI3_PSEAE Putative uncharacterized protein OS=Pseudomonas aeruginosa PE=4
SV=1
MKYAVVIPARHASTRLPGKPLLDLCGVPMIVRTYRQCIQAVDAEHVLVATDDERIRAVCEGEGIRTLM
TSSRCLTGTDRVAEVAGQVSAEIFINVQGDEPLFNPDDLRKLIDAAQASPEAIINGYCGIADETTFRNP
SVPKVVFRPDGRLLYMSRAAIPTTKQGEFSRAWRQVCAYAFPREALRAFAARPTKTSLEEVEDIEILR
FLELGWEVKMIEMSDQSISVDNLEDVERVLDAILQRKPRTASVGGDQHAP
>B__37 tr|A4C037|A4C037_9FLAO 3-deoxy-manno-octulosonate cytidylyltransferase OS=Polaribacter
irgensii 23-P GN=kdsB PE=3 SV=1
MKIIAMIPARYSASRFPGKLMKDLGGKSVILRTYEAAIKTGLFEAVYVVTDSEIIQENIANAGGDVIMS
TAEHQCGSDRIAEAVAFLEVDIVINVQGDEPFIDKVSLSKLIDVFKNDSEKEIDLASLKVQITAKEEIEN
PNNVKVITDINNMAIYFSRSVIPYQRDQDIAVKYYKHKGVYAFRKQALLDFYKTPMTPLEAAEKIEA
IRYQELGKKIKMVETTIESVGIDTPEDLEKAKRYLNLL
>B__38 tr|A8V8Z5|A8V8Z5_NEIME SiaB OS=Neisseria meningitidis GN=siaB PE=4 SV=1
```

```
MEKQNIAVILARQNSKGLPLKNLRKMNGISLLGHTINAAISSKCFDRIIVSTDGELIAEEAKNFGVEV
VLRPAELASDTASSISGVIHALETIGSNSGTVTLLQPTSPLRTGAHIREAFSLFDEKIKGSVVSACPMEH
HPLKTLLQINNGEYAPMRHLSDLEQPRQQLPQAFRPNGAIYINDTASLIANNCFFIAPTKLYIMSHQDS
IDIDTELDLQQAENILNHKES
>B__39 tr|B7R841|B7R841_9THEO Cytidylyltransferase, putative OS=Carboxydibrachium pacificum DSM
12653 GN=CDSM653_494 PE=4 SV=1
MYRGKTILALIPARGGSKGVPRKNIRLLLDKPLIAYTIETALEVDFLDKIIVSTEDLEIAKISMEYGAEV
PFLRPYELATDEAKSIDVVLHAMDWMEKHDETFDLILLLQPTSPFRNSEDIKTAFDIFFKKNAKAVVS
VCEAEHSPLWMNTLNDDLNMKDFIRKDILNKNRQELVKYYRINGAIYIAEWNYLKQNRTFFGNETY
AYIMPKERSIDIDTEMDFKFAEFLIKLQKAN
>B__40 tr|Q8F885|Q8F885_LEPIN CMP-KDO synthetase-related protein OS=Leptospira interrogans
GN=spsF PE=4 SV=2
MNGTRSTHKVYAFIQARTGSSRLPEKVLLEFPSGSGKTLVDRIYDRILTVLPKEQIVYLIPKEDKELRY
FLNQRNYLFFAGDLLDVRRRYIEAAHFFNADSILRLTGDNPFYDTIHLDQLLQSFQFFESDLSYVSGL
PLGMGGEIFTRKALEWTPNTLEERHKEHVSLHIKENSDRFRITKLSSLLSEKEKLILPKLRLTIDEPKD
FETTSNIFNILNEQNPIFGARECINLFEIDPNVFAGNQNVEQIRFQTLPSQRTKKFRIGVFAGDPKDFGS
GHFERSRILFALLATVPYETFWLREFPKEEDLDLLIVDSRDISIPEYSKTKVLLLDHFGSDRKKFHHYD
LLPHSDIEDRFSLDQILIPPGLFNLDRKQIDNSYILCYAGNIDHHTTFSLDSFLESLSLRENIPKIVRIGGT
STTFKQIEFIPRVSKFKFQNLLATCSGFVSYFGQSLFEAIFLDKKVCTYSISPVHSSLSLLCEKKYGIPFA
GDLNSTSLENEVKLQFSSQPVSGKGYPKLLQEIERILSEFDLSD
>B__41 tr|B9P351|B9P351_PROMA Acylneuraminate cytidylyltransferase OS=Prochlorococcus marinus
str. MIT 9202 GN=P9202_1704 PE=4 SV=1
MKTVAFIPARGGSKGLPRKNIKIFNKKPLIYWPINAALNSKYIDEVVVTTDDVEIADIAKNFGAYIPFI
RPKDLAGDLSTTEETLQHALLKYEEIKGKIDIAIFMTCTDVFRKPEWIDDAIKKLMNNPNLESVFSGH
KTHKNFWEKDNKSNWIRIKDSMKIYSSRQIREFIVREDTGIICASRAYLWRNGRRIGDKVDIILNDDS
FTGLDIHSEEDFLIAEFAHKIRNKINLK
>B__42 tr|Q5ZXI0|Q5ZXI0_LEGPH CMP-N-acetylneuraminic acid synthetase OS=Legionella
pneumophila subsp. pneumophila (strain Philadelphia 1 / ATCC 33152 / DSM 7513) GN=neuA PE=4 SV=1
MRILAVIPARAGSKRLPGKNTRLLAGKPLIAHTIVAALQSSCCEEIVVSTDSKQIADVAVQYGASVPW
LRSEDLATDTSDVIHTVIDLLFKFQQMDVFFDSVLLLQPTSPFRKPETIRHAVEIHKVTGKSVVSVSPIS
LKPSWCRSIDSQGNLVKPELFQDLEIYCNENPIYKLNGSIYIATAKQIIENKSFYSEPTKPLLLNSISESI
DIDTPIDWALTEKLMELNQEALV
>B__43 tr|Q8F5R2|Q8F5R2_LEPIN CMP-N-acetylneuraminic acid synthetase OS=Leptospira interrogans
GN=neuA PE=4 SV=1
MTDQKILGLIPARGGSKGIPRKNIKLIAGKPLIVWTIEAALKSKHLTSIVVSTDDPEIAEIAEQSGASVP
FLRPAELATDYSSGIDPVLHALDNLPGFDYVMLLQPTSPLRTSADIDDCIEFTMRKKTNSVVSVCEAQ
ENPFWMFRLDDDSFKMTKLLNGEDVARRQDSPKVFTLNGSIYLSEVNYFREKKKFITDDTLAYLMS
KENSIDIDDMIDWKLAEILLLDKV
>B__44 tr|Q8F5P2|Q8F5P2_LEPIN 3-deoxy-manno-octulosonate cytidylyltransferase OS=Leptospira
interrogans GN=kdsB PE=4 SV=1
MKTIAVLPARMASSRFPDKPLVKISGLEMIEHVRRRVEMSSSVDEVVVATCDEIIKQRVESFGGKAVM
TSDVHRGCIDRVAEAALYVEGDIVIVVQGDEPLILPAMLDDLVKPMLNDSSIYCTNLVTKIVDEEEFQ
SPNAPKVVVDKNWDLLYASREPIPSRKKYPNEDYLKFKQLGVIAFRNDFLQTFAALAPTPLEIIESVD
MNRAVEHGYKVRMVLTEGIMIGVDVPGDVSRVESVFKTDLLLSKYLS
>B__45 tr|Q2WB83|Q2WB83_MAGSA CMP-N-acetylneuraminic acid synthetase OS=Magnetospirillum
magneticum (strain AMB-1 / ATCC 700264) GN=amb0088 PE=4 SV=1
MSNRILALICARGGSKGLPGKNVRPLAGRPVIAWSVEAALGSSLIDRVVVSTDDPAIAEVARAAGAE
VPFLRPAELASDTASLYDVIFHALEALDEEPSHVVLLQATSPLRIAADIDGCIRLCLDHGAPAAASLCE
PGKSPYWMFLLDPDGTVRPVIPHDASGGRRQDLPVAWAPNGAVYVAETAWLRRERNFWKAGVTLG
YVMPLERSVDIDSLLDFRVAEMLMSDRLGPKAG
>B__46 tr|C4PM35|C4PM35_CHLTZ 3-deoxy-manno-octulosonate cytidylyltransferase OS=Chlamydia
trachomatis serovar B (strain TZ1A828/OT) GN=kdsB PE=3 SV=1
```

MFAFLTSKKVGILPSRWGSSRFPGKPLAKILGKTLVQRSYENALSSQSLDCVVVATDDQRIFDHVVEF
GGLCVMTSTSCANGTERVEEVVSRHFPQAEIVVNIQGDEPCLSPTVIDGLVSTLENNPAADMVTSVT
ETTDPEAILTDHKVKCVFDKNGKALYFSRSAIPHNFKHPTPIYLHIGVYAFRKAFLSEYVKIPPSSLSL
AEDLEQLRVLEIGRSIYVHVVQNATGPSVDYPEDITKVEQYLLCLSKASF
>B__47 tr|Q8F5Q2|Q8F5Q2_LEPIN CMP-N-acetylneuraminic acid synthetase OS=Leptospira interrogans GN=neuA PE=4 SV=1
MNLAVIPARGGSKGLPGKNIKNLCGKPLIAWSIEAAKRAKKIDRFIVSTDSEEIAAVAREWGCPVLKR
PDELATDETKTISVLSQISKEIPEALNFILLQPTSPIRDVGLIDECLDLYEQGNYSNLATGYWCKYQEF
GKHNNMRRQDYKGFFYDDGNVYILSKKLVSEGLWFGDHICRHVISKRQNFEIDDEIDFVILEALLNK
YGTIN
>B__48 tr|Q2WB38|Q2WB38_MAGSA CMP-2-keto-3-deoxyoctulosonic acid synthetase OS=Magnetospirillum magneticum (strain AMB-1 / ATCC 700264) GN=amb0133 PE=4 SV=1
MTRIIGLIPARMAASRFPGKPLFPILGRPMIEHVFERAKLFGRWDALAICTCDEEIRAFAQSKGYPVIM
TSDKHTRALDRVAEAATKCGVDVADSDIVLNVQGDEPMMHPDMIAATIKPMEERSEVRGTMLAMD
IVDEAQFRNPDALKIIHDLSGRVLYTSRQPIPHCKTFGPELEAKRIYGIFGFKWDFLKLFTELPPSPLEI
KEACDSNRLYDYGHHQHIAPYPFRPSFSVDSPHDIGIVEAAMKDDPLWGTY
>B__49 tr|Q5LD42|Q5LD42_BACFN Putative lipopolysaccharide biosynthesis OS=Bacteroides fragilis (strain ATCC 25285 / NCTC 9343) GN=BF2274 PE=4 SV=1
MKKNKILIIIPSRFASTRLPEKPLVKIAGKEMVLRVAEIANYVCNKVEGCNYIVATDHEKIVNFCKENN
IAVMMTSENCKSGTERCWDVTTKIAEKPDFIVNLQGDNPLCPPWFIEQLIEAWKNDKEGQVFTPSLH
LSWEEYDRMKESKKITPYSGTTVEVDKFGYALAFSKAMIPVIRNEEKVRKILDKSPVRRHIGLYSYT
YDALKKYFEVEASPYELPEGLEQMRFLHNRIPVKMIDVDYRNRKSMSGVDSPEDIERAEKIIAEFGE
FNLSPE
>B__50 tr|Q3AHE8|Q3AHE8_SYNSC Demethylmenaquinone methyltransferase-like OS=Synechococcus sp. (strain CC9605) GN=Syncc9605_2247 PE=4 SV=1
MKKYAIFLPAKGNSERVASKNTRLLDGKPLFLHTLEKLLQLGDDYEVYLDSECPEIFKMASHLEGFK
PLIRDPALATNATDGNKLFLNEVFSCNHQVIAQHLCTSPFIENNTIKNCFSKVGQMSQDGSCWDSSFL
VRKEKFYMWKNDLPLYSIDSIPNSADIESTTVETMGLYVCSREAALDTERRIGRNPCLIEAKPLEAVD
VNYEEDLEMAQLLAAGKREQERNLLRNLSSLLSSPIISDILDSLGLESQIIRGLNQKCRNNAKIFGPAK
TMHLRRKSKEDRFDEVYDALKHYETLVPGDVICVQNDCQEFAYFGELNANLAIRSGAIGAVIGGMT
RDQEAVCSLGLPVFAKGITCQDVKNRAVYSSMNNRICIDGVFISPGDMMFADNEGTICIPKKHFKTVL
ELAKESIRKEGEIISDIATGINVSALRAERGDF
>B__51 tr|Q3AHH6|Q3AHH6_SYNSC Putative acylneuraminate cytidylyltransferase OS=Synechococcus sp. (strain CC9605) GN=Syncc9605_2217 PE=4 SV=1
MAVPRGALALIPARGGSKGIPGKNLLTVGGVPLVCRSIRAALASNGVGRVVVSTDDEAIAAAAETEG
AEVIRRPAEIAGDTASSESALLHALDVLEKQGPLETELVFLQCTSPFTTGAQIDAVLAALHGRGSNSSF
SVTPWHGFLWRADGRGINHDPALLRQRRQDLEPAFLETGAIYAMAITAFRRCGSRFCPPTSPVVLQE
VGPEIDTPEDLALCRSIAAQKGE
>B__52 tr|A7NHE7|A7NHE7_ROSCS Acylneuraminate cytidylyltransferase OS=Roseiflexus castenholzii (strain DSM 13941 / HLO8) GN=Rcas_0776 PE=4 SV=1
MLKRVVNRTARSAILHDVIVATTIEQEDETIVELCKKHGWPCFCGSINDVLDRYYKAARDYRVDVIV
RITSDCPLIDPCIIDLVISKFLEKNPLDYASNTVQRSFPRGLDVEVVSFEALERAWHEDCNPLWREHVT
PYIYLHPEKFALTAVVNNKDYSYMRWTVDTPEDLDFVRRVYDHFGHDNFSWKEVLAVLDEHQDWL
AINSHIQQKVI
>B__53 tr|Q2W0L5|Q2W0L5_MAGSA Spore coat polysaccharide biosynthesis protein F OS=Magnetospirillum magneticum (strain AMB-1 / ATCC 700264) GN=amb3806 PE=4 SV=1
MKIIATIEARMTSSRLPGKVLLPAQGRPMLARMVERLKMVPSLDGIVVATTVNATDDPIEALAAELGI
GCWRGSEDDVLKRVLDAAHAFEADVIVELTGDCPLIDPAIVETCVQAYRAAGVDYLSNVQERSFPIG
MDTQVFATAILDDVARRTDDPTDHEHVSLYIYRHPELYSLKNVAAPPELRDPELRLTLDTPQDYQLID
QVFAALLPQGPGFALGDILALLQARPELRKINDHVAHRWV
>B__54 tr|Q2WB80|Q2WB80_MAGSA CMP-N-acetylneuraminic acid synthetase OS=Magnetospirillum magneticum (strain AMB-1 / ATCC 700264) GN=amb0091 PE=4 SV=1

```
MSRAVAIIPARGGSKGILRKNLREVGGAPLVVHTIRQALESSSVDRVYVSTDDTEIRDVSQAAGAHVI
HRPVDISGDTISSEAVIHHALPVMRQEGDFDLVVFLQATAPLRAKGDIDGAVATLVREEADSLLSVSRS
HRFLWRQGTQGPEAINYDPAHRPRRQDLEAQFVENGSIYVFKPWVLERYDNRLGGRIALYEMTASA
AFDIDDEIDLAVIDLLMTRRFG
>B__55 tr|O25093|O25093_HELPY CMP-N-acetylneuraminic acid synthetase (NeuA) OS=Helicobacter
pylori GN=HP_0326 PE=4 SV=1
MRAIAIVLARSSSKRIKNKNIIDFFNKPMLAYPIEVALNSKLFEKVFISSDSMEYVNLAKNYGASFLNL
RPKILADDRATTLEVMAYHMEELELKDEDIACCLYGASALLQEKHLKNAFETLNKNQNTDYVFTCS
PFSASPYRSFSLENGVQMAFKEHSNTRTQDLKTLYHDAGLLYMGKAQAFKEMRPIFSQNSIALELSP
LEVQDIAHFRRFRISQAQIQPFEKRMPVKILCDCFLTSGLGHVRRCEKILSFIEKLGVEASLYLHKQNN
ISAFLEGVGGNDFLITDSYCLNSKDFYLLKEKAKSLMVIEDTEHAKGFYPKNTKILNFTLNALKHYH
HLSKDYQYYLGVGFYPVDARFIYDRPINTENKEVLITLGGSEQKTLKEIVKILENKNVNLHIISPYTPK
NPPKNTHYYSPLNPLEFSSLMKSCACAISAAGQTLYELALSQTPSLILPIASNQIIQSKEFESLGIFKQTS
LKTLAKDFENLQIQKNQAWAKNLVFGDKLEGALREFLEI
>B__56 tr|Q3AN76|Q3AN76_SYNSC 3-deoxy-D-manno-octulosonate cytidylyltransferase
OS=Synechococcus sp. (strain CC9605) GN=Syncc9605_0180 PE=4 SV=1
MAIRKCVVAVPARLQSSRLPNKVLADIGGKPMIQRVLERCSEAQGVEAVVLCTDSTELQTLAEGWGF
PVLMTSESCNSGSERIASVAHPLMALGWGDADPVAEETAVINVQGDQPFIEPAVIDAMAEEFRSQDPV
PAVVTPVYGLKPESVHNPNVVKTLLAHDGRALYFSRSAIPHVRDIAEADWHQHTTYWGHVGMYGF
RGDVLAAWDQLPASPLEDLERLEQLRLIEAGLTIATFRVEGTSLSVDTAEQLEQARAMV
>B__57 tr|Q3ARI1|Q3ARI1_CHLCH Acylneuraminate cytidylyltransferase OS=Chlorobium
chlorochromatii (strain CaD3) GN=Cag_1132 PE=4 SV=1
MQTVAIIPARGGSKGLKYKNIYPVAGKPLLAWTIEQARASQFVDKVFVSTDSEDIADIAKEYGAEVIE
RPADIAGDKATSESAILHALNVIQAEHHITVSAVVFLQATSPLRKQGDIDGAIELFRRENADSLISVTK
ADDLTIWEQRKSGEWASVNFDYRNRGMRQDRPAQFIENGSIYMFTPETLHRFNNRIGEKLVAYEMEF
WQTWEIDTLNEIELVEFYMKRKGLM
>B__58 tr|Q3ANB5|Q3ANB5_SYNSC 3-deoxy-manno-octulosonate cytidylyltransferase
OS=Synechococcus sp. (strain CC9605) GN=Syncc9605_0141 PE=4 SV=1
MAIRKCVVAVPARLQSSRLPNKVLADIGGKPMIQRVLERCSEAQGVEAVVLCTDSTELQTLAEGWGF
PVLMTSESCNSGSERIASVAHPLMALGWGDADPVAEETAVINVQGDQPFIEPAVIDAMAEEFRSQDPV
PAVVTPVYGLKPESVHNPNVVKTLLAHDGRALYFSRSAIPHVRDIAEADWHQHTTYWGHVGMYGF
RGDVLAAWDQLPASPLPPSGSRAPRCRWTPPNSWSRPGR
>B__59 tr|Q1QP14|Q1QP14_NITHX Acylneuraminate cytidylyltransferase OS=Nitrobacter hamburgensis
(strain X14 / DSM 10229) GN=Nham_1201 PE=4 SV=1
MPNEVVIITQARMTSTRLPGKILMRAGERTLLETHLDRLADVGVPVIVATTTNLSDGPVVELCAALD
VPVFRGSENDVLERYQRAARAHEARHVIRVTSDCPLICPDVICAGLDVYLGLKSESIYLSNGERRTYP
RGMDFEIFSRRSLDFAAEHARLPSEREHVTGCIRTSMPDIRHEHCTDEEDLSDWRITVDTVDDFKLA
KLLIEDHHAAEMDYPGLKALLKAHPELMKINSHIQQKPI
>B__60 tr|Q2W9F3|Q2W9F3_MAGSA Spore coat polysaccharide biosynthesis protein F
OS=Magnetospirillum magneticum (strain AMB-1 / ATCC 700264) GN=amb0718 PE=4 SV=1
MTAVVIVQARMGSTRLPGKILKGLGGMSALAQCLRRCKAIPGIDRVVCAIPEGEAEAPAVAEAQRCG
ALVVRGPSDDVLKRYALAARAAGADIVMRVTSDCPLIDPDLCGRLLAKLRAEGLDYCCNNLPPSWP
HGLDAEVFRATALFEAEDKATEPFDREHVTPWLRRAPHLKRGNIAREEGDLSHQCRWTLDYPEDYA
FLAALFERLPEAIVSMEAVLAVLAANPALADINAMRRGVRAMPVPAKGASVSDRR
>B__61 tr|C4PPI9|C4PPI9_CHLTJ 3-deoxy-manno-octulosonate cytidylyltransferase OS=Chlamydia
trachomatis serovar B (strain Jali20/OT) GN=kdsB PE=3 SV=1
MFAFLTSKKVGILPSRWGSSRFPGKPLAKILGKTLVQRSYENALSSQSLDCVVVATDDQRIFDHVVEF
GGLCVMTSTSCANGTERVEEVVSRHFPQAEIVVNIQGDEPCLSPTVIDGLVSTLENNPAADMVTSVT
ETTDPEAILTDHKVKCVFDKNGKALYFSRSAIPHNFKHPTPIYLHIGVYAFRKAFLSEYVKIPPSSLSL
AEDLEQLRVLEIGRSIYVHVVQNATGPSIDYPEDITKVEQYLLCLSKASF
>EU_Acyrthosiphon tr|C4WWF5|C4WWF5_ACYPI ACYPI002677 protein OS=Acyrthosiphon pisum
GN=ACYPI002677 PE=2 SV=1
```

```
MTVRNSHWVALFSQLFVLSASVVLFMIPIYLLWSHITYEQNHLSNYKPHVVALILARGGSKGIPRKN
LAVIQNTTLLRRSLDTINDCGLFHDVWVSTDDKEIAAEAELAGAKVFHRSAETASDEATSLSAVKEFS
MYHQKVDIFAIIQCTSPFMTAEYLTRAYKMAVNQKFESVFSVTRTHKLRWKLDSDGQLQPSNFDVT
KRPRRQDWDGEYVENGMFYFTYKNLIVNNKLQGGKLGVVVIPLNRSLEIDTQFDLQAARLLAPLLD
KPKNVTNRTNEIIKKS
>EU_Aedes tr|Q16KU9|Q16KU9_AEDAE Cmp-n-acetylneuraminic acid synthase OS=Aedes aegypti
GN=AAEL012868 PE=4 SV=1
MRLVLILTLLSAHPAFGFLQEKSVTCPTSPESISNDSVIALILARGGSRGIPLKNLAKLDDRQTLLSRAL
HTALSTDGFHSVWVSTEDDRIAQAVERDFPHDLVRVHLRPPEVAQDHTSSIESVREFLDHHPRVQNV
ALVQCTSPFLGVRYLDEALQRFQSLRVDCVFSVTRSFKLRWRKEKDGRVNALNFDPRKRPRRQDW
DGELVEAGMFYFARRKLLLEGRFQNNNCEVVVIDERDSLEIDSLYDLELARKIIGSTVE
>EU_Anopheles tr|Q7Q9Y7|Q7Q9Y7_ANOGA AGAP004527-PA (Fragment) OS=Anopheles gambiae
GN=AGAP004527 PE=4 SV=4
TIHQSGVFDSIWVSTDHEQIAYEADRHGAAVFARSAMHAQDHSTSLDAVREFLDAHPEIHNFALIQCT
SPFLRAHYLVTAAHQFHSRMNAGMENAPACTFSVVRSYKLRWKEDVETGHLKAINFDPFARPRRQD
WPGELVETGMFYFADRRLVMEMGSFQSETCTVVEVDEQDVLEIDNWRHLTLARILLGME
>EU_Aq [rspp] Aqu1.200879
MIALVCELALAGRVDAVIVAGDDERILEAARAEGAQAILTSSEHRSGSERIAEAISLLGLDPDVVVVNI
QADEPLLPPTLIDQVAGLLVDRPEFAMATLCEPIEKARDIFDPAVVKVVFDAQGRALYFSRAPIPWQR
EAWSSTASLRPAADEDEDALDRLPATAGVRHYRHIGLYAARAX
>EU_Arabidopsis tr|Q9C920|Q9C920_ARATH AT1G53000 protein OS=Arabidopsis thaliana
GN=AT1G53000 PE=2 SV=1
MSVCSSSSSSQKTWIVNGILAGTAIAAAIGARAYLGRSKKFRSRVVGIIPARYASSRFEGKPLVQILGK
PMIQRTWERSKLATTLDHIVVATDDERIAECCRGFGADVIMTSESCRNGTERCNEALEKLEKKYDVV
VNIQGDEPLIEPEIIDGVVKALQVTPDAVFSTAVTSLKPEDGLDPNRVKCVVDNRGYAIYFSRGLIPYN
KSGKVNPDFPYMLHLGIQSFDSKFLKVYSELQPTPLQQEEDLEQLKVLENGYKMKVIKVDHEAHGV
DTPDDVEKIESLMRERNMS
>EU_Branchio tr|C3Y778|C3Y778_BRAFL Putative uncharacterized protein OS=Branchiostoma floridae
GN=BRAFLDRAFT_123265 PE=4 SV=1
MLIPDDVYKVLNLMGFLSWLGLAVTVSGMVWWRYKKPDLPRPIKLNLAIPVLFVLLSLFMVGVSF
VSAPVECASGLAILASALPVYAVFVHWENKPTWFLSFEGYYLSQRLTPYERMMNEERGLGMADPRA
EIHEKVQQAGIPLIQENDYQIVENGLVGTGTFGKVYKGKYDDDEVAVKEIQMGRKVKLTDHQIREIQ
IHGLLPGHKNVASLLAVSIGKNLRMICNFIDGRSLDEIIFEEDEIPKLLKDALPFPNLSQCIYFEPVQIPA
HEEPYVALQIAEGVAHLHDCGIIHRDLKPENVMVDTKRHAYICDLGTAKVEQHNNLSVIESMIKPGS
PIFLAPECSVDGLDADYSSDIWSLGCTLAELFAREEIWETEDIEFELKRNLKDRKEPDGLKDVREPFN
SLVRKLLSYDPKNRITAFDAVRELQKTFDWKKYEQYKDLTSQNILENRRDALPFPNLSQCIYFEPVQI
PAHEEPYVALQIAEGVAHLHDCGIIHRDLKPENVMVDTKRHAYICDLGTAKVEQHNNLSVIESMIKP
GSPIFLAPECSVDGLDADYSSDIWSLGCTLAVTVVELHNNNNNKKLHSDDINGFAGHTIGYADWSSS
PTFSPSVHKRCTMNDSLSTINLEHLGNSASNVIGTSHVTLPTIPQAPSGTGIPVSIANIPVTLSSGLTTTI
PMLFPLERPMVNNLTSVQNTPVGLDKSLTLDITSLRQMSSDLLTTSLTLIAATTTLTTSTAALPPINTHV
VNQQNILQGGNTSTVNVTTVPMSSISVFPNSNNLTVTTCSVLESLLQPTIQNHVPVGSGERSLEAVAR
QLEEELNEPIKTDFWVFTRTAIQARTQFGPLQGKIVKKSDVESESDWTNMWEIFRDERSSHFIDARDE
NEANWMMFVKRARTSLEQNLVAHQCGGDIFFTSCKDIAEGDELLMWFAGNYAKLTGVTSKPEQSY
KCCSCERQFADLGALGRHTKYAHPDMTGRKWKCDLCERAFTSSSKLQVHIMVHTKVKPHKCNYCE
KTFTDPSNLRTHLTIHTGVKKHACSVCNKTFRQKAHLLSHMVTHTGEKKMKCQFCDKMFSRQSDV
KQHMYMHTRDREVKCEECGKIFWRLQHLKKHMKSHTGERNFPCDRCNKAFFTKYHLNRHKKAC
KGRNAQQNTDVVRVVISEMAPNGGPTQNVDVTRGKTSKKQQRRQNTKTRASALQFPTDLLGGAE
QHGVQMVDSNDLICNGNNNGGIVMPVPIVPNIVNTTGGLMPQVIIQGQQPFGNVKMTNNGKSAPEL
RYKRMGFEPTPELRAEWKRRFGKEEPHIATLILARGGSKGIPMKNNKKLGGLELIGWVIRAAIDSDI
MDSVWVSTDHDDIAATARKCGAQVHRRSAEVSQDTCNSWVTIDEFAKYHPEVDIIVDIQATSPCLHP
FHLHDGMRLTLKDGYDSVFTVVRRHGFRWTDPKWGIQPFPLNFDPAKRPRRQEWNGEIVENGAFY
MMTRDVEQQGLMQGGKMAYLVMEPEYGVDIDTDLDWEIAEQRLFRFGYHGRTPKTVKLVVVNID
GTLLDGQVQVSPTGEELRSFKTTDIVGVQQLRENGIEFRVVAEGDNEVQRALAGKLSAKLEENCTD
KLTVIDAWRQELGLEWKDIGYMGYDISDMSCIKKAGMSACPSDAYAAIQKHSHFISTYRGGQGALR
EFCEQVLEKGNMAAYVDVAEQTWDDFQRKRSEGMAKKSRTRSLSQSQSVGSDVPSI
```

>EU_Branchio2 tr|C3ZUW2|C3ZUW2_BRAFL Putative uncharacterized protein OS=Branchiostoma floridae GN=BRAFLDRAFT_115660 PE=4 SV=1
MDGDCATTEDLCMQRFGKKKPHRACLVLARGGSKGIPLKNIKPLAGTPLVGWVLRAAIDSEAFDSV
WVSTDHDEIAAVSREFGAQVHRRSPEVSRDASTSLETIQEFIRAHPEVDIFGNVQATAPCVHPFHLRK
AMKAMTEDGFDSLFAVVRRHAFRWKEVKEGEVTAPLNLNPEKRPRRQDWDGELIENGSFYFCTTA
LVQKGLLQGGKVGYLEMAAEYSVDIDIDIDWPIAEQRVLRFGYFGKEKRQAPCLVVVSAEGALTEA
QVHLSPTGEEFRSFNVHDINGMKQLANRGVEVRIISEDESEVHRQLAAKIGCKLEEGCEDKLAIVDS
WRKDLGLDWLQVAYIGSDASDSECIKKAGVNGCPLDAQYPAKNYSRFVSKRRGGQGAIREFCEHV
EITIEKANNENKPRPE
>EU_Branchio3 tr|C3Y777|C3Y777_BRAFL Putative uncharacterized protein OS=Branchiostoma floridae GN=BRAFLDRAFT_123266 PE=4 SV=1
MPMVQCMLDPCDVATCPGNPRARCEHNFCGGCNAVFYDKTGAMADCNGDGCPDDKPMARCMM
DPCDVTSCSANPDATCVADYCGGCNANFFDRSGEPAWGSKGIPMKNIKMLGDLELIGWGIRASLDS
DIMDSVWVSTDHDDIAAVASRCGAQVHRRSSQVSHDKCNSWTAIDEFVRNHMDVDIIVHIQATCPC
VHPFHIQEGVRMMLQDGKDSVFSVYRRHQLRWSTPRYGMQAYPLNFDVYDRPRRQEWPGELIENG
AFYMFTRDVEQQGVLQGGRMGCLVMDSEYSIDIDTDQDWQLAEQRLLRFGYHGRAPKKMKLVVV
NIDGTLLDGKVQVSSTGEEYRSFKATDIAGIQMLQRNGIEFRVVAQGESQVQKELAKKLSLKLEENC
EKSGMSGCPSDADTEVQKHSHVVSIYRGGQGALREFCEQLLLKCGVEPTLEARNETGNFKVSSRVI
QKDQVSSQPEMPS
>EU_Branchio4 tr|C3Y2M7|C3Y2M7_BRAFL Putative uncharacterized protein (Fragment) OS=Branchiostoma floridae GN=BRAFLDRAFT_124744 PE=4 SV=1
MDKIINITPQLESIWERRFKTKTPHKCALVLARGGSKGIPMKNIKMLGGTPLIGWVLRAAVYAEVFDS
IWVSTDNERIANVARDYGAQVHYRSEEVSKDTTSSWETVGEFLQHHPEVDIIANIQATSPCVHPFHIQ
EGMRMLLEDSYDSVFTISKKHGFIWKEPPQKDMQPININLDLMNRPRRQDYDPEPVENGAYYMYTR
DIHAQGKTQGGRIAWLEMGGEYSVDIYTDLDWDIAEERLLRFGYHGKHPKIVKLVVIAAEGVLLDN
HVEINVRGEELMSYHLSDVNAIKHMQNNLVEVRIIAEGENVVQRRLAASLGCKIHENTKNKLQVLD
QWREELGLTWRNVGYMGSENTDTQSVKAAGMSACPADAHDDVQLVSHFISKYKGGQGGLREFSE
QVLKVNKMSTYKDRVDKTTVYANVQRNLRRQILSMAPTERIPHIFTLILLQVVISNQQAVVLPTVRIP
TAISSSFTGRGLDDQAVARAANTINWRSKFYRFLQGRLNGDELNEDDLVAVEARVLAALKQPKLNAI
SAPALRHLLTSADIARDVPSVVYILENIDLVKFFDILQIAASTTAVGQSRQPDRSVIVPSRPVNGLPISTP
LRDAILNIGMKHMQGDLDEAWQDSLPKLSPYLPGITVGHVNRVKESDMVADKLLELLQNIRPRLSN
AVRANVVRYFKRVKAGNWSTEENVINSLGVFVGDISLNDVDNITIDTWKLVFNSHRLEDIDTTLARG
ILDKLLYANVTVTEPIYNVQKLGKLTCVLPEALGYRANSLSQDVWRELARVLHKCNTGVTGRPGNQ
LSRPGRKLAQRIAKSLIPAPASASVTDLRNVGPAIVGLSVSDLRQLQGRRVFDAISELSQAQGLSRAQ
AQALLQKYVDGKGIVGNSPDSVVNDLGPLISDLSLGSVKNVFPSQQDIPTLATLPWSQAQAANLIKR
IKESRQLTPALVGTFGTLLRGIRCPDMDDVALQDIVEVVVTLYDYVDQLPRGVRVCLANATQVAMNI
TLKSGRVTGSTGAVATVPGHVIAEMRYEILASIPRPDCSDVFRKIGKLDMSHFPLWYQRNITALALTC
LGKQKGQQYRLDATDVDVLGRFLCHIDDFFKITEDAVLDAINILKECQCFTYRQREAVKDLIKTAFG
NNARWWTSSELGQAGPLLHLLDGDIQRISQDALLDAVDQVLPVPQDNRDIINANRSTDRDCLAVKR
NNNLYRRVITAVARGRTSPIPPRNRRQTLQAPLLCDQVQLLGDAVVQLTEQQVTYMENREFLDCLEV
FGAARANNTDWDSYMLRILRIKTFAAAGGTVQQLTPDIMARVGRIAVTFSPGDLAALNY
>EU_Candida tr|C5M2S3|C5M2S3_CANTT Putative uncharacterized protein OS=Candida tropicalis (strain ATCC MYA-3404 / T1) GN=CTRG_00362 PE=4 SV=1
MEFTAIVVCGKGKALSPFSQIRSTGIPKALLPIANKPMVHYVLDWCLKANFAKVIVLADEDENDNKL
IQASVTKFQEEYEKKENRPAQVNVIGYNCDNNGSILHYLYKIYNDEEQSNALNKNFIILPCDMITNLP
PQVIIEAYRNKQETDLGLLIAYKNQLDIEDKKGKIFPRNYTIYSDNDASGNNGPVLLDMYNTQDIEFH
QALTIRTQMCWRYPAATISNTLLNSGIFIGSCKEIFDIFQAQKERFTESYFAKRSLLKIVRDLSRRSWRH
ATAQENIGLFIVPQQATFFRCNNLPVLMEANRYFMKLQAMNKSASSSVKQDKQSGAHVGNDSLVGE
STELGEKTNVKKSVIGSNCKIGRKNKITGCLILDNVVIGDDVTLENCIIGHDVIIQAKSKLTNCNVEST
NEVVSGTQTKNTNLLCLTLEGLVEGEEVIDSSEEIDDDDSDEDESYDDEDESEFEDEYTGNEDGLFAY
>EU_Ciona tr|Q0E670|Q0E670_CIOIN Putative CMP-sialic acid synthetase OS=Ciona intestinalis GN=cmas PE=2 SV=1

MEHSNSAINHLAALILARGGSKGIPMKNIVNVGGLPLICWVLRAAVDSNAFDSIWVSTDSDEVAEVA
SSYEVVKIHRRSDEVSKDNTSSMESTQEFLNYHPEIDAIGLLQATTPCIQPSQLLSAAEMIKFGGFDSV
FSVVRRHFFRWKEVKQGKDGDVTHPLNFDPSHRPRRQDWAGELCENGGFYFAKTSVVRQGLFQGG
RTGYQEMPHEHSVDIDTPFDLVVADYVINKYGYKGK
>EU_Drosophila tr|B4LGF7|B4LGF7_DROVI GJ11503 OS=Drosophila virilis GN=GJ11503 PE=4 SV=1
MNAPIYWFLTFVLHSTNDVHALILARGGSKGITYKNLVQIDGLSILSRAITTISNSSCFKHIWVSTDDE
KIALEAKKYGAIVHIRPAKYALDGTSSIEAIQEFLEGHKAIDNFALFQCTSVFLKEKYIEEAYQKFKVH
DCVFSVKRSHNLRWKYVEEQILPDNFNLKARPRRQDWKGDMVEAGMFYFSTRKLAMKGLLQNKK
CDVVEIAAEDGLEIDSYQDLAIARCIINSNI
>EU_Ixodes tr|B7Q984|B7Q984_IXOSC Cmp-N-acetylneuraminic acid synthase, putative (Fragment) OS=Ixodes scapularis GN=IscW_ISCW021121 PE=4 SV=1
MRVAGLVLARGGSDAVKKKNARLVCGLPLLSWVLRPMRHCKTLDEIWVSTDEQEIEDIANAEGCSV
LRRSEDFAQAESPSILAVQEFVRQAPGVEVIALVQCTSPCLAPSYLDEAVSLVTSDKYDSVFSIARDYK
WRWTELSQGAHLSLKRFVLGPADGTTRPLNFDPAQRLCRQEWAGEIVESGHFYVTRAEFVREGLLQ
GGR
>EU_Micromonaspu tr|C1N3H9|C1N3H9_MICPS Predicted protein OS=Micromonas pusilla CCMP1545 GN=MICPUCDRAFT_52178 PE=4 SV=1
MGCAASKADEPARKVRVLGVIHARCDAKDSKAFPHGKPLVDVAGAPLVWRTFLKVRGAKCVSRCV
VATDDDRIARVVAALGGEVVRVSGKWGSHTAAAVWETCERLRFWSDGGGGGGGGGAKKNQKNL
TATVDGGGAESSPSSPFDVVVCVDADEVGIDAHLVERLADAVAGSDAVMGTLACDRPGPPSGGGGG
GGGGGDGDERRRRGVGDRRKVGARSYRVSFLPKLSRRSLSNDEDRSSPRGHLAEEAWPCREEREA
AACGRTVKREVVAESVPRVGTREELARETARARELPTPPKVRTAGVSARGGASRKNTV
>EU_Micromonassp tr|C1E3I1|C1E3I1_9CHLO Predicted protein OS=Micromonas sp. RCC299 GN=MICPUN_57690 PE=4 SV=1
MISARTAARCSRGRRYASRRRLPTGVARHPSRHRRAFSPTRLLVPFRAARLAPPRPERAYRVVVERGS
ERHATSSLRRSRPLTFDSRRSRTQPRVLGVIPARYNSRRFAGKPLALLGGKPLVMHVYRNACKAKCL
DACVVATDDDRIAKAVRDNGGEVIMTDAACETALERCLEVVRKLRRRRADGKKGTTWTAASGEFD
VVVCIDADEPFVQPHHIETVADLVINSDAVVGTLVRRGLNEEDVASADTVKVVTDDRDYALLFSREP
VPRRGSSAGGGGVVRSAYDPGVAYKVRVGIQACGTDFLEKYASLPGCGGEGSIEQSRVADKENTSPR
RTTKNGDSCAIPPYEPPDEPEQNRVVFAGYKVKVDVVESTFVPGVKTPEDLRHLNAGLGAGLVNDL
QLTFLVA
>EU_Mus sp|Q99KK2|NEUA_MOUSE N-acylneuraminate cytidylyltransferase OS=Mus musculus GN=Cmas PE=1 SV=2
MDALEKGAVTSGPAPRGRPSRGRPPKLQRSRGAGRGLEKPPHLAALVLARGGSKGIPLKNIKRLAGV
PLIGWVLRAALDAGVFQSVWVSTDHDEIENVAKQFGAQVHRRSSETSKDSSTSLDAIVEFLNYHNE
VDIVGNIQATSPCLHPTDLQKVAEMIREEGYDSVFSVVRRHQFRWSEIQKGVREVTEPLNLNPAKRP
RRQDWDGELYENGSFYFAKRHLIEMGYLQGGKMAYYEMRAEHSVDIDVDIDWPIAEQRVLRFGYF
GKEKLKEIKLLVCNIDGCLTNGHIYVSGDQKEIISYDVKDAIGISLLKKSGIEVRLISERACSKQTLSAL
KLDCKTEVSVSDKLATVDEWRKEMGLCWKEVAYLGNEVSDEECLKRVGLSAVPADACSGAQKAV
GYICKCSGGRGAIREFAEHIFLLIEKVNNSCQK
>EU_Oryza tr|Q688Y2|Q688Y2_ORYSJ Os05g0561500 protein OS=Oryza sativa subsp. japonica GN=OJ1115_B06.12 PE=2 SV=1
MPICPPSSESSPAPGLGGRALIFHGLALGAAAAAAAAAAYLYRRPGGFRGRAVGIIPARFASSRFEGKP
LAPILGKPMIQRTWERVMLASSLDHVVVATDDERIAECCRGFGADVIMTSESCRNGSERCCEALQKL
NKHYDIVVNIQGDEPLIEPEIIDGVVMALQRAPDAVFSTAVTALKPEDASDTNRVKCVLDNQGYAIYF
SRGLIPFNKSGKVNPQFPYLLHLGIAGFDSKFLKIYPELPPTPLQLEEDLEQLKVLENGYRMKVIKVD
HDAHGVDAPEDVEKIEALMRARNIQ
>EU_Ostreolu1 tr|A4RSI4|A4RSI4_OSTLU Predicted protein (Fragment) OS=Ostreococcus lucimarinus (strain CCE9901) GN=OSTLU_36315 PE=4 SV=1
MARIVAFVPCKGTSSRVESKNTKLLDSKPLFLHTLEKLLFECPFIDKVVLDSESDDVFKVAEFLASDN
PKLVFLKRDEALATNKTDGNTLFMNEVERFDAEIYVQALCTSPFIHPETIKKGVESVMSGVYDSACT
VRDDKLYTWTKAGPSYDIQNIPNSVDLPTTTVETMGLYNITRDAAKLTRRRIGNKPHMLHVTPEEAV
DINYEVDFALA
>EU_Ostreolu2 tr|A4S882|A4S882_OSTLU Predicted protein OS=Ostreococcus lucimarinus (strain CCE9901) GN=OSTLU_27757 PE=4 SV=1

MSANFVLGVVSAPYEDGEDDQFAGKALMIVHGKPAFARAAESMKRCARLDRVVVATDDYRIVETA
KEYGIETVLVAQDARRTASTYARQAAKLTGGGWDFVVTLKVDECLIDADSVDACVMEMEANADE
WCVASTCVAATDPARFESETRPRCVFDKQGFALYLTRAVIPAIGSKPRETMTGELDLTPTKRETKNRL
DAAAMHWHVLGCSCYDAMYLRTLVDDSNDTTPLQRMENIEALNTLENGYKIKVCPVNNQTPALR
APEDVAVLEEVLRARVVQNAASAVKKKRASPKSALPRHVKRDKSRAARVRAEPSASEVSDAPSPSA
GSPSATDDVPSPRAGDRTAAS
>EU_Ostreotau tr|Q00UE9|Q00UE9_OSTTA Cytidylyltransferase family (ISS) OS=Ostreococcus tauri GN=Ot16g01010 PE=4 SV=1
MGFPLALDVIVVSVARAIAARWRTVLALGVGAVAFAKTDHEASINDRAFRTRARRARRNFDKFRRA
RRLAGTTTETSDDADVTRSDDAQNDVKRSSAGTRRVKTRGLRTNFVLGVAHAPRVDRNGRYEGKP
LALIHGKPAFARLAETMKRCTRIERVVVATDDYRVAETAKEYGIETVLVAPEVARTSTVYAREAAKAT
GGGWDYVCVVDVEECLLDADSIDACVCEMETNVDEMCVMSSCVTAIDPAGRESGGEERNAECIRP
RCVEDVNGFAMYISRAAIPVNGEARNGELVDVSTPTKRAIARDFQAWSWGIVEATCFDAVYLRMTG
NEQKEDTPLSRIENIDALNALERGYKIKVCHVNHQVPPLRDPKDVSKLDELLRAKVDKLKRASMKS
PAKKGSRNPSVRPNVGSASEAPSTSGTPAPLTDTP
>EU_Physcomitrella tr|A9U265|A9U265_PHYPA Predicted protein OS=Physcomitrella patens subsp. patens GN=PHYPADRAFT_154565 PE=3 SV=1
MGEHGSSSSLAPLLVHAAVAAGAITIAAALNHYWFRARLKARPRVVGVIPARFKSSRFEGKPLVHIM
GKPMIQRTWEQAMKCTSLQAIVVATDDDRIADCCRGFGADVVMTSENCQNGTERCNEALEKLAAK
YDIVVNIQGDEPLIEPAIIDGIVLALQMCPDAVYSTAVTGLKPEDASDPNRVKCIVDKNGYALYFSRGL
LPSNKKGIPNPDFPYMLHLGVQCYDAKFLKVYASMPPSPLQLEEDLEQLKVLENGFKLKVIKVDHE
AHGVDTPGDIAKVEAVMRAHHMD
>EU_Physcomitrella2 tr|A9RVY8|A9RVY8_PHYPA Predicted protein (Fragment) OS=Physcomitrella patens subsp. patens GN=PHYPADRAFT_35819 PE=4 SV=1
IPARYKSPRFEGEPLVHISDIPMIQRTWEQAKQFTSLKAVVVAIDDGRIADFYRAFGADVMTSENCLS
GTERCNEALEKLAAEYDIVVNVQVYQCDASACNWIKNIVDKKGFALYFSRALIPSNKKETLNPVFPY
MRHLGMQCHNAEFLRAYASMPPSSLQLEEDLE
>EU_Picea tr|A9NQZ3|A9NQZ3_PICSI Putative uncharacterized protein OS=Picea sitchensis PE=2 SV=1
MPESASGSGQCKFWILHGLLAGAAVAAATVGFRLVARSNRARNRVVGIIPARFRSTRFEGKPLVEILG
KPMIQRTWERAMLAATLDKVVIATDDEKIAECCLGFGADVVMTPESCRNGTERCNEALRKLKKKY
DIVVNIQGDEPLIEPEIIDGVVKALQGAPDAVFSTAVTSLNPEDALDPNRVKCVVDKHGYAIYFSRGL
VPFNKTGKVNPNFPYHLHLGIQSYDAKFLGIYPQLHPTPLQLEEDLEQLKVLENGYKMKVIKVDHE
AHGVDVPSDIAKIEAFMRERNIS
>EU_Pongo sp|Q5R6R5|NEUA_PONAB N-acylneuraminate cytidylyltransferase OS=Pongo abelii GN=CMAS PE=2 SV=1
MDSVEKGAATSVSNPRGRPSRGRPPKLQRNSRGGQGRGVEKPPHLAALILARGGSKGIPLKNIKHLA
GVPLIGWVLRAALDSGAFQSVWVSTDHDEIENVAKQFGAQVHRRSSEVSKDSSTSLDAIIEFLNYHN
EVDIVGNIQATSPCLHPTDLQKVAEMIREEGYDSVFSVVRRHQFRWSEIQKGVREVTEPLNLNPAKRP
RRQDWDGELYENGSFYFAKRHLIEMGYLQGGKMAYYEMRAEHSVDIDVDIDWPIAEQRVLRYGYF
GKEKLKEIKLLVCNIDGCLTNGHIYVSGDQKEIISYDVKDAIGISLLKKSGIEVRLISERACSKQTLSSL
KLDCKMEVSVSDKLAVVDEWRKEMGLCWKEVAYLGNEVSDEECLKRVGLSGAPADACSTAQKAV
GYICKCNGGRGAIREFAEHIFLLMEKVNNSCQK
>EU_Ricinus tr|B9RU62|B9RU62_RICCO Cmp-2-keto-3-deoctulosonate (Cmp-kdo) cytidyltransferase, putative OS=Ricinus communis GN=RCOM_1632840 PE=3 SV=1
MSICSSNSSSSSSGSKSWIVHSIVAGAAIAVGIGARAYLSRSRKFRSRVIGIIPARFASSRFQGKPLVQILG
KPMIQRTWERAKLATTLDHIIVATDDEKIAECCRGFGADVIMTSESCRNGTERCNEAIGKLEKKYDV
VVNIQGDEPLIEPEIIDGIVKALQEAPDAVFSTAVTSLKPEDAFDPNRVKCVVDNQGYAIYFSRGLIPY
NKSGKINPQFPYLLHLGIQSYDAKFLKIYPELQPTPLQLEEDLEQLKVLENGYKMKVIKVDHEAHGV
DAPEDVEKIERLMREHNVL
>EU_Ricinus2 tr|B9T7L7|B9T7L7_RICCO Cmp-2-keto-3-deoctulosonate (Cmp-kdo) cytidyltransferase, putative OS=Ricinus communis GN=RCOM_0282830 PE=3 SV=1

```
MSSSGSSSTSTKSWIVHSIVAGAAIAVGVGAGAYLGRFRKFRRRVIGIIPARFDSSRFQGKPLVHILGKP
MIQRTWERAKLATTLDHIIVATDNEKIAECCRGFGADVIMTSESCRNGTERCNEALGKLEKKYDVVV
NIQGDEPLIEPVIIDGIVKALQAAPNAVFSTAVTSLKPEDAFDPNRVKCVVDNQGYAIYFSRGLVPYNK
SGKVNAQFPYLLHLGIQSYDAKFLKIYSELQPTPLQLEEDLEQLKVLENGYKMKVIKVDHEAHGVD
TPEDIEKIERLMRERNVL
>EU_Sacco [sko_genscan] scaffold9909|GENSCAN_predicted_peptide_4|392_aa
MDTDKRHFAALILARGGSKGIPLKNIKILAGQPLLSWSLRAAIDSKCFDSVWVSTDHDEIARVATEW
GGQVHRRSADVSRDSSSSLDTIKEFALHHPEVDVIGLVQCTSPITHPWMLQEPSAMMRERGFDSVFS
VTRKHLFRWKPSPNTAGELVGADNLDPNNRPRRQDWDGELYENGQFYMFSRQLLDQGLLQGGKV
GYFEVGPEYSVDIDSEIDWPIAEQRVTKYGYFGVESKKDIKMVVFDVDGVITDNQVHINEKGEEFRS
YNRCDLEGVKQLKETGIIVRLLSQDTSMADMKLINEKMGCLLENGAKNKCSVLGKWVDELGLEWS
QVAYMGNDISDVPAMRKCGLSAAPADARTEARYAARFNATDRGGRGAVRQLCDHIMHLISTK
>EU_Sacco_2 [sko_genscan] scaffold14414|GENSCAN_predicted_peptide_3|924_aa
MDNLRDQSHSRKLDNNSIDLYYPSPIGYIFVDFAFNGDLEYTLPAAVISYGLKNATESQLTVDRLSQN
GIALTIIGAARAVPYSPIQMKYTLLCKIVPKCLVVRLQNPYFETMHVTDFVMSSPPKRTRSSVSNDIRE
DKRHFAALICARGGSKGIPLKNIKILAGQPLISWALRAAVDANVYDSIWVSTDHPEIARIASEWGAQI
HNRNPEAARDTSSAMETILEFTDRHPDIDIISIIQCTSPLIHPWMLQEAVRMIRNEGYDSVFGVTRKHL
FRWTEAKNGAETKAENLDPKNRPRRQDWDGELFENGQFYMFTRELLDQGLIQGGKIGYYEMEPQY
SVDIDTDLDWPIAEQRVTKFGYFGTEKRKDIKLSVFDIDGVVTDNQVHITDKGEEFRSYNCSDLEGT
ELLKKAGIEVRFVSEAKNTAGIKVITEKTGCQVETNVKDKLEVLTKWVDEMGLKWSQVTFMGNDV
SDVPAMRKSGLGAAPADGQTEARYAARFTATNRAYMRFVTVVTIPSTQLINPKIIYFNPGCCVVAQN
NLPLLNVEEKVVLFIHLDQTNSLTKTNEMSAQTLAAPYTTVQEGDKNKSMSYNERQWHCHSGGTA
SQTNSLTKTNEMLAQTLAAPYTTVQEDDKNKSILYNYPAVLKLLNQKLPFGLAEIARCFRKNIHDNS
DKFLFNLPEFTQMSMQFYSPPSVATKWMNVWQRERIIWWRKFAGYPSKFHLTDIKMTETQQTVDIQ
YEFPWGVDSVEIITNRGDVDLVELERQSGINMQGRYGRKVVTPVVIETTAGLDRGMLAYLLDAYQE
KERSDTRGKLKERKILRLHMKITPIKVVVHPIKNTRELREISEHLAKELRLSGVNTLYVCDAMTLDQ
HLVRFDEMGVPFTVIINEGTLKTGVIGMRSRDTALREQMHITEVKTLLLEHLKKE
>EU_Sacco_3 [sko_genscan] scaffold3004|GENSCAN_predicted_peptide_26|457_aa
MNGVILQPKMRKDKRHFAAVIPARGGSKGIRLKNIKLLCGQPLLSWVLRAIIDSNEFDSIWVSTENAD
IARVAAEWGAQVHHRSPEAAQDNTLAVDTMTEFAEHHPEVDILAMIQCTSPLLHPWMLHEPVKMIR
EHKYDSVFAVSRRHLFRWKEIQKPGELTVPENLNPSYRPNRQAWPGELYENGQFYMFTRELLDKRLF
QGGKVSYYELPEASSVDIDTDLDWSIVEQRVTKFGYFGLEKKKNVKLVVFDVDGVITDDEIHISDKG
TVQHKTFPVLVIKVVSLHHVVFDMGFQYLNVSRVKINYSLHISVLLRMLIICSFVFEGEEFCSYHRSD
LEGIKLLKKAGIEVRLLTENSNAPSVRLISEKIECMLEIGVKNKVELLTKWVEELNLDWQQVTYMGN
DETDIPAMRKSGLGAAPADAQTETRYAARFIATNRGGRGAVRQFCDYILRLSEK
>EU_Strongylo tr|Q0E669|Q0E669_STRPU Putative CMP-sialic acid synthetase OS=Strongylocentrotus
purpuratus GN=cmas PE=2 SV=1
MSDSSDSNRKSGHFACLILARGGSKGIKLKNIKALAGQPLIAWVLRAAIDSGEFDSVWVSTDHADIA
RISKEWGAQVHMRSPHTARDQATSIEAMQEFLKEHPEVKFVANVQCTSPCLHPSHLQRTCHMIRNL
GYDSVFAVNRRHLFRWTETPIDQAVSTKAENLDPAKRPRRQDWAGELYENGSFYFATRELLMAGLF
QGGKVGYCEMQPEYSVDIDTDIDWPIAEQRVIKFGYFGKTKPQGVKLVVFAADGVLIDNQINFTGTG
EEVRSFSLSDSIGIKHLREKGVEVRVIADEENAVTTKLAARLGIAMATESKNKVDLLNEWRKGLDLD
WSQVAYMGQDTPDLDCMKTVGIGGAPVDAQNEITTTAKFVAKSKGGKGAGREFCDHLLLVMEKA
KSAAEDRKI
>EU_Techamonas [ttrprot] AMSG_03137T0 | AMSG_03137 | Thecamonas trahens ATCC 50062 3-deoxy-
manno-octulos
MNISASVYTAAVMRSLARAFTTATVSRATSHAPLHETESSASSALAAALRVVAVIPARAGSTRLPNKP
LANLTPDLETKWTVIDAVVSAAMDAGELDDVIVATDSPDIAAAADAAGAHAVLTPADLPTGSDRVH
HALRQLHASRHPSVRSLAGVVNLQGDEPFVDPHLIDALAAELRAAAAATAHSHASAPDIVTAAVELP
DGIAPPDLDMVSVVVDHASRALYFSRAALPAAAPHADTASHPSPPPFLAHLGIYGYTLPALDAFVAA
PRPWIEQREMLEQLRALYLGLHIHVVTTTPPNTLFPWPPIAIDTPDDLHRARSWLASRNQPATSGGPA
PASITS
>EU_Trichoplax tr|B3SEY8|B3SEY8_TRIAD Putative uncharacterized protein (Fragment) OS=Trichoplax
adhaerens GN=TRIADDRAFT_34782 PE=4 SV=1
```

```
MSNLKKSVIIIPARLAARRLPNKPLIDIEGVPMIIRVYNKALEANIGKVVVAGCNDELRNLINSYNFNY
ISTDPTLPSGTDRVYAAYKNLNEKYDNVINLQGDMPHINSMTIKKVNDILNDEQDIDIATAASKINDT
ENDINDINTVKVIKSYQNYALYFTRYPVRFDENYKHIGIYGFKVNSLEKFVCLPQSPLEKNEKLEQLR
ALENGMKIKVAITNDPNISIDTEQDLLKSKNYLK
>MICROMONAS tr|C1EJI1|C1EJI1_9CHLO Udp-n-acetylglucosamine 2-epimerase/CMP-N-
acetylneuraminic acid synthetase OS=Micromonas sp. RCC299 GN=MICPUN_64942 PE=4 SV=1
MPACPNTEADKASLYAPDKCTDMKTTLRRKICVVTSNRSDWSKLKLVAINLRKLCTSQDNQEASDIQ
VDIICLGSHLLHELGATKNIVKEDFPNAYELHTLVAGDSVESMTDSVGFGIVKLTSLLCALKPNIVLVH
GDRFDAFCAAIAANMLNLTIAHVEGGELSGTVDGTLRHAITKLSHLHFTCTPEAARRIRGMGENPAS
IFVTGCPSYESLFAVSATCWEDEKMDQFFNGTPFKLKPNKFILVIMHPVTNDLEESNTLYGSLLSCLFS
RKTPTVMFYPNVDPGNKSMIQTLHKHQKADPASTSWLRLVTHMPHAKFTALMRHASAMVGNSSA
GIRESCVFGIPTLNLGSRQEGRRVPANVTTLVKPSIRSIDCWFDNELGKRYAQSTMYGFPDSAKRIAH
HLSRIDTSAGQLKQFWEPRYALLPPLQPRQYVSRTQAILADSTSSPTSSPIGRCKILGLITARGGSKGIP
GKNIIDLNGKPLIQYTIEAALSSKQLDRVILSTDSDEIAEVAQNCGCEVPFRRPSELAADDSSHLACIV
HALNILRETECFVPDFVVILQPTSPFRKSIDIDSCINIMLTSSCDMVLSVCESSLNLSKNFYFAADGTLS
PFAESTAEIDYTPRQKLLKTYAENGAVYVLRTQSLLYPPDNAPNVGSFRSADTKGYEMPVERSLDIDN
PFDLHVARLLMAKPF
```

Epimerase-2 alignment, phylip format

```
 114 335
Brancoma1 mKvCvAtCnRaDySkLaPiMfCLrDdMeLrVvVmGtHlIdDyGsTyRmIqQdGfEv-----
dGyLhTiVrGeDeAaMaEsVgLaLvKlPdVlVrLkPDlLIvHgDrFdAlSlAtCAaLmNiRiLhIeGgEvsgTiDdSiRhSiSkLaHyHaCcTdRa
RrRlLsMcEdnDrlILsGcPtYdKlLsCnVrnyIvAlQhPvTtfElMlDaVlDfGvVtVvLfPnIdAgSkEmTrTiRkNfSlCkHiPyDqFiIlMaNaGcL
iGnSsAgIrEaGaFGtPvInLgKrQiGr---EaGaNvLhVkDsKiLhAiNlQyKkQfPrFsHiYgdGhAcQrImKfi
Brancoma2 MRVCVATCNRADYSKLAPVMFALRDDMELQVVVMGSHLIDDYGSTYRMIQQDEFEV-----
DGYLHTIVRGEDEASMAESVGLALVKLPDVLVRLKPDLLIVHGDRFDALSLATCAALMNVRILHIEGGEVSGTIDDSIRHSIT
KLSHYHACCTDRARRRLLSMCEDNDHILLAGCPTYDKLLSCNAKNFIVALQHPVTTFDLMLDALLEFNVKTLVLFPNIDAGS
KEMTRMIRPNFSLAKHIPYDQFIILLANTGCLIGNSSAGIREAGAFGTPVINLGNRQIGR---
EAGENVLHVRDGKILHAVNLQYQKQFP-CSYIYGDGHAVQRIMKFI
Danio     LRVCVATCNRADYSKLAPIMFGIKSHFDLEVVVLGSHLIDDYGNTFRMIEQDDFDI-----
GSKLHTIVRGEDEAAMVESVGLALVKLPDVLQRLAPDILLVHGDRFDALALATAAALMNIRILHLEGGEVSGTIDDSIRHAISK
LAHYHAVCTLSAERHLISMCEDHSRILLAGCPSYDKLLSAYKRDYIVALQHPVTTYELMLDALISFNKKTLILFPNIDAGSKE
MVRVMRQNFRAVKHVPFDQFIQLVAHAVCMIGNSSCGVREAGAFGTPVINLGTRQTGR---
ETGENVLHVRDNKIYHALELQFGKRYP-CSKIYGDGNAVQRILKFL
Takifugu  -------------------MFGLKSHFELEVVVLGSHLIDDYGNTFRMIEQDDFDI-----
GSKLHTIVRGEDEAAMVESVGLALVKLPDVIQRLHPDILVVHGDRFDALALATAAALMNIRILHVEGGEVSGTIDDSIRHAISK
LAHYHACCTRMAEQHLIAMCEDHSRILLAGCPSYDKLLLSHQKDYIVALQHPVTTYGLMLDALLSFNKTTLILFPNIDAGSK
EMVRVMRPNFRAVKHIPFEQFIQLVNHAGCMIGNSSCGVREAGAFGTPVINLGTRQTGR---
ETGENVLHVRDNKIYHALELQFGKRYP-CSKIYGDGNAVPRILKFL
Tetraodon LRVCVATCNRADYSKLAPIMFGLKSHFELEVVVLGSHLIDDYGNTFRMIEQDDFDI-----
GSKLHTIVRGEDEAAMVESVGLALVKLPDVLQRLQPDILLVHGDRFDALALATAAALMNIRILHLEGGEVSGTIDDSIRHAIS
KLAHYHACCTRMAEQHLIAMCEDHTRILLAGCPSYDKLLLTHHKDYIVALQHPVTTYGLMLDALLSFNKTTLILFPNIDAGSK
EMVRVMRPNFRAVKHIPFEQFIQLVCHAGCMIGNSSCGVREAGAYGTPVINLGTRQTGR---
ETGENVLHVRDNKIYHALELQFGKRYP-CSKIYGDGNAVPRILKFL
Gallus    LRVCVATCNRADYSKLAPIMFGIKAEFELDVVVLGSHLIDDYGNTYRMIEQDDFDI-----
HTRLHTIVRGEDEAAMVESVGLALVKLPDVLNRLKPDIMIVHGDRFDALALATSAALMNIRILHIEGGEVSGTIDDSIRHAITK
LAHYHVCCTRSAEQHLIAMCEDHDRILLVGCPSYDKLLSAKNKDYIVALQHPVTTFELTLDALISFNKRTLVLFPNVDAGSKE
MVRVMRPNFRAVKHVPFDQFIQLVAHAGCMIGNSSCGVREVGAFGTPVINLGTRQTGR---
ETGENVLHVRDDKILHALQLQFGKQYP-CSKIYGDGNAVPRILKFL
Homo      LRVCVATCNRADYSKLAPIMFGIKTEFELDVVVLGSHLIDDYGNTYRMIEQDDFDI-----
NTRLHTIVRGEDEAAMVESVGLALVKLPDVLNRLKPDIMIVHGDRFDALALATSAALMNIRILHIEGGEVSGTIDDSIRHAITK
LAHYHVCCTRSAEQHLISMCEDHDRILLAGCPSYDKLLSAKNKDYIVALQHPVTTFELTLDALISFNKRTLVLFPNIDAGSKE
MVRVMRPNFRAVKHVPFDQFIQLVAHAGCMIGNSSCGVREVGAFGTPVINLGTRQIGR---
ETGENVLHVRDDKILQALHLQFGKQYP-CSKIYGDGNAVPRILKFL
Mus       LRVCVATCNRADYSKLAPIMFGIKTEFELDVVVLGSHLIDDYGNTYRMIEQDDFDI-----
NTRLHTIVRGEDEAAMVESVGLALVKLPDVLNRLKPDIMIVHGDRFDALALATSAALMNIRILHIEGGEVSGTIDDSIRHAITK
LAHYHVCCTRSAEQHLISMCEDHDRILLAGCPSYDKLLSAKNKDYIVALQHPVTTFELTLDALISFNKRTLVLFPNIDAGSKE
MVRVMRPNFRAVKHVPFDQFIQLVAHAGCMIGNSSCGVREVGAFGTPVINLGTRQIGR---
ETGENVLHVRDDKILQALHLQFGKQYP-CSKIYGDGNAVPRILKFL
```

```
Xenopus   LRVCVATCNRADYSKLAPIMFGIKAEFVLSVVVIGSHLIDDYGNTYRMIEQDDFDI-----
QARLHTIVRGEDEASMVESVGLALVKLPDVLNRLNPDIIVVHGDRFDALALATSAALMNIRILHIEGGEVSGTIDDSIRHSITK
LAHYHACCTRSAEQHLIAMCEDHDRILLAGCPSYDKLLSVNNKDYIVALQHPVTTFEFTLDALLSFNKKTLILFPNIDAGSKE
MVRVMRPNFRAVKHVPFEQFIQLVAHAGCMIGNSSCGVREAGAFGTPVINLGTRQTGR---
ETGENVLHVRDNKIIHALQLQFGKRYP-CSKIYGDGNAVPRIVKFL
Saccosus2 MRVCVATCNRADYSKLAPIMLGLKEDFQLQVIVMGCHLIDDYGSTYRMIEQDKIKI-----
DAKLHTIVRGEDEAAMVESVGLALVKLPDIFLRLKPDILIVHGDRFDALSTATAAALMNIRIIHIEGGEISGTIDDCIRHSITKLA
HYHVCCTERARGRLQAMCEDNERILLAGCTSYDMLLKSNTKNYIIALQHPVTTFDLTLEALLEFGKKTVLLFPNIDAGSKDM
TRILRPNFSTAKHIPFEEFIVLVANCACIIGNSSAGVREAGAFGTPVINLGSRQTGR---
ETGENVMHCRDEKILHALELQYQKQYP-PSYIYGDGHAVSRIIKFM
Paractus1
--------------------------------------------------------------------------------------------------------------------------------------------------------------------------
------------------------------------------
YSLMVDALMEFNKRVIMLFPNIDAGSKDIVRIMRPMFYPVKHIPFEEFIILVANAGCMIGNSSAGVREAGAFGTPVVNLGSR
QTGR---ETGENVLHCRDQKIHHALEIQYQRQFP-PSYIYGDGHAVPRIIKFL
Strontus1 FRVCVATTNRADYSKLGNIMQAIKDDLELSTIVLGCHLIDDYGSTYRLIEKDGFTI-----
DSRLHTIVRGEDEAAMVESVGLAMVKLPDIMLRLKPDVVIVHGDRFDVLSLAACAALMNIRIVHVEGGEVSGTIDDSIRHTI
SKLAHYHVCCTERAHKRLLAMCEDNDRILLAGCPSYDKLLSTDVVDYIVALQHPVTTYSLMVDALMEFNKRVIMLFPNIDA
GSKDIVRIMRPMFYPVKHIPFEEFIILVANAGCMIGNSSAGVREAGAFGTPVVNLGSRQTGR---
ETGENVLHCRDQKIHHALEIQYQRQFP-PSYIYGDGHAVPRIIKFL
Saccosus1 IKVCVATSNRADYSKLSPIMIGLRDDFELQVIVTGSHLLDDFGTTIKNIERDNIKI-----
NWKLHTIVRGEDEAAMVESVGLSLSKLPDILLYLQPDLLVVHGDRFDALALATAAALMNIRIVHVEGGEVSGTIDDSIRHAIT
KLSHYHLCCTEKARARLIAMCEDVDRILLSGCPVYDNLLKIDTLTYIISLQHPVTTFNFAMDALLEFGKKTLVLYPNVDAGSK
DMTRIIRENFVTAKHIPFEEFIVLMANCGCLIGNSSCVVRETCAFGTPVVSLGSRQTGR---
ESGENVFACPDHTILQALELQYQKEYP-RSFLYGDGNSVRRILDFM
Symsation -KIAVATCNRADYSKLAPVLRGLKQDFQVSIVVMGSHLIDDYGNTYRFIERDGNQI-----
DSMLHTLVRGETEGAMVESMGLAMIKLPDILNRLQPDIVMVHGDRFDAMSVAISAAVMNVRVLHLEGGEVSGTIDDVIRHS
ITKLAHYHICCTKGARLRIESMCEDSSRVLLAGCPAYDELIHTDVSKFIICVYHPVTTFGLLLDALVHFDVRTVILFPNVDAGS
KELVRCIREKISCYKHVPFSEFVYLMGNCGLMIGNSSAGIRESNVFGTAVINIGTRQRGR---
QSGANVVHVKNEDLLWHIRSQFGKVYP-RDYIYGDGRAVERIVKFI
B_DELTA   -KLCLVITTRGNYAKVKSIIQAAENDVELQIIVGGGAILAKYGNIADSLTTMGVNV-----
DRRIHFLVEGETPVTMAKSAGLAVSEFTTAFENLQPDVVMVIADRFECLSIAMTASYMNIPVAHMEGGEVSGSIDESIRHAI
TKLSHVHFPATQEAADRIIRMGEAPETVFPVGATSLDVIAGLDLDNYLTVIQHPVTTVNQTLNAIHELGMGTIWIWPNMDAG
SDGISKGIRAPMHFFKSLPIEHYAPLLKNAQAIVGNSSSGIREAAFLGTPCVNIGSRQSGR---
DRSVNIIDVGYAEIKAAIQKQLDHGRYEENHMWGDGKAGPKILDVL
B_ALPHA   RKICVVVGSRANYSSIKSAMRAIQDHLELQLIVAASAVLDRYGSVVNLIEKDGFRP-----
HARVTMLIEGETPATMAKSTGLGLIELPTLFEQLGPDVVLTVGDRFETMATTLAAAYMNIPVAHTMGGEVSGTIDESIRHAV
TKFAHIHFPASQGAAERIIKLGELPRHVHMVGCPRIDLVAEILGRSFALVSQHPVTTITLTLEAVREQGLAAIVLWPNADAGSD
DISRGIRDRMHFFKNLPIETYVNLMRSAACLVGNSSSGIREGAYIGTPVVNIGTRQHMR---
DRGDNVIDVGYKQISDAIARQVEHGRYAMDPIYGDGTAGTKIADIL
B_DELTA_2 RRICVVTGTRADYGLLYRLMKEIEGDLQLQVVATGMHLSPEFGLTYRDIETDGFTI-----
HERVEMLLSSDTPVGIAKSIGLGVIGFADAFERLRPEIIVLLGDRFEMLAAAQAALVGRIPVAHIAGGDTTGAFDEAIRHSITK
MSHCHFVTNEAALHRVRQLGENPDAVHLVGSPGIDQIRRLTLLSNLLVTFHPATLMQALFDALDRLGVGIILTKPNADTGGR
RLSDMIDDQAKAFTSLGQLRYLSLIPQVDAVVGNSSSGLYEAPSFKKPTVNIGDRQKGR---
LQASSVINCAPDDILQAVRAAFGLDCSTAVNPYGDGNASSRIAAIL
B_CYAERIA RKICVVTGTRADYGLLRWVIDGISKSLQLQIVATGMHLSHEFGLTFKEIENDGYNI-----
DLKVEMLLSSDSSVGITKSIGLGIIGFADAFKELNPDIVLLLGDRFEIFAAATAALISKIPVAHCHGGETTGAFDESIRHSITKM
SHLHFVAAREYQERVLQLGEDPENVHLVGGLGIDNIKKLSLLDNYLVTFHPVTLITELLRALSNLVTGIIFTMPNSDTGSRPL
MNLINEYASSYTSLGQIKYLSTLKYVDAVIGNSSSGLLEAPTFKTATINIGDRQKGR---
LMASSVINCDPISIQNAINHLNSHAFKQTVNPNGSGGSADKIIKVL
B_SPIES_5 RKICVITGTRADYGLLRWLILEISKSLDLQIIATGMHLSPEFGLTYKEIENDGFLI-----
HKKIEILLSSDSTVGVSKSIGLGLIGFSEAFADLNPDIILVLGDRFEILAATVAALISRIPVAHLHGGETTGAFDEGIRHSITKMS
HLHFVAANEYRNRIIQLGEDPSNVFLVGGMGIDGIKKSNLLKNLLITFHPVTLMIELLLSLETLSIGLIFTMPNADTDGRIIFELV
KSNAWYFTSLGQTRYLSCLQFVDAVVGNSSSGIIEAPSFKIGTINIGDRQKGR---
LRAKSIIDCEPIEIIDAFKRLYSSDFQKTVNPYGEGGASEKIVRIL
B_AQUICAE -----FTGTRAEYGLLKPLMYEIKKDLELQIIASGMHLSPEFGLTYKEIEKDGFTI-----
DEKVEMLLSSDTPFGISKSIGLGVIGYTDALERLKPDITVVLGDRFEALAFVIASFTLKIPIAHLYGGETTGVLDEGYRHAITKL
SYLHFTSTEEYKKRVIQLGEEPERVFNVGALGIDNIKKLKLLSNLLITFHPETLFKELLKALDELETLLIFTKANADTEGRKINY
LIDEKTVVFTNMGQLLYLSTMQFVDAVVGNSSSGIIEAPSFKIGTINIGDRQKGR---
IRAESIIDCKPESIKKSINKLFSQNFRQVNNPYENKNTARKIKNIL
B_EPSON_2 KNICVVTSTRAEYGLLYWLMKEIENDLQLQLIVTGSHLSPEFGMTIKEIEKE-FKI-----
DKKIEMLLSSDSKIGISKSMGLLQISIAEAFEELKPDFLVVLGDRYELIPIVSTAVVFNIPVVHLHGGEITGAIDEYFRHAVTKM
SYLHFTSTEEYKKRVIQMGEEPNRVFNVGAFGVENIFRLKLLNNLLITYHPETLFKEILDVLDELKTNFIFTKANSDTNGRIIN
KMIDNKAISFDSLGQIRYLSALKHVDAVIGNSSSGIIEAPSFKTATINIGNRQKGR---
IKAKSIIDCEPESIKKAFEKLYSKEFQEVKNPYQEKYPSKKVVDVL
B_GAMMA_7 RKISVVTATRAEYGLLRGLLEDINAALELQLIVTGTHLSPEFGLTLRQIEEDGFCI-----
NKKVEILLSSDTAVGVSKSMGLALISFAEVFDELKPDILVVLGDRYELIPIVSAANIARIPVAHLSGGELTGAIDELIRHSVTKM
SQLHFTAMEEYSRRVVQMGELPSRVFTVGEIGLDNLKRIQLSSNLLITYHPETTFELILKALDQLDTLLIFTNANSDVGGRSV
NMMIDEKSIAFTSLGVKRYLSALQYVDAVVGNSSSGIVEAPSFRIATINIGERQKGR---
VRASSVVDVDVEQILKALEAIYSDDFKATVNPYGKGDSVQKVIAVL
```

```
B_DELTA_4 LKICIVTGSRAEYGLLYWLLKDIAADLKLQIIATGMHLSPEFGLTYRQIEVDGFTI-----
DAKVEMLLSADTPVAVTKSMGLGVIGFADALDRLAPDIIIVLGDRFEIFAAAQAAMVARIPLAHIHGGETSGAYDEGIRHAISK
MAQWHFVATEPYRQRVVQLGEAPQRVFNVGAPGLDHLSRTQLLTLFVVTYHPVTLMKELIAGLRKFEASIVFTYPNADSG
GRALRQIINQRIRAYTSLGQQRYLSLIRQADVIVGNSSSGLIEAPALKTATVNIGDRQKGR---
LKASSVIDASETDIAKAIRHALSPEFRITQSLYGCGDASASILRQL
B_ALPHA_4 KRICVVTGTRAEYGPLFWVLKEIDAHLELQLVVTGMHLSPEYGSTWKTIEADGFAI-----
AAKVEMLLSGDTGVAIAKSMGLGTIGFADAFERLKPDIVVVWGDRFELMAAVQAAVVARIPIAHIGGGDVTGAFDDAIRHAI
SKMAHLHFPIIADSARRLRQLGEDPERIHLTGNASLDHLRRTIFLENVLVTYHPVTLWAEMLAALESLGVGIVMTAPNADND
SRELMVALEPNAILRTSLGSQLYMSTARLCDAVVGNSSSGLLEIPSLGVATVNIGTRQKGR---
PRADSVIDCPPAAIRAAMDRAMAMDCSAVVNPYGDGQSARRIVQVL
B_DELTA_6 RRICVLSGTRADYGLLYWLMKELQTYVDLQIIATGMHLSPEFGMTYKFIEADGFLI-----
NEKVEMLLSGDSVTALCKSMGLGVMGISEALARLSPDLLVVLGDRFEALAGAQAAMIHRIPIAHLHGGEATGLIDEAIRHSIT
KMSHLHFVAAEAYRNKVIQLGESPDRVFNFGATGIDNIKRLPLFNVFLVTYHPVTLMADLLSAIEEFSACVILTKSNADTAGR
VINEMIDERAMSSVSLGQKLYLSVMSIANVVVGNSSSGIIEAPAMNIPTVNIGPRQQGR---
AKAPSVLDCNESEIVSGIQTVLENSFMADKHVYGDGNTAVKVAEVL
B_EPSILON RKICIVSATRAEWYLLRNLCLEIQNDLELQIIATGAHLSPEFGLTYKEIEKE-FKI-----
AKKIPILLASDDKISLCKAMALANIGFSEAFESLKPDIVVILGDRYEMLSVASVCLMMNLPIAHLCGGELTGAIDDSIRHSISKM
AHLHFVSTQTYKNRLLQLGEEKNRVFNVGSLASTIMKNINFLDIYLITYHPVTLIDILLKTLDELNASLIFTKANADENGLLINEI
LKKKAKLFDNLGSQKYLSIMKIAKAMIGNSSSGISESPFFKTPCINIGSRQKGR---
LRTQNIIDCEFENLDQAFEKLESKDFKQFKNPYENDKNPNIIIKTL
B_FIRES_2 RKIAVITGTRADYGIYYSVLKAIENHLELHLIVCGMHLSPEFGMTINEIEKDGFKI-----
DDKIDTILSSDSGEAMAKSIGITLMGLTQSLDRIKPDVLLILGDRGEMMAGALAAIHMNIPVAHIHGGEVTGTVDESIRHSITK
LSHIHFPANEDSRERIIKMGEEKKNVYVVGAPGIDYIKNTEYLSIFILTQHPVTTIEETLSAIAELGVQTIISYPNSDNGGREIIK
VIEDFLKVFKNLSQVEYLSLLNTADIMIGNSSSGIIEAPSFKLPVINIGTRQQGR---
LRACNIIDVSYKEILSAIDKVLYNEEFKCENPYGDGHSGERIADIL
B_FIRES_3 RKIAVITGTRADYGIYYSVLKAIENHLELSLIVCGMHLSPEFGMTIEEIQKDGFKI-----
DDKIDTILSSDSGSAMAKSIGITLMGLTQSLERINPDILLILGDRGEMIAGALAAVHMNIPVAHIHGGEVTGTVDESIRHSITKL
SHIHFPANEDSKRRIIKMGEEEKNVYVVGAPGIDYIKNTKYLSIFIMTQHPVTTIEETLSAIAKLGVQTIISYPNSDNGGREIIKV
IEDFLKVFKNLSQVEYLSLLNTADVMIGNSSSGIIEAPSFKLPVVNIGTRQQGR---
LRACNIVDISYKEILDAIDKVLYDEEFRCENPYGDGKAGERIADIL
A_EURY    RKIAVVTGTRADYGIYLPVLKAIQRSLDLSLIVTGMHLSETFGHTVDEIEKDGFSI-----
DAKIPLGLLEDSGASMALDVGICILGLTDALKKIKPDILLVLGDRGEMLATTIAGIYMNIPVAHLHGGEVSGTVDESIRHAITKL
SHIHFPATEESAERIRNLGEDEFRIYVVGAPALDTILSETFVPIILVVQHPVTTIRETMDAVVELGEQTVVIYPNADAGGRKIIE
TIEAFIKIFKNIRHVDYLSLMRTTNVMVGNSSSGIIEAPSFGLPVVNIGTRQTGR---
QRGQNTIDVDYDEIIKAIKVGLYDKDFKCISPYGNGHAGTAIAEIL
A_EURY_2  RKILVLTGTRAEYGLLRSSMEAIQNHLTLSIVATGMHLSPQHGMTVEKIREDGFSI-----
DREVLMQLSGDSETAMAKSLGIGTASLADAFESLDPDVVLLLGDRDEALAGALAASHMNIPVAHVHGGDSAAMIDESIRHA
ITKFSHIHFPASERSAERIKKLGEESWRITIAGAPGLDDILAGEYEDLLMVLQHPVTTMAATLDAVESTEAQAVIIYPNSDAGS
NQMIDEIEADVRTFRNMPRKEYLGLMAATDVMVGNSSSGIIEAPSFDLPVVDIGPRQRGR---
ERTQNTFSAEHEQIREKIYQCLNEGVLDQDNPYDYGGAGSRICERI
A_EURY_3  KKVAVVTGTRAEYGILKPLIEKINDDLELQLMVTGMHLLKKFGYSIKEIENDGFPI-----
ASKIRMYDEDSLGSYHGVSLGRAVSEFTREFVYLNPDIVLVIGDRLEALAPVLSASTLNIPIGHIHAGDSTGHIDEQIRFAISR
FSHLLFAPTEKCVERLTKMGEEPWRAYNVGALGLDSILSYKPLTVAIIIFHPVIHIESIMKAVIETKINTVVIYPNNDLGSKKIIEII
KENIKLFENLEHNIYISLMYHANLMIGNSSSGIIEAPSLGLPVINVGSRNTGR---
EHGDNVLFVKPKEIVEAINKALYDIDFIKNNPWGDGKTSERLVNIL
B_GAMMA_9 RKIIYVTGTRADYGLMREVLKRLHQSIDLSICVTGMHLDALYGNTVNEIKADQFSI-----
CGIIPVDLANAQHSSMAKAIGHELLGFTEVFESETPDVVLLLGDRGEMLAAAIAAIHLNIPVVHLHGGERSGTVDEMVRHAI
SKLSHYHFVATEASKQRLIRMGEKEETIFQVGAPGLDEIMQYKTSTICLLIYHPVVQFQSVIQAALATNLQIICLEPNSDTGG
HLIREVIQPDVRIIKHLHRPEFIDCLANSDVMLGNSSSGIIEAASFNLNVVNVGSRQNLR---
ERSDNVIDVDVDAILTGLREALNKPKIKYFNCYGDGKTSERCYQLL
EU_Riinus RKICVFTGTRAEYGLLRPLMQAIATRATLQTLVTGAHLAESTGATWREIANDGLPI-----
DERVEVLLDGGGDESICTSIGLGVMRYAEALKRLAPDMLVILGDRYEAIAAAVAATVCKVPIAHIHGGELTGAMDDAFRHSIT
KMSYLHFTSTEAYRKRVIQLGEQPARAHNVGALGVENIRTLNLYDYLLVTFHPATLLKALLNALDAFPHVCVFTGANADPG
GAGLNRLLA------------------------------------------------------------------------------------------------------
B_ALPHA_2 KRIVYLTGTRADFGLMLPTLRAIDRSLELELLVTGMHLSDRFGRTEREVEAAGLKI-----
GRRIPVPIDDDSGHGMGVSTGLITCAVADYLAETACDVLLLLGDRGEMLAAATAGLFADVPIVHVAGGDRSGSVDESIRHAI
SKLAHIHCVSNEDARQRLIRMGEDPDRIFDVGAPGLVGLKKPDRATFVLVLFHPVVQWRAMFDALAGLPFRYVALMPNAD
HGTTSIRTEIEGQLTTIDHMPRADYLALLAECRFLIGNSSSGIVEAATFGTPVVNVGDRQFGR---
LRSANVFDAPPGAIGRAIEKAIGFDPQGLRNVYGDPHADIRICEIL
B_ALPHA_3 REILFVTGTRADFGKIEPLALAARDRFKVSFLVTGMHMLDRYGLTKIEVHRVQ--------
GATVHEFLNQREGDPQDTILAKSIIGFSDIFAELRPDLVVFHGDRIEALACALVCATNYIRSAHIEGGEVSGTIDEVFRHCNTK
LAACHFVSSEAAAKRVMTLGEPADRIHVIGSPELDFHARPSGVTYGVTLFHPVTSAADLFGALEASGRNFVVIAPNNDPGS
REIFAVLE-RFRLIPSMRFAHFSELMKHAACLVGNSSAGVREAPFLGIPSLDIGTRQTNR---
AEAPSLFSADAREKIAAFLATEWGKRYPPHTAFGEGRAAERFLEVL
B_CYAIA_4 KVLLFVTGTRADFGKMEPLAREAFNNFKVIFFVTGMHMMREYGLTKEEVHKNK--------
DIQIFEFSNQKYGDKLDTILSNTVRGFSNYVKEINPDIVIIHGDRIEAIACSLVCSTNNIISAHIEGGEVSGTIDEVFRHCNTKLC
TFHLVSSNEAKKRVRQMGEPEKNIFVIGSPELDIHGRKSGVDYGICIFHPVTTAENLFKSLSISNRNFVIILPNNDPGSIYICN
EID-NFRIIPSMRFNYFSELMKNSSLIIGNSSLGVREAPFLGIMSINIGTRQNKR---ALTQSIYNCSGPEIVDAIGKFWNKKTT-
SHKGFGSGNSRKKFLKFI
```

```
B_FIRUTES RKILFITGTRADYGKIKPLMRKVEESFELHIFVTGMHMLSKYGSTWKEIEKDGF--------
KNIYKFINQQYNFHMDIALSNTIIGLSNFVNELRPDMIVVHGDRLEALAGAIVGAFNNIRVAHIEGGEVSGTIDESIRHAITKFS
HIHFVCNEEAKKRVIQLGEKPESVYVIGSPDIDVMLSDTLPSYAIFIYHPVTTIKEVIDALIESGKNYVVIYPNNDEGSNIILKEY
EKRFKIFPSLRFEYFLTLLKHADFMIGNSSAGIREAGIYGVRVIDIGIRQKGRYDISKFKNIIHVNEEEILKAIQKVQLVDKF-
YISAFGDGKSAERFINIL
B_BACES_9 KKILFLTGTRADFGKIKSLIQILEAHFEPYIFVTGMHLQEEYGYTFLEVQRCGF--------
RNIHTFQNHTHETTMDLTLAKTITGLSAYVKDCQPDMIVIHGDRVEALAGAIVGSLNNILVSHIEGGEVSGTIDELIRHATSKM
SHIHFVSNDQAKKRLLQMGELHESIFIIGSPDVDIMFSKSLPDYAVVMFHPVTTANDFVHVLLADTHNYVVVFPNNDLGSHA
IIQAYEKRFRVFPSLRFEYFLTLLKNAQFIIGNSSAGIREAPYYGLPIINIGTRQQNR---
AAHTHIINVDYLHIQEALSSIDSHKVQPSKNDFGQGNSAKLFLASL
B_DELA_11
KKILFITGTRADFGKLCSLIDKVSEHFEYCIFVTGMHTLSRYGYTVDEVMKKYSSFRLEGGFRNVHVFMNQVHGESMDMV
LGNTIFGLSRYVSEYQPDMIVVHGDRVEALAGSIVGSLRNILVAHVEGGELSGTIDELIRHAVSKMAHLHFVANGDSRRRLM
QMGEAEETIYPIGSPDVDLMFSPNLPSYAITLLHPVTTAGAVVDAMLDSDDNYIVIYPNNDHGSDIILDEYDGRIAVYPSLRV
ESFLVLLRAAKYLLGNSSAGIRETPCYGVPSINIGSRQDGR---
FCCSSIINVPGDAIVKALADVRKMVPHPPRFEFGRGNSAEQFIKIL
B_BETA    KRILCITGTRADFGKLKPLLAYIENHLELHLIVTGMHMMKTYGRTYKEVTRENY--------
QHTYLFSNQIQGEPMGAVLGNTITFISRLSDEIEPDMVMIHGDRLEALAGAAVGALSSRLVCHIEGGELSGTVDDSIRHSISK
LSHIHLVANEQAVTRLVQMGEKRKHIHIIGSPDLDVMASSTLPSYGISMFHPVTTAAQYFKALELSGQNIISIYPNNDTGTESI
LQELL-KFIAFPSIRFEYFLVLLKHAKFMVGNSSAGIREAPLYGVPSIDVGTRQSNR---
HMGKSIIHTDYKNIFDAIQQACSLGKFEADDTFNGGDSTERFAEVI
B_BETA_2  KRILCITGTRADFGKLKPLLAYIENHLELHLIVTGMHMMKTYGRTYKEVTRENY--------
QHTYLFSNQIQGEPMGAVLGNTITFISRLSDEIEPDMVMIHGDRLEALAGAAVGALSSRLVCHIEGGELSGTVDDSIRHSISK
LSHIHLVANEQAVTRLVQMGEKRKHIHIIGSPDLDVMASSTLPSYGISMFHPVTTAAQYFKALELSGQNIISIYPNNDTGTESI
LQELL-KFIAFPSIRFEYFLVLLKHAKFMVGNSSAGIREAPLYGVPSIDVGTRQNNR---
HMGKSIIHTDYKNIFDAIQQACSLGKFEADDTFNGGDSTERFAEVI
B_BETA_3  KRILCITGTRADFGKLKPLLAYIENHLELHLIVTGMHMMKTYGRTYKEVTRENY--------
QHTYLFSNQIQGEPMGAVLGNTITLISRLSDEIEPDMVMIHGDRLEALAGATVGALSSRLVCHIEGGELSGTVDDSIRHSISK
LSHIHLVANEQAVTRLVQMGEKRKHIHIIGSPDLDVMASSTLPSYGISMFHPVTTAAQYFKALELSGQNIISIYPNNDTGTESI
LQELL-KFIAFPSIRFEYFLVLLKHAKFMVGNSSAGIREAPLYGVPSINVGTRQSNR---
HMGKSIIHTDYKNIFDAIQQACSLGKFEADDTFNGGDSTERFAEVI
EU_Mionas RKICVVTSNRSDWSKLKLVAINLRKLIQVDIICLGSHLLHELGAT-KNIVKEDFPN-----
AYELHTLVAGDSVESMTDSVGFGIVKLTSLLCALKPNIVLVHGDRFDAFCAAIAANMLNLTIAHVEGGELSGTVDGTLRHAIT
KLSHLHFTCTPEAARRIRGMGENPASIFVTGCPSYESLFAVSATWFILVIMHPVTNYGSLLSCLFSRKTPTVMFYPNVDPGN
KSMIQTLHSWLRLVTHMPHAKFTALMRHASAMVGNSSAGIRESCVFGIPTLNLGSRQEGR---
RVPANVTTLVKRSIDCWFDNELGKRYA-QSTMYGFPDSAKRIAHHL
B_BACETES LRLLIIVGTRPEIIRLAAVINKCRRYFDCILAHTGQNYDYNLNGVFF----HDLGLQAP--DVYMDA-----
VGDDLGSTMGNILNASYKLMSHLRPDAVLVLGDTNSCLS-
VISAKRLHIPIFHMEAGNRCCLPEETNRRIVDIISDMNLCYSEHARRYLNASGVAKERTYVTGSPMAEVLSENLSAEYILLSA
HREENFASLFEGINAMAYDMPVLYSCHPRSRNRLESSGFSRVIRHAPLGFHDYNCLQMHAYAVVSDSGTLPEESSFFSFP
AVCIRTSTERPEALDKGCFILAGIDASLLQAVDTAVEMNRNGPVPDYMDRNVSTKVVKLI
B_BACES_7 LRLLIIVGTRPEIIRLAAVINKCRRYFDCILAHTGQNYDYNLNGVFF----HDLGLQAP--DVYMDA-----
VGDDLGSTMGNILNASYKLMSHLRPDAVLVLGDTNSCLS-
VISAKRLHIPIFHMEAGNRCCLPEETNRRIVDIISDMNLCYSEHARRYLNASGVAKERTYVTGSPMAEVLSENLSAEYILLSA
HREENFASLFEGINAMAYDMPVLYSCHPRSRNRLESSGFSRVIRHAPLGFHDYNCLQMHAYAVVSDSGTLPEESSFFSFP
AVCIRTSTERPEALDKGCFILAGIDVSLLQAVDTAVEMNRNGPVPDYMDRNVSTKVVKLI
B_BACES_2 LKVMTIVGTRPEIIKLSRVMAELDKYTEHIMVHTGQNFDYELNEIFF----QELRIRKP--DYFLDA-----
AGKNAAETIANVIRKSDELMDQVKPDALLLYGDTNSCIS-
VISAKRRKIPIFHMEAGNRCRVPEEINRKIVDHLSDINMPLSEHARKYLLAEGLRPETVIKIGSPMTEVLIYHKAEEYFIVSTH
REENFSDLLSSLNAIVYHKKVIVSTHPRTRKKLESIGFPMIEFMKPFGFMEYIKLQQNAFCVISDSGTITEESSILHFPAITIRQ
AHERPEGMDEGTLIMTGLNDRILESIEIVTSQYAEGSIPDYASDNVSKKVVRII
B_SPIES_3 FKVMTIVGTRPELIKMSRVIAELDRNFKHILVHSGQNYDYELNQVFF----EDLEIRKP--DHFLNA-----
AGESAAATIAQVLLKADEVFEKEKPDALLLYGDTNTCLA-
VISAKRRKIPIFHMEAGNRCRVPEELNRKVVDHLSDINLVLTEHARRYLLAEGIKPETIIKTGSHMDEVLKYYKAKEFFIVSSH
REENLKKLLESLNAICYGLPVIVSTHPRTRKRLEDFKEPLITFLKPFGFFDYVQLQMSAFCILSDSGTITEEASLLDLPAVTIR
NTHERPEGMDVGTLIMSGLSERILESVRIVVDQYRAGVVDDYSAGQVSKKIVSIV
B_ALPHA_7 LKVMTIVGTRPELIKLCCVISEFDKYTHHILVHTGQNYAYELHQVFF----DDMGIRKP--DYFLEV-----
AADNTAKSIGLIIEKVDAVLEKEKPDAVLFYGDTNSCLS-
AIAAKRRKIPIFHMEAGNRCRVPEEINRKIIDHISDVNITLTEHARRYLIFEGLPPELIFKSGSHMPEVLDRFMSKLYFLISSHR
EENLKELLSSLQTLIYNFPVIFSTHPRTKKRLEDLEEDKIRFLPAFGFTDYVKLQMNAFCILSDSGTITEEASILNLPALNIREA
HERPEGMDAGTLIMSGFKERVLQSVKAITEEHENNIVPDYAAGLVSKKILRIV
EU_Rius_2 -----------------------------------------------------------------------------------------------------------------------------
MEAGNRCRVPEELNRKVLDHLSDINMVLTEHARRYLLAEGIPADRIIKTGSHMQEVLEYYMPNEYFLLSTHREENLQDLLD
TLQALVYDMPVVVSTHPRTRQRLEKLGVNRIHFLKPFGFFDYIKLQKEAFCILSDSGTITEEASLLNLSAVTIRNAHERPEG
MDEGTLIMCGLKERVLDAVNIVTHQHKKTIVPDYLGGLVSRK-----
B_BACES_6 LKVMTVVGTRPEIIRLSRVLMALDESIEHIIVHTGQNYDYELNQIFF----EDLGLRKP--DYFLEA-----
AGKTATETVGNILIKIDPLLEQLQPDAFLVLGDTNSCLC-
AIPAKKRQIPIFHMEAGNRCRVPEETNRKIVDHTSDVNLTYSDIAREYLLREGLPADRIIKTGSPMFEVLHHYLPQKYFVVSS
HREENFKGLIESLNLIAYNYPIIVSTHPRTRNMIDKMKVPEVQFLKPLGFHDYNALQMRSYAVLSDSGTISEESSTLNFRALN
IRDAHERPEAMEEASVMMVGLSERIMQGLVQLQSQEIGESVADYSMPNVSQKMVRII
```

```
B_DELTA_8 MKVMTVLGTRPEIIRLSRVLAKLDQHCEHKIVHTGQNYDFELNEIFF----QDLGVRKP--DFFLNA-----
AGESATETIGNIIKTVDKVMGEFLPEAVLVLGDTNSCLS-
VIAAKRRKVPIFHMEAGNRCRVPEETNRKIVDHTADINLTYSSIAREYLLREGLPPDRVIKTGSPMFEVLTHYRKQDYFVVS
AHREENFDKLVESLNLVAFDIPVVVSTHPRTQNRIDKRGSSNVRLLKPLGFSDYNHLQLHARAVLSDSGTITEESSIMNFPA
LNIREAHERPEGFEEASVMMVGLDIRIMQALRILETQPRADQVADYSMPNVSDKVVRII
B_SPIETES LKVSTIIGTRPEIIRLSRVLAKLDQYCDHIMIHTGQNYDYELNEIFF----NDLEIRKP--DYFLNA-----
AGTSGAATIGNVIIKVDELLAQVQPDAVLVLGDTNSCMA-
VIPAKRRKIPIFHMEAGNRCRVPEEINRRIVDHTADINLTYSNIAREYLLREGLPSDMVIKTGSPMFEVLNYYLEGKYFVVSA
HREENFAKLIDIINTIAFKFPVIISTHPRTQKKINVSNVPLVQLLKPLGFKDYNKLQLSAKAVLSDSGTITEESSILNFPALNIRE
AHERPEGMEEASVMMVGLEERVLQVLQILEKQPKGENVSDYSMPNVSEKVVRII
B_SPIES_2 LKVSTIIGTRPEIIRLSRVLAKLDQYCDHIMIHTGQNYDYELNEIFF----NDLEIRKP--DYFLNA-----
AGTSGAATIGNVIIKVDELLAQVQPDAVLVLGDTNSCMA-
VIPAKRRKIPIFHMEAGNRCRVPEEINRRIVDHTADINLTYSNIAREYLLREGLPSDMVIKTGSPMFEVLNYYLEGKYFVVSA
HREENFAKLIDIINTIAFKFPVIISTHPRTQKKINVSNAPLVQLLKPLGFKDYNKLQLSAKAVLSDSGTITEESSILNFPALNIRE
AHERPEGMEEASVMMVGLEERVLQVLQILEKQPKGENVSDYSMPNVSEKVVRII
B_SPIES_4 LKVSTIIGTRPEIIRLSRVLAKLDQYCDHIMIHTGQNYDYELNEIFF----NDLEIRKP--DYFLNA-----
AGTSGAATIGNVIIKVDELLAQVQPDAVLVLGDTNSCMA-
VIPAKRRKIPIFHMEAGNRCRVPEEINRRIVDHTADINLTYSNIAREYLLREGLPSDMVIKTGSPMFEVLNYYLEGKYFVVSA
HREENFAKLIDIINTIAFKFPVIISTHPRTQKKINVSNAPLVQLLKPLGFKDYNKLQLSAKAVLSDSGTITEESSILNFPALNIRE
AHERPEGMEEASVMMVGLEERVLQVLQILEKQPKDENVSDYSMPNVSEKVVRII
B_GAMMA_8 LKVMTVVGTRPEIIRLSRVIVACDKHFQHILVHTGQNYDYELNEVFF----HDLGIRKP--DYFLNA-----
AGKTSSATIGNVIIAVDQLLEETQPEALLVLGDTNSCLA-
VLPAKRRKIPTFHMEAGNRCRVPEEINRRIVDHTADINLTYSTIARDYLLAEGLSPDLIIKTGSPMFEVLHFYHEQKYFVVSA
HREENFLDLIHTLNAVAYQYPVIVSTHPRTRKRIEELNIPLIQLLKPLGFLDYNKLQLMAKATLSDSGTINEESSILNFPALNLR
QAHERPEGMEEAAVMLVGLQERVLQGLAILDEQRRGQLVADYSMPNVSAKVIRIL
EU_Enzoon LKVVTVVGTRPEIIRLSRVMAELDKHCDHILVHTGQNYDYELNEIFF----RDLGIRTP--DYFLNA-----
AGVSGAQTIGNVIIAVDRVLEDVKPEALLVLGDTNSCMA-
VIPAKRRRIPTFHMEAGNRCRVPEEINRRIVDHTADINLTYSSIARDYLLREGLPPDMVIKTGSPMFEVLNYYRDGE----------
------------------------------------------------------------------------------------------------------------------------
B_ZETA    MKVVTVVGTRPEIIRLSRVIACLDRSVDHKLVHTGQNYDYELNEIFF----QEMGIRKP--DYFLNA-----
VGASPVQTIGRILETIDPVLEQEAPDAMLLLGDTNSCLA-
VIAAKKRKIPIFHMEAGNRCRVPEESNRKLVDHLSDINMPYSSHAREYLLREGLPADRIIKTGSPMHEVLHHYSGNEYFVVS
CHREENIRRFVHVLNSLAYDRKIIVSTHPRTRARLESLGVSHVALLKPLGFFDYVHLQMHAYAVLSDSGTITEESSILNFPAL
NIREAHERPEGMEEGSVMMVGLDAQILQALSILETQQRGEQVADYAPLNVSEKVVRII
B_ACTIA_5 ARVMILVGTRPEIVKLSRIIAALERAVDVCLVHSGQHYDYELNQVFF----DELGIRKP--DHFLDA-----
VGASAAETIGRVIARSDAVFVDESPDALLLYGDTNTTLA-
VIAARRRHIPVFHLEAGNRCRVPEEINRRLVDHLSDINLPLTEHARRHLLAEGLPAQRIFVTGSPMKEVLDHYAPLDFLVVSA
HREENLIGLLETLNALAYRVPIIVSTHPRTRDRLDALEAGLVRFCRPFGFADYIALQRAAQCVISDSGSLTEEASLLGFPAVMI
REAHERPEGVDHGVAVSCLPRDRVLAAVDLVVDAAQG-IVPDYDVDDVSRRVVRII
B_BACES_4 MKITIVAGARPNFMKIAPITRAIDAAISYRLVYTGVENDNSLDASLF----ADLHMKAP--DAYLGV-----
NGNNPTELTAGIMIAFERELTENPTHVVLVVDDLTATMSCAIVAKKQNIKVAHLVAGTRSSMPKEINRMITDGLSDYLFTAGM
VANRNLNQTGTENETVYYVGNILIDTIRYNRNRIYILLTLNRHVLLQELMETLLKKA--
MPIVAPLHTYVRDAIKALGIPNLHIMPTQSYLSFGYLMNQAKAIVTDSGNVAEEATFLGIPCITLNTFAEHPETWRTGTNELV
GEDAALGACMDKLMNGEWKQTLPERWDGRTAERIVQIL
B_ACIIA_2 -HFLHVVGARPNFMKAAPLIRALEQRSRQTLVHSGQHYDRNMSTVFF----DQLGIRKP--DVNLQV-----
GSGSHAQQTAAIMSRVEPVLLNQRPDAVIVYGDINSTVAVALVCAKLGIKLIHVEAGLRSSMPEEINRLVTDQLADVLFTPSL
DGDENLHREGIPDNKVHFVGNIMIDTLVRLLPLEFGLVTLHRPSNLAPLLFALDRIA--
LPLLFPVHPRTLQHMQEFSIHHLQILEPLPYIDFLSLQQRAALVITDSGGIQEETTYLGIPCLTVRENTERPVTVTLGTNLLVG
SDHRMESEARKVIAGNKKCSIPPLWDGHTSDRIASIL
B_ACTIA_3 LHAVVIAGARPNFIKVKPVLDALAADARSSFVHTGQHYDEAMSDVFF----ADLGLRQP--DYHLEA-----
GSGSHAVQTAAVMTAFEPLLARLAPDVVVVVGDVNSTLACALVAAKANVPVAHVEAGLRSTMPEEINRIVTDRVSDLLFAP
SPEGVVHLLSEGARPESVHLAGNVMVDTLLACRERRYGLVTLHRPSNLDGLLTALGEIA--CPLVFPVHPRTASRLAGRL-
DGVRALGPAGYLDCVALQMGARLVLTDSGGVQEESTVLGVPCLTLRESTERPITVTEGTNRVVGCSAAIMAGAFEVLDRP
PPPRCPDLWDGHAGRRISAVL
B_ALPHA_8 PKVLCVVGARPNFMKIGPVIRALKGAMAAPLVHTGQHYDEAMNDRFF----ADLGLPRP--DVNLEV-----
GSGSHAQQTAEIMRRFEPVLDDLAPQALLVVGDVNSTIACALVAAKKGVTVVHVEAGLRSAMPEEINRVLTDQISDLLFTTE
RDALANLTREGIDPARVHFVGNVMIDTLVMNRPRVYALVTLHRPSNLARLLGGLAEIA--
TPLVFPIHPRTRARIEAAGLARTVLLPPAGYLEMLGLMAGARLVITDSGGVQEETTALGVPCITVRENTERPITVSEGTNTVV
GTDALLLAAWRDVLASGGKTRIPEFWDGKASERIASIL
B_DELTA_9 -NVFLVAGARPNFMKIAPLYRELTREVTCRIVHTGQHYDYEMSQTFF----DDLQLPQP--DYFLNA-----
GSGSHAEQTARVMVTFEELCAKELPDLVIVVGDVNSTLACSIVAKKAGIRVAHVEAGLRSSMPEEINRMVTDSIADLYFVTE
TSAVRNLFNEGKPPEQIFLVGHVMVDNLLHQMVQHYVFMTLHRPSNFRGIAFAVNEIA--
IPIIFPVHPRTRKMMESFGIKGVTCLPPLSFSESLYFWKDARVVMTDSGGLQEETTALGVPCVTIRENTERPITVEMGTNVL
AGTDDRILTRVFDAAEGRGKRKVPPFWDGKAAERIWDVI
B_BETA_4  KKVYLIAGARPNFMKIAPIVRALQQQLTYKIVHTGQHYDREMNDVFF----EELGIPQP--DIFMAA-----
GGGSHSQQTAKIMVAFEEYCQTEPPDAVLVVGDVNSTLACSIVAKKLHIPVAHVEAGLRSCMPEEINRLVTDSITDWFFVTE
PSGQQHLLQEGKPASAIHYVGHVMVDNLLYQVEQAYGVVTLHRPSNLERISLTLKQIA--
LPLVFPAHPRTQNNLKKFNIPNILLMGPQAYMPFLHLWKDATLVLTDSGGLQEETTALGVPCITIRENTERPITVNEGTNILAG
TRERILAAVDDILGGRGKQRRPHLWDGNAARRIVEIL
```

```
EU_Saoeca GNFTIVAGARPNFMKIAPIIHEIQKLVSFRLIHTGQHYDKKMSGDFF----EQLDIPQP--HANLGA-----
GGGTQAEQTAAIMVAFEKELMENRPDLVLVVGDVTSTLSCSIAAKKLQIDVAHVEGGIRSGMPEEINRMVTDSITDHFFTTS
EIANQNLRNSGFSEDKIHFVGNTMIDTLLAQMPKQYFVMTMHRPANLKAMIDAILDGT--
LPIIFPVHPRTAKNLQAIGIAPNLNMNPLGYLEFNYLVKNAKGVITDSGGITEEASVMNVPCITLRDNTERPETIDLGTNELVG
TNEKLKPYLDKIMSGDWKKKGIPLWDGKTAERIVKIL
B_DICALES -KVVSIVGARPQFIKLAPFSAELRKNIKEVILHTGQHYDENMSELFF----KELEIPEP--DYNLGI-----
GSGSHGEQTGRMLIGIEEVLVKEKPDVVIVYGDTNSTLAGALASAKIHIPLAHVEAGLRSKMPEEINRIVADHLSDILFCPTE
TAVENLKREGIE-KGVYLVGDVMFDALMHFSKLKYYLITVHRAENLKNIFSAISELD--
KEVIFPIHPRTKNRLKELGLGRVRIIDPVGYLDMIELEKNALAILTDSGGVQKEAFWLRVPCITLREETEWVETLKYGWNILV
GSNERILEAIKNIKNG----KEISFENDYASPKMREVL
A_EUREOTA -RIASIVGVRPQFVKASVVSRELRKKNEEILIHTGQHYDYQMNKLFF----EELNIPEP--EYHLDI-----
GSGSHGYQTGEMLKKIEEVLIKEKPDLVLTYGDTNSTLAGALAASKLHIKTAHVESGLRSAMPEEINRIVTDHCSDILFCPTE
NAVENLKNEGIT-KNVYLTGDVMVDSLLYNRKIEYLVATIHRASNLKNIVNAFSELK--
ETIVFPLHPRTDKFLKKYGLSSVILIEPLGFFEFIKLMNNAKMILTDSGGVQKEAYVLKVPCITLRENTEWIETVNDEWNVLV
GTNEKIVKMVKEFRPFLEK-HRDRFGNGDASKNILLII
A_EURTA_4 -KIVSIVGARPQFIKCAPLSRLIREKHEEILIHTGQHYDTGMSDIFF----DELKIPKP--NYNLGV-----
GSNSHGVQTGKMLIEIEKILLRESPDLVLVYGDTNSTLAGDLAASKLHIKTAHIEAGLRSSMPEEINRVLTDHTSDLLFCPTE
TAVLNLKKEGIT-TGVYNVGDVMLDSLKYNIRIEYIVATVHRASNLSSITNAFCHTG--
VSIVFTVHPRTEKYLKQYGLEKVKVIPPLGYLEMLKLMVHAKKILTDSGGVQKEAYMLGVPCITMRENTEWVETIEDGGNV
LVGTDEKIMDAILNFKGVPV--KGNFFGNGNACAEICKTL
B_ACIERIA -KLVTVVGARPQFIKSGPVSLAIEKHIEEILVHTGQHYDSEMSQVFF----EEMNLRTP--KYNLEV-----
GSGNHGEQTAQILARCEKVLMDEKATALMVYGDTNSTLAAALAAVKLHIPVFHVEAGLRSEMPEEVNRVLTDHISDLLFAP
TDTAVENLKAEGIT-KGVELLGDVMYDAIQQHLQTKYALMTMHRASNLGQILSAISEIA--
IRVVWPVHPRARKRMEDFGIEGITLISPASYLDMMMLVSNASLVLTDSGGLQKEACWMRVPCVTLRDETEWVETVASGW
NTLAGADDQILMAARKALTSKPS-ETPGAQHAGASERIAASI
A_EURTA_2 -RPAIIFGTRPEIIKLSPVIRAFIKNIQPILIHTGQHYDYEMSSIFL----EELELPEI--DYHLEV-----
GSGTQAEQTGIAMIKIEKVLMNEVPDVSIVQGDTNTVLAGALASVKLLIPVAHVEAGLRSTMPEEINRILADHSSEVLFAPTL
EAKKNLEREGIR-ENVFVVGNTIVDAVLQNSKIEYILVTAHRKENLKKLVDILTSLP--
MQVVYPVHPRAEKRLKEFGLSGVMLLKPLGYLDFLKLEKNAFIVMTDSGGIQEEAIILNVPCLTLRYNTERPETVKAGGNIL
VGVEDLALRYVKKLIEDEAFYAKNPFGDGRSGERIVSIL
B_THEOGAE -RVLSLVGARPQIIKEAMLHREFKEKIEEILVHSGQHYDYNMSDVFF----EVLEIRQP--HYNLNV-----
GSGTHGEMTGKIMIEFEKVLLREKPDLVLVYGDTNTTLAGALVAAKLKIPVAHVEAGLRQDMPEEINRIVTDRVSQILFCPSE
LAVENLRKEGIT-EGVYFTCDVMYDLFLKMKPL-YIVCTIHRDFNLREILEQLRRLS--
YEVVFPVHPRTAKRIREFGLNEILVIEPADYLNMMGLIEKSRFVITDSGGLQKEAYWCGKRAIVVMPDTGWRELVEAGWNV
LSEPD-EITEKSEYIDNHVSL-PENVYGEGNASEKLAEVI
B_THEAE_2 -KVLSLVGARPQIIKEAVLHRRFKEKIKEVLVHSGQHYDYNMSDVFF----EILQIRKP--DYNLNV-----
GSGTHGEMTGKIMIEFEKVLLKEKPDLVLVYGDTNTTLAGALVAAKLKIPVAHVEAGLRQDMPEEINRIVTDRVSQILFCPSK
LAVKNLEREGIT-EGVYFVGDVMYDLFLKMEDR-FILVTLHRDFNLRKILEQLKRIS--
KKVVFPIHPRTKNRVKEFGLEGMVVIDPVDYLNLMGLVKRCWKVVTDSGGLQKEAYFAGKRAIVIMPDTGWRELVEAGW
NKLASEE-NLFDVTMEEDCSEY--PSGLYGDGNAAGKIVEVI
A_EURTA_6 -KIAIILGTRPEIIKLSSIIRELQNFIEYFIIHTNQHYSKNMDEIFF----KELNLPTP--KYNLNV-----
GSGTHGEQTAKMIDGIEKILISENPDVVIVQGDTNTVLAGALSASKLHIKVAHVEAGLRSNMPEETNRVLTDHISNYLFAPTE
IAKHNLLKEGIN-
KNVFVVGNTIVDATIQNIEIEYFLLTLHRAENLTNIVNAIIKATYNKKIIFPMHPRTEKKLKEYNLSKIEIIEPVGYLEFLLLEKNAK
LILTDSGGVQEEACILGAPCITLRDNTERPETIYIGSNILVNADNKILDGIGVMANKKIN-GNNPFGDGNSGKAIVKIL
A_CREEOTA TKVVSIVGARPNFVKLAAVAETFDREFEHTVIHTGQHYDYEMSKVFF----EQLRLRDP--DIHLGV-----
GSGSQGYQVGEIVKKAEEHLKSINPDVVVVYGDTNSTLAGALAAAKAGYPVAHVEAGLRSKMQEEINRRVVDHVSALLFA
PTPSARDNLLKENVP-GKVFLTGDVHVDVLQKWIESQYVMATIHRAENLAEVIKCLKELA--
TKVVFPIHPRTRSRAEEAGLPNLVATKPLGYVDFLRLVTGSKLVVTDSGGVQREAYLLGKPAIVLRDRTEWIELVQAEWVRL
ADANARLVQEYKKLLEKPPAPQPGLLGDGKAAERITKTI
A_KOREOTA -RLIISVGTRPEIIKMAPLYEALSKLIELLLVHTGQHYDWEMSGIFF----KELGVEEP--DINLNI-----
GSDDQVSQTSAVMKEIGKIIESYEPDGVIAVGDTNSVLGTAIAASKMEVPFIHVESGLRSSMPEEINRRISDHLASLNFAPTS
RAFSNLMEEGISPKTTFLSGNTIVDSVIKVLRKSLVTLTLHRKENLIGVLKALEEL-
DDITFVWPIHPRTQKALKTFDLKNVRILEPLGYLDFLGLLISSDLVMTDSGGVQEEAATLKRPCIILRENTERPEIIEMGFGEI
AGTNTAIISLVRKYLYQANLITPNPFGDGSSSKIIASVI
B_BACES_3 KKIMLVFGTRPEAIKMAPLVKEFQKYFQTIVCVTGQHRE--MLDQVL----HLFHI-RP--NYDLNIM---
KQGQDLYDVTARVLIGMRDVLAEVRPDLLLVHGDTTTSTAAALSAFYQQIPVGHIEAGLRTPWPEEMNRQVTGRIATYHFS
PTLLSRQNLLNEGVKGDFIIVTGNTVIDSIYMVVDRRLVLITGHRRENFISMCQAIKALSPNVDFVYPMHPNVRKPIHEVFGN
NLFFIEPLEYLSFVYLMEKSIIVLTDSGGIQEEAPGLGKPVLVMRDTTERPEALEAGTVKLVGTNDKIVNEVSALLDDSDYYA
VNPYGDGKACSKIVNFL
B_BACES_8 KKILLTFGTRPEAIKMAPLVKELQEKFEILVCVTGQHRE--MLDQVL----HIFDI-SP--DYDLNIM---
RQGQDLYDITSRVLLGMRTILIETKPDIVLVHGDTTTSTATALAAFYQQIPVGHIEAGLRTPWPEEMNRQITARIATYHFAPTIL
SKQNLINEGVKEDFIFVTGNTVIDALYWVVNKKIVLITGHRRENFIHMCHAIKTLAPDVDFIYPMHPNVRKPINDVFGINIFFIE
PLEYLSFVYLMDKSFLVLTDSGGIQEEAPGLGKPVLVMRDTTERPEALEAGTVKLVGTDYKIVKEVSKLLESKTYYAVNPYG
DGCACSRIVTLL
B_BACES_5 KKILLVFGTRPEAIKMAPLVKALQRDFETKVCVTAQHRQ--MLDQVL----EVFDI-IP--DYDLNIM---
APNQDLYDITTKVLLGLRDVLKDFCPDTVLVHGDTTTSMAASLAAFYRQVAVGHVEAGLRTPWPEEMNRQVTDRICTYYF
APTGKSKQNLLQENIDAKKIFVTGNTVIDALLMAVDISYILVTGHRRENFLHICKAIRELAPEMDIVYPVHPNVQKPVYELLSD
NVYLISPLDYLPFIYAMQHSTLLLTDSGGVQEEAPSLGKPVLVMRNTTERPEAVEAGTVKLVGTDEAIVSNVTELLRNKELY
THNPYGDGHACERILSAL
```

```
B_GAMMA_4 -KVLTIFGTRPEAIKMAPLVKVLATDIESKICVTAQHRE--MLDQVL----ALFEI-VP--DYDLNLM---
RPGQDLFDVTSGVLLGLRDVLNDFKPDLVLVHGDTATCLGATLAAFYMQIPVGHVEAGLRTPWPEEANRVLVSKLATWHF
APTQRNKESLVQEGINPERIFVTGNTVIDALQWVVKKQFVLITGHRRENFENICSALKKLAPETHFIYPVHPNVQTPVNRLL
GNNVHLIKPLGYEAFVYLMQHSYLVLTDSGGIQEEAPGLGKPVLVMRDTTERPEAVEAGTVKLVGTSDSIVRELQALLNNE
SKYAHNPYGDGNACDRVLAVI
B_DELTA_7 LNVMFVFGTRPEAIKMAPLVKELQKQFKTIVCVSAQHRQ--MLDQVL----QLFDI-RP--EYDLDIM---
KPGQDLFDITSNVLLGLKPVLTKEQPDILLVHGDTTTTMSAALAAYYCRIPVGHVEAGLRTPFPEEMNRRTAGALTDLHFAP
TETARRNLLSEGVADTSIFVTGNTVIDALFSVIGENLILVTGHRRENFESICRAIARIAPDVEILYPVHPNVREPVRRLLGSNV
HLIEPVDYLPFVYLMAQSYLIITDSGGVQEEAPSLGKPVLVMRDTTERPEAIGAGTVKLVGTRESIYREAARLLDAADAYAL
NPYGDGRAAERIAKCL
B_GAMMA_2 KKVLIVFGTRPEAIKMAPLVQQLCQDFVAKVCVTGQHRE--MLDQVL----ELFSI-TP--DFDLNIM---
EPGQTLNGVTSKILLGMQQVLSSEQPDVVLVHGDTATTFAASLAAYYQQIPVGHVEAGLRTPWPEEGNRKLTAALTQYHF
APTDTSRANLLQENYNAENIFVTGNTVIDALLAVREKHLILVTGHRRESFERICQALITTAPECQILYPVHPNVREPVNKLLKS
NIVLIEPQQYLPFVYLMDRAHIILTDSGGIQEEAPSLGKPVLVMRETTERPEAVAAGTVKLVGTNQQICDALSLLLTDPQAYA
HNPYGDGKACQRIADIL
B_GAMA_11 KKVLTVFGTRPEAIKMAPLVNALASDFEAKCCVTAQHRE--MLDQVL----ELFEI-TP--DYDLDLM---
KAGQTLNEVTAGIIQKLKPVLQEFKPDVVLVHGDTATTFAASLAAYYEQIAVGHVEAGLRTPWPEEANRRLTGVLTKYHFAP
TETSKQNLLKENFEPSNIIVTGNTVIDALLMVKEKDLILVTGHRRESFERICEALATVASDVQILYPVHPNVREPVNRILGDNI
FLIEPQQYLPFIYLMDKAYIILTDSGGIQEEAPSLGKPVLVMRDATERPEAVEAGTVKLVGTDKRIVSSLNELLCDDLAYAHNP
YGDGKACHRILKAL
B_GAMMA_5 IKVLSVFGTRPEAIKMAPVIETLKNDFDSRVCVTGQHRQ--MLDQVL----ELFEI-VP--NYDLNIM---
KPGQDLTDVTTGILQGLRDVFSQFKPDYVLVHGDTATTLSTTIAAYYHQVKVGHVEAGLRTPWPEEGNRKVTGSLANLHF
APTSTSQNNLLAENIPADTIVVTGNTVIDALFMVRDKNVVLITGHRRESFERICQAVSELAIDVDFVYPVHPNVREPVNRFLS
PNIFLIEPQDYLPFVYLMDRSDIILTDSGGIQEEAPSLGKPVLVMRDTTERPEAVEAGTVKLVGTDSRIINEVSILLTDKQAYA
HNPYGDGAASQRILDAI
B_GAMMA_3 NRFLVVFGTRPEAIKMAPLVEGLKKNLNFRICVTAQHRE--MLDQVL----ELFNI-IP--DYDLDIM---
NQTQTLSTVTSSILEKIQPVIDDYKPNVIFVHGDTATTLAASLAAYYNQIDIAHIEAGLRTPWPEEGNRKLTAALAKYHFTPTQ
ATKENLLRENIDPSKIYITGNTVIDALFLVNKKEVILITGHRRENFENICAAISSLAPDVQFVYPVHPNVREPVNRLLKNNIHLI
QPLDYFSFIFLMKNAYLILTDSGGIQEEAPSLGKPVLVMRKTTERLEAVQAETVKLVGTTELIIDSVNELLTDIEAYAHNPYGD
GRAVERILNVF
B_GAMMA_6 NRFLVVFGTRPEAIKMAPLVEGLKKNLNFRICVTAQHRE--MLDQVL----ELFDI-IP--DYDLDIM---
NQTQTLSTVTSSILEKIQPVIDDYKPNVIFVHGDTATTLAASLAAYYNQIDIAHIEAGLRTPWPEEGNRKLTAALAKYHFTPTQ
ATKENLLKENIDPSKIYITGNTVIDALFLANRKEIVLITGHRRENFENICAAISSLAPDVQFVYPVHPNVREPVNRLLKKNIHLI
QPLDYFSFIFLMKNAYLILTDSGGIQEEAPSLGKPVLVMRKTTERPEAVQAGTVKLVGTTELIIDSVNELLTDIEAYAHNPYGD
GRAVERTLNVF
B_BETA_6  KKILLAFGTRPEAIKMAPLVRMLQTRVDARVCVTAQHRQ--MLDQVL----ALFDI-AP--AYDLNVM---
RQSQTLADVTTGILQTIGAVFDDFDPDIVLVHGDTTTTLAVSLAAFYRYLPIGHVEAGLRSPWPEELNRRVTDAVSSWHFAP
TERAQHNLFSEGVPTQGVVLTGNTVIDALQDVKRMDVVLITGHRRESFAHFCDALRTLAPGVRFVYPLHPNVQGPAHALL
DPNVHLIAPQEYLSFVFLMSRAHFIITDSGGIQEEGPALGKPVLVTRDTTERPEAIQAGTARLVGTDERIVGEASRLLDDDD
AYATNPYGDGHASERIAHAL
B_GAMA_10 MKVLSIFGTRPEAIKMAPLVRALAAEIDSRICITGQHQS--MLQQVL----DMFEL-KA--DYSLDVM---
RPDQTLNSLTAALYAAIDPILDEMKPDKVLVHGDTTSAMVAAMSAFHRRIPIGHVEAGLRTPWPEEMNRRCIDLISDHLFAP
TAESRRNVLGERLQ-
GISFVTGNTVIDALHLTAQRDVLVVTGHRRENFLNICKALGELADDIQIVYPVHPNVLGPVTEHLGPNVHLIKPLDYLSFVRL
MQRAHVILTDSGGVQEEAPSLGKPVLVMRDVTERPEAVAAGTVRLVGTEDAIIRGVNALFDDDALWAANPYGDGKASARI
VDAL
B_ALPHA_6 KKVLCIVGTRPEAIKMAPVILALKQQAVVHVLATAQHRQ--MLDQVL----NFFGI-VP--DSDLNVM---
RPNQDLTELTARLLVGAGEVLDRERPDVVLVQGDTTTVMAVAIAAFYRRIALGHVEAGLRTPFPEEANRVIAGRLARWHFA
PTEGAKQNLLREGVSEDRIFVTGNTVIDALLNTAER-
IVLITAHRRENIRRICAALKELAPDVHLVYPVHPNVKEVAETMLAQNVQLCAPLDYAPFVSVMRRSHLIISDSGGVQEEAPAL
GKPVLVLRDETERPEAVAEGVVKLVGTDDYIVSQVQLLLDDADAYGVSPYGDGMASQRVIEIL
B_BETA_5  -KILIVVGTRPEAIKMAPVILALKKEANVRVLATAQHRN--MLDQVN----EFFGI-DP--DIDLNIM---
RPNQALTTLTARLLPELDDVLQAEKPDAVLVQGDTTTVMTVALACFYHHIPVGHVEAGLRTPFPEEANRVITGKFARWHFA
PTEGSRQNLLKEGVADSKIIVTGNTVIDALLMSASK-
LVLVTSHRRENFRNICRALQTLAPDVQFLYPVHPNVKDVAHEFLAHNFTLCEPLDYASFIAAMKRAYIILTDSGGIQEEAPAL
GKPVLVLRDETERPEAVEQGVVKLVGPNDAIVQETQCLLDDEFAYGISPYGDGKAAERIVQVL
B_DICES_2 KKISLVFGTRPEAIKMAPVAKIIKESFDLQIILTAQHRE--LLDQVI----EIFGL-KS--DYDLNIM---
QEKQTLTHITVKVLQGLDIIWQKDPPDMVLVHGDTTTTFAASLAAFYKKLPIGHVEAGLRTPYPEEMNRHLTGVLADLHFAP
TQRAKDNLINERVPKENIFITGNTVIDALLFVHRNNFILVTAHRRENLKNIVLALDEILEDFYVVFPVHPLVREQVYSVLKKRAI
LIPPVDYVTMVYLLDKCYLVLTDSGGLQEEAPSLGKPVLVLREVTERPEAVEAGTVKIVGTSESIVREVRRLILNKEEYAINP
YGDGKASERIRDIL
B_THEAE_3 IRVLSVFGTRPEAIKMAPLVKKLEEEVESLVCVTAQHRQ--MLDQVL----EVFDI-KP--DFDLNIM---
KERQSLADITVNALSGLYDLIEELKPDIVLVQGDTTTTFAGALAAFYHRIPVGHVEAGLRTPFPEEINRRLTGVLSTLHFAPTK
RNRENLLRENVM-
GKIYVTGNTVIDALRYTVKEHYILLTSHRRENLENICKAVRRIVEDVRVIYPVHPAVREIVFPMLEERVFLIDPVNVIDMHNLM
ARSYLIMTDSGGIQEEAPALGRPVIVLRRETERPEAIEAGVAVLGGVEERIFELAKKLLLDREEYAVNPFGDGRASERIVKAI
B_GAMMA   SKVLFVFGTRPEAIKMAPLVIEFKNNIEVKVCVTGQHRE--MLDQVL----DFFEI-EP--DYDLNIM---
KQKQSLGSITCSILTRLDEILASFMPAHIFVHGDTTTTFAASLAAFYQNIKVWHIEAGLRTPFPEEGNRQLTSKLAFFHAAPTL
QAKDNLLRESVKEKNIIVTGNTVIDALLIGIKKTIILVTLHRRENLRTICDDIKQLADDIEIVFPVHPRIREVVNEKLSVNIKLVEP
LAYPGFIWLMNNAHFILSDSGGVQEEAPSLQKPVLVARDTTERPEVIENGAAMLVDPRNNIYSSCKKLLSDERLYAGNPFG
DGKASKKILDYF
```

```
B_CYAIA_2 PRVTIVLGTRPEAIKLAPVIQEFQACLETRVVLTGQHRE--MVTQVM----DLFSL-KA--DQDLNLM---
APRQTLTHVTCAALQGLRDDFQAYPPSLVLVQGDTTTAFAAALAAFYEQIPVGHVEAGLRTPFPEEANRRLISQVAQLHFA
PTQQSESNLLASSVV-
GKVMVTGNTVIDALLRMAERPVILATVHRRENLQSIAEGMLQVLPDTALLLPLHPTVREPLQALLGPRVVLTEPLDYDRLVA
AMKGCTLLLTDSGGLQEEAPALGKPVLVLRRTTERPEAVEAGTARLVGTDGTIAEEAHRLLSDPQAYAVNPFGDGQASARI
LEAA
EU_Paella PLVSIVLGTRPEAIKLAPVIMSFQKCLRTRTILTGQHRE--MVNQVM----KLFQL-SF--DKDLALM---
EPNQTLTHLTATILEGLRKEFLVHRPSLVLVQGDTTTALASAMAAFYEQIPVGHIEAGLRTPFPEEGNRCLISQITKLHFSPTI
QSAANLQASGVI-
GEIHITGNTVIDALLSMSQKSVILATVHRRENLNGIAYGFFKLLSDVALLLPLHPAVRKPLNDILKPRAFLTEPLAYDQLVAAM
RSCSLVLTDSGGIQEEAATFGKPVLILRRTTERSEAVEGGTARLIGTDTDILNEASLLLKDDSAYFNNPFGDGKASDRILTAA
B_CYAIA_3 IRVAVVLGTRPEAIKLAPVIQQLQLSFETHVILTGQHRE--MVKQVM----QLFDL-DA--DYNLEIM---
QRSQTLTDITWRSLQGLEGLFQQLRPDIIIVQGDTTTAFAAALAAFYQKIPVGHVEAGLRTPYPEEANRRLISQLTQFHFAPT
TQAVDNLRNSGVL-
GEVHHTGNTVIDALLSVVKCPLILATVHRRENLTQIAESFLKILPDTALLLPLHPTVREPLQKILGPRIFLTEPLDYAELVAAIDR
CYFVLTDSGGLQEEAPSLGKPVLVLRKTTERPEAITAGTAKLVGTEAEIVEAATELLNNINTYAINPFGDGHAAEYIVQII
B_CYAIA_5 KRISVIIGTRPEAIKFGPLILAFLKTIDLRIISTGQHYE--LVDQVN----ELFKI-VP--NKNLKIM---
VPGQSLTKITNEVLIGLKEDFNEYPPDLVLVQGDTTSAFSAALAAFYEKIPIGHIEAGLRTPYPEEANRRIISQIASIHFAPTKIA
FENLKKESVL-
GEVYLTGNTVVDSLLFISEKQIILATVHRRENLKQIAKGLKKILLDYILILPMHKSLREPLEEILGERAILTESLSYNSLVGTLKH
TKLLLTDSGGLQEEAPTFGVPVLVLRDSTERPEAIKAGTAKIVGSNNKIFKEANNLLTNQKEYAINPFGDGKASERIVKYC
B_DELA_10 -RILIVFGTRPEAVKLAPLIHDLRQRADVEVCLTGQHRE--MVTQVV----SFFGV-DV--DHDLEIM---
RPNQTLSDVAARTLTGVDRILESRRPDWVIVQGDTSTCLATALAAFHRKVRVAHVEAGLRSPFPEEMNRVLTTPIASLHLA
PTSRAKANLRAERVPEDRIRVVGNTGIDALLLAVQTRLVLVTGHRRESFEELCEAIRDVADDVDVVYPVHPNVREPVFRILS
SNVHLVEPVEYPALVWLAQRSRFILTDSGGIQEEASALGKPVLVMRDVTERQESVEAGVSRLVGTSEVIREACNSLLRDEA
TYRVDLYGDGKASARIGDAL
B_ACTERIA LRVGIVYGTRPEAIKLAPLVLALDADFEPVIITTGQHRD--MLDEIN----ELFGL-RP--RHNLDIM---
RPGQRLSAMASRIVGELGDPLLDELVDVAVVQGDTSTAFAAAYAAACERIPVAHLEAGLRTPFPEEINRRLITQLADLHFAP
TADAAGNLLAEGVRSDDVYVTGNTVIDAMHLVLDRGTVLLTMHRRESMGRVAAAVAELCPTLRFVIPLHPEVRRVFRSHLS
SQVLLCEPLRYSEFIRLMHRAVLVLTDSGGVQEEAPTLGKPVLVLRDRTERPEGIAAGCARLVGTDALIVKEVGRLLDDPEA
YGIFCYGEGDAAARCLEAL
B_ACTIA_2 LRVGIVYGTRPEAIKLAPLVLALDADFEPVIITTGQHRD--MLDEIN----ELFGL-RP--RHNLDIM---
RPGQRLSAMASRIVGELGDPLLDELVDVAVVQGDTSTAFAAAYAAACERIPVAHLEAGLRTPFPEEINRRLITQLADLHFAP
TADAAGNLLAEGVRSDDVYVTGNTVIDAMHLVLDRGTVLLTMHRRESMGRVAAAVAELCPTLRFVIPLHPEVRRVFRSHLS
TQVLLCEPLRYSEFIRLMHRAVLVLTDSGGVQEEAPTLGKPVLVLRDRTERPEGIARAAP---
GWWPTRHSSSKRSADCSTTPGVRASSATAASAGCLPPT
B_ACTIA_6 RTVLLVYGTRPEAIKMAPVVHELGRTLRPVVAVTGQHRS--MLDQVN----DLFGI-VP--DYDLGIL---
RDRQSLAGVTTRSLGGLESVMAEVRPDVVVVQGDTTTAFTGALAAFYQGLPVVHMEAGLRTPFPEEINRRLTSQLADLHL
APTPSARANLLAEGIRPESVLVTGNTVIDALLHVTAAPVLLVTAHRRESLARVGAALARLAPELLIVLPVHPVVRETVLPPVE
PNILVADPVDYAAFAHLMKAATVVLTDSGGIQEEAPSLGKPVLVLRDNTERPEGVQAGTACLVGTENRIVAAVDRLLDDPVA
YAINPYGDGQAARRTVAAI
EU_Cow_2  TRVLSIFGTRPEVIKFYPVLKEMDRRLLSITCVTGQHRQ--MIDPLL----SLFDI-AT--DIDLNLM---
TQGQTLSNLSARLFTALAPVFEAVRPHLVLVQGDTTTAMIAAMCAFYYRVPVGHIEAGLRTPFPEEVNRRIVSVVGNLHFAP
TNYAAEALRAERIDMQTVFVTGNPVIDAILLIREKPHILLTAHRRENIAAICRAVKRIIKDVNVWYPVHPAVCVPVRQELGERV
HLLDPLEYDVFVHFLDRMYMVMTDSGGIQEEVTALAKPTIVLRETTERPEGVDAGVTKLVGIHDTIVKEANVLLTDEQTFKC
FPFGDGTAAKQIVDIM
EU_Cow_3  TRVLSIFGTRPEVIKFYPVLKEMDKRILSITCVTGQHRQ--MIDPLL----SLFDI-AT--DVDLNLM---
TQGQTLSNLSARLFLALAPVFEAVRPHLVLVQGDTTTAMIAAMCAFYYRVPVGHIEAGLRTPFPEEVNRRIVSVVGNLHFA
PSSYAADALRAERIDMNTVFVTGNPVIDAIMLIREKPHILLTAHRRENIAAICRAVKRIIKDVHIWYPVHPAVCVPVRQELGER
VHLLDPLEYDVFVHFLDRMYMVMTDSGGIQEEVTALAKPTIVLRETTERPEGVDAGVTKLVGINDTIVKEATLLLTDDQTFK
CFPFGDGTAAQQIVDIL
B_ALPHA_5 RHVMVIVGTRPEAVKMAPVIKALYARFRCTVVSTGQHRE--MLRQTL----TSFGL-SV--DVDLDIM---
QPDQTLASLTGAAIGACEKIFQAARPDVVLVQGDTTTVLSASLAAHYAQIPVGHVEAGLRTPFPEEMNRKLVTSLATLHFAP
TERSARQLRKEGVPPSKVFVTGNTVVDALEELRGRTFVLVTCHRRESLNVIVEAIAMLAPDRTFFFPVHPNVRALVMPRLN
ENVVLADPVPYADILFCLSSAELVLTDSGGLQEEAPSFGVPVIVLRRTTERPEGVRAGFSRLVPIEENIVSLASSWLRSHRK-
RPNPYGDGNAAARIVDIL
B_ACTIA_4 TEVHLVAGTRPEAIKLAPLVPALRAQMTPVFVASGQHPT--MVHQAL----DAFGL-EP--DVTLSID---
RGSGSQAELMAALTMKLEKHWQQRTPAAVVVQGDTTTVLAAAMVAFWAKLPIAHLEAGLRSPFPEEGNRKLVGQISRLHL
APTARARANLEREGTPAADIVVTGNTVIDAVLGIAARGLVLVTAHRRESLDRVLNAVALLLSDVEVVLPAHPAVAEQVRAVLG
PRVLVTEPLAYPVLVGALAASTLVLSDSGGIQEEAPSFGVPVIVLREVTERMEAVDAGCAILVGTDDAVLGNACRLLDDPDE
RKGNPFGDGRAAERSAAAI
EU_Oscusl IKIAIVFGTRPEAVKMAPVIQAVARSLSAILISTGQHKQ--MLEQVL----RQFSL-QDKIQHELALM---
KPNQQLAELTSSAVRAVDGVLRSSKPDAVLVQGDTTTAFITSLAAFYLKIPVGHIEAGLRTPFPEEVNRQCISVMATYHFAPT
EHAAKNLYDEGRR-TNVFTTGNTVTE--------PLLLLTAHRRENILNIFTSIEKLLPDVVVIYPIHP-
MSDHAPPTTTRRLLIVPPLDHADLLFMMKESFFVMTDSGGIQEEAVTLGKPVLVLRDTTERPEGVLAGAAKLVGHGESIYT
EAASLLKDPDSYSKKTYGDGNAAGNIVAKE
EU_Ossl_2 IKIAIVFGTRPEAVKMAPVIQAVARSLSAILISTGQHKQ--MLEQVL----RQFSL-QDKIQHELALM---
KPNQQLAELTSSAVRAVDGVLRSSKPDAVLVQGDTTTAFITSLAAFYLKIPVGHIEAGLRTPFPEEVNRQCISVMATYHFAPT
EHAAKNLYDEGRR-TNVFTTGNTVVDALYAILKTPLLLLTAHRRENILNIFTSIEKLLPDVVVIYPIHP-
MSDHAPPTTTRRLLIVPPLDHADLLFMMKESFFVMTDSGGIQEEAVTLGKPVLVLRDTTERPEGVLAGAAKLVGHGESIYT
EAASLLKDPDSYSKKTYGDGNAAGNIVRIL
```

```
EU_Cow    PLALVFVGTRPEVIKMAPLVHEMQENFDTIVVSTGQHRE--MLAQTL----RGFDL-KI--DLGLSVM---
QHNQNLGALFAAVVAQSTAVIASLRPEVVLVQGDTSTALACATAAYYERVPVAHVEAGLRSPFPEEMNRRAIDSFATFHFAP
TEYAAEAIRRENTCARNIVVTGNTGVDAVLARLNAPIILVTSHRRENLESICRAVKRIVPVVRVVFPVHPNVQKVVFAHLKDR
ITLTDPLDYDVLTYLLQEVTIVLSDSGGLQEEAVSIGKPVVLLRQTTERPEGVYAGTVRLAGTNDRVVSDTQNVIDNLVEYG
QNVFGDGAASGRIANVL
EU_Naeria PRVVIVLGTRPEAIKCAPLISHLKSNLELIVLSTGQHSE--ILKQTL----AVFDQ-HV--DIDLELM---
EPNQSISSFFTLAYQRISLEFERQGPSIVLVQGDTTTSLAAALAASYLKIPVGHVEAGLRSPFPEELNCKMIDSFARLLFAPT
DFSKAALVREGICPQHVHVTGNTGIDSFYSLMSVPVILVTMHRRENFKEMCNAIKTISKNVMIILPVHPNAKSVVNEVLSDN
VQLVDPIAYDIFPHVISEADIIVTDSGGIQEESASIGKPVILMRDTTERPEGIYIGTIKKIGVNFHIVKAMTDAIKDSQNSSKHIF
GDGKASMRISKIV
EU_Naia_2 PRVVIVLGTRPEAIKCAPLISHLKSNLEVIVLSTGQHSE--ILKQTL----GVFHQ-HV--DIDLKLM---
TPNQSISSFFTLAFQKISLEFEKQGPSMVLVQGDTTTSLVAALAASYLKIPVGHVEAGLRSPFPEELNRKVIDSFASLLFAPT
EFSKAALVREGICPQHVHVTGNTGIDSFYSLKSLPVILVTMHRRENFKEMCNAIKTISKNVMVILPVHPNAKSVVHEVLSDN
VQLVDPIAYDIFPHVISEADIVVTDSGGIQEESASIGKPVILMRDTTERPEGIYIGTIKKIGVNHHIVKAMTDAIKDSQNSSKHIF
GDGKASMRISKIV
EU_Naia_3 PRVLIVLGTRPEAIKCAPLISHLKSNLEVIVLSTGQHSE--ILKQTL----GVFHQ-HV--DIDLKLM---
TPNQSISSFFTLSFQKISLEFERQGPSMVLVQGDTTTSLVAALAASYLKIPVGHVEAGLRSPFPEELNRKMIDSFASLLFAPT
DFSKAALVREGICPQHVYVTGNTGIDSFYSLKSTPVILVTMHRRENFKEMCNAIKRISKNVMVILPVHPNAKSVVNEVLSDN
VQLVDPIAYDIFPHVISEADIVVTDSGGIQEESASIGKPVILMRDTTERPEGIYIGTIKKIGVNLDIVKAMTDAIKDSQNSSKHIF
GDGQASMRISRIV
EU_Spello ------MGTRPETIKMFPIIRALNTHINPIICVTGQHQE--MVEPLL----RLFDI-KP--HVNLNVM---
KSGQSLNSLTERIVGRMTKAVTHYRPDWLLVQGDTTSAFVASVVAFHEKIAVGHVEAGLRTPFPEETSRRLIAGLASYHFV
PTQHAANNLKAEGIPSENVYITGNTVIDALQWVASLPLLLVTMHRRENLISVCKAIKRIVPDVHIVLPVHPTVQETVTQFLGK
RITLLGPLSYEFFAQLLKRATLLLTDSGGVQEEGTAFSKPILVLRNDTERPEGVTAGVAKLIGTDDNVFRHVQELLTDKQAYK
TFPYGDGTAGSKIVDIV
B_CHLROBI KKLLLAAGGKPGALLLAPLYNALKKNFKPVAVFAAAQSAEPLSRELA----TCFGIGEA--GHTITL-----
GEDSPGQQLAAVITGMEPIIAREQPVLVMVCGSDNAALGAALAATKLGVPVAVADAGLRCSDADEINRILIDSIADLHFISEH
SGEYNLINEGVADEKVFFSGNLSIDTLAVLMEQEYALVLLNPAGALEMTLRLLKKMS--
TTVLVPLAAGLDELMKQHKLEGLVMIAHPGHAGLLTLLRDSVLLITDSEELQAESTVMNVPCLTMMDNTARPATIEIGTNVLV
GVDEDIMSRIHDILHSGEHAKIPEKWDGAAASRIVEVL
EU_Phella KAVRKLKLMKPEIIRLSLILPKLDRYERHVLVHTGQNFAERLSGIFF----SQLGLRKP--DYVLQD-----
SAASLGTQLSAMFSQMEDILEREKPDTILLLGDTNSALC-
AILAERLGYPVVHMEAGNRC---------------------------------------------------------------------------------------------------------------------------------
-----------------------------------------------------------------------------------------
A_EURTA_3 FVLAVVTATKPDFYKQAPVVAAAEANVPCFVIHTGQHYDDVLGHGL-----EEYGI-EPHIAADLGIR------
GDLTQKTAEMMLAVKTLAERLEPDTTMVHGDTHAAGVFPQAMFATNQQVAHNEAGLRGPFPEQYDTFVGSAASIYQFAP
VALNREHLEREGYPHERMPVVGNSVVDAIEMKRDHGWLRVDIHRRANFTAIVEGVIELVYNVNFVELNA--
TKRALENYGY---
RLADERENFLFTGLWKKHAHVFTDSGSMQEELNEIAALCLTARFSTDRPETVDANTNLLVPPTAFVADMITHVAETDDVRG
QALYGD--VGEEIVEFL
```

ROK domain alignment, phylip format

```
 112 225
A_CRERC_2   LfLgIdIgAtWtRaLlIdEhGdIivKtvGvGsIgPlDlKsGwVvNSpNsPtRrFpLvEpLgKpIvVaNdCvAaVwGeYvFkY--
nMvYiTlSTgVgVgAiVnGnlLlGkDgNaHeLgHaViRrCgCgGrGHfEAyVgGaHikArKfInLwLdAlAaGiAtVvaAyDpElLiVgGSiDiIsR
ElPaRlRdYlSl--RePeIlKasFGdDeVaVgAaAlAyKtPdTl
A_EURY_2
IAGAADIGGTNTRVGLIREDGKIVFPTLGIAAAGPLNIREGILDHPPNIPFDFVPIVAPLNLPVIFQNDCRAAVLGEVCAGGAR
TVVYITISTGIGGGICTNGKVITGRGGNAGEIGHFPVLTCTCGLSGHWEGYASGRGIRYAGFRDALAQVNGRGLSSVIVAYD
PDCIILDGTVPDLLDQALVYTDRYL----DLPPCIFSPLNGDAPLIGAAMAVFHPKMI-
A_THAEOTA      YKLGVDLGGTKTEAILLDDSLNVLVPTLGICTPGAISKKTGLIKN-
SNTCLIGKSLKEDLKKTIVMENDANCFVMAESKMGAAKLVFGVIMGTGVGGGITVNGKLHSGRTNIAGEWGHHTLNPCYC
GKTGCVETYISGPALIGKTWKSEFLENFGYSLANVIDILDPDAIVLGGGLDFLYTEGKKSVYEKVFSDLVDTPILKNELGDSA
GVYGAALLN-------
B_ALPA_19      LRLGLDLGGTKTEIIALDDEGRILRPSVGVAMPGAISPASGLVKN-
ANSWLNGQRLDHDLGRPVRVANDADCFALSEATDGAAASVFGVILGTGVGAGIVVNGRLLAGPNAIAGEWGHMPLPDCY
CGLKGCVETFCSGPGLQAQASLDRHADRLARALAVVINILDPQVIVLGGGLPHLYQALPRLWTPWVFSDRVDTRLVAPRH
GDSSGVRGAAWLWPPSSTLE
B_BETA_2       LRFGIDLGGSKIELIALDRDGREIVPTVGVGTPGAVSPASGRMKN-
CNSCLNGQPLREDLGREVRVANDANCLALSEATDGSAASVFGVILGTGVGGGVVVHGRLLQGANAIAGEWGHSPLPACY
CGKKGCIETWLSGPGLPCSATLARYEERLARALAGVINLVDPDVIVLGGGLARLYDTVPRLWPRYVFSDRVDTKFVPPKYG
DSSGVRGAAWLWE------
B_ZETA         -RIGIDLGGTKTELIALDKTGAARRPTIGIGTPGAISPRTGRMKN-
CNSCLNDQPLQQDLRRPVRLSNDANCFALSEAIDGAAAVVFGVILGTGVGGGIVVNGQLLEGVNSIAGEWGHNPLPPCY
CGRKGCIETWLSGPAMACMLTLERYCERLARALAGVINILDPDAIVLGGGLSLLYEQVPRLWSSYIFSDSVSTRLLPPAHGD
SSGVRGAAWLWGR-----
```

```
B_DELTA_2     LRLGVDLGGTKIEIIALDDGGNELEATVGIGTPGAVSRATGLLRN-
SNSCLNGKPIARDLGRPVRIANDANCFALSEYRDGAARVVFGAIVGTGTGAGIVVGGEVLEGVNAIAGEWGHNPLPRCYC
GRYGCIETFLSGPGMSCEHTLLRYEDRMARSLAHVINILDPDVIVLGGGMRRLYRNVPRLWGAYVFSDRVDTRLVPPRFG
DSSGVRGAAWLWG------
B_ALPHA_2     IRFGVDFGGTKIEVAALNAAGDFVKPNLGLGIPGSISPRTGLIRN-
ANSYLNGRPFGENLARPVRIANDANCLALSEAADGAGAVVFAAIVGTGCGGGVVVDGKIINGHNGIGGEWGHAPLPDCW
CGRKGCLETWIAGPAFKASAALDRYVDRLARSLAVVCDLIDPDIIVLGGGMDALYERLPAAIAPHVFSDIFETPVRKAVHGD
SSGVRGAVWLWPPEV---
B_GAMA_20     MRMGVDLGGTKIELVALSDEGNELVTTVGVGIPGVVSPYSGLVKN-
ANSWINGHPLDVDLQREVRVANDANCFAVSESVDGAAAVVFGVIIGTGCGAGVAINGKVHGGGNGIGGEWGHNPLTRCF
CGNPDCIETFISGTGFIAVAAFDRYMDRLARSLAHVINMLDPDAIVLGGGMAAIYPRLPALLAHYVVGRECHTPVVQNLYGC
SSGVRGAAWLWEK-----
B_GAMMA       -YYGLDIGGTKIELAVFNEKLEKLVPTVGLGVPGFVNQQTGLAEI-
ANIVADNKPILCDLGREVRAENDANCFALSEAWDTENQTVLGLILGTGFGGGFVLNGKVHSGQVGMAGELGHLQLYQCG
CGNKACLDNYLSGRGFSAVDFVNLFVELAAISIGNIITAFDPHMIVLGGGLDYLYEALPKALPPHLMRKAKVPPIKKAKHGD
SGGVRGAAALFLTK----
B_GAMA_10     -YYGFDVGGTKIEFGAFNAQLERVVATVGLGIPGMEDADNGCVLT-
VNVAAKGKPLRADLGRAVKVENDANCFALSEAWDDELKSVMGLILGTGFGGGLVYEGKVFSGRNHVAGEIGHMRLLGCG
CGNKGCMDNYLSGRGFKAVEHVERFMELLAICFANIFTANDPHVVVLGGGLDLIYEEMPKRVPKHLLSVAKCPKIVKAKHG
DSGGVRGAAFLNIK-----
B_GAMMA_7     -YYGFDIGGTKIALGVFDSTRRLQVPTVGIGIPGMPETEDGTLYA-
ANVAASGKPLRADLDRDVRLDNDANCFALSEAWDDEFTLVMGLILGTGVGGGLVLNGKPITGQSYITGEFGHMRLRRCG
CGQMGCIENYLSGRGFQAHAHVERYLDLLAVCLGNILTIVDPDLLVIGGGLTAITTQLAERLPRHLLPVARAPRIERARHGDA
GGMRGAAFLHLTD----
EU_Oryza      -YYGFDIGGTKIALGVFDKDLRLQVPTVGIGIPGMPETDDGTLYA-
ANVAASGKALRADLERDVRLDNDANCFALSEAWDDEFRLVMGLILGTGVGGGIVINGKPITGRSYITGEFGHIRLTRCGCG
QHGCIENYLSGRGFQAREHVERYLDLLAGRLSRFRAMDKPRVVVLTGGIDGLWEEHRVEDVATPEGFARDPDLVQAQLQ
QPEIAPNPAHLALAKLEEA
B_ALPHA_7     -LIAFDIGGSRIRAARVFAPDHLELPMVAISIAGVIDPASGRLTS-
ANLAVNGRRLAGDLGRPVWIGNDADCFVLTEARRGVARNVFGIILGSGVGGGLVLDGALVAGAGGITGEWGHGPVLRCG
CGQTGCLDTVGGARGIAAAETVSVWLELVSGPLAVVLNVVGSSVVPVGGGLRDLVAALDRAVRQRLLRPTTEPVLRPA-
IHPEPGLVGAALAGLETFG--
B_Alpha
AALAIDLGGTELRAALVDRDGKILVPTVGVGAPGPLDPLAGIAVGPPTLGWQDVPLADILGLPVRLENDANAAALGEWRFG
AGHSLVFVTVSTGIGGGVVADGRILHGRRGLAAEIGHMTIERCVCGVVGCFEAIASGTALLALALLEEEARWLGVGFTNLL
HLYSPDVLVVGGGIDLMHPVIEATIRQRAMRAYRDVPVVQAQLGRHAGLVGAASLVLFDDGSL
B_ALPA_14
VALAFDLGGTELRAALVDEEGNLLVPTIGIGAPGPLDPEAGVVIAAPTLGWNEVPLADILQLPVRLENDANAAAVGEWRYG
AGRSIVFVTVSTGIGGGVIADSRILHGRRGLAAEIGHMTIERCFCGAVGCFEAVASGTALLALELLNAEARWLGIGFTNLLHL
YSPDVLVMGGGIDLLHDTIMATICDRAMPAYRDVPIVAAQLGRHAGLIGAASLILGSEAQA
B_Alpha_2
VALAFDLGGTELRAALVDEEGNLLVPTIGIGAPGPLDPEAGVVIAAPTLGWNEVPLADILQLPVRLENDANAAAVGEWRYG
AGRSIVFVTVSTGIGGGVIADSRILHGRRGLAAEIGHMTIERCFCGAVGCFEAVASGTALLALELLNAEARWLGIGFTNLLHL
YSPDVLVMGGGIDLLHDTIMATICDRAMPAYRDVPIVAAQLGRHAGLIGAASLILGSEAQA
B_Alpha_3
VAIGIDLGGTQVRAALVDEQGRILEPTVGVSAPGPLDTVAGVASNIPTLGFVDFPLKAELPFPVDLENDAIAAAIGEWQFGA
GKNLVYYVTVSTGIGGGVVSDGRVVRGRKGMAAHVGHMSVELCPCGNRGCFEAYGSGTAFLANQLVDEEAEILGRGFTSL
IHIFSPDIIVMGGGLDRLQPGIQAYITQWAMPAFRDVKVVLAALDQNSGLVGAAALAFLTGKVP
B_Chldiae
CVIGIDLGGTKIGIGVLNVSGTLILKTIGIGVAGQIDEETGVVRFAPNLGWHQVTLRKNLEIPVKVVNDVRAITWGEWLYGAG
KDLICLFVGTGIGSGIVCQGKMQKGDNNTFGEVGHMTIPRCTCGNNGCFEAFAGGWGILALLILEKVKQALIAGCINLVNAF
NPACLILGGGVPEILSFIDKGIRETALKTADKLQIKTALLGKNVGIIGSGAVILDVLKNN
B_DELTA_4
YAVGVDLGGTKIAVALVDDRGEVLYLTIGIGVAGQIARDDGMVRFAPNLGWRNIPLGEQLRLRVVVVNDVRAAAAGEWAFG
AGKDLICMFVGTGIGGGIVAQGRMLHGCGNSAGEIGHVVVPLCHCGRRGCMEALAGGWAILARQLIDRAAEALSVGAVSL
VNAFNPCRLILGGGVPELIERVREGIRHHALVTAESVTVVPASLGDDAGVIGAAVLAMQSPA--
B_DELTA_9
PTLGIDLGGTFARAAVVDEVGKLIVALCGVAAAGQIHKDSGVLSVAPNLGWRNVPLGALLGQPVRVVNDLAAAAWGELHA
GAGRDMLVVFVGSGVGSAIIAGGRLVDGGGGVAGELGHIKVRRCGCGELGCLEAYAGGHNLAAGEVYARAAQFLALAVA
NQVTMLNPARLVLGGGVPGLRRRVEEGVRAWSSTTSEGLLIADAELGDDSGLIGAALLVK------
B_Cyaeria
KILALDFGGTKLAAATLISGQQKWRQLIGISFGGPVDAKTGIVRLSHHVGWENVPLREILNVPVRVDNDANVAALGEYKFG
AGQSLLYITVSTGVGGGWILNNRLWQGTQGMAGEIGHTVVPLCLCGKRGCVERLASGPYILCQNAIKLAGWAIGVGIGNA
ANLINPQRFILGGGVDLFWHQVRQISRQTALPE-VDFEIVSAQLGDEAPLWGAVALAETGLEN-
B_Fircute
FCIGLDVGHNRLKVVVMNLAGEIKIPTLGIGMPGLLDPDTGMVNFSPDFGWEHVDLLGRFGFPVLIENANRVMALGERWF
GAGRDFLCVNLGHGIGSALVFDGEIYHGNSGSSGEIGHITLPLCECGNHGCLEALASGRAIEAEDILDNAIEYIGIAIAGVVN
LFDPELIIFEGGLSYLLPRLKETIRRHQMHLARNVQLLKGNLGEDITAVGAATLLLQDLLNH
B_CHLROBI
WAIGIDLGGTAVKAAIVSRKKGILVPTIGFGAPGAVDIEAGTLSYPPNLGWTTFPLRSELSVPVVIENDANAAAYGEAVYGAG
RDFLMVTLGTGVGGGIVLNRKLYRGPNGTAGEIGFMIVPAVHAGIHGTIEGMIGKERILSLAVWNHVGAILGTGFACVTSLM
DIRKFVIGGGITLIFEPAYRQLLRSTLPSMDGLELVPAELGNSAGIYGAAALCFS-----
```

```
B_Cyaia_2
QVIGIDLGGTAIKLARFNRSGDVLIPTVGVGLPGPMDANARVARVCINLGWDDVPLADWLQRQVTLANDGNCAVVGEAWL
GAARDVVLLTLGTGVGGGVILGGALFTGHNGAAAEPGLIGIPPCNSGNQGSLEQFASIAALQAKAIWQQYGTTLGVGIASL
VYVFTPQRVLLGGGLAHFLPDVRREVEARVQAVSEGLSIEACALGNGAGRLGAARLALQRLSG-
B_Cyaia_5
QVIGIDLGGTAIKLGRFSEDGTCSVETIGVGTPGPADAQGRIAQIAINLQWENVPLADWLNKPTVIENDANCAGIAEAWLGA
GRNFIMLTLGTGVGGAIFLDGKLFVGHRGAAGELGLITLPMCKSGNPGSLEQYTSIKAIEALQFWQEYGQNLGIGLTSLIYV
LTPQAIVLGGGVEFFLPTLKAEIEQRVMSTSAGLQILPAELGNSAGMVGAARLAWKKFGNG
B_Aquicae -KKGIDIGGTFIKVLWEDGKKEKH---VGIAVAGFTSLN-
GIVYKSPNIALDKTDFRKVLNIPFVVGNDVNLGAFGEWYYDYR-
ILLLVAVGTGLGTGLVYKGEVFFGVCGSALEMGHHTVELCNCGRKGCWEAYCSSYAIEALKVVESFKDYLITGLMNGVHIF
NPDRIVLAGGVKEFLRDVPEKVKEISEELPLCLKISFSRAGEFMGARGALAYALKYL---
B_Aquae_2 RYLGIDIGGTFIKFVY--
KKGDDIVYILGVAVAGLIDKKTGVLTASPNIPLENFPFKDELKIPVYIENDASLAAYGEYLYGAGKILICLTLGTGLGGGAVING
KLLTGVSGSAMEIGHTTILPCHCGRKGCLESYVSSYGLTAMRSMERFSEYLSVGLMNIVHIFNPDRIVLAGGIPAVVDMAVS
NLKNIAFHLPRDLTVKRAELKEFSGAYGALGYAENESR--
B_FIRMIC
YYVAIDIGGTNIKYGLVDQEGQLLMPTVAISSAGMVDPDKGEIFYGPQINYAGTQFKKEITIPCEIENDVNCAGLAEAVSGSG
KVTLCLTIGTGIGGCLIMDRKVFHGFSNSACEVGYMHM------
QDGAFQDLASTTALICMEGIDRMVDYLGKGLANICYVANPEVVILGGGIAILKPKIRTALKEALVPSLEKTRLEFAHHQNTAG
MLGAYYHFKTKQS--
B_FIRIC_6
EYIAFDIGGTQIKYGIVSETGTVLVPTIGISTAGIVDVNKGIVTGADHIGYSTIPIINRLKVPVSIDNDVNCAAFGEKWNGSGRN
FIMLTLGTGIGGAIFIDGELYRGHSFSAGEWGNMLI-------
EGAFEEVASISGLEVTQAVEVFFKHLAIGISNLAYIFNPEMIVIGGGINQFLKEVKEEVEKYLQKEINNCEIELAQNGNCAGMI
GSIYHFLHHHK--
B_SPIRO_2 HYLAIDIGGTSTKYSLSDSSGVFFISTVAICIPGFVDLKGNVLRV-
NAIGFVNYPLKERLGVSTEIENDANCVALAEKFKGNAINFIAITLGTGIGAGIFANGKLLRGNSFMSGEVGFMITFNC------
KWESIASVSALHAKNEVDRFFENLSFGIFNLTFILNPEKILIGGGIPDLIDRIYEKLENLWSLEMNLVALEPTKFNNESGKIGAL
FHYFTCKNQN
Brancoma1
SALAVDLGGTNLRVAIVGQQGEILEPTIGISTGGRVNPYDGVVLHTEILGWNSIDLRTPIHLPVWVDNDGNCAALGEKKFGK
GLDFITLIVGTGIGGGIFLNNELVHGANFCAAELGHISVPDCTCGSSGCVESYASGLALKAEKVVERACEALGSAICTLLHTV
NPSHIILCGHLPHYVDGVREVIQRRALPSANNIQVMVSDLEE-PALLGAASIVLEFATRR
Brancoma2
SALAVDLGGTNLRVAIISQQGEVLELTIGISTGGRVNPHEGMVLHTKILEWSSIDLRTPIHLPVWVDNDGNCAALGEKKFGK
GKDFITLIVGTGIGGGIVLNNELVHGANFCAAELGHISVPDCMCGSSGCVESYASGIALKAEKVVETACEALGLAVCTLLHT
VNPSHIILCGHLPHYVDGVREVIQRRALPSANSIQIMVSDLEE-PALLGAASIVLEYATRR
Danio
SALSVDLGGTNLRVGIVSMKGKVVQLNVGVSTGGRVNPQDGVVLHTKLIEWSSVDIRTPLHLPVWVDNDGNCAALAERK
FGHGKNFVTIITGTGIGGGIIQHNELIHGNTFCAAELGHIVVPECMCGGHGCIEAYSSGLALKAETVLHTAGTALGLGIVNILH
MINPSLVILSGVLVHYETPVRQVIGQRALLTAQGTKVMVSDLED-PALLGAASMVLDYTTRR
Takifugu
SALAVDLGGTNLRVAIICMRGNIVQANVGVSTGGRVNPQEGVVLHTKLIEWSAVDLRTPIHLPVWVDNDGNCAALAEKKF
GHGKNFVTVITGTGIGGGIIHQNELVHGSTFCAAELGHIMVPECSCGSRGCIESIASGMALKANIVLNKASTALGMGIINILHI
MNPSLVILSGVLSYYQAPVQRTIMERALFSAQSVKVVISDLEE-PALLGAASMVLDYATRR
Tetraodon
SALAVDLGGTNLRVAIICMRGKIVQANVGVSTGGRVNPQEGVVLHTKLIEWSSVDLRTPILLPVWVDNDGNCAALAEKKFG
HGKNFVTVITGTGIGGGIIHQNELVHGSTFCAAELGHIMVPECSCGSRGCIESYASGMALKANIVLNKASTALGMGIINILHT
VNPSLVILSGVLSYYQAPVQRIISERALFSAHSIKVVTSDLEE-PALLGAASMVLDYATRR
Gallus
SALAVDLGGTNLRVAIVSMKGEIVQLNVGISTGGRVNPREGIVLHTKLIEWSSVDLRTPIHLPVWVDNDGNCAALAERKFG
HGKNFVTLITGTGIGGGIIHQHELIHGSSFCAAELGHIVVPECPCGSQGCIEAYASGIALKAESILRTAGTALGLGVVNILHTM
NPSLVILSGVLSHYVNAVKDVIHRQALSSVKTVDVVVSNLAD-PALLGAASLVLDYTTRR
Homo
SALAVDLGGTNLRVAIVSMKGEIVQFNVGISTGGRVNPREGIVLHTKLIEWNSVDLRTPLHLPVWVDNDGNCAALAERKFG
QGKNFVTLITGTGIGGGIIHQHELIHGSSFCAAELGHLVVPDCSCGSHGCIEAYASGMALKAQSILRTAGTALGLGVVNILHT
MNPSLVILSGVLSHYIHIVKDVIRQQALSSVQDVDVVVSDLVD-PALLGAASMVLDYTTRR
Mus
SALAVDLGGTNLRVAIVSMKGEIVQFNVGISTGGRVNPQEGVVLHTKLIEWNSVDLRTPLHLPVWVDNDGNCAAMAERKF
GQGKNFVTLITGTGIGGGIIHQHELIHGSSFCAADVGHLVVPDCSCGSHGCIEAYASGMALKAQSILRTAGTALGLGVVNIL
HTMNPSLVILSGVLSHYIHIVKDVIRQQALSSVQDVDVVVSDLVD-PALLGAASMVLDYTTRR
Xenopus
SALAVDLGGTNLRIGIVSMTGEIIQPNVGISTGGRVNPREGVVLHTKLIEWSSVDLRTPLHLPVWVDNDGNCAALAERKFG
QGKDFVTVITGTGIGGGVIHNHELVHGSSFCAGELGHIMVPDCMCGSRGCVEAYASGIALKASNILKTAGTALGIGVINILHT
INPSLVILSGVLNQYVNVVKDVIRQRGLASIQNVDVVVSSLSD-PALLGAASMVLDYTTRR
Saccosus2
SAIAVDLGGTNLRVSIVSKAGEIMKPTVGVSTGGRVNPHEGLVMHTKSLGWSAIDLRTPIHLPVWVDNDGNCAALGERKF
GHGKDFLAIATGTGIGGGLVLNGRLVHGSNFCAAELGHIAVPACMCGSSGCVEAYSSGLALKAEKIIHLGGEALGTAVISILH
LLNPTLVVLCGVLPTYVDIVRETIENKALPSARGVKVQVSLLKE-PALMGAASLVLEYATRR
```

Paractus1
SALAVDMGGTQLRVALIAADGEIKRPTVGISTGGRVNSKEGLVMHTKAIGWDEIDLRTPIHLPVWVDNDGNCAALGERKFG
HGRDFITIATGTGIGGGIVLDKRLIHGTNFCAAELGHIKVPACQCGSHGCVEAYSSGMALKAKKILDTGAQALGTAITSLLHV
LNPKLVILCGVLSVYLEPVREAVKAHSLPSASEVEIVSSQLVE-PALLGAASLVLEYATRR
Strontus1
SALAVDMGGTQLRVALIAADGEIKRPTVGISTGGRVNSKEGLVMHTKAIGWDEIDLRTPIHLPVWVDNDGNCAALGERKFG
HGRDFITIATGTGIGGGIVLDKRLIHGTNFCAAELGHIKVPACQCGSHGCVEAYSSGMALKAKKILDTGAQALGTAITSLLHV
LNPKLVILCGVLNVYLEPVREAVKTHSLPSASEVEIVSSQLVE-PALLGAASLVLEYATRR
Patirme_3 --
LIHGTDFCAAELGHTKVPPCQCGGTGCVEAYSSGLALKAQDIMDRGAKAVASVVVNLLHILNPSRVVLCGVLPAYIETVRNI
VKKEALPSAAHAEI------------------------
XENOTELLA --
KAEAVLDRGAKALSAAILTLLHTVNPSLIIITGVLPVYIDKVQDTIKKAALPAAQQVDIKVSKLKE-PALLGAASIALDFMTRR
Symsation
LALAVDLGGSFIRVALVDDRGNIETEMVGVSTGGRVNSETGEILFTKVLGWGGVALKTRLGLPCYVENDGNCAALAEVHF
GSHKDMVVLHFGTGIGGGIIQDGCLLNGSSYSAGEFGHIVVPDCMCGNSGCVEAYAGGWALYAVQHINRAVRAVASALLTI
YSSYNPPVAILAGPLPVYFDGVKNKLEERSAILGRNFTLLQSDMTE-MSLKGAATLVLNNPSRA
B_GAMMA_6
QYIGIDVGGTHVKYGVINSDGEELFDTIGVSFPGHINPHNGHAAKAGALYLDDVNLMELFDLPLVVENDANCAALGEMWR
GAGQNLVCITIGTGIGGGIIVGRELYRGAHFHAGEFGVMPV-----
GNNGSMHKIASTSGLHLREAVNDWARYLSRGVYSVISMFDPGVMLIGGGIEKLYPLLTRHLETFEMWEALQVPIQPCQLG
NQAGRLGAVWLAQQKLARS
B_Baces_6 MKLSIDLGGTNIRIAQVE-
KGNCLVPCIGIGVPSIVDPEKGIVYNVANISWKEIHLKEILKVAVAINNDSNCFTLGESLYGEGKNMVGVTIGTGIGAGVVIGR
RLYGGQYMGAGEIGSFPY----------
FEHYCSSFLFAALEIWKEFGRHLGNLIKAILFAYAPQAIVLGGGIPFFKNAMEQTMQSFPYKIIDNVSVVAS-
HQKDSSLLGAAALLE------
B_GAMA_19 SVLCLDLGGTKLMLAQVE-
GKTLLYPVISIGIPGMVDMQSGTLLEVLNIALTATQLAQQLEMDVVVNNDANLFALGEAVLNRNQDMLGITLGTGVGAGVIF
NGQLYSGKHCAAGEIGSLSY------
RDGIIEHYCSGQYFQALQAFAHFGEHLAHMIAQTLLVYDPKDIILGGSVPFFIEALNQKLQSLVYGPQADLTISASQHHN-
AALIGAAQWFLQQKDSV
B_FIRC_17
QYLAIDVGGTTTKYGLVSATGALSQPTIGLALPGVIDSQQGLVKASATLFLEGLVLGTQLTVPILIENDGNAAALAEHWRGNL
ANSAMVVLGTGVGASLFLNGQLYHGSHHVAGEPSFMVTPIMREQTAAGLSAVATINAMEAAVILRTFTRGVAAMIYNMQTV
LDLEKVIIGGGIPRVIKEIRDDIEAYQQ---RLPVVEPAKYRNAANLIGAVAPLVVRG---
B_Bacetes YFIGVDIKKFAINIGLINFKGDMVIPYICINISGRVNPESGYSFS---
MNFSELPLAEVLGYSVCIDNDTRAMTYGEYMQGCVKNIIFVNVSWGLGIGIIIDGKVYTGKSGFSGEFGHINVILCHCGKKG
CLETEASGSALLCIEIVEEIGQKLGKQIAGLINLFNPELVIIGGTLDYITQPIKTAVRKYSLNLVKDSVIITSKLKDKAGIVGACML
ARSRMFES
B_GAMA_14
HYLSIRISRGEIFLALRDLSSKLVLPLIAITLPGIIDTENGVVHRMPYYDVKEMPLGDALGVPVYIQHDISAWTMAEALFGASR
DVIQVVIDHNVGAGVITDGHLLHAGSSSLVEIGHTQVKRCYCGNHGCLETIASVDSVLAKDIISGVGTHVGRILAIMVNLFNP
QKILIGSPLDILFPAIADSIRQQALPAYRNTVVESTQFTNQGTMAGAALVKDAMYNGS
B_GAMMA_9
HSIAVRIGRDYIQLSLYDLGGNSLHEFIGVALPGLVNPETGVVEYMPNVAINELPLGATIHVECFVGNDVRGIALAEHYFGAS
QDSILVSVHRGTGAGIIVNGQVFLGYNRNVGEIGHIQIEQCQCGNFGCLETVATNPAILAKQALVRVGNQLGKAIAITVNLFN
PQKIVIAGQIEIVFPAIQRNVENQSLKTFQHLPIVSSQVYK-QPTMGAFAMIKRAMLNG
B_Tenutes
KILGIDLGGTSAKVGVISQNGDLEITNVGITAPGFVDHNKGIVVMAPNIGWFNYDLKTEAKKPVYVINDVNAAALGEYRKGS
GLSGLFYWLGTGIGGAIICDGKLISGSHGFAGEFGHGGSLKCNCGLNNCIEKVCSATTIELKNSLLEVYDELFNHMSLLIHA
LDPEVVVIGGGGNNLLELFQLGVKNKLTDSYDIVDFKLALLKNDAGMIGAAFYALEQSLKT
B_Aciia_4
VVLTADLHPRKAAIAVIDLNGRVLLSLVGVSLPGRVDSQTQRLIFAPNLHWPDFDLRQALRMPVEMENAANACLTSELWFG
RMQNAVLITVSEGIGGGILANGQLVTGQNGMAGEFGHISLPRCGCGQRGCWETFASCKAAHAVGALTEQARQIGRGLRL
VIASLSPELVLIAGEVNLVAPALRKEMEAQWLGG-TMPRIEPT-FSDAARLRGAAAMLLQRRASA
B_ALPHA_3
LAIGFKLMVGSVECVVTDLATSPLLTLIGISMPGVIDNDQATCVRSHRFNWNNVPLASILKVPVWLEDDTNAYAIAQQLFGV
GRNMAVLAIGVGISCALIIEGKLYRGANGAAGKFGHTLHRLCECGKRGCLMAYHSQTSMAVGDILREAGIGIGKALANLVN
VTDPEVIVVGGEAEAFFEPLRSTLAAHTF---RASPPLLPDWEDNSWARGAAALVTQKFFDF
B_ALPA_15
YAVGIHITPLGINAALINLSGDVIREAVGLALPGPFGVESMSFVGPTTMGWQDVALRERLGLPAFFETDMAAAAMGERLYG
LGAEYYYLYFGVGLGGVMVHEGSVQRGAWGNAGEVGHIPVEACPCGNRGCLERYLSLEALTEADWVAEVAPIFHNAVAV
IENLFDPETVILGGLATDLLERLAGSTGGLHNSVSIAPRLMVARGGQHSVLRGAAALAVSGVLSP
B_ALPA_21
LSFGVKIGRRSADLVLMDFVGQIRQIYLGIAAPFELWNWAEEVGADGAMVWRGVDLQAEIPYPVYLQNDATSACGAELVF
GVGPDFVYFFIGSFLGGGIVLNSSIFVGRTGTAGALGPLPV----
RGRNGTLQEIASIFVLPLEIWIQETAKALAQAIVAAASVIDFSAAVIDGGFNWVRERVVRATIKEAAELDVMPEIIEGMVGPQA
RAIGGASLPIFARYLI

B_ALPHA_6
YFLGLKIGRRSAELVLIDFLGQPVRIYMGIAMPFQLWNWVQYVGAQSEMAWRERDIQTELGLPVHLRNDATAACGAELVF
GTGEDFLYFYFAYFIGGGLVLNGHLYAGRTGNAAGVGPMPV----
PDRSGRMRNVASMSVLVLSDWLDGAAAGLASAILSASALVEMETVMIDGWMPEIRAEMVRRTHAALHRLDEPPQIREGT
VGAQARALGAAAIPLSQRYLV

B_BETA
FGIGVRLDRMRIETALVNFAGDVIHDTVGVAQPYNLGSWLREIDLADAFAWAEVDFARELELPVFSENDGNAAAIAELFYG
CGRDFVYLFLGPAIGGGIAVDGDCLRGVTGNAGDFAMMPVLPSVPKPSGAWDTRASLNGLAVDEWIDDCVDALAPALRA
VLAVLESPVVVLDADTAGLIDTLIARLHTALAALAGTPVLVRGTFGPDAGAIGAATLPMFFNFSP

B_EPSILON SKLYIDIGGTYLRSELLKNGKT--FKEIGISFAGQVD--
HGKIVSSPNIAVKEYDIKKYIPVSLKIDNDLNCAMLAEKE---
DINMALLYIGTGMGSAVLEQGEIVRGERNLAYEIGHVPFFRCGCGKDNCLELFSSGSGLYAKKIYQNFQEGLFRAAATLVTL
ANPKVLVLGGGVHFLKEKVEKKIGKYALASNEGLEIKLSQL-KNASIEGAKQL--------

B_Diclomi
NALAIKVGVTHTYLSKIDFAMNIKFLTVGIGIPGIVDNTFKNVVVAPNLNWKNLPLGDMIEIPVKMDNEANMAVVAEGMLGTK
INIVYVYVGEGIGTGLILDGKLYRGRYNTAGEFGHMTVLKCKCGNLGCWERYASLGSESDEKVLKKYINELAIGLINIINGLN
PDVIILGGPLEFIRQEIKKIVEEKSITSDGKVRIELTSFDYPAELVGAGIWAFWDIFEG

B_Theae_2 ---------------------TE---
IVIGVPGSVDKTHKKLAFAPNLRWRDIDVEKYFVFEVYLENDANLAALAEMMRNKHFNIVYILVREGIGGGIIIEGKLYKGSF
NAAGEIGHMKM-PCFCGRVGCWEANTSISHCLAKEVLDEFTGILIDGIVNLVNILSPEIVIVGGEGESVFEVIVSETRRQVH---
KEVSVEKGSLNKEVVLEGTSILSSMMISER

B_Acter_4 RILAFEIGTFFTRYVVFE-
DGRMGIPGIAMSVPGFIDVSKQVAVTAGALMLYKHEIGKELPVPTWMENDANCAAMAEKLSGNAVDFALITIDTGIGGALFL
GGGIRRGKDWRAGELGMMI-
PNYETGGFNTMQNYLSTIVLRVRKIVDKWIDYLAIAIFNTAAATDPECILLGGGIQQLLPMVNAALDRIPQWGDFRTSVKRC
RHTNNAGLIGAYYAFETEVGGL

B_GAMA_17 YVIGVSINTDSAVVCIADLGCNVLLRTMGFAIAGFFLEN-
RQINAPEPLDWSLIDLQPILGMPVWLENNATTAAIGESLVGVGANFIYLSFNFGFGAGVVINGKPYFGSHGNAGEITLYND--
----
EESINRPALGVDTWLARVKPTLDRLVNALAGLFDPQAVVFGGQLPELGRRLIAATAFWGAHRYPRPQLLLSETNGDAAAIG
AALVPLKERFFV

B_GAMA_23
YSVGVSVMVDGFAVVLIDFAGQVRFPLVGLSMTGPRIGDGTRVNPPLSLEWMQVELDRFVQLPVWMDNDAHCAALAEAV
YGIGRDLVYLSISDGFAAGVIAAGNVRRGAHGNGGELGRISA-------------
TGMARPTLQIDAWLDAVEPTVTLAVAAIIALIDPRVIVFGARLTDLAQRLIARIAFEPAPRRPYPTLQVGQVTAHATVLGAAML
PFKETLF-

EU_Blistis
YFVGIDLGATNAKAGVVNDEGELLQPLIGVGSPGTIDFDNGVVIKASNFTWDHVPLANLIGVAAVLDNDANAACAAECWIG
AGKNMVMITLGSGIGAGVVVNGRVIHGGSGWAGEPGHAIYTPCGCGQKGCFEKTICRSIP---------------------------
PPCEPSRSSLPKLW-----TLRRPATR--------------------------------------

EU_Ecpus2
LFVGLDIGGTNLKAGVIDTGGQLLERLTGVGCPGQIDREAGVVIGASTFAWHNVPLANLVGRPVTVLNDASAAASAE--
FASRGTIAVLTLGTGIGLGVVCAGRVVTGCRGLV-
EGGHMIVRLCACGQRGCLEMYASASAV--

B_FIRC_15 LAVGLDITKNHFGLLLTNLKGEILFDYIGISFPGIVNLEKEIVSYSHMLGLQTL-
LFTEVSYPCCFLNDANAGAYAEGINKEFQRFFYLSLSNTVGGAIFNCDELIQGENFRCGEVGHITVVPCYCGKLGCLDVYC
SAKNLEILKIWDKYTTYLSVAINNIHMVLDCD-
IILGGYVSYLENDIRQKVLQRNTFSEDGMFVKTCNYKVGAAAFGAALKVIEAFIKQ

B_ALPA_13 YFLGIEIGVEHISAVEIDLGGNIVEPFIGVAVPAQMDKH-
GFVRLAPLLRWENVQLAELVSVPVVAENDGNAFAIGASYGRNDKVTLFLVMESGVGGGIIANGSLFRGANGLAGEIGHLRI-
-----
EPSSLEEVLGLEHIGAVSIAEEWARALAFGLIQACRVIDADRIILGGSVAALYPLMAARVAHHIQLAQPLPSIGVNEEETVGPA
FGAACILHQRFLSL

B2_Actino
RIGALEAGGTKMVLAIGTREGKIFMPTLGVGAFGPTCVNYGKILSTPKQGWVNYDFLGALGVPIGYDTDVNAACLGEALF
GSARNVVYLTVGTGIGAGVLLGGKLLHGM--LHPEAGHIPIV-CQY-
HVSCLEGLASGPSIEKQEVWELEATYLAKALAVYVLCYSPQRIILGGGVTQLFPLIRQKLLENLNGYIIDSYVVSDGCSGNQ
GILGALALGLQSLDS-

B2_Baetes
LLGGLEAGGTKFVCAVGTGPDDIRFPTLGIGSFGPVDPDYGYITTTPKPGWAHTDVAGTLNVPVAFDTDVNAAALGEQRW
GAGRTFVYLTIGTGIGGGVIVNGRRHHGR--QHPEIGHLFIH-CPF-
HGDCLEGLASGPAIPNHPAWDEVAQYLAFGLANLILTLSPQRLILGGGVTHLFPRIRRHVAACINGYVLDTFIVPPALGDRAG
VLGALALAEEVMSGL

B2_CHLORO
LYGGIEAGGTKWVCAIGTGPDDIRFPTIGVGSFGPVDLNYGSITTTPKPGWAHTDVVRTLGRPIGFDTDVNVALLGERQW
GAARVAVYITVGTGIGGGAVVGGKLVHGL--IHPEMGHMRLI-CPY-
HSDCLEGLACGPALADHPAWELEADYLGQALANLLCILSPERIIIGGGVPQMFPLVRAATQRWLNGYLPDRLIVPPALGQR
AGVLGAIALAMHTAGGV

```
B2_Thullum
VYGGIEAGGTKFVCAVGTGPDDIRFPTIGIASFGPVDPNYGYITSTPKPGWANTNFAGYIGVPVGFDTDTNVAALGEHRW
GAAVTFIYLTVGTGIGGGGGLVNGKLIHGM--LHPEMGHLAIW-CKF-
HKDCLEGLASGPAIADHPAWQLEAKYLAFGLVNYIYTLSPQRIIMGGGVSHLFPLIRQRVQHLLNGYIIDQYIVPPKLGTRAG
VLGAIALAQDTVGD-
B2_GAMMA LFGAIEAGGTKFVCAVGSGPDDL-
FPTIGIGSFGPVDLQYGFVTSTPKPGWADTEFAGVFDLPVGFDTDVNGAALGEQRWGAAQNFIYITMGTGIGGGVVAGGK
LIHGL--VHPELGHMFVQ-CPY-
HGDCFEGVAAGPAIADHPAWDLQARYVATALASLVCSFSPQRIILGGGVAHMLNSVREKTVQVLNGYIINEYLVLPGLGDKA
GILGAMVLAEQAFDNR
B2_SPIRO -
LGGIEAGGTKWVCAVASSPDCILFPTLGIGCFGPVDVGWGFITTTPKPGWRDTGVAGVFGVPVGFDTDVNAAALGECSY
GAAQNVVVYLTVGTGIGAGVVVDGRPVHGL--VHPEAGHVLVR-CPF-
HGDCLEGMASGPALPEHEAWRMEAFYLAQGVMNLVLAVSPERVILGGGVGHLFPLVRSELLRLLAGYVVEDYLVPPGLG
QRAGIVGALVLAARTAP--
B2_FIRMI
KIGALEAGGTKMVCAIGDENGNIIIPTIGVGCFGPVDLNYGYITSTPKLSWRNFDILGNLNVPIGFDTDVNASALGEATYGIT
KVSIYITIGTGVGVGVFINGQLLHGM--LHPEAGHILLS-CPY-
HQNCFEGLASGPAIDCDEVWKLEAYYIAQALYNYTCMFSPNKIVLGGGVKQLYPLIREEFKKIMNGYILENFIIAPSLNDNQG
ILGCLELANREMRLL
B2_FIMI_2
VLGAIEAGGTKFVCGIGNERGEVLFPTLGVGSFGPIDPIYGYITTTPKPHWGNYNLIGKLDVPMGFDTDVNGAALGESIWG
AAKNCLYITIGTGIGAGALVGGKLVHGL--SHPEMGHILVT-CPY-
HGDCLEGLAAGPALVDHPAWEMEAHYLAQALMSYILILSPQKIIMGGGVDQLFPLIRTKLQQLLNGYVIEQYIVSPGLGDNA
GLCGALALAKEKLNS-
B2_FUOBAC
IIAAVEAGGTKFICGLGTEDGKIIIPTMGVGSFGPIDPVYGYITKTPKPYWSDYNLIGELDVPMEFDTDVNGAALAESWWGA
GENVMYITVGTGIGAGAVVDGKMLQGL--THPEMGHIFLR-CPF-
HKDCMEGMAAGPAIDRNEVWDMEAYYLAQAVVNYTLILSPQRIIMGGGVKQLFPLIRKYVLEFLNGYVIEDYIVYPGLGDE
AGFVGSIALGKIALENN
B2_FUBACT
KIAAIEAGGTKFICGIGNENGEILIPTMGIGCFGPIDPVYGYITKTPKPHWSDYDIVGELDVPMEFDTDVNSAALGESLWGAG
QSVVYITVGTGIGAGAVINGKMLQGL--THPEMGHIFVK-CPF-
HKDCLEGLAAGPAIERNEVWEMEAYYLSQALVNYILILSPQKIIMGGGVSHLFPLIRKMVRETLNGYVIDNYIVYPGLKENAG
LMGSLALGRLALENR
B2_PLNCTO
LWGAIEAGGTKFVCVVGYGPGELLFATLGVASFGPVDLHYGQITTTPKPGWQNADILGPLGIPIALDTDVNGAALGEHRW
GAAQDFVYITAGTGIGGGGMARGRLLHGM--VHPEMGHLGLA-CPF-
HGRCWEGLCSGPAIPDDPAWDLTIRYMAHALANITYVLSPRKIILGGSVEALFQQLRLRLREVLAGYIIETFVVPPTLGDDA
GICGAIALAAQQIEPP
B_DELTA_5
FRGAVEAGGTKFVCAVGAGLDDIRFPTLGVASFGPVCLDYGRVMTTPKPGWSGFDLLGALGRPIGFDTDVNGAALGERE
WGAARDFLYVTIGTGIGVGGLAGGRLLHGL--THPEMGHIGLV-CPF-
HGDCWEGLCSGPAMADHPAWEHEAAYVAEALATVTYALSPRRIILGGSVEGFFAKVRERFVATLGGYLVAEYIVPPGLGEL
AGVAGAYCLACAAPGGR
B_GAMMA_8
YLAGIEAGGTKFFTTIGDFDGNVITDTIGLACFGPIDINYGYITNTPKIAWQNFDIVNAVSGPIGFNTDVNAAAICEKLWGCAQ
NLLYLTVGTGVGGGIICNNKLVQGA--MHPEIGHLLIS-CPF-
HGNCLEGLASGTAIDDHIAWQFEAEYLAKAIVNYICSFSPERIILGGGVTILFDMIRKNVTKYLNNYLMTKFIVPASFGDNTGV
KGSLALALETFNNS
B2_ALPHA
VFAGVELGGTKVMVGFGSGPDDLSIPTIGVATFGPARLDWGRILPTPKPGWTGAVIAPRLGVPVAFDTDVAGAAMGEGHW
GAAQDHAYVTVGTGVGVGLIVNGMPLHCA--LHPEAGHIKVV-CPF-
HGDCLEGLVSGPALADDPVWDLVADYLAQAMATLCFVAAPRRIVIGGGVPTVLAATRLRLRDELGGYLIETFLVPPALGDR
SGVLGAIALARALHDTS
B_ALPA_27
RIAAIELGGTKVMVAFGSGPDDLSIPTIGVASFGPIRLDWGHILKTPKPGWSHADVAARLDRPLALDTDVNGAAVAEGLWG
AAKDYAYVTVGTGVGVGLVVNGAPTHGL--LHPEAGHILVS-CPF-
HGDCLEGLISGPALKDDPVWALVADYLAQLVANLALIASPRRVIIGGGVPQLLEQTRTRLQTHLAGYLIDAFVAAPGLGANS
GLLGAVALGLRHDAIL
B2_GAMA_2
LYAGIELGGTKTICLIGTGVDSIRIPTIGIGAFGPVNIDYGCIESTPKPGWSHTSVVPFFSCPINLDTDVNAAAIAEHQHGNGK
NFIYITVGTGIGGGALIEGAPVRGN--SHPEMGHIALA-CPY-
HQNCAEGLASGSALPEHPAWDMQASYLAEFFHSLTLLFSPQRIIVGGGV-
QLLARVRTALYKKLNGYVLESYLCLPELAGNAGPLGSLMLAYPEYRRH
B_FIRIC_2
LYGSLEAGGTKFVCAVGDENFNVVFPTLAVGSFGPIDIDYGFITTTPKPNWANVDLLGALNVPMYFTTDVNSSAYGEMVNN
AGGNLVYYTIGTGIGAGVIQRGEFIGGV--GHPEMGHYYVV-CPF-
HKGCLEGYAAGPSLLNNPVWDVQAYYIAQAAVNATVTFRPDVIVFGGGVQHMLDRVREKFTSLLNGYLVRDYIVTPAVGN
GSATLGNFVLAKEVSK--
```

EU_Acoeba
VFAGVEAGGTSFSVGLARSAESIFFPTLGIASFGPVDLDYGYITTTPKPDWGNVDVLSRFNVPTAFETDVNAPAVAHLA---
QFSCAYITVGTGVGVGIAVEGKPVHGL--LHPEMGHIFVT-CPF-
HGACVEGLVASRALDDREEWDHVAYYLAQLCAALVLTVSPHRIVMGGGIPTILPLVHKHVLALLNGYIIGNYITLSPFGGNAG
LVGACELARLALPHN

EU_Dilium
KFLGVEAGGTSICLSIAIKPSNVIIPTIGIASFGPIDLDYGYITTTPKPNWGQTNILGWFKCPKGFDTDVNGAAISETFHGLHK
SCAYITVGTGVGVGCVVNGSCIHGL--VHPEGGHSFCT-CPF-
HGNCIEGLVSTGAIDDHPVWLIIGNYLAQLCANITCLLSPQVIVLGGGVSILYPIIRTETIKILNGYIIDQYIVQSPFASHAGSIGS
LELARRALIKN

EU_Polium
VYAGVEAGGTGFTLAIASTPDNIVIPTIGVASFGPIDLDFGFITTTPKPMWGNTNILGWFQCPKKFDTDVNGAAISETFHGR
HESCAYITVGTGVGVGVVANEKPIHGL--VHPEGGHIFTT-CPF-
HGNCIEGLVSTGAIDDDPVWQIVGHYLAELCATITCIMSPQVIVLGGGVTILYPIIRDELLKILNGYIVHKYIVQSPFGSNAGIV
GALELARRAHLEK

EU_Phtora
RFAGVEVGGTTWVAAIAEHPENILVDTIGIASFGPVDLNYGYITSTPKPNWGNTDVVGVFDVPIGFDTDVNAPALYEVAYGG
HGSAVYITVGTGVGVGVCTNGSAIHGF--MHPEGGHIIVV-CPF-
HGDCIEGMVASGSIDDDPVWDTIAHYLANLCINVTFITSPDVIVIGGGIEKLFDLIREKFVARVNKYG-----------------------------

EU_Neella
YIAGVELGGTSCVAAIAENPTTIVVSTLGIASFGPVDLKYGYITSTPKPGWKYVEIVGVFGTPIAFDTDVNAPALAEMAAALN
TTVIYITVGTGVGVGVTVEGSPVHGL--LHPEGGHIIVC-
CAYGHECCVEGMIDSQAIDNNPVWKVVGYYLGALCLNITYLLSPNLIILGGGIRILYTETRHWFKELLKEYLLASYIKEPFHG
SKAGIVGALELARHKC---

EU_Milla
RFAGVEGGGTTWVCAVAEDPDAIVFPTLGIATFGPLCLDYGTITHSPKPGWTRVDVLKRLDCPVAIDTDVNAPAASELAFTA
ADNLAYVTVGTGVGVGIIVGGEPVHGL--VHPEAGHARVA-CPY-
HEDCVEGMCSAAAVDDDDAWDATAHYVAGLCANLILTTSPQRIVLGGGVRALVSKIRVHVKAQLGGYVLRELIVSSRRGN
DAGVVGALTLAKGAAEKA

EU_Miassp
RFAGVEGGGTTWVVAISEVPENIVFPTIGIATFGPVDLDYGYITHSPKPGWADVDVLGILDCPAGFDTDVNAPALSELSEGA
EDNLCYVTVGTGVGVGVVCGGQPVHGL--SHPEAGHIRVG-CPY-
HADCVEGMANASAIDDHDAWDAAAHYLAGLCSVLVMTASPQRIVLGGGVKTLVTKVRAQLKRQLGGYVLVEFIVSSKLGN
DAGIVGALAVAEKAKADW

EU_Ecrpus
IYAGVEGGGTTWRVAIADHPTNITFVTLGIGTFGPVDPRYGYITSTPKPGWNMVDVVGYLSVPCKFDTDVNAPALAEFMW
GAKKSCAYITVGTGVGVGLVVNGQAVHGL--MHPEAGHLCLVDSVF-
GGASVEGLASTVALDSDPVWEATAHSLAGLCASLVLVVSPERIVLSGGVTLLYDKVRKWTRELLNGYIVDDYITPSSFGQN
AGMVGSLTLAHIAYEEA

EU_Mosiga
VYAAVEGGGTSWRAAIAVRPDNIVFKTLGIGTFGPIQPDYGCITSTPKPFWKNAKVVQTLGIPHLFDTDVNAPAYAEFLHAT
NPSLAYITVGTGVGIGLVINGQPVHGL--LHPEGGHMLIV-CPF-
HKDCVEGLVSAPALDDDFIWDAAADALANACANLLLTVSPQAIVISGGVACLFDKVRARTLELLAGYVPEEVIRPSTWGNN
AGIMGALHLSKTALERS

EU_Osccus
KLLGIEGGGTTWIARAIEVEGGSSFETIGVATFGPLELNYGYITTTPKAGWQDVDVLGSLQVPLAFETDVNAPAMLEHRHV
HLVSCCYVTVGTGVGVGVVCNGLPVHGM--LHPEAGHMFVT-CPF-
HGNCVEGMVGSGALDDDDIWEHAAHYLAGMCVNLILTLAPERIVLGGGVECLFSKIRANVRDILQGYLLRHFIVPPAWGYQ
TGLTSALYLAERALQRE

EU_Osus_2
KLLGIEGGGTTWTARATSVDAGTVFNTIGVATFGPVELNYGYITTTPKPGWEDVDVLGALDVPLAFDTDVNAPAALEHRNV
HRASCCYVTVGTGVGVGVVANGRPVHGM--LHPEAGHMHVT-CPF-
HGNCVEGMCGSNALDDDEIWDQCAHYLAGLCANLILTLAPERIVLGGGVECLFSKIRAKTRELLNGYLLRHFIVPPAWGYE
TGINSALFLAENALKRE

EU_Teonas
VVAAVEAGGTTFVVAVFAAPVTPLMATVGVASFGPVDLNYGFITTTPKAGWANTDVVGTFGVPVAFETDVNAAALAERMA
GAGAAFVYVTVGTGVGVGVALDAEAVHGL--LHPEAGHMCVV-CPF-
HAWCVEGYVASGGLDDDVGWDAVAWYLAQLVINVTLVLSPHTVVLGGGIRSLYARVRSHVHTLLNGYLIEAYLRPPRL-
AEPGLTGAALLV-------

EU_Pacium
SKAALEIGGTFIQVGIGTKTDNLNFDTIQIASFGPLCLNFGSITSTPKLKWQNFPIATRLKKPFAIDTDVNACAMAEFMLGNH
NSLAYITIGTGVGVGIIVNGQCVHGM--LHPEGGHILVV-CAF-
HGDCLEGLCTNVAIDDHPIWELVGFYLAEACQNILYLLSIEKIVLGGGVKLLYPIIDKHLRRLVNKYV-ENYIVEPQV-
EDVGLIGALLLQ-------

EU_Temena RIACIELGGTSIRLAIGI------
IETVGIASFGPICLDYGFITTTPKVSWKNFPLLKRVTQRIGFDTDVNAAACAEYNFGNHKSLAYITVGTGVGVGLIVDGKCV
HGL--THPEGGHVLIV-CKS-
HGNCVEGMVTNHALDSHEIWNSVAYYLAQLCLNLTLISSPEVIVIGGGIQPLLGLIKQNFIKLLNQYVIDDYIVKPFF-
TDSGLVGSMVV--------

```
EU_Naria_
LIVGIEGGGTSYKVAIGLNIEHAMITEIGIANFGPIDVNVGCILPTPKVSWRSFNIVEYFVKHVQFDTDVNGPAMAEYQLVKG
SSLAYVTIGTGIGVGLVINGSTVSGL--MHPEAGHIYTF-CTF-
HTECLEGMAASPSIDDDEIWDIEAHYLAHLCVNLILISSCQVIVMGGGISILFDKIREKTVQLLNGYHIKNIIKPSVYGEHAGIK
GALHLSQL-----
EU_Tronas
YACGIELGGQTASIAICEKVGEIYIKTIGIASFGPLDLYKGSIGNTPKPNWGFYPLVKKIDCKVSMETDVNAPAYSEYLHLKEQ
SVGYVTIGTGVGVGVFCDGKPLHGR--MHPECGHIMAT-CPF-
HGACFEGLISAQALDSDPVWDIYIEYVAQLTVTMSYVYSLDAFIIGGGIEWIFDKILARSQQLINNYI---
IISKPFHGADAGLVGACAVAINPDVFA
EU_Tras_2
YACGIELGGQTAAFAICENLGSLYIKTIGIASFGPLDVYKGSIGNTPKPKWGNYPLVASIEAQVVLETDVNAPAYSEYLHLNS
KSVAYATIGTGVGVGVFCDGKPLHGK--MHPEGGHFKPC-CPF-
HGDCVEGMISAVALTDDPVWDCFTEYAAQLSANCALLYSLDYMVIGGGIEYLIEKIQKRTKELLNGYI---
KVIKPFYGGDAGLVGATAVALHPDVFT
EU_Trnas2
YLCGIELGGTSSSAAIIDEDGKVLITALGIASFGPLNVESGTIGKSPKKGWYYFHVKAEFDIPIAMETDVNAPAYSEFIEFSKK
SLAYLTIGTGIGLGLYSDGSIYHGR--LHPEFGHTYIV-CHI-
HGDCAEGLISASAINDHPIWDLYVEYVSQIVANAALAYSLDVFVIGGGVGFLYDRIYSRASELINDYI---
LVVRPHFDRDAGLIGATVIARRKFNKI
EU_Phylum ---------------------
ILIDSLGIACFGPLGVNYGRILASPKAAWRGVDLLTPLALVTRVDTDVNAPALAEYYTSASSSIAYVTVGTGVGVGLVVHGQ
PVHGR--MHPEGGHVAVN-CPY-
EGVTVEGLASSVALDDHEVWDHAVNALANLCVTLLLTLSMEKIVMGGGIALLLPRIQARTVELLNGYLMSTLIATSSFGDDI
GLIGAMVLAQSSLQLN
EU_Thsira ---------------------------VGIATFGPAGVNYGRILESPKREWRGVDLITPIDALVGFDTDVNAPALAEFR-
HRYKSLSYVTIGTGVGVGLIVNSQPIHGL--LHPEGGHVSI--
YSWGSEKPEDGSSGGDSNDDHEVWNHASNAIANLCVSILLLTSCQKIVLGGGISILFDKIRHRVWVLLNGYLLADLIVESS
WELGSGLVGAFALALDA----
EU_Sopora IVAGVELGGTKSIAVIGR-GREIIVPTIGIASFGPISVADGTMLPTPKPHWAGAEIVAPLD-
KVAFHTDVTGAALGEGAFGAAVDFLYVTVGTGVGMGIIAGGRPVTGV--MHPEAGHIRVA-CPF-
HGDCLE-----------------------------------------------------------------------------------------
B_Acter_2
AGIGAEVGVDHLAVCAVDLRGRVRVRYLAVAVPGLVARDGRTVVRAPNLDWHDADLGALLDLPPTVDNEANFGALAELWL
GDGTDFLHVSAEIGIGAAVVFDGRLCAVPAGSPVSWGMCRSPGAPAVGAG---
AWNSTPVRDVRRALREAGTALGIALTGAVNLLDPEAVVLGGALARRVPPARSRCPSWVR--------RAA--
ARSRALRGPAVLDDPGAVAE
EU_Crdium FSAGLNIDRDHLTLVVLDLSGQIRAEILGIALPDELG--
TLAFPGPLTYAWARTDAADLFAGPTFVENDAVAAAIGEAQFGHGLSFVYVLISVLLGGGIVIDRT----------------------------------
----------------------------------------------------------------------------------
B_FIRC_14 VSLSLYLEGFEIYWFITDHCGNKI---IMIGIASNIN--
SGKITSHMEYELYDIDVIKYLEIPINVENDMKVAARGYWARHENKAVASIYMGNGMGSNMVIDGKIWSGTSNFAGEIHYLPI-
------KKPMY--------EFNDIDTNIVQYYGKIIQSYIALINPNLIVLYS--SYIIDEIKFYCKCRIPEN-AMPKIIISEY-------
GLSKMANELMD--
B_FIRIC_4 LGLAIFLEKNETNYTIFNCLGEVK---ISIGVPGSVD--
NGRIFYIPGYKFQNFNLKSHLSMPVVIENDMNAAVLGYYKNTGNYSLVYLYSGNGPGAGIMVNGDVVRGSTFFSGEISFVP
Q----YDNKNFLQALRSE---DSNNPEEYNIDAITRLIATCIAIINPHGFIFCDEVQFVIDQIVKSCPQYIPAE-HIPKITVSNW-
KEDYLYGLKSLGLDLMITR
```

cytidil domain alignment, phylip format

```
 71 152
B__ IPARLGSTRLKEKPLKNLLGKPLIRWVVEGLVKTG--ERVILATDSERVKEVVEDL-
CEVFLTPSDLPSGSDRVLYVVRDLLIINYQGDEPFVYEEDIKLEDVKVVRSPIPKHVGIYGFRKETLEQLRLLENGIIKVLITE
NYYHGVDTEED
B__4
IPARLGSTRLPEKMLADIEGEPLIVRTWRQAMQCCRASRVVVATDSVKIAEVLTTYGAEVVMTSPEARCGSERIAEAARQF
VVVNLQGDEPLISHETIDLNQVKVVRSPIPRHVGLYAFKAEVLEQLRLLEHGYIRCMVTHDDQPGVNTAED
B__2 IPARIGSTRLPQKALAEIAGKPMIVHVAEQAKKAA-
FGRIIVATDHNNIAKVVTAYGHECIITCRDHKSGSDRIYEALTHIVILNVQGDLPTITPHEIISNIVKIIRATAPHHIGIYAYRREALE
QLRALEHNMIDVEIVDTIPLGVDTQRD
B__3 IPARLGSTRLPNKPLADICGKPMIVHVADRAAAAK-
LGRTVIATDSEEIFKVVAAHGHEAIMTRGDHESGSDRIYEALAKLAVVNVQGDLPTIDPDTIRRNVVKIVRATAPHHIGLYAYR
RSALEQLRALEAGMIDVEIVKTVPLGVDTQAD
B__11 IPARMAATRLPGKPLLDIGGLPMVVHVLRRAETAG-
IGRVAVATDTPEIAAAVTAHGGEAIMTRADHPSGSDRVFEALDRLTIINLQGDFPTIRPDIIRDNVVKAVRATAPHHIGLYAYRR
KALEQLRALEGGMIDIMIVDDVPRGVDTAAD
B__9 IPARMQATRLPGKPLADIHGEPMIIHVWRRSVQAG-
LGPVVVACSEAEVFDAVHAHGGQAVMTDPDHPSGSDRVWEAVRKLAIVNVQGDLPTLDPQIIRANVVKAVRATVPHHIGL
YAYRRDSLEQLRALENGMIDCALVDTVPLGVDTPAD
```

richoplax IPARLAARRLPNKPLIDIEGVPMIIRVYNKALEAN-
IGKVVVAGCNDELRNLINSYNFNYISTDPTLPSGTDRVYAAYKNLNVINLQGDMPHINSMTIKKNTVKVIRYPVRKHIGIYGF
KVNSLEQLRALENGMIKVAITNDPNISIDTEQD

B__13 IPSRLSSTRLKQKPLQLIGSITLIERVFKQVNQAG-
LEHTYVATDSEEIASVITKVGGKVIFTDSAIPTGTDRTYEAFKLIYIVNVQGDMPFIEPSSILKSNVTVVRSLIPYHVGMYGFR
KNALEQLRALENGMIGTCLVENVPISVDTEED

B__5
LPSRWGSSRFPGKPLAKILGKTLVQRSYENALSSQSLDCVVVATDDQRIFDHVVEFGGLCVMTSTSCANGTERVEEVVSR
HIVVNIQGDEPCLSPTVIDGHKVKCVRSAIPLHIGVYAFRKAFLEQLRVLEIGRIYVHVVQATGPSVDYPED

B__46
LPSRWGSSRFPGKPLAKILGKTLVQRSYENALSSQSLDCVVVATDDQRIFDHVVEFGGLCVMTSTSCANGTERVEEVVSR
HIVVNIQGDEPCLSPTVIDGHKVKCVRSAIPLHIGVYAFRKAFLEQLRVLEIGRIYVHVVQATGPSVDYPED

B__61
LPSRWGSSRFPGKPLAKILGKTLVQRSYENALSSQSLDCVVVATDDQRIFDHVVEFGGLCVMTSTSCANGTERVEEVVSR
HIVVNIQGDEPCLSPTVIDGHKVKCVRSAIPLHIGVYAFRKAFLEQLRVLEIGRIYVHVVQATGPSIDYPED

B__8
IPARYASSRFPGKPLAKIGDKTMIEWTYRNASRSSVLSELVVATDDVRIHEVVQKFGGRSVMTSSDHPSGTDRIIEVANQFII
VNIQGDEPGIEPELIDGNRVKVIRSLIPRHLGIYGYDRDFLEQLRAIEAGYIGIYLSKEAGLSVDTPAD

B__19 IPARYSSTRLEGKPLKMIEGHTMIEWVYKRAKKSN-
LDALIVATDDERIYNEVINFGGQAIMTSKNHANGTSRIAEVCEKMTVINIQGDEPLIEYEMINSNNVKVVRSVIPKHIGIYGYK
RDFVEQLRVLENGYIKVLETTHSLIGVDTQEN

B__37
IPARYSASRFPGKLMKDLGGKSVILRTYEAAIKTGLFEAVYVVTDSEIIQENIANAGGDVIMSTAEHQCGSDRIAEAVAFLIVIN
VQGDEPFIDKVSLSKNNVKVIRSVIPKHKGVYAFRKQALEAIRYQELGKIKMVETTIESVGIDTPED

B__7 IPARYHSSRFPGKLLQEINGITVIERVYRQALLAE-
PKSVIIATDHDEIADRAIQFGAEVVITSHTHQTGTDRIAEVIAKGVIVNVQGDEPFIRPKLIQQNVVKVVRSAIPRHIGLYAYRA
AFLEQLRILWSGFIRVDEACEPLQDINTKED

B__12 IPARYASTRLPGKPLQDIAGQPMIQRVWNQARKSA-
ASRVVVATDDERILAACQGFGAEALLTRAEHNSGTDRLEEVASRLIVVNVQGDEPLIPPALIDQNVVKVARAPLPRHIGIYAY
RVGFLEQLRALWHGVIHVADARNMLPGVDTPED

B__14 IPARFSSTRLPGKPLVDINGKPMIVHVLERARESG-
AERIIVATDHEDVARAVEAAGGEVCMTRADHQSGTERLAEVVEKCVIVNVQGDEPMIPAVIIRQNAVKVVRATIPRHLGIYG
YRAGFIEQLRVLWYGEIHVAVAKVPGTGVDTADD

B__18 IPARYQSTRLPGKPLADIGGKPMIQWVYEQAMQAG-
ADRVIIATDDERVEQAVQAFGGVVCMTSPNHQSGTERLAEVVAKMIVVNVQGDEPLIPPAIIRQNAVKVIRATIPRHIGIYAYR
AGFIEQLRVLWHGEIHVAVALAPPAGVDTPED

B__26 IPARFASSRLPGKPLADIAGKPMIQHVFEKALQSG-
ASRVIIATDNENVADVAKNFGAEVCMTSVNHNSGTERLAEVVEKLIIVNIQGDEPLIPPVIVRQNAVKVLRSVIPRHIGIYAYR
AGFIEQLRVLYNGEIHVELAKVPAVGVDTAED

B__21 VPARLASTRLPNKPLADLGGKPMVVRVAERAREAG-
AQQVLVASDAQRVLDAVREHGFDAVLTRADHPSGTDRLAEVAAKLIVVNVQGDEPLIDPQLVRDNYVKVVRAPIPRHIGLYA
YRARFLEQLRAMWHGEIAVRVTEAPEAGIDTPAD

EU_Oryza
IPARFASSRFEGKPLAPILGKPMIQRTWERVMLASSLDHVVVATDDERIAECCRGFGADVIMTSESCRNGSERCCEALQKL
IVVNIQGDEPLIEPEIIDGNRVKCVRGLIPLHLGIAGFDSKFLEQLKVLENGYMKVIKVDHDAHGVDAPED

omitrella
IPARFKSSRFEGKPLVHIMGKPMIQRTWEQAMKCTSLQAIVVATDDDRIADCCRGFGADVVMTSENCQNGTERCNEALEK
LIVVNIQGDEPLIEPAIIDGNRVKCIRGLLPLHLGVQCYDAKFLEQLKVLENGFLKVIKVDHEAHGVDTPGD

mitrella2 IPARYKSPRFEGEPLVHISDIPMIQRTWEQAKQFTSLKAVVVAIDDGRIADFYRAFGAD-
VMTSENCLSGTERCNEALEKLIVVNVQ-------------NWIKNIRALIPRHLGMQCHNAEFLE----------------------------

B__10
IPARHASTRFPGKPLAIIAGRTMIEHVWRRCQEAQAFDEVWVATDDDRIRAAVEGFGGKAVMTSPACATGTDRVAEVALG
RIWVNVQGDEPLVDPATLQRHVVKAVRSLVPGHIGLYGYRREVLEQLRALEHGIIRCAKVTSHTVAVDLPGD

B__56
VPARLQSSRLPNKVLADIGGKPMIQRVLERCSEAQGVEAVVLCTDSTELQTLAEGWGFPVLMTSESCNSGSERIASVAHP
LAVINVQGDQPFIEPAVIDANVVKTLRSAIPGHVGMYGFRGDVLEQLRLIEAGLIATFRVEGTSLSVDTAEQ

B__58
VPARLQSSRLPNKVLADIGGKPMIQRVLERCSEAQGVEAVVLCTDSTELQTLAEGWGFPVLMTSESCNSGSERIASVAHP
LAVINVQGDQPFIEPAVIDANVVKTLRSAIPGHVGMYGFRGDVLPRCRWTPPN-------------------

B__30
IPARMGSSRFPGKPLALIHGVPMVGHVAFRTAMSRCLSATYVATCDTVIEDYCKEASLACVMTGDHHVRCSTRTAEALLKII
VVMVQGDEPMVLPGMIDANEVKVVREPIPKQVCIIPFRRDYLDMMRILEHGEVRMVPTDARTWSVDTPED

B__44
LPARMASSRFPDKPLVKISGLEMIEHVRRRVEMSSSVDEVVVATCDEIIKQRVESFGGKAVMTSDVHRGCIDRVAEAALYVI
VIVVQGDEPLILPAMLDDNAPKVVREPIPKQLGVIAFRNDFLDMNRAVEHGYVRMVLTEGIMIGVDVPGD

B__48
IPARMAASRFPGKPLFPILGRPMIEHVFERAKLFGRWDALAICTCDEEIRAFAQSKGYPVIMTSDKHTRALDRVAEAATKCI
VLNVQGDEPMMHPDMIAADALKIIRQPIPRIYGIFGFKWDFLDSNRLYDYGHQHIAPYPRPSFSVDSPHD

B__20
IPARHASTRLPGKPLLDLCGVPMIVRTYRQCIQAVDAEHVLVATDDERIRAVCEGEGIRTLMTSSRCLTGTDRVAEVAGQVIF
INVQGDEPLFNPDDLRKSVPKVVRAAIPRQVCAYAFPREALEILRFLELGWVKMIEMSDQSISVDNLED

B__36
IPARHASTRLPGKPLLDLCGVPMIVRTYRQCIQAVDAEHVLVATDDERIRAVCEGEGIRTLMTSSRCLTGTDRVAEVAGQVIF
INVQGDEPLFNPDDLRKSVPKVVRAAIPRQVCAYAFPREALEILRFLELGWVKMIEMSDQSISVDNLED

B__25
VPARYASTRLPGKPLADICGKPMIQHVYERACKVPYIDDVIVAVDDRRVAEVVESFGGKVIMTSTQHDSGTDRLVEVMGKY
IYINIQGDEPLIRSEDIALNTVKVVRSPIPKHIGVYAYRKETLEQLRLLDAGIIRVFEVPETGPGVDTPEC

B__6 IPARLKSSRFENKVLEDIFGLPMVVRCAK---
NANLVDECVVACDDESIMQTCQKFHIKAVLTSKHHNSGTERCLEAARILRVLNLQGDEPFLEKEVILANLVKVVRSLIPGHIG
IYGFHKEILEQLRALYYQKIAVKIVQSESVGIDTQED

B__49
IPSRFASTRLPEKPLVKIAGKEMVLRVAEIANVCNKVENYIVATDHEKIVNFCKENNIAVMMTSENCKSGTERCWDVTTKIFI
VNLQGDNPLCPPWFIEQSGTTVEKAMIPRHIGLYSYTYDALEQMRFLHNRIVKMIDVDKSMSGVDSPED

romonassp
IPARYNSRRFAGKPLALLGGKPLVMHVYRNACKAKCLDACVVATDDDRIAKAVRDNGGEVIMTDAACETALERCLEVVRKL
VVVCIDADEPFVQPHHIETDTVKVVREPVPVRVGIQACGTDFLEQNRVVFAGYVKVDVVETFVPGVKTPED

Ostreolu2
VSAPYEDDQFAGKALMIVHGKPAFARAAESMKRCARLDRVVVATDDYRIVETAKEYGIETVLVAQDARRTASTYARQAAKL
FVVTLKVDECLIDADSVDATRPRCVRAVIPHVLGCSCYDAMYLEALNTLENGYIKVCPVNQ-TPALRAPED

Ostreotau
AHAPRRNGRYEGKPLALIHGKPAFARLAETMKRCTRIERVVVATDDYRVAETAKEYGIETVLVAPEVARTSTVYAREAAKAY
VCVVDVEECLLDADSIDAIRPRCVRAAIPGIVEATCFDAVYLDALNALERGYIKVCHVNQ-VPPLRDPKD

romonasspu
IHARCDSKAFPGKPLVDVAGAPLVWRTFLKVRGAKCVSRCVVATDDDRIARVVAALGGEVVRVSGKWGSHTAAAVGAKK
NQVVVCVDADEVGIDAHLVERDAVMGTRPGPPRKVGARSYRVSFLEEREAAACGRVKREVVAESVPRVGTREE

B__15
ILARQNSKGLPLKNLRKMNGISLLGHTINAAISSKCFDRIIVSTDGGLIAEEAKNFGVEVVLRPAELASDTASSISGVIHATVTL
LQPTSPLRTGAHIREGSVVSARQQLPPNGAIYINDTASLIANNCFFIAPTKLYIMSQDSIDIDTELD

B__17
ILARQNSKGLPLKNLRKMNGISLLGHTINAAISSKCFDRIIVSTDGGLIAEEAKNFGVEVVLRPAELASDTASSISGVIHATVTL
LQPTSPLRTGAHIREGSVVSARQQLPPNGAIYINDTASLIANNCFFIAPTKLYIMSQDSIDIDTELD

B__38
ILARQNSKGLPLKNLRKMNGISLLGHTINAAISSKCFDRIIVSTDGELIAEEAKNFGVEVVLRPAELASDTASSISGVIHATVTL
LQPTSPLRTGAHIREGSVVSARQQLPPNGAIYINDTASLIANNCFFIAPTKLYIMSQDSIDIDTELD

B__33
IPARAGSKGIKDKNLQLVGGVSLVGRTILAAQESGMFDQIVVTSDGENILKEATKYGAKPVARPESLAQSDTRTIDAILHCTA
ALLQPTSPLRNALDIRNKSVVSARQKLPANGAIYINDIASLFEEKRFFIAPMRFYLMPYRSIDIDSTLD

B__23
IPARKGSKRIKWKNIVPLAGSPMLEYTVKCALNSKYIDRVIVSTDSYYIKKLAKKMGADTFIRPKNLATDDAKTIDVLLHAYLV
LLQNTSPLRKSWQVDEDSLVSIKNKFRINGAIFINKIDKNFNSDTILTNNQLPYIMKETSIDIDTIED

B__24
IPARGGSKGIPHKNIMKICDKPLISYSIEAAKKSKYIDYILVSTDDVDIKEVSLNYGAKVFLRPNEISTDRAKSIDVVLHGYVVL
LQPTSPLRTSKDIDDDSLISVRQELPFNGAIYINKVYMLQNKKEFIDENTIPFIMDKKSIDIDNMID

B__39
IPARGGSKGVPRKNIRLLLDKPLIAYTIETALEVDFLDKIIVSTEDLEIAKISMEYGAEVFLRPYELATDEAKSIDVVLHALILLLQ
PTSPFRNSEDIKTKAVVSVRQELVINGAIYIAEWNYLKQNRTFFGNETYAYIMPERSIDIDTEMD

B__43
IPARGGSKGIPRKNIKLIAGKPLIVWTIEAALKSKHLTSIVVSTDDPEIAEIAEQSGASVFLRPAELATDYSSGIDPVLHAYVML
LQPTSPLRTSADIDDNSVVSVRQDSPLNGSIYLSEVNYFREKKKFITDDTLAYLMSENSIDIDDMID

B__45
ICARGGSKGLPGKNVRPLAGRPVIAWSVEAALGSSLIDRVVVSTDDPAIAEVARAAGAEVFLRPAELASDTASLYDVIFHAH
VVLLQATSPLRIAADIDGPAAASLRQDLPPNGAVYVAETAWLRRERNFWKAGTLGYVMPERSVDIDSLLD

ICROMONAS
ITARGGSKGIPGKNIIDLNGKPLIQYTIEAALSSKQLDRVILSTDSDEIAEVAQNCGCEVFRRPSELAADDSSHLACIVHAFVV
ILQPTSPFRKSIDIDSDMVLSVRQKLLENGAVYVLRTQSLPNVGSFRSADTKGYEMPERSLDIDNPFD

B__42
IPARAGSKRLPGKNTRLLAGKPLIAHTIVAALQSSCCEEIVVSTDSKQIADVAVQYGASVWLRSEDLATDTSDVIHTVIDLSVL
LLQPTSPFRKPETIRHKSVVSVEN---LNGSIYIATAKQIIENKSFYSEPTKPLLLNSESIDIDTPID

B__32
IFARGNSKGLPGKNIKILGTKPLLAHSIDVARSVREISKIFVSTDCKAIADVAQQYGAEVIWRPEELATDNSPEWLAWQHALF
VSLPATSPLRSTEDVQNDIVVTARQDAPMTTVAYVTRPEYILGNERLFAGVVRSIVVPERAIDIDDIYD

B__35
IPARGGSKGVPAKNLAAVGGVPLVARAVRACLDAPLVTHVAVSTDDPGIAAVARGAGAEVVLRPAAIAGDTATSESAVLHAV
VLLVQCTSPFISREDIEGDSAVTVRQDRPETGAAYAMSAAGFREAGHRFFGRTALVHTDARVLEVDDPHD

B__51
IPARGGSKGIPGKNLLTVGGVPLVCRSIRAALASNGVGRVVVSTDDEAIAAAAETEGAEVIRRPAEIAGDTASSESALLHAEL
VFLQCTSPFTTGAQIDANSSFSVRQDLEETGAIYAMAITAFRRCGSRFCPPTSPVVLQEVGPEIDTPED

B__54
IPARGGSKGILRKNLREVGGAPLVVHTIRQALESSSVDRVYVSTDDTEIRDVSQAAGAHVIHRPVDISGDTISSEAVIHHALV
VFLQATAPLRAKGDIDGDSLLSVRQDLEENGSIYVFKPWVLERYDNRLGGRIALYEMTSAAFDIDDEID

B__57
IPARGGSKGLKYKNIYPVAGKPLLAWTIEQARASQFVDKVFVSTDSEDIADIAKEYGAEVIERPADIAGDKATSESAILHAAV
VFLQATSPLRKQGDIDGDSLISVRQDRPENGSIYMFTPETLHRFNNRIGEKLVAYEMEWQTWEIDTLNE

```
EU_Aedes
ILARGGSRGIPLKNLAKLDDRTLLSRALHTALSTDGFHSVWVSTEDDRIAQAVERHDVRVHLRPPEVAQDHTSSIESVREF
NVALVQCTSPFLGVRYLDEDCVFSVRQDWDEAGMFYFARRKLL-LEGRFQNNNCEVVVIDRDSLEIDSLYD
_Branchio
ILARGGSKGIPMKNNKKLGGLELIGWVIRAAIDSDIMDSVWVSTDHDDIAATARKCGAQVHRRSAEVSQDTCNSWVTIDEF
IIVDIQATSPCLHPFHLHDDSVFTVRQEWNENGAFYMMTRDVE-QQGLMQGGKMAYLVMEEYGVDIDTDLD
Branchio2
VLARGGSKGIPLKNIKPLAGTPLVGWVLRAAIDSEAFDSVWVSTDHDEIAAVSREFGAQVHRRSPEVSRDASTSLETIQEFI
FGNVQATAPCVHPFHLRKDSLFAVRQDWDENGSFYFCTTALV-QKGLLQGGKVGYLEMAEYSVDIDIDID
Strongylo
ILARGGSKGIKLKNIKALAGQPLIAWVLRAAIDSGEFDSVWVSTDHADIARISKEWGAQVHMRSPHTARDQATSIEAMQEF
FVANVQCTSPCLHPSHLQRDSVFAVRQDWAENGSFYFATRELL-MAGLFQGGKVGYCEMQEYSVDIDTDID
EU_Mus
VLARGGSKGIPLKNIKRLAGVPLIGWVLRAALDAGVFQSVWVSTDHDEIENVAKQFGAQVHRRSSETSKDSSTSLDAIVEF
IVGNIQATSPCLHPTDLQKDSVFSVRQDWDENGSFYFAKRHLI-EMGYLQGGKMAYYEMREHSVDIDVDID
EU_Pongo
ILARGGSKGIPLKNIKHLAGVPLIGWVLRAALDSGAFQSVWVSTDHDEIENVAKQFGAQVHRRSSEVSKDSSTSLDAIIEFI
VGNIQATSPCLHPTDLQKDSVFSVRQDWDENGSFYFAKRHLI-EMGYLQGGKMAYYEMREHSVDIDVDID
Branchio3
ADYCGGSKGIPMKNIKMLGDLELIGWGIRASLDSDIMDSVWVSTDHDDIAAVASRCGAQVHRRSSQVSHDKCNSWTAIDE
FIIVHIQATCPCVHPFHIQEDSVFSVRQEWPENGAFYMFTRDVE-QQGVLQGGRMGCLVMDEYSIDIDTDQD
EU_Ciona
ILARGGSKGIPMKNIVNVGGLPLICWVLRAAVDSNAFDSIWVSTDSDEVAEVASSYEVKIHRRSDEVSKDNTSSMESTQEF
AIGLLQATTPCIQPSQLLSDSVFSVRQDWAENGGFYFAKTSVV-RQGLFQGGRTGYQEMPEHSVDIDTPFD
B__47
IPARGGSKGLPGKNIKNLCGKPLIAWSIEAAKRAKKIDRFIVSTDSEEIAAVAREWGCPVLKRPDELATDETKTISVLSQINFIL
LQPTSPIRDVGLIDESNLATGRQDYKDDGNVYILSKKLV-SEGLWFGDHICRHVISRQNFEIDDEID
B__16
IPARGGSKRIPRKNIKIFCGQPMIAWSIQAALNSACFDRIVVSTDDAEIAEVARELGAEVFIRPHVLADDHTGTLPVIRHAQV
CCIYATAPFITTKDILQTYAFTASQDLEDAGQFYWGKVDSWMQELPFFTVNSCPVLLPHRVQDIDTVED
B__31                                                        ----------------------
MIAWSIEAAIESGCFDKVMVSTDDAEIAAVARKYGADVFMRPVELADDHAGTLPVIRQAQVCCIYATAPFVRPEDLYQAYAF
SVSQDLEDAGQFYWGRSEAWLKEIPIFAGNAVPILLPHRVQDIDTPED
B__41
IPARGGSKGLPRKNIKIFNKKPLIYWPINAALNSKYIDEVVVTTDDVEIADIAKNFGAYIFIRPKDLAGDLSTTEETLQHAIAIFM
TCTDVFRKPEWIDDESVFSGRQIREEDTGIICASRAYLWRNGRRIGDKVDIILNDFTGLDIHSEED
Branchio4
VLARGGSKGIPMKNIKMLGGTPLIGWVLRAAVYAEVFDSIWVSTDNERIANVARDYGAQVHYRSEEVSKDTTSSWETVGE
FIIANIQATSPCVHPFHIQEDSVFTIRQDYDENGAYYMYTRDIH-AQGKTQGGRIAWLEMGEYSVDIYTDLD
B__55
VLARSSSKRIKNKNIIDFFNKPMLAYPIEVALNSKLFEKVFISSDSMEYVNLAKNYGASFNLRPKILADDRATTLEVMAYHIAC
CLYGASALLQEKHLKNDYVFTCTQDLKDAGLLYMGKAQAFKEMRPIFSQNSIALELSLEVQDIAHFRR
B__28
IPARGGSKGIPRKNIRLLNGKPLISYAIEVAKKSNLIDKVVVSTDDIEIGNIAKKYGAEVIMRPDHISSDEVPLDPVIHYTIVVTV
QPTSPLLSIFTLENDTVLTGRQYLPESGAVFATKRKCI-TPNNRMGENITIYVVGEESIDIDSYTD
B__50
LPAKGNSERVASKNTRLLDGKPLFLHTLEKLLQLGDDYEVYLDSECPEIFKMASHLEGFKLIRDPALATNATDGNKLFLNEVI
AQHLCTSPFIENNTIKNDSSFLVSADIEETMGLYVCSREAALDTERRIGRNPCLIEAKLEAVDVNYEED
Ostreolu1
VPCKGTSSRVESKNTKLLDSKPLFLHTLEKLLECPFIDKVVLDSESDDVFKVAEFLASDNLKRDEALATNKTDGNTLFMNEI
YVQALCTSPFIHPETIKKDSACTVSVDLPETMGLYNITRDAAKLTRRRIGNKPHMLHVTEEAVDINYEVD
```