

## **DESCRIPTION of ADDITIONAL DATA**

### **Supporting Information Index**

- 1) Materials and methods for genomic Southern analysis, Figure 1**
- 2) Amino acid sequences of genes listed in Table 2**
- 3) Accession numbers for sequences used in Figure 4**
- 4) Accession numbers for sequences used in Figure 5**
- 5) Expression of *C. parvum* phosphoglucomutase in *T. gondii***
- 6) Table of primers used for RT-PCR experiments, Figure 6**
- 7) Phylogenetic tree of aldolase**
- 8) Alignment files for phylogenetic trees in Figure 4**
- 9) Alignment of 1,4-alpha-glucan branching enzyme sequences used in Figure 5**

## 1) Genomic Southern Materials and Methods

Primers used to generate *C. parvum* probes for genomic Southern analysis.

Name	Size	Forward Primer	Reverse Primer
aminopeptidase N	1040	5'ATGGGCAATTTATTTCGGAAAATGTG	5'ATTACCTGTGTAGTTATGAAAGTATTC
glucose-6-phosphate isomerase	1040	5'ATGCCAGAACTTTATGAACTTC	5'GCCTGACAATATGGGAGAACTGCA
glutamine synthetase	1443	5'ATGAATAACTTCAAAGTTTCC	5'GTCATAATACAGTTCAAACCTC
leucine aminopeptidase	1572	5'ATGGTAAATCCAATACATCCAG	5'AGTAACTAAAGTCTTGACGCC
pteridine transporter (BT-1)	699	5'GATAAAAAACAGACAGAAGATCTC	5'CAACCTAAAAAATGAACTTG

The PCR products amplified to be used as probes were purified using a Q-Biogene Gene Clean II Kit and probe DNA was labeled using the Invitrogen Random Primers DNA Labeling System and P32 labeled ATP (Amersham). Hybridization was performed at 42°C in a 50% formamide, 10% 50x Denharts, 25% 20xSSC, 0.5% 20% SDS, 1% boiled salmon sperm hybridizing solution. Five post hybridization washes were applied 15 min each at 42°C, from least to most stringent: 2x SSC, 0.5% SDS; 2xSSC, 0.1% SDS; 1x SSC, 0.1% SDS; 0.5xSSC, 0.1%SDS; 0.1xSSC, 0.1%SDS. Kodak Biomax MS film was used to capture the images.

Gene Name	Expected size (bp) BamHI	Expected size (bp) EcoRI
aminopeptidase N	6118, 5245	NA
glucose-6-phosphate isomerase	21848	NA
glutamine synthetase	NA	2639, 808,83
leucine aminopeptidase	6453	NA
pteridine transporter (BT-1)	16,603	NA

## 2) Amino acid sequences of genes listed in Table 2

>222\_2 tryptophan synthetase beta subunit

MTDNTYLNNFPDSNGYFGKYGGKYVPEVINNAMKEIEDAYNKISKSEDFINELKKIRKEFQGRPTPIYYAK  
NLTKKYGGAEIYLKREDLNHTGAHKLNHCMGEALLAKYMGKKKLI AETGAGQHVALATAAAAYFGLECEIH  
MGEVDVKKEYPNVIRMKILGAKVVCVEFGDKTLKEAVDSAFEAYIKDIDNTFYAIGSVVGPFPFKMVRDF  
QSI VGI ESREQFLEMNGDILPDIVTACVGGGSNAMGIFSGFISDKQVELICVEPLGKGNKIGEHAASITY  
GSEGIMHGFNSIMLKDEEGNPSKVHSIASGLDYPSVGP E IAYLNSIGRTKTVCITDQEAINGFFELSRTEG  
IIPAIESSHAIGYVLKIAKEMKGKKILINLSGRGDKDLDFV VQNYGYGNN

>239\_15 acetyltransferase EAK87438

MISSFEVRKATIDDYFELRNLICDVTRCTETLSREQAEERFRYNTYHPYCLVDTENGRIVGYAGFYIIPHL  
GRKNDSRIEHV IISKEYRNRGLGRLLCKQI IEDAKNKFNCGRIDLTVESHIAKKLYSSLEFEKVNTEVMRN  
SFLDLTPKSD

>246\_15 alpha-amylase

MVINGNEWYKNCNWHYHIYPIGLCGVERINSDKIVRVNKL EELGSKSWIDHLKLNLNIGGIYIYIGPVFESEAHG  
YD T D L L S I D K R L G S N D F K N L V K I Y H S N G I K V I D A V F N H V G R N F F A F N D I K I N G K H S K Y C D W F K G L D F S  
K K S Q K G D N F T Y T P W E G Y Y E L V T L N H E N Y E V K K Y L F D A V E F W F K E Y E I D G L R L D A A D C I S I E F W R I F R D F C K  
T K F G E N F A L L A E I H G D Q R I W V R P N P S A S I K Y E Y T N Q P F D G V T N Y V L W D A I W K S H R F N N L E L L A N V I S Q O N  
D L F S S G W M Y N F V D N H D V T R I A S Q I E I E D D L L T V Y I I L F M L N G S P S I Y Y G S E F R F K G V K G K R R A D F Q L R P K  
L T I K D L E W L N S D K N D G F L V L I K F I S Q I R G H P I I G K I L T K G N Y S L L L N T K T L L V F E R R I K E E F I F V A I N I D T  
K S V E D F E V D W K E K D G Q W R D I L S P S D T Y Q S K K G K I K I N V L S N W A I I I S S C I P K T S Y N Q Y K I K K F K V L Y N I P  
T L D R S E L I E N

>32398905\_Cryptosporidium parvum 1, 4 glucan branching enzyme

MSTRGV LQKY N M K S L K E Q K I D T T L S G K Q K T E C K V K S S S K F T G E V L I T S K N I S S P T Q S L E K S D S L P V S S S D  
N S R I E S S V P T A A N S P T S Y S S S M A S I P D E R G D I Y E A L D R V P D D G K I L D H L H L L N K A L H P E P F G V L G Y Q K F  
G D D G Q Y E Y Y V R A W V R N A R R I Q I K A I D Y S F S S I G T N N T P V E M E Q R Y I D G N P S W M F E K A F K T L K E I P T C R  
T H F C G Y D E N L E N S N I E N K R I G N Q S N K D D Y S T D K F S D V G S E A N N K V N K N T I V S K K E I T A D C N S I V E K T Y S  
N V I V N T L N K S K T N L V P N K E S H K G N E M P K L E L N Q S N L M S S N I P R N K N I R S E A C S S D E A C N C Y V H D N I N S Q E  
C V R K L Y E L L V E Y N G D D S G K V Y A I R D T Y S F G L L L S N G M E L F Q S G S C W H V D N I L G S H I I E Y N G V K G V R F S  
V W A P H A K Y V R V V G D W N S W D G R V N P M R F R H G V G I W E L F I P H L G P G E K Y G Y E I H S Q S N D V F V K I D P Y S Q E Y E  
V P P R Y A S I I S A C D D S Y K D E T D R F R W E D Q E W I K R R E Q L G S K G E M R R Q P M S I Y E V H L P S W M R R E N G D Y L G Y R  
E I A E R L V E H V K N L N F T H V E F L P L A Q H P F E G S W G Y Q V T G Q Y A P Y S R L G T P D D F K Y L V N E L H K A K I G V F I D F  
V P A H F C K D A W G L V Y Y D G T P T Y E Y G D P R E G E H K Q W G T A V F N F R R N E V R S F L L G A A Y H W L R R Y H I D G L R I D A  
V S S M L Y R N Y L R P N G E W I P N E F G G D A N L E A V S L L Q E L N W V I H K E F P G V F T M A E E S T A W Q G V T H K D G G L G F D  
A K W D L G W M N D T L S Y L Y T P P D N K S S K H N K L T F R G L Y M S H E H W V L P L S H D E V V N G K G S L L D K C G F T G A P Y M D  
R I R T L K T L F G Y Q V G M P G R P L L F Q G A E I A Q G R E W K E N R S V D W H E G E E D V R K K V C I F L S D L L A V Y K N N V S L H  
A G D D E S W N F Q W D C E N S Q D C I I A F L R K Y K E W N D V V V V C N F S S R R Y N H Y P I G V P H G K E W L V M L N S D D W K Y  
G G A M F G P G N N S T V H A S H G G R V G W D Y C L W I D I P E F S C M Y L K P L F N P N E N D S

>32398951\_Cryptosporidium parvum 1,4 glucan branching enzyme

MSELRNEKVEWSQIKYEELPNYIKKQEKMGAMFYKEFGICIFRLWVVHCKNVWL VNSGTNESFLMENDD  
EINIKRVIMKNVNMNERYHFLLET SNGEKITRRDPYARYTDYD TDSSYITD T L G N F N K S I R E Y L K K H E F N  
H L I I Y E L H V E S Y I R R F S E Y F P L Q S E A K F S K N P S F F R Q I A D Y G L D N I K D L N F N C L E L M P I V E Y C G E W G Y N P  
R L L L S I N S H L G S V D D F I Y L V Q K V H E K D L F I V V D L V L H H G A S K L N S L W N F D G Y N H Q G G I Y F E G G D T G W G A  
K F N F H K K E V Q D M L Y E A C R V L L G E Y G V D G I R F D S V H N M P N W L L K N I T F R L K E E F P G R F L I A E V V P E N P N Y L  
K E C G F D S C W I H S S Y D I I S Q F R N Q K T Q Q E W N S G Y S K S L I Q H S G F D S T S Q C I L T M L G N H D Q A G N R C N G G N  
P S G D D R I G R Y I V D Q F G G R Q N W D A R A Y C R L L Y S L S C I S F G I P M V F M G T E N L Q G K W S A K D K A Y N Y D W G L I K  
N N D Q V T Q Q M R K L V K D I N L L K T K E K E V F A S N Y N N F S H F N V N T S N F V V V F I R K S K D A F L C V V N M D K T E W L E  
N N Y Q V K I D Q N I I Q S F G R E L K Q E F N S Q S E E Y G G W D G S Y T S P K K G Q I K Y S I Q D S A E N N V S Y M L N V P K Y S V T I  
Y R F C N L

>258\_55 superoxide dismutase

MSIRFIFNNSKLVNKFCLFSGMPFELPLPYDKKALEPVISPETLDYHYGKHHAGYVTKLNSLIKSTE  
FENETDLMKVIMKSSGPIYNNASQIWNHTFYWSCLRSPSETNKPTSKVSKLIEESFGSFESFESFTANAT  
GHFGSGWIWVVIDPINNQKLIKIVQTHDGDNPEKLGKPVLTCDVWEHAYYIDYRNNRGSYVDQFFSIINW  
DFVESNLS

>261\_4 DNA-3-methyladenine glycosidase

MKFNYTERELEVLRDEKMKKIEEIGYIDREYIDDLFQGLFYIIIGQOVSOKAQSIVNKAKTTLKSID  
PETISNYSLEEIRKVGVSLLKATFIKGIKAEKINKEIDLNLLHEKDDEEVCEELTKLNGIGVWSAEMAMIF  
CMNRKNVFSFSDIAIKRALKMIYGHKEITKEIFEHYRELFSPYCSIVSLYLWEISNGDYKDKLSKKNVKLY

>265\_16 RNA methyltransferase

MDDIKINKDPKDLPEFYLVLSNISKRQNFQTLRSACFGVSEVLVVGKLMFTFGNKGTLPKLSLTQYEN  
IDQVVDIIKEKEMDLVGIKISDESRIYPHPFKRSTAFLLGNEGTLGSKYIKLCDYLIHIPLYGNGTASL  
NVAIAGSIVFHHFAIWAKFTESKKGAKYIIQDDSNNSNNNNNNNNHNDISNVKGIKIRHYLFPSELGTFFS  
F

>266\_37 peroxiredoxin

MSTLVRKLPNFTAFAVMAADGSFKKVSLSDYRGKYVVLFFYPNFTVCPSEILAFNQAKDFEKLGVQLL  
AVSVDSQYSHAARRRTPLEQGGIGPVNFPILSDSSHSISKNYGVLLEEEGIALRGLFIIDKEGVVRSEVIY  
DLPLGRSVEETLRVIDALQFTETYGEVCPANWKKGQKMSATHEGVSSYLKDSF

>267\_6 hypothetical protein A EAK88787

MKLLGVNSFALRQFVEGYRGSYIPRMSPEFLRNVNYYIENNPTLVDGYADFKHIFIPNFTEAKQSIVK  
ITNENEKYIKTGYISRRDEEIPVLSRWFPKDSPPASQLIKSKYLDIILYSKEQCEKESIMNCCLODILDD  
REKNPDWYIISIKAQNESFEVPMEPITILRNLTIEEGSGVPLKREKYLESVFVKEHAIVSS

>268\_40 Leucine aminopeptidase

MVNPPIHPECIPMSEFNESLKFVSVLKDKEYDESLKGDCLVLCVFNKDIMQSIPSFYESLDTILNGILKDII  
ETANQESSLKDENNTIHNLPKYPFKYMLISFGENKKTSPVFFNKMITQMNEFQKKNKNIQSFTVVLGF  
DNVSSDIMYFCTQFPIDNTIEKRFKGNIVGSEPKPKTKIERIHFVLRSEDTIKDCNSRVLDAQSVSRGL  
LFARDLTSAPPNYCDPVNMAKEVISMASVGLGKILQPKCEELKMGAFLLAVAQGSKSPAQFVHLYTKPK  
GEIKKRIALVGKITMDTGGYNIKHQMIHFMKGDMGAAAVFGTALSGLIKPENIEVHFISAICDNLVSR  
DAYLPGCIITASNGKTIEVGNDAEGRLLADALVYACNLKQVETVIDLATLTGANRYLFEGRYASVLGND  
ELFQMIQTCGKTVEYKFWQLPLDPEFQEMVKSEIADLNNTSEGKAPISTSAAFLECFVDKNVYAHIDIAG  
CSGMGTFNGFGVKTTLVTIIKELSSK

>269\_185 G6PI

MPELYELPSYSALKEYANKQKCIHLRDLKNEVRNSCLTVNFGEIFMDFTRQNLDEEGFELLIKLAAESNL  
MEKIKLQKGGIINSTEKRAVLHTALRSKSNIPITLSSGQNVLDVNEVNRRIKFKANAIKRGELLGSTGK  
ILKDVICIGIGGSYLGPEFVYEALRTTQEGFEASMGRRRLFLANVDPIDIRRATEGLHPETTLVIVSKTF  
TTAETILNAKTIKEWLHKALKSETAVSKHLAAVSTNIKATSDFGIPVDHVFGFDWVGGFRFVCSAVGLVP  
LSIHFGANLVQEFLDGCWMDQHYETAPVSKNLPVLLGLLSVYNSTFMDKSCVAVLPYCQALCKFPAHVQQ  
LLMESNGKSSSIDGNMLHDKIKTGAIFFGEPGTNAQHSFYQLLHQGRNTTNEFIGFAKSQCDSDQILGDP  
SNHDELMCNFFAQPDALAIGKTQRELNLEDPENLIPKFLQGNRSSISLLLPICSAYYIGOLLALYEHR  
AVEGFILNINSFDQYGVELGKVLAKDIRNIIISLKTNSPEIIONKEDKIPGPTRRLIDLRLVSKSLTN

>271\_217 glycerophosphodiester phosphodiesterase

MPGIYLIIFYMMVRKASLHLVAHRGASYFVPEESLAAAYSLAAKHGVNFIKCDVSTKDGVLILRHSPNLS  
ESTNIDDLIEFSLRRTEILYDNNNDAITGIFSWDLTYEIEIKRIKCTHEKKYNNRNPSTLNGIYDILTDE  
FLKFSVSGFDGNFPGLYIEIKYPTLHERKIDNITDKLLDALDRFGINKKTDVYIQSFEPSNLEYIRARS  
EVNLVQLLRDSENNFVYAIEDVENGGMKYVTCDEYISKFANVVSPYKEDLIDRTSTSNNTLPSSSHFV  
SKAHELGLIVVPYTFRPESILSYFKDIYEEYDYFIKTLKADGLFTDDSESIMQYVNNNTYTRLYE

>272\_300 hypothetical protein C EAK89015

MIFLSLIGAFEPFISIDNLKFIKVLNQLTGTGKIKFIKHAETNINLVMKGRCSVGNISDLKDYADLKEQSN  
PGNHQOISTNSMLEVLLKMKNFITENSKNQMSNFFLIGITSLDIFPTNDHEYVFGQSDPEKGIASISAY  
RLYQNNKLNMSRILKVMHEYGHLMKLGHCXHVCLMSVICISISELDNRTLYCCENCLNQIREKSSTSES  
KNLFEF

>273\_158 ATPase of AAA class

MKPVIGNGHLEETNLEKNLLNKRLFLDKIKSFGNKQTEKEIIDNLDRLFDSDSKDLFTDTLIDLNTVLSSA  
ENKAKKFDQHSRESDFLNTDSSSNIYHDDSVIPLLENLGVQVFLNRDNCDSHRNLWDSLGGYQDVKRQIEE  
HILLSFKYPDVLDKIVNGTRAQSNNSNRPKLILFEGPPGTGKTTSAKIIGNSIQVPLIYVSLNIVSKWYG  
ESETKLAQIFDIAKKFNEGCIIFIDEIDTLASSRDKTFMHEGSKKILSVLLRKLDFDFTLNSKTLICAT  
NRRRDIDEAFLNRVDTTVLFNLPNENERELIFKQYAKHLTFEERMVLAKMSKKLSGRSIRHVCLEAERWA  
SMILKKKEKGEYQRDEIELPTIEIYKEALKKRVS

>273\_362 glutamine cyclotransferase

MFLGLGLFLSCLIIKWNFSLSQDSLHSVQLPNIYSVKILQEFPHYHIPENHIFSENDSSYKYNKNPFTQGL  
MFLNSTILIESAGLIEGSFIKLIIEFSPMKNLQHKILPKRNWEGGIATLNHGIYQLTWTSTRVMFGYSLDLSV  
VKRYSIPLTGWGLTSDNESRIWATSGSDQLFELNVPDFDSSDNKVTIKKVIMLRCLGKPMHFVNEMEYIPE  
TKTIWGNIFQSOMIVEFSPETGECISIANLKSIDYPSKSTLLEHFDILNDVNLNGIAYHPSYQEKKISNLNG  
PVLVFTGKRWPRIYKIELTKIPIRARKNEQINESGDFDKYFEFYSSSIENPSSISNEKFEM

>274\_159 alcohol dehydrogenase

MLILTPIIYYNMDSFREDIISDLLDSNVGRAILVFS SAVKQAMRVESMVKKNGIQIEFSELCESDANALST  
ILLQMKSYKPDICIIGMGGGHCMEDIAKVLRLVLYEDPNTTLRSLAMGTNNNDKEKSSKKAMIHRRGSLIKKLV  
CIPTTCGSGSEVTSTAVLRNDDGRQMI VSGVAFLPDISVIDSSFILTIPMFVASITGMRALHGLEYSISN  
SSNHYSMCAIQLRILFNSLTKAVIEKDISLQEIHKAAISGVAISATDVGLGTILSRISIEVSTLPHG  
LIDSIVITQVLNFIKKSQT VGPLIANLSVELGISNPEDSNSNKIQALLNRIDEIKKTLLLPDSLSSIHSS  
LSNFKSKGWFDPIGVKLSVFSQCIEIDNQEFINHVS GSF SISDKETRAYVTEENALRTIPILSKKSVIDNL  
NFNTLQMNLD RMISRALSDEAIHTNPATVTKEDLSTILKETWG

>274\_402 aminopeptidase N

MGNLFGKCGCRDLTICNLNVL RPAHSQDRIYRKEYKVPNFLIDHVNLDINIKDDVTVVSSVIIMRRNPSS  
FRGDLSLDGDCLKLVSVKLVILEKSLYKGYFQPDGPDGKLVISSNLLPNKDQQTLETVEIFPDRNTK  
NMGLYYSAGVYSTQCESDGFRRITFFLDRPDVMCKFRVRIEGDKTKSPVLLSNGNLLKGDVQGSENRHF  
IFVDPPFKPCYLFVAVAGVLRLEDKFI TKS GKT VRLVFYSEPKYVDRLRLAMESLKLAMKWDEDRFGLEY  
DLEIFNIVAVESFNFGAMENKSLNIFNCSCLLASENITPDYFFTNILSIVGHEYFHNVTGNRVTCRDFWQL  
TLKEGLTVYRDQEF SRDCIDRLSEQLGDIEVLRNYQFQEDSGPLAHP IRPDSIVSTNNLYTSTVYRKGAEV  
VRMYETILGREGFRKGM DLYFARHDGQAVTCDDFR KAMEDANNYNFTQFERWYDQAGTPEVEVVSIDHNKA  
EGTCSITLRQRCSP TPEQPKRPFYIPVIVGLIGKDSCKEIRQSETLILKEQQQTFILDGVWETPVASILR  
GFSAPVNLRFKTPRSDEELAF LFAFD TDEYNRF DAAQ TLYKKILIQASTNSSQEISPSSVIESILSQTILS  
CFESFISKIKTSKNPNGTVSPMVASYTLRIPSYSQVLASIAEPNDFV VESFNALTVAFA SRFKTQIVDLF  
QVLTEDLSALPSQSPFIHGKGVNVEAIAIRLLNILLDFIAKLDPSLGTKLAFEQHNRFDYMTSKLGAISA  
LQIADHKSSECIQALD SLLATSDDVSTLNQYFSIQASCLIPDNVERVINIYHSNPQFIKYRENPTIFSSL  
VGTFA SNFVAFNRKDGLGYSFVADAIILMDKVNPMASASVSRAFTKVLKLDEGRNLLKENVVRILOSPGL  
SKDTSEILKSIDF

>276\_106 glutamine synthetase

MNNFKVSSSFVCRSFD ELYSAIQTENIEYLNYSFCDTFGSLHHITISSNSFKSGKDIQKGI AFDSSSVRGMQ  
FGDFSDMIVVPDFEKVWIDPFFDRKTLHVACFVLFHTGEPSPGCVRSISKRAQEALSLSGIADKCFIGPEI  
EFFVFESVSYNNSPNHSYFCLDGDEAYWNSGNATFYPHNRPNLASRRPLKQAYCAPYPVDRDLSLRSEIL  
ELENIGVTVEKHHHEVATCQHEIGVHCSTLVQSADIVESIRYLIKGI AHRNNKTATFMPKPLGNDNGSGM  
HINISLWKNEKNIFFDPESSY NLSNQALYFIGGILSHAKAIMAFTNPTTNSYKRLATGYETPTKLSYGAG  
DRSSSIRIPLSGF SCTKTQRIEFRI PDSSACPHLAFSAILCAGIDGILHKIHPAAYLNTIKSQSNEYNLPK  
SLTEALQNL EEDYEFLIKDGVFSHEFILNYIKLKRDEVVLVESFPTPKEFELYD

>276\_511 Pteridine transporter (BT-1)

MDDKKQTEDLTAKNQLFSLKEFLKPIIWGPAIYIFTYTAGPDYDDAMFFYFTNRLGFSPTFMGSLRLTYG  
IAGIIGIVLYRIILKKT PPREILLWTTLFSIPIYILPLALVTGLNLMGISNRMFALSGGFLIEAIAEIQL  
LPLLVM TAKFCPKGLEGSVYAVMMSIRSLGIGVSKVISAGLAYS LGITAFNFSNLGLLIWISSAFLLLPLF  
FLNLVSYHTFSIFQVHFFRL

>Cpnew84 aspartate ammonia ligase

MEVKKQSQFKTVLADIEFIKITFQQLLKESLDLIRISAPLFLEKESGLNDDLSGYEEKVSFTFNNKVLIV  
QSLAKWKRYALKKYELNGLYADMNAIRKSEELDNLHSIYVDQWDWEMVIDTNKGKTKDEILVDIARI IHNN  
IYNLERLYWNMKKEKQDIVNIPNENEMIKELYIISSEELNMYPNLSSNDREREICKKYGSVFIKQIGKK  
LSNNTVHDLRAPDYDDWEYNGDLIYWSNINLNGPIELSSMGVRVNKESLIKQLEICNSTERLKLKPYCKMLLN  
NELPETIGGGIGQSRCLMLILRKEHISQVQCSYWDEKDXRYFIIIFLKKIL

>264\_49, asparaginyl-tRNA synthetase

MIKPEKKFVTEYSVTGRLRIASILDGEDGGVSYIGKKVTVGGWARTVRKQCSDTLLFISLNDGSTSSNLQC  
VVEKTVKGFEEGLKATAGCSFKITGTIVKSPAQQQSVELLNLTGDELKICGLCDASKYPLAKKHHSKEFL  
REVAHLRPRSQFFSSVMRIRNSLAIAIHEYFQKNGFMYIHTPIITAADCEGAGEMFQVTTVLPPEKNNIS  
NIPASKIEGSQDMAVDYKDKDFGKASYLTVSGQLALENFACMSDVYTFGPTFRAENSHTTRHLAEFWMIE  
PEMAFADLSDNMKLGEGLLKYTVEYVLINMPDLLYLDKNIENGLVERLKVICKKEEFARISYTEAIEMLPK  
HDKEFTVPVSWGMDLGSEHEKYITDVELEKRPCIYNYPKDIKSFYMKLNEDGNTVAAMDILVPKIGEVIGG  
SQREDDIEKLENAIKSRNMDPAPYWWYNELRKYGSVPHSGFGLGFERLIMMVTGVENVRDVI PFPYPNHC  
EF

>gi|18996773|gb|AAL83208.1|AF426177\_1 inosine-5-monophosphate  
dehydrogenase [Cryptosporidium parvum]

MGTKNIGKGLTFEDILLVPNYSEVLPREVSLETKLTKNVSLKIPLISSAMDTVTEHMAVGMARLGGIGI  
IHKNMDEMESQVNLKVKKNWISNLEKNESLTPDQNLDKESTDGDGKDTKSNNNIDAYSNNENLDNKGRLRVGA  
IGVNEIERAKLLVEAGVDVIVLDSAHGSLNIIRTLKEIKSKMNIDVIVGNVVEEATKELIENGADGIK  
VGIGPGSICTTRIVAGVGPQITAEKCSSVASKFGIPIIADGGIRYSGDIGKALAVGASSVMIGSILAG  
TEESPGEKELIGDVTYKYRGMGSVGMKSGSGDRYFQEKRPENKMPVPEGIEGRVKYKGEEMGVVYQLVG  
GLRSCMGYLGASIEELWKKSSYVEITTSGLRESHVHDVEIVKEVMNYSK

>gi|44804693|gb|AAS47699.1| thymidine kinase [Cryptosporidium parvum]  
MAKLYFYYSAMNAGKSTVLLQSSFNQERGMKTLFTISKDCRFEEGSICSRIGLSEKAHTFTPDLLKLLD  
IINQENNEKIDCVLVDDESQFLTQFQVRELCIVVDKLNIPVLCYGLRTDFKGNLFEGSKYLLAWADKLTE  
IKTICRCGKKATMTIRLNSNGEPVFSGEQILIGDNSIYTSVCRKHIIINCEEYNF

>gi|10444017|gb|AAG17668.1|AF274310\_1 lactate dehydrogenase  
[Cryptosporidium parvum]

MIERRKIAVIGSGQIGGNIAYIVGKDNLADVFLDIAEGIPQKALDITHSMVMFGSTSKVIGTNDYADI  
SGSDVVIITASIPGRPKDDRSELLFGNARILDSVAEGVKKYCPNAFVICITNPLDVMVSHFQKVSGLPHN  
KVCGMAGVLDSSRFRTFIAQHFGVNASDVSANVIGGHGDMVPVTSSVSVGGVPLSSFQKGLITQEQID  
EIVCHTRIAWKEVADNLKTGTAYFAPAAAAMAEAYLKDRAVVPSCSAFCSNHYGVKGIYMGVPTIIGK  
NGVEDILELDLTPKQKLLGESINEVNTISKVLDNAPAAGA

>gi|32765705|gb|AAP87358.1| malate dehydrogenase [Cryptosporidium  
parvum]

MRKKISIIAGAQIGSTIALLLGQKDLGDVYMFDIIEGVPQKALDLNHCALIGSPAKIFGENNYEYLQN  
SDVVIITAGVPRKPNMTRSDLLTVNAKIVGSAENVGKYCPNAFVICITNPLDAMVYFFKEKSGIPANKV  
CGMSGVLD SARFRCNLSRALGVKPSDVSAIVVGGHGDEMIPLTSSVTIGGILLSDFVEQKITHSQINEI  
IKKTAFFGGGEIVELLKTGSFAFYAPAAASAVAMAQAYLKDSKSVLVCSTYLTGQYNVNNL FVGVPVVIKNG  
IEDVVIVNLSDDKSLFSSKSVESIQNLVQDLKSLNL

>hypothetical protein B (276\_500), CAD98502.1

MIFQYETQSFVSLVLLCDRIFLMHHARRSSLNADHTAIMVIFPSILMNLIIIDKKDINEDLTVELSSRQS  
NHCISVLEINIGSKVNVGVKNSGKGTATVTKIMRKETNKDGITKSESIINLSSIEEIGNKVNRRRSIS  
EINLNNFQYAVTIKLDTEIHKEEYKHDYPLIDLLIALPRPKVFQKVLQNAVTIGVGRIIFVCTDKSEK  
YLNSSKLLKESIDEIVQLGLEQASKTLCPDVYVYASWTFQHLRNYFFCKESMIGIVADVLGNSKITEI  
GLQSHGTPIILAIGPEGGWTKELTDLITMGFKVVNMGDRILKVETALISLYSKVSFGRTQRN

>gi|44804700|gb|AAS47700.1| uridine kinase uracil  
phosphoribosyltransferase [Cryptosporidium parvum]

MSNISLEKLIKDIYGDQALTPSSNSNVFVIAVAGGSASGKTSVCTRIFSELGDKRVTVIETDSFYKTP  
VLEEGQTMADYNFDHPNSVDFELLYNVLLSLKNGEGVHIPNYCFKQHKRLETGRKVS PASIIIVEGIFIL  
FHPKIRHLINMSIFVDTDDDIRLVRRIRRD TIERGRQIDDILNQYEKTVKPSYDEFIYPTRRYADIVIPH

YPNEVAVDLVVQHLRYKLMDDLKRIYSNLHIIPSNQIRHMHSIIRNKDTSVVDFVFWSDRLIRLVVEN  
ALSHLSFTGQTIETPIGELYDGVQFNKDKLCAVSIVRGGESMEIGLSAVCKDIPIGKILLEFQNPKTEL  
DAQFDKPKIIYCKLPDDIASRNVFILDPIILGNFGVFSAIKYLLSKGVLQRKIIIVLSLIVAHNAIHRICK  
EFPEVTLITTEIDRDVNSDGFVIPGLGCFADRYFGTE

>gi|44804760|gb|AAS47705.1| calcium-dependent protein kinase 1  
[Cryptosporidium parvum]  
MLNIEQNADECAKRVGANMVPVLHNVTEECNKAEEKEIVPEVEEAKSESNNCLDSDDYLEGHIYAAMCT  
KCAGELQQLNPTEPYRDPKKVSGKEQISRGLIIESKSFVDANKNIKFSKRSDKNEYAGLCSSPEVTTTPN  
GERETSTDSNIKNTESTKVSHGIFDRTCLIQEHALVNRNINDFYELNLGNLGRGSYGSVVKAIKQSGAQ  
RAVKIILKPKLENINRLKREILIMKRLDHPNIIKLFVFEEDTNYLYFVMEICTGGELFDRIIKRGHFSER  
YAAVIMRQVFSIAIYCHSNFMRDLKPENLLFSDSSPNSLLKVIDWGFAAKCPKTHKFTSVVGTPIYYVA  
PEVLYGSYSKLCDLWSAGVILYILLCGYPPFHGKDNVEILRKVKIGQYSLEHNSWKYVSDSAKDLIKRLL  
MTDPNKRISAQDALNHPWIKSQISSPNTADATYFTNDVCNSLLARFRDFQRQSKLKKLALTCVAYHLNDA  
DIGALQKLFSTLDRNGDGVLTINEIRSALHKIQNVSQLGDDIDNLLMELDTDGNRIDYTEFIAASIDHK  
LYEQESLCKAAFKVFDLMDGRISPQELSRVNIITFLQEAFAQSTIDSLLKEVDINQDGYIDFNEFMKMM  
MGDKHEQKLETKVQKIEDEGSGNKKLSKGGIISDIFGATSIFKLLNN

>gi|44804774|gb|AAS47706.1| calcium-dependent protein kinase 2  
[Cryptosporidium parvum]  
MMGQGGONALNGVCSRRYEKDSIEINKNKEVRRKQSNKRGNFPAHNMTHGKPEYLKEAVDRLINHGNIVGN  
SPDRYLAKERKSQASCSTSDPTSPLAYRNGNDNYFEQYECDLIELTSEMALSNESSRFESLSEYESIPKL  
SPSTTVMGITDVSATNTSNSNSATPTESEKSGPKLSLTEGKFRREGLIPACKGSIHSDYIIDSGRIGKG  
TYGSVKSGTNRLTGCIRAIKTIPLTRVEALDNFMKEINILKNLDHPNIVKLYETYQDKENIYLVMEELCSG  
GELFDRIISQGSFDEIYAANLMKQVLSTICYCHDHGIVHRDLKPENFLFLNKNYNAPLKIIDFGLAARVN  
NEDTSLNTRAGTPYYVAPEVLQKQYDKQCDMWSLGVILYILLCGYPPFHGSNDSIILHKVQKGVYAFKEE  
DWKHVSFLAIDLIRKLLTYNPSERITARDALNHPWITRFADDILFLSPRNYFYSNDGFMKTGIVSSDYID  
SSINNLKALECQKPRNCNLMGSRNGYIKGGLRYCRSKRRSISSSSGLNLLSNFRAFHKYNRFMKVALT  
VIAQQMTESQISNLKEAFILLDANCDGTLTPQEIIITGLKNSGITELPSDLLAILNDIDSQDGSIDYTEF  
IAATLDSKQYSKEQVCWAAFKVFDQDGNKGITANELLNVSFYNSEQGSAGINDKALSDVKNMIKEVDVDG  
DGEIDFQEFLEMFRRSNS

>gi|44804783|gb|AAS47707.1| calcium-dependent protein kinase 3  
[Cryptosporidium parvum]  
MTVATSSRRNSRRSSLSESSLTSHTSSTKPKNSYIQKNSSVSGGNGTSHNCYSHKGRSGSGFGSSSSYE  
NRAYLVSGGGSCIGNNTSMGNANGSLLAPNVQLDNLPSERVGNGSPNWALSADVCSLKDFKELIECKNEH  
AESVLRKYSGIKSVNGGRGIRQVLASIVRGSRTSKNRSSSVDSGSCGMVSSSTSTTTSSAALVSNLNRKN  
TTGGDINNTRIKIEDSIDNGTIINDNNGSSTSNRECNILTGQPIQKAILTORDIMELRGKIDKVLSTED  
SFVRSALQSFQDFDGDGNLTIEELLDLLTTLGEHLALPPINRKVVANEISIRLNSKLTVDSSASESKM  
LISFPFFLKYLYSVLTTIRQKHFSSVKINGELQQNKAIRKHLVHEDDINDLYTFHYQLGTGTGEVFLVT  
ENYTRQRRVCKIVDKVKCRRLDEIDHEVEILKQLDHPGLVHIYEVYEDRLNLYFIFLFCAGGNLFHSLQ  
DAIQSGFRFSEFHVSRIIQQILLAIRYLHLKRVVHKNLKPQNILLTNSFGSGNTSVKLVDFGLAEIFSSD  
EINLPELTTSNVSPSHFHSPLKSLPLSSSSSSSYITQDHIKTHNQAQPAQIQGHHTNTSLSSISNQRA  
SEEGYLFASVDEVGTGIHKSNSHNSPSSNSLSLAQNLHQEESLASYPKYLDFTAPELLKGETFSYSLDVW  
SVGVIMYSVITGRHPFSGRSRAETRHNICFGIPAISEITCVSNACKSLGKLEKNPRERITVDEALSHD  
WFSISHSSYCEVDIGMPLLAHLKVYTRQSDLRHILHMLSHQLALDTTQINMATSVFKSLDTDNDGVLSI  
SELGSLQKLGISSRESSMIKAMDIDNGIISYSEFITACHIWRRNEIRQLKSFFMKIDRNNQGKITRE  
EFKQLLKAQKSKLLSNITKFSSTNDVNYRTILPSSNQGVYTEWDSVLDEIDTNRDGLIDWKEFCTFIVD  
YFNTRSRESLRACYNSSISNISSTFSIDPPSVAGLSRRRDSGICPVQEANQHPPHPSQLD

3) **Genbank or other accession numbers for sequences used in Figure 4.**

**Leucine aminopeptidase**

1. 23509661 Plasmodium
2. TA09045 Theileria (Sanger GeneDB <http://www.genedb.org>)
3. 46227265 Cryptosporidium
4. 15235763 Arabidopsis
5. 1703282 Solanum
6. 23042880 Trichodesmium
7. 17227733 Nostoc
8. 15607055 Aquifex
9. 24216140 Leptospira
10. 15611584 Helicobacter
11. 27364860 Vibrio
12. 17547134 Ralstonia
13. 13476527 Mesorhizobium
14. 27887802 Fusobacterium
15. 29840173 Chlamydophila
16. 16080258 Bacillus
17. 19074788 Encephalitozoon
18. 19909972 Coprinopsis
19. AY581147 Dictyostelium
20. 19113905 Schizosaccharomyces
21. 21355725 Drosophila
22. 7705688 Homo
23. 29832567 Streptomyces
24. 21674006 Chlorobium
25. 5870602 Leishmania

**G6PI**

1. 15595075 Borrelia
2. 16752000 Chlamydophila
3. 46227265 Cryptosporidium
4. 4835820 Toxoplasma
5. 23509563 Plasmodium
6. 15239045 Arabidopsis
7. 1169797 Oryza
8. 37960111 Phytophthora
9. 15804618 Escherichia
10. 23136626 Cytophaga
11. 17510665 Caenorhabditis
12. 17737445 Drosophila
13. 18201905 Homo
14. AY581147 Entamoeba



15. AY581146 Dictyostelium
16. 6319673 Saccharomyces
17. 7437359 Trypanosoma'
18. 21673816 Chlorobium
19. 15964236 Sinorhizobium
20. 15806745 Deinococcus'
21. 29832844 Streptomyces
22. 19173604 Encephalitozoon
23. 23125809 Nostoc
24. 18030019 Spirochloa
25. 13924524 Giardia
26. 18030022 Trichomonas
27. 15644137 Thermotoga
28. 16080187 Bacillus
29. 15669801 Methanococcus

**Tryptophan synthetase beta subunit**

1. 34557127 Wolinella
2. 46226438 Cryptosporidium
3. 22958844 Rhodobacter
4. 28900440 Vibrio
5. 33456993 Cycloclasticus
6. 16079321 Bacillus
7. 14520675 Pyrococcus
8. 11499193 Archaeoglobus
9. 15606106 Aquifex
10. 32417954 Neurospora
11. 24215988 Leptospira
12. 15642912 Thermotoga
13. 29345943 Bacteroides
14. 34557707 Wolinella
15. 15963781 Sinorhizobium
16. 17546702 Ralstonia
17. 1174780 Zea
18. 17231286 Nostoc
19. 33239640 Prochlorococcus
20. 15805965 Deinococcus
21. 29832719 Streptomyces
22. 28898734 Vibrio
23. 15612264 Helicobacter
24. 19703662 Fusobacterium
25. 29840324 Chlamydomonas
26. 14520795 Pyrococcus
27. 18313363 Pyrobaculum
28. 15606594 Aquifex
29. 11498839 Archaeoglobus
30. 15643305 Thermotoga
31. 29350072 Bacteroides

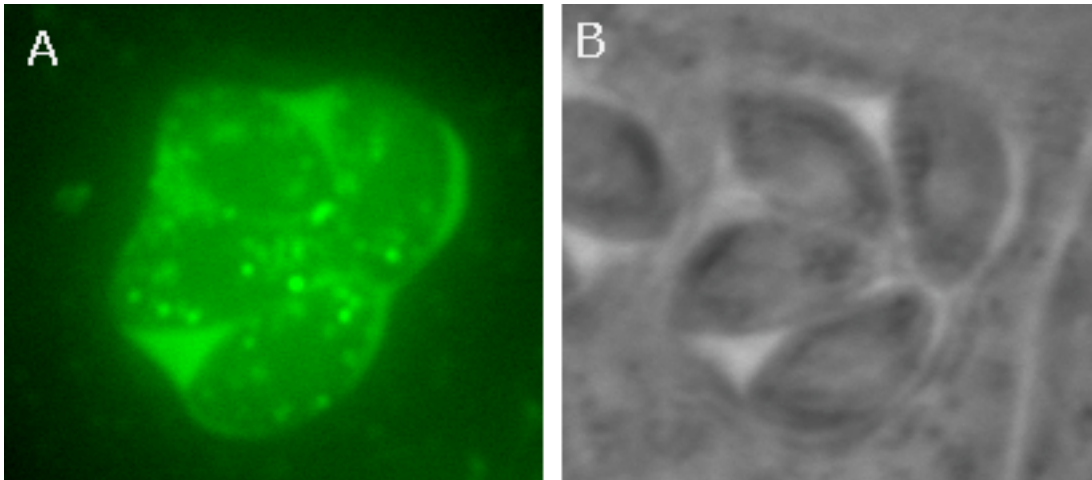
4) GI accession numbers for the sequences used in Figure 5, 1,4-alpha-glucan branching enzyme sequences are listed in the order in which they appear on the tree.

45548447 Rubrobacter  
29829347 Streptomyces  
45916434 Mesorhizobium  
22986676 Burkholderia  
23468472 Pseudomonas  
22968855 Rhodospirillum  
15806848 Deinococcus  
15618386 Chlamydomophila  
22971471 Chloroflexus  
15606119 Aquifex  
22999765 Magnetococcus  
32398905 Cryptosporidium  
995313 T\_gondii  
32472075 Pirellula  
23021493 Clostridium  
17228208 Nostoc  
45509261 Anabaena  
23126496 Nostoc  
23474508 Desulfovibrio  
18309045 Clostridium  
19704191 Fusobacterium  
16080150 Bacillus  
32470978 Pirellula  
45916029 Mesorhizobium  
45548449 Rubrobacter  
22968918 Rhodospirillum  
22986678 Burkholderia  
23471845 Pseudomonas  
23475793 Desulfovibrio  
22968087 Rhodospirillum  
22971470 Chloroflexus  
23124761 Nostoc  
46135186 Anabaena  
23130645 Nostoc  
32398951 Cryptosporidium  
XXXXXXXXT\_gondii  
23050680 Methanosarcina  
23123689 Nostoc  
23135626 Cytophaga  
29346181 Bacteroides  
28829250 Dictyostelium  
171569 Saccharomyces  
16416077 Neurospora

4557619 Homo  
17554896 Caenorhabditis  
32564391 Caenorhabditis  
28573410 Drosophila  
30680140 Arabidopsis  
15227650 Arabidopsis  
2801805 Gracilaria  
1169912 Solanum  
29245217 Giardia

5) **Expression of a *C. parvum* GFP reporter construct containing the N-terminal extension region of phosphoglucomutase in *T. gondii***

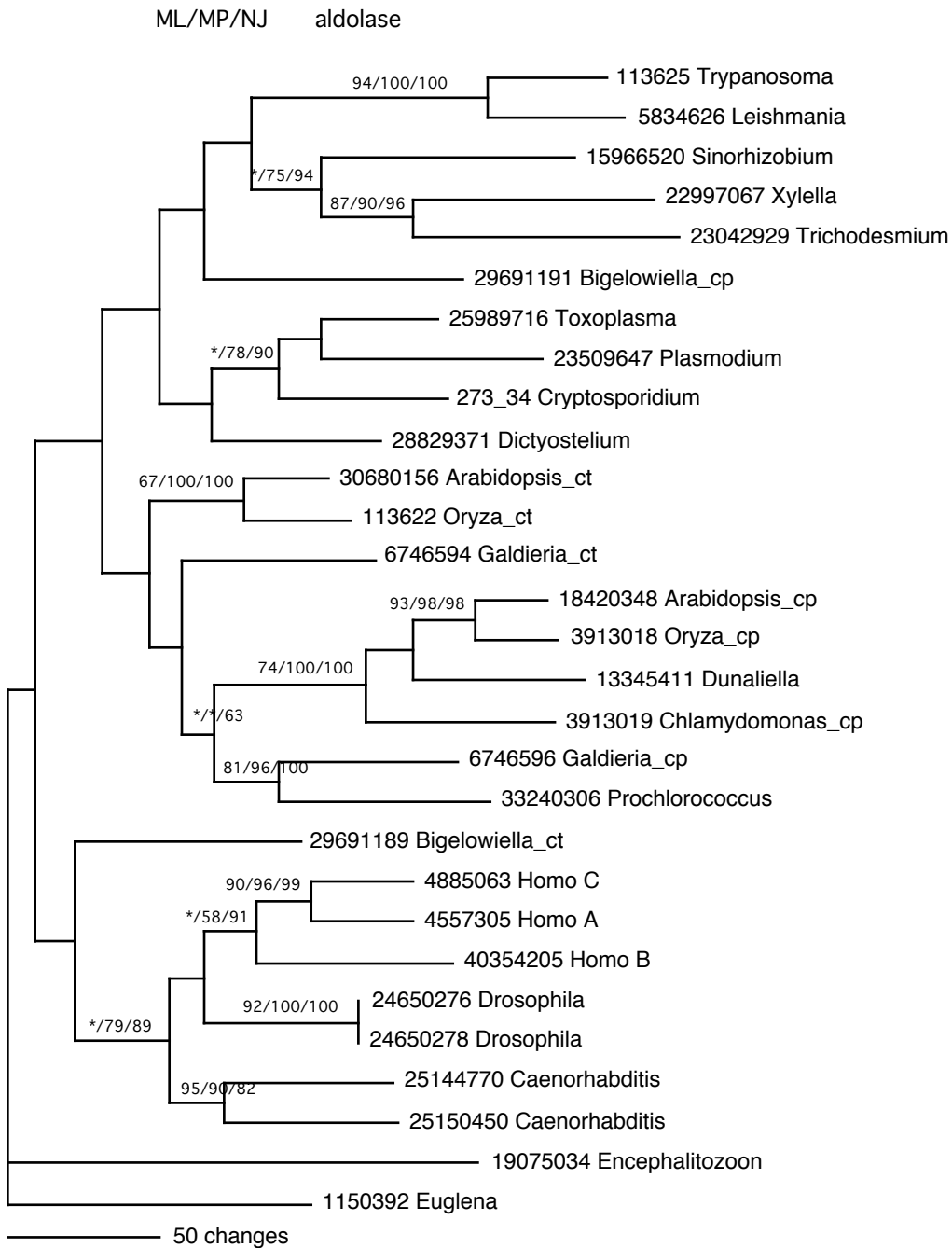
The *C. parvum* phosphoglucomutase N-terminal extension targets a GFP reporter to a dense granule secretory pathway in *T. gondii*. The secreted protein is retained in the vacuolar space surrounding the parasites. A) Fluorescent image illustrating the location of the tagged protein in a single vacuole containing 4 *T. gondii* parasites. B) Phase contrast image of the same vacuole.



6) **Primers used for semi-quantitative RT-PCR experiments**

<b>Gene Name</b>	<b>Forward Primer</b>	<b>Reverse Primer</b>
glucose-6-phosphate isomerase	5'GCACTAATGCTCAACACA GTTTTTAT	5'GTTGGCCAATATAATAT GCACTACAG
leucine aminopeptidase	5'ATTATCACTGCCTCTAAT GGAAAAAC	5'ATTCACATTCTTGTCT ACAAACTCG
biopterin transporter (BT-1) family protein	5'TTGCAGGTATAATTGGAA TTGTTCTA	5'CAAGACTTCGTATTGAC ATCACAACC
calcium dependent protein kinase	5'GAAATTAGCTCTTACATG TGTTGCAT	5'ACTTCCTTTAACAGCGA GTCTATTGT
dihydrofolate reductase-thymidine synthetase	5'AACCTTTATCAAAGGAGT TGTGACTT	5'ATCTCCATTTTAATAGT GGGATATGG
tyrosyl-tRNA synthetase	5'TCTGCAATATTTGTTGAA GATACTCC	5'GTCAGTATTAAAATGGT CACGAATTG

## 7) Phylogentic tree of Aldolase



## 8) Multiple sequence alignments used to create phylogenetic trees in Figure 4

### Leucine aminopeptidase (Fig. 4a)

CLUSTAL X (1.81) multiple sequence alignment

```
_23509661_Plasmodium_falci
Theileria_annulata_sanger
268_40_Cryptosporidium_par
_15235763_Arabidopsis_thalia
1703282_Solanum_cp
23042880_Trichodesmium_erythr
17227733_Nostoc_sp
15607055_Aquifex_aeoli
24216140_Leptospira_interro
15611584_Helicobacter_pylori
27364860_[Vibrio_vulni]
17547134_Ralstonia_solan
13476527_Mesorhizobium_loti
27887802_Fusobacterium_nuclea
29840173_Chlamydomytila_caviae
16080258_Bacillus_subt
19074788_Encephalitozoon_cuni
19909972_Coprinopsis_cinere
Dictyostelium_discoide
19113905_Schizosaccharomyces_p
_21355725_Drosophila_melano
_7705688_Homo_sapien
29832567_Streptomyces_averm
21674006_Chlorobium_tepid
_5870602_Leishmania_major]

ASQLIAAPSNYCNPVSLSNAAVELAQKLNLEYKILGVKELEELKMGAYLS
AKELTTAPPNYANTVTVSEFLKMLSSVGLKVKVLDYDDCKKLMKFCYLS
ARDLTSAPPNYCDPVNMAKEVISMAKSVGLEKILQPKCEELKMGAFILA
GKELVNSPANVLTTPAVLAEALNLASMYSMNTANILNEEQCKELKMGSYLA
GRELVNAPANIVTPGALAEAKIASTYSITVNIILDAEQCKELKMGAYLG
ARELVAAAPANSCTPITMAETAQTLAKEFGLTLEILEKEDCEKLMGAFILG
ARQLVAAPANSVTPITMAETAQQIAQDYGLQIEILEQEDCEKLMGAFILG
ARNLVNEPQNVINPITLAEAEAKKLAEEFLECKVYDEKQIQEMGMALYS
VRHIAHL PANHFTEPEFVSRSEIAKDNLKTI VTFDEPQLKKEKMGGILS
VRDLVNTPPMIGTPVYMAEVAQKVAKENHLEI HVHDEKFL EKKMNAFLA
CKDLGNMPPNIANPAYLASQARRLADDEYIETTKIIGEQEEMKLMSSYLA
TRDLGNLPSNICPTPYLAN TARQIAKDFKLVKVEVLGRKQIEALKMGAFILA
ARDLVNEPANILGPVEFAARAKELE-TLGVKVEILTEKEMKLMGMSLLG
VKDLVNEQANVNLNPKELADRATKLGKEFGFNVEILDEKKAQKLMGNAYLA
TRDLVNGNADEVTPQKLANIAKGMMAKEFPM DVKILNKDAILKEKMGLLAA
ARTLVNMPGNMLTATDLAS YAEELAAKYDFECEILEKSEMEELGMGGILA
ARFLGDT PANLMNPTLFVEYATKYLDKQKVTFEVDFKSFMERKSMNLLLG
ARTLM EYPANMMP TLFTEERVKQEFAGIPVEIIVRDEAWAAEKGMMNPLS
ARVLMETPANLLPTNFVQHVSSQMKELIVEMIVREEQVVKDQKMGFWG
ARSLMECPAN YMTSLQFC HFAQELFQNSVVKVVFHDEKWI DEQKMNGLLT
TRQLQEMPSNLLPTAFAQNVVEVLCKSGVNVVEVKVEGWAESQSMHAFILA
ARQLMETPANEMTPTRFAEIEIENLKSAS TEVHIRPKSWIEEQAMGSPLS
ARDLINTPPNDLNPESFAAVAQAAAKEHGIKVQVLDEKALTKGGYGGILG
ARDLVNLPGNHLSAEDLAEAAIEAGKRGGEVTFVDFK KIVELGMGGLLA
ARNLGNLREDEGVPEFYAEAWAKKYMIEPGKVRKVLRGQQLLEEAGLHLMYNI
:
:
:

_23509661_Plasmodium_falci
Theileria_annulata_sanger
268_40_Cryptosporidium_par
_15235763_Arabidopsis_thalia
1703282_Solanum_cp
23042880_Trichodesmium_erythr
17227733_Nostoc_sp
15607055_Aquifex_aeoli
24216140_Leptospira_interro
15611584_Helicobacter_pylori
27364860_[Vibrio_vulni]
17547134_Ralstonia_solan
13476527_Mesorhizobium_loti
27887802_Fusobacterium_nuclea
29840173_Chlamydomytila_caviae
16080258_Bacillus_subt
19074788_Encephalitozoon_cuni
19909972_Coprinopsis_cinere
Dictyostelium_discoide
19113905_Schizosaccharomyces_p
_21355725_Drosophila_melano
_7705688_Homo_sapien
29832567_Streptomyces_averm
21674006_Chlorobium_tepid
_5870602_Leishmania_major]

VGKGSMPYKNF IHLTYKSGDVKKKIALVKGKITFDSSGGYNL KAMIDLKMF
VCQGSKFPPKFLHAYVYKPGEVKKT LAFVKGKITFDAGGYNVKNMIHYMKL
VAQGSKSPAQFVHLYKPG EIKKRIALVKGKITMDTGGYNIKHMIHF MKG
VAAAANPPHF IHLIYKPGPVKTKLALVKGKITFDSSGGYNIKTLIELMKF
VAAAENPPYF IHLCFKTRERKTKIALVKGKITFDSSGGYNLTKI ELMKN
VAQGSDLPPKF IHLVYKPTPRR-KLAI VVGKITFDSSGGLNLKVG IEMMKI
VALASDLPPKF IHLTYKPTPKR-KLAI VVGKITFDSSGGLN IKGDIETMKI
VGKGSATPPRF IHLIYKPGPKPKI ALVKGKITFDSSGGLNIKPYMR TMKM
VCEGSDKAKMILLEYTPPITK KKLAIIGKGLTFDSSGSIKPKDMHEMKY
VNKASLAVNPPRLIHLVYKAKK KIALVKGKITFDSSGGLSLKPKYMTMKA
VGRGSKNESMMSII EYKGDSEAKP IVLVKGKITFDSSGSI LKPGMDEM KY
VTKGSQEPQF IVLRYEGPAKQAPVVLVVGKITFD TGGISLKP GMDEMKF
VAQGSFPRGAVM QWNGPKDGP IAFVVGKVTFD TGGNSMKPKGMDMKF
VARAAYYRPMVI IMRYKGSKYT-YGLVKGKITFD TGGLSLKP SMLTMRC
VAKGSVDPCF I VLT YQGSKDH-TVLI GKGVTFD SGGLDLKPAMLT MKE
VNQGSTEPPKMI VLKYQGEWEDVVLVVGKITFD TGGYSIKTGI VGMKS
VSQGSAQEPKLL VARYRGDAVD--IALV GKGVCFD SGGISLKP RMRMKG
VTRGTSEPAKFL EIH YKGDKNAPLAFVVGKITFD TGGISLKP GMKLMRG
VAKGSDEPLKFL EHLHYRGADGKDSIVYVVGKITFD SGGISIKPNMGLMKG
VNAGSDIPPRFLEVQYIGSKDDGWLGLVVGKVTFD SGGISIKPNMKEMRA
VGKASCEPPIFL ELSYGAER-PIVLV GQGITYDAGGLCKKELFNMRG
VAKGSDEPPVFL EIH YKANEP-PLV FVGKITFD SGGISIKANMDLMRA
VGAGSAAGPRLV KLSYTSKAKKHLAFVVGKIT YDSSGSI LKPHNETMKC
VNGKSEQPPTF VILDYKPGKAKK TIALVVGKVTFD SGGISLKP GMDEMKS
VGRGSRYP EYL VVLE YIGRSSATTAI VVGKVTFD CGGLNIKPSMETMHS
* .: :*:* : * * : * :

_23509661_Plasmodium_falci
Theileria_annulata_sanger
268_40_Cryptosporidium_par
_15235763_Arabidopsis_thalia
1703282_Solanum_cp
23042880_Trichodesmium_erythr
17227733_Nostoc_sp
15607055_Aquifex_aeoli

DMSGCAAVLGCAYCVGTLKPNVEIHFLS AVCENMVSKNSYRPGDIITASN
DMGGFATVFS AAYALGKLPNVELHFVGLCENMLDANSYRPSDVVTASN
DMGGA AVFGTALS VGLIKPNIEVHFISAICDNLVSRDAYLPGCIITASN
DMGGSAAVLGAAKAIGQIKPGVEVHFIVAACENMISGTMRPGDVL TASN
DMGGA AVLGAAKALGEIKPGVEVHFIVAACENMISGAGMRPGDIVTASN
DMGGAGTTFTG IKAIAQLKPDVEVHFISAVTENMVSGHAIHPGDFLTASN
DMGGAATLGAAKAIAQIKPNVEVHFISAVTENMISGKAMHPGDILTASN
DKSGCAVGLIMRAIAQLKPDVEVHGLIGAAENMPDGNAYRPDDVIKAKN
```

24216140\_Leptospira\_interro  
15611584\_Helicobacter\_pylori  
27364860\_[Vibrio\_vulni]  
17547134\_Ralstonia\_solan  
13476527\_Mesorhizobium\_loti  
27887802\_Fusobacterium\_nuclea  
29840173\_Chlamydomphila\_caviae  
16080258\_Bacillus\_subt  
19074788\_Encephalitozoon\_cuni  
19909972\_Coprinopsis\_cinere  
Dictyostelium\_discoide  
19113905\_Schizosaccharomyces\_p  
\_21355725\_Drosophila\_melano  
\_7705688\_Homo\_sapient  
29832567\_Streptomyces\_averm  
21674006\_Chlorobium\_tepid  
\_5870602\_Leishmania\_major]

DMCGAATAIHAIGIAELGLGVPVIAAIGVAENMPDAAAIAKPGDVYATYN  
DKGGGSAVIGLLNALAKLGVAEVHGIIGATENMIGPAAKPDILISKE  
DMCGAASVFGTMMKALAKLNLVNVIGVLAGCENMPGSNAYRPGDILTTFMS  
DMCGAASVGLTLRAVAEMGLKLNVI AVVPTCENMPSGIATKPGDVVTSMS  
DMGGAASVGLMHALAARKAKANVVGIGLGVENAVDGHQORPGDIVTSMS  
DMGGAATMIGTMCVAKMKIKKNVTCVVAACENSIGPNAYRPGDILTAMN  
DMAGAATVGLGILSGIAALELPVNVTAIVPATENAI DAAAYKMGDVYTGMS  
DMGGAASVGLGAMETIGELRPEQNVLCVIPSTDNMISGGAMKPDVIVSLS  
DMLGAASVLSVFGLAADMGKIKNMLVPLVENLPSGATKPGDVHVGMN  
DMGGAATVVSAAALAIKALQPLINLVVTTPLTENMPGSPATKPGDIYAMN  
DMGGAATAVSAAMFGVASLGLVNLITITPLCENMPSGKATKPGDILTAANG  
DMGGAAVMLSSIIYALEQLSIPVNAVFTVPLTENLPSGSAKPGDVFMRN  
DMTGAAVVVATCRVAVAGLRPLVNIIRGLIPLCENVVGCSFRPGDMVKSMN  
DMGGAATVCSAIVSAAKLNLPINIGLAPLCENMPSGKANKPGDVVRAKN  
DMSGAAAVFAAVVAARLGEVNVNVTGWLALAEENMPSGATRPGDVLRYMS  
DMSGAAVVI AAEAAASLGLPLRVVGLVPATDNMPGSAQKPGDVITTFMS  
DMMGAAATFCTLKAIAKALQPLINVVAAVGLVVENAIGPESYHPSIIITSLK  
\* \* \* \* \*

\_23509661\_Plasmodium\_falci  
Theileria\_annulata\_sanger  
268\_40\_Cryptosporidium\_par  
\_15235763\_Arabidopsis\_thalia  
1703282\_Solanum\_cp  
23042880\_Trichodesmium\_erythr  
17227733\_Nostoc\_sp  
15607055\_Aquifex\_aeoli  
24216140\_Leptospira\_interro  
15611584\_Helicobacter\_pylori  
27364860\_[Vibrio\_vulni]  
17547134\_Ralstonia\_solan  
13476527\_Mesorhizobium\_loti  
27887802\_Fusobacterium\_nuclea  
29840173\_Chlamydomphila\_caviae  
16080258\_Bacillus\_subt  
19074788\_Encephalitozoon\_cuni  
19909972\_Coprinopsis\_cinere  
Dictyostelium\_discoide  
19113905\_Schizosaccharomyces\_p  
\_21355725\_Drosophila\_melano  
\_7705688\_Homo\_sapient  
29832567\_Streptomyces\_averm  
21674006\_Chlorobium\_tepid  
\_5870602\_Leishmania\_major]

GKTIIEVGNTPDAEGRLLADALVYAEKLGVDYIVDIATLTGAMLYSLGTSY  
GKTVEVLNTPDAEGRVTLADSLYAAKLDPDYLVYATLTGAAMTALGVK  
GKTIIEVGNTPDAEGRLLADALVYACNLKVTVIDLATLTGANVRLFEGRY  
GKTIIEVNNTPDAEGRLLADALVYACNQGVKVDLATLTGACIALGTSM  
GKTIIEVNNTPDAEGRLLADALIYACNQGVKEIIDLATLTGAIIVTALGPSV  
GKTIIEVNNTPDAEGRLLADALVFAEKLGVDAIIDLATLTGACVVALGNDI  
GKTIIEVNNTPDAEGRLLADALVYTDKLGDAIVDLATLTGANVIALGDDI  
GKYIEIDNTPDAEGRVTLADVLSYASEL-KDKIIDMATLTGACMVALGEY  
GITVEVQNTDAEGRVLVGLDVSYVGGKFDYMLDLATLTGAIISLGH  
GKSIEVRNTPDAEGRVLADCLSYA-QDLSVDIVDFATLTGACVVLGGEFT  
GQTVIEVNLTPDAEGRVLCDALTYV-ERFEDCVVDVATLTGACVIALGHHI  
GQTIIEILNTPDAEGRLLCDALTYV-ERFRAVIDVATLTGAVIALGHIN  
GQTIIEVNLTPDAEGRVLADALWYCNDRFQKFMVNLATLTGAIMVALGHDY  
GKTIIEVNTDAEGRLLADALTYIVRKEKNEVIDAATLTGAMVVALGEDV  
GLSVEIGSTDAEGRLLADALTYIYALKYCNTRIIDFATLTGAMVVALGEDV  
GKTIIEILNTPDAEGRVLADGITYAKQHG-SVLDVATLTGCVVVALGTF  
GKSVEINNTDAEGRLLADALVYAQ-EANTYIVDVATLTGAMMIALGDFA  
GKSVEVDNTPDAEGRVLSDAIYTYSTEYKHTLIDVATLTGAMVIALGEVY  
KTVEVDNTPDAEGRLLGDALHYACS-FKTHIIDIATLTGAIDVALGQHYA  
GLSVEIDNTPDAEGRLLADALVHYVSSQYKKAVIDEASTLTGAMVVALGNV  
GKTIIEVQNTDHDVLDALVYLAQN-FCKCIDIGTCSGYMRQSLDEAA  
GKTIIEVQNTDHDVLDALVYLAQN-FKVIILNATLTGAMVVALGSGA  
GKTIIEVNLTPDAEGRVLADALWAASE-EKDAIVDVATLTGAMMIALGNRL  
GITVEVGNTPDAEGRLLADALFYAKKEYNDVIDLATLTGACVIALGNSV  
GLTVEVNLTPDAEGRVLADVLTYYQKDKKTRIIDLATLTGAIIVGLGSR

\_23509661\_Plasmodium\_falci  
Theileria\_annulata\_sanger  
268\_40\_Cryptosporidium\_par  
\_15235763\_Arabidopsis\_thalia  
1703282\_Solanum\_cp  
23042880\_Trichodesmium\_erythr  
17227733\_Nostoc\_sp  
15607055\_Aquifex\_aeoli  
24216140\_Leptospira\_interro  
15611584\_Helicobacter\_pylori  
27364860\_[Vibrio\_vulni]  
17547134\_Ralstonia\_solan  
13476527\_Mesorhizobium\_loti  
27887802\_Fusobacterium\_nuclea  
29840173\_Chlamydomphila\_caviae  
16080258\_Bacillus\_subt  
19074788\_Encephalitozoon\_cuni  
19909972\_Coprinopsis\_cinere  
Dictyostelium\_discoide  
19113905\_Schizosaccharomyces\_p  
\_21355725\_Drosophila\_melano  
\_7705688\_Homo\_sapient  
29832567\_Streptomyces\_averm  
21674006\_Chlorobium\_tepid  
\_5870602\_Leishmania\_major]

AGVFGNNEELINKILNSSKTSNEPVWVLPINEYRATLNSKYADINNISS  
AAFYSNDEDLAELYTNSARNSGELMWRMPLLEKEYKDYLNKSVADYANVYN  
ASVLFNNDLQFQMTQTCGKTVYEFQWLPDPEFQEMVKSEIADLNNTSE  
AGIYTPSDKLAKEVIAASERSGKLRWMPMEESYWEMMKSQVADMVNTGG  
AGAFTPSDGLAREVVAAEASGKLRWMPMEESYWEMMKSQVADMVNTGP  
AGLWSPNDNLAAEITAAAEKAGEKMWMPLEEKYFEGLRKAMHADMKNTP  
AGLYTPDDALAGQLEQAASESGEKIWRMPLEEKYFEGLRKAMHADMKNTP  
AGLFTNAPDFAEI KKTAKRTGERVWELPMDDLRKKIKNTVAVLNTGG  
AGVMSNSDVLNLLKEASISSDERIWEMPLWEEYSEDLKSDIADIRNVAG  
SAIMGHNEELKNLFTSGLSEGLLAKLPFNRHLKLLIESKIADV CNISS  
TGVLSNHNPLAHELVNASEQSSDRAWRLPMADYHEQLKSPFADMANIGG  
TGVYARSDALANALLAAGQSLDTGWRMPLDEEYQELLKSNFADMGNIGG  
AGLFSNNDLAEERLAAAGQATQERLWMPPLGAEYDKLIDSKNADMKNIGG  
TGVFTNDDKMARKVIDASENWNVEYFQWMPMFDIYKKNLKS SYADMNTGV  
AGFFSNNDVLAQDLSEASAE TSEALWRLPLVEKYNKALHSDIADMKNIGS  
TGAMTNDQSFYQVADAAQECGEAIWQLPITEKDKKRVKSSMADLSNSPG  
IGYFTADDDLSKIIHQSGIDANDPVWRMPLSOLYLPMSKSNVADLKNAEG  
SGVFASSDNLWQLYEAGQIEHDMRWRMPLDDEFQPQIHSNDQNTGG  
GCFTTDSLWDQLNECGNISGERLWMPPLIPEYRQMETSVADLINSAGR  
TGAFVQGEELWKNLE TASHDAGDLFWRMPLFEHAYLQKLTSSNADLCNVSR  
CGVFTNSEILWQIQIKHASMHTGDRVWRFPPLWNFYSKAVRAGSLDVQNYGI  
TGVFTNSSLWLNKLF EAS IETGDRVWRMPLFEHYTRQVVDCLADVNNIGK  
FGVMANDDAFTAVVEAAEEVGEDSWMPPLPEHLRKGMDSP TADIANMGE  
AGLFSNDEKLAESIFEAGQSSGKVVWRLPLWDEYDELKSDVADVHNTGG  
AGLFSPSAMLANSLMAAGTQCGEELWMPPIGDEHKDMMKGIADLINAAP

_23509661_Plasmodium_falci	SVASSIVASLFLKEFVQNTAWAHIDIAGVSWKPKGFGVRLTTEFVLN
Theileria_annulata_sanger	GKCGSINGALFLKEFVDD-KWLHVDFAFPKGTGFGVLTIVNFVLT
268_40_Cryptosporidium_par	GKAPLSTSAAFLECFVVKVYAHIDIAGCSGFGNGFGVKTLVITIKE
_15235763_Arabidopsis_thalia	RAGGSITAAFLFKQFVSEVEWMHIDMAGPVWAATGFGVATLVEVQON
1703282_Solanum_cp	RDGGAITGALFLKQFVDEVQWLHLDIAGPVWNAATGYGVSTLVEVVL
23042880_Trichodesmium_erythr	RPGGAITAAFLFKQFVKNTPWAHLDIAGPVWGATGYGVRTLNVNLS
17227733_Nostoc_sp	RPGGITAAFLFKQFVKDTAWAHLDIAGPVWGATGYGVRLLDVWVLS
15607055_Aquifex_aeoli	RYGGAITAAMFLEEFVGEIKWVHLDIAGPAWGGTGFVVRTCLEYIMK
24216140_Leptospira_interro	RAGGSLSAAKFLERFVEPIAWAHIDIAGTAWGPTGYGVRLLDLVEK
15611584_Helicobacter_pylori	SRGGAITAGLFLNEFIRDDKWLHLDIAGPAYGASGAGVRACTAFVEE
27364860_[Vibrio_vulni]	RPGGTITAAACFLSKFAKK-NWAHLDIAGTAWGSTGRPVSMVLVQFLN
17547134_Ralstonia_solan	RPAASVTAACFLARFTEK-DWAHLDIAGTAWGATGRPVPLLRFLMD
13476527_Mesorhizobium_loti	RYGGAIITAAQFLQRFVKDTPWAHLDIAGTAMWGSFGVRLDLRVL
27887802_Fusobacterium_nuclea	RWGGSTNAAKFLEEFIDDTKWVHLDIAGTAWGATGQVFRVTSYIKD
29840173_Chlamydomydia_caviae	NRAGAITAALFLKRFLEDVAWAHLDIAGTAYYASGFGVRLIYYIEK
16080258_Bacillus_subt	REGHAIMAGTFLGEFAESTPWVHLDIAGTATGATGVMARLTLAER
19074788_Encephalitozoon_cuni	GHGGSATAAIFLSEFVG-FKWAHFDIAGVMDGATGCGVPVLIEMIEK
19909972_Coprinopsis_cinere	RPAGSATAALFLKPFVNGIKWAHLDIAGSMEGMTGRPVRLVEFTR
Dictyostelium_discoideum	SGGACCAAGFLKEFITAQSWSHLDIAGVMSGMTGKPTRTLIEFAK-
19113905_Schizosaccharomyces_p	AGGGCTAAAFIKFLAQLSFAHLDIAGVMDGMSGRPVRTIIEVARK
_21355725_Drosophila_melano	GRGRPCKAAAFLEFVFP-GQWMIIDATNMVMGAGRPTRTLIEFIAQ
_7705688_Homo_sapien	YRAGACTAAAFLEFVFP-PKWAHLDIAGVMTGMPTRPTRTLIEFLLR
29832567_Streptomyces_averm	RMGGGLVAGLFLREFVGEITWAHLDIAGPAFGGTGTAVRTLVRAL
21674006_Chlorobium_tepid	RGAGTITAAKFLKFDIGKHWAHLDIAGPAFGATGFGVRLLDLKLK
_5870602_Leishmania_major]	GRAGSCTAAAFLEFVFPVQWAHLDIAGVADGVTGFGVQLLDVFLRH
	.. : :

**Glucose-6-phosphate isomerase (Fig. 4B)**

CLUSTAL X (1.81) multiple sequence alignment

15595075_Borrelia	ETHLKIFQNLSDLEANLIEKYKEVLSNRKVLHHLTRGQGDVIDNKENMR
16752000_Chlamydomydia	DAILAALISLAERGLHESMLAMQSEMRPALHTATRAWTDSSFGAEDIA
269_185_Cryptosporidium	EEGFELLIKLAAESNLMEKIKLQLTEKRAVLHTALRSKNIPITSSGQNVL
4835820_Toxoplasma	LETLQHLVNLAAHERQVPAMVKRMFTENRAVLHVALRMPGSEPVVDGKDV
23509563_Plasmodium	EKTLNKLVEYAEVEELKKKVEKFTTENRSVLHTALRIPNTHKIIDNKV
15239045_Arabidopsis	VETMDKLLNLAKASQLTEKISRMFTENRSVLHVALRAPKDAVIADGMNV
1169797_Oryza	GETMEKLFKLAEEAAKLEKIEKMFTEENRSVLHVALRAPRDEVISNGVNV
1169798_Oryza	GETMEKLFKLAEEAAKLEKIEKMFTEENRSVLHVALRAPRDEVISDGNV
1346073_Zea	GETMEKLLKLDAKLEKIEKMFTEENRSVLHVALRAPRDAVIDSGVNV
37960111_Phytophthora	SKTLDDLFLAETAELKQAIATEDRAVMHMLRAPKTEKIVDGHADV
16273471_Haemophilus	QTTLSHLRQLAQECALDSAKEAMFTENRAVLHTALRNRNTPVLDGKDV
15804618_Escherichia	EETLAKLQDLAKECDLAGAIKSMFTENRAVLHVALRNRNTPILVDGKDV
23136626_Cytophaga	DETISLLIALAEAGLQAIEDMFTKRSVLHVALRNRNTPILSDGKDV
17510665_Caenorhabditis	DESFQLMRLAKSRGVEESRNAMFTENRAVLHVALRNRNTPILVDGKDV
17737445_Drosophila	DEVWDLTLAKVRRVNAARDAMFTENRAVLHTALRNRNTPVLDGKDV
18201905_Homo	EDVMRMLVDLAKSRGVEAARERMFTEGRAVLHVALRNRNTPILVDGKDV
Entamoeba	EETMKLLIKLCEAVHLKEKIEAEFTEKRAVLHTALRNRNTPVLDGKDV
Dictyostelium	KETMDLLFELARASKVEEMRNSMFTEDRAVLHTALRNRNTPSAVIVDGENV
28923732_Neurospora	EETLDKLVRLAEAEAGVEKKRDAMFTEDRAVHVALRNRNTPQEMVVDGDM
6319673_Saccharomyces	DEIIAALIELAKEANVTGLRDAMFTEDRAVHVALRNRNTPMYVVDGNNVA
7437359_Trypanosoma	DEIKDVLFLKVEERGIRAFMRALFAENRSVLHIALRNRNTPIFVNGHDVM
15677249_Neisseria	DETMSLLFELAREAGVPERMRQMFTEENRAVLHVALRNRNTPVVDGEDVM
15612158_Helicobacter	DTTLLKLLFELAKDCSLKEKIEAMFTEKRAVLHTALRSLDAEILLDNMEVL
21673816_Chlorobium	ARTMELMLDLVRRSGIEKKRQMFTEHRSVLHTALRRPGYTMTIDGNDVA
15887753_Agrobacterium	EETVTLLEQLAREGGVEAKREEMFTEDRAVLHTALRNRNTPVLDGKDV
15964236_Sinorhizobium	DRIIDGLEALAKAAKVEEKRDAMFTEERAVLHTALRNRNTPVLDGKDV
15806745_Deinococcus	DETLRLLRELAQATGVEARRDAMFTEGRAVLHTALRAPDAVIEVDGKDV
37521261_Gloeobacter	DETLRLLSVLAEESDLRGRIEAMFTEQRSVLHTALRAPGATVIEDGENV
29832844_Streptomyces	DETLRLLRELAQATDVFGLRDAMFTEGRAVLHTALRAPDAVIEVDGENV
29828312_Streptomyces	DETLALLQELATATDVFGLRDAMFTEENRAVLHTALRAPDAVIEVDGENV
15608086_Mycobacterium	RETLALLIDLARTAHLEERRDQMFSEDRVLHTALRPLDPAELVVDGQDV
17546438_Ralstonia	PETLALLQLADEAGVTLTRDAMLTEHRSVLHVALRGHEDDYRANGAAM
15676249_Neisseria	EDTLQLLNLDAADLEGKMRALRSEGRAALHTALRPLGADAVVDGRDVL
27573133_Neisseria	EDTLQLLNLDAADLEGKMRALRSEGRAALHTALRPLGADAVVDGRDVL
19173604_Encephalitozoon	EETVLDGYLEKMKD--FGEIKDGMFTENRKLHVALRDKVLRMLVGHGDAKL
16330760_Synechocystis	DLQPKFAKAFQDMVALEKGAIANPDEQRMVGHYWLNRN----AAPNDGIR
23125809_Nostoc	SLQPKFDKAFADMAELEKGAIANPDENRMVGHYWLNRN----DAPTPELT
18030019_Spiroplasma	--YAKNFEIALAEFAAIEKGEIVNKSERNRPVDHYNHRLT----QLVPGKSL
13924524_Giardia	LYKQAHAAAFDLMRDRIEKLENVNVSEDRPVDHYNLRME----CVK-GRSL
18030022_Trichomonas	EYKPKVAQALELIKRMENGEIVNHSSEDRMVDHYNLRME----KLVKGKSL



15644137\_Thermotoga  
16080187\_Bacillus  
15669801\_Methanococcus  
23473348\_Desulfovibrio

LFEPNISGGLTDEDVKSVEEKVTS-----AVRNFVE----NPDFAKLD  
EHELTYLRDFVKTAHHNIHEKTGA-----SDFLGWVD----LEHYDKEE  
EISLEDINKVDFANAYSNLMEKLD-----VGF RDVI  
SNSAFLNWT HARHGRISGTD PASLGEAANLLGLELAAGR L PFI MPYADG-

15595075\_Borrelia  
16752000\_Chlamydomphila  
269\_185\_Cryptosporidium  
4835820\_Toxoplasma  
23509563\_Plasmodium  
15239045\_Arabidopsis  
1169797\_Oryza  
1169798\_Oryza  
1346073\_Zea  
37960111\_Phytophthora  
16273471\_Haemophilus  
15804618\_Escherichia  
23136626\_Cytophaga  
17510665\_Caenorhabditis  
17737445\_Drosophila  
18201905\_Homo  
Entamoeba  
Dictyostelium  
28923732\_Neurospora  
6319673\_Saccharomyces  
7437359\_Trypanosoma  
15677249\_Neisseria  
15612158\_Helicobacter  
21673816\_Chlorobium  
15887753\_Agrobacterium  
15964236\_Sinorhizobium  
15806745\_Deinococcus  
37521261\_Gloeobacter  
29832844\_Streptomyces  
29828312\_Streptomyces  
15608086\_Mycobacterium  
17546438\_Ralstonia  
15676249\_Neisseria  
27573133\_Neisseria  
19173604\_Encephalitozoon  
16330760\_Synechocystis  
23125809\_Nostoc  
18030019\_Spironucleus  
13924524\_Giardia  
18030022\_Trichomonas  
15644137\_Thermotoga  
16080187\_Bacillus  
15669801\_Methanococcus  
23473348\_Desulfovibrio

EFFQSELEKIYNFAKQIHSNGNIKSSNGKFKNVVQIGIGSSSLGPKALYS  
VRSRVEAQR LKDF LTKVRS-----QFTTIVQIGIGGSELGPKALYR  
NDVNEVNRRIKFKFANAIKRGELLGSGTKILKDVICIGIGSSYLGPFEVY  
DEVHAVLRRI RVFSEKVRSGEIRGHTGKKLVNVISIGIGSSYLGT E FVHL  
EDVHGVLKKIEKYSDDIRNGVIKTKNTKFKNVICIGIGSSYLGT E FVY  
PEVWVLDKIK E FSDKIRSGSWVGATGKPLKDVIAIGIGSSFLGPLFVHT  
PEVWGVKDKIKQFSETRSGSWVGATGKALTNVSVGIGSSFLGPLFVHA  
PEVWVSKDKIKQFSETRSGSWVGATGKPLTNVSVGIGSSFLGPLFVHT  
PDVHEVLDAIRAFTSNVRDGLVGATGKQLKNVISIGIGSSYLGPFEVFE  
PEVNAVLA KMDFCQR IISGEWKGYTGKAITD VVNIIGIGSSDLGPVMVTE  
PEVNAVLEKMKTFSEAIISGEWKGYTGKAITD VVNIIGIGSSDLGPVMVTE  
PEVNAVLEKMKGFCEKVRSGAWKGYTGKSIDIVNIIGIGSSDLGPVMVTE  
PDVNRVLAHMKFEFCNEIISGSWTGYTGKAITD VVNIIGIGSSDLG PLMVTE  
PDVRAELAHMKFEFTNMVISGVWRGCTGKQITD VVNIIGIGSSDLG PLMVTE  
PEVNVLDKMKSFQCRVRSQDWKGYTGKAITD VVNIIGIGSSDLG PLMVTE  
PGVNAVLNKMGKFAEGVRNGSIKGYTGKFEFTD VVNIIGIGSSDLGPVMVTE  
PSVRKVLDMKSFSEVRSGQWKGYTGKAITD VVNIIGIGSSDLG PLMVTE  
NTVNEVLQHMKEFSEQVRSGEWKGYTGKLTNIINIGIGSSDLGPVMVTE  
PEVDSVLKMKFSEQVRSGEWKGYTGKAITD VVNIIGIGSSDLG PLMVTE  
PLVNVLEQMKKLSKVRSGEWKQSGKPIRHVNIIGIGSSDLG PMMACE  
PKVNRVLRQMGFEFAHEVRSGSWLGYTNQVITD VVNIIGIGSSDLG PLMCT  
KSVRSVLKRMRAFSDSVRSKRLGYTNQVITD VVNIIGIGSSDLGALMVCT  
SEVSDVLDQMKAFCKKVISGEWKGYTGKAITD VVNIIGIGSSDLG PFMVTE  
PDVNGVLAAMGKFADAIRSGSLKGTGKAITD VVNIIGIGSSDLG PVMATL  
PDVNAVLEAMGKFADIRSGALKGTGKAITD VVNIIGIGSSDLG PVMATL  
PEVHEVLDRMATFADAVRSGEWLGYTGKPIKNIVNIIGIGSSDLG PVMAYE  
PEVHAVLDRMAEFADRVRRGEGWRYTGRRIRTVVNIIGIGSSYLGPDMAYD  
PAVHAVLDKMSFTEAERVRTGEWTGHTGKPIKNIVNVGIGSSDLG PAMAYE  
PAVHAVLDRMADFADRVRSGEWTGHTGSRIRTVVNIIGIGSSDLG PAMAYE  
TDVHAVLDAMGAFDRLRSGEWGTGATGKRISTVNIIGIGSSDLG PVMVYQ  
PDVLRVRAQMRDFAQRVHSGTWTGHAGQRTD VVNIIGIGSSDLG PVMVCR  
PEIRRELNRALKFAHSLDDGYSYQGITGKRIITD FVHIGIGSSDLG PAMCVQ  
PEIRRELNRALKFAHSLDDGLYQGITGKRIAD FVHIGIGSSDLG PAMCVQ  
DEVYDELMKIKAFVDFD SGRVCGVTGKLEIVVNIIGIGSSDLG PVMVCD  
AEITEPLRQIKAFVADVHQGNIKPPTAPKFTDLLAIGIGSSALGPQFVAQ  
QEI VQTLEQIEAFAEKVQTGAIHPPRASRFTDIISIGIGSSALGPQFVAE  
EHS LKLWEKVNTFVSSIRN-----DYKHIFNGIGSSYLGPYMLIQ  
AKSIALWNTVKAFKAKCIDA-----KDYDHVIYNGIGSSYLGPYMLLT  
AHTLAMWEEAKKFAEDVMTGVIKTSAGKKYESIIFNGIGSSYLGPMLLII  
RSWIDS VKSLEDWII-----NFDTVVVLGIGSSGLGNLALHY  
FARIQ--KSAEKIQS-----DSDVLLVVGIGSSYLGARAAIE  
YD--ENLDKYKSLNG-----YENVVVIGMGSILGTMAYY  
--MTARLQELEPLLR-----RFEH MVVLGIGSSALGARALQK

.: \*:\* \* \*

15595075\_Borrelia  
16752000\_Chlamydomphila  
269\_185\_Cryptosporidium  
4835820\_Toxoplasma  
23509563\_Plasmodium  
15239045\_Arabidopsis  
1169797\_Oryza  
1169798\_Oryza  
1346073\_Zea  
37960111\_Phytophthora  
16273471\_Haemophilus  
15804618\_Escherichia  
23136626\_Cytophaga  
17510665\_Caenorhabditis  
17737445\_Drosophila  
18201905\_Homo  
Entamoeba  
Dictyostelium  
28923732\_Neurospora  
6319673\_Saccharomyces  
7437359\_Trypanosoma

SIKNGYFISNIDPDESEEVLSINVD ETLFII VSKSGNTLETKANM QFLI  
ALRHVHFISNIDPDNGAEVLDITDCAKALVVVSKSGTTIETAVNEAFFA  
ALRRLRFLANVDPIDIRRATEGLHPETTLVIIVSKTFTTAEITLNAKTIK  
ALAQIHFLANVDPVDVWLAERGFDEETLVVVISKFTTTAETMMNARSVR  
AMKNVRFLANVDPNDVNRAIQNL DQYDTLVI IISKFTTTAETMLNARSIK  
ALQQLRFLANIDPVDVARNISGLNPETTLVVVSKTFTTTAETMLNARTLR  
ALQQLRFLANVDPVDVARSIKDLDPETTLVVVSKTFTTTAETMLNARTLK  
ALQQLRFLANVDPVDVARSIKDLDPETTLVVVSKTFTTTAETMLNARTLQ  
ALQQLRFLANVDPVDVARSIKDLDPETTLVVVSKTFTTTAETMLNARTLK  
SLRNLRFLANVDPVDVARATSGLNPEETLVVVSKTFTTTAETMLNARTLR  
ALRNMHFVSNVDGTHIAETLKKVNPETTLFLVASKTFTTQETMTNAQSAR  
ALRNMHFVSNVDGTHIAEVLKKNPETTLFLVASKTFTTQETMTNAHSAR  
ALKHVHFVSNVDGTHIVETTKLNPETTLFIIASKTFTTQETITNAESAK  
SLKNVHFVSNVDGTHVAEVTKKNLNAETTLFIIASKTFTTQETITNAETA  
ALKHSHFVSNIDGTHLAEVLKKNYETTLFIVASKTFTTQETITNATSAR  
ALKRVVYVSNIDGTHIAKTLAQLNPESLFI IASKTFTTQETITNAETA  
ALKTLHFVSNVDGTHMVEALKKNPETTLFVICSKTFTTTAETMLNAHSAR  
ALKRAHFVSNIDGTHLAEVTKHLCPETTLFIVASKTFTTQETITNAQSAR  
ALKTLRFVSNVDGTHIAEALASDPETTLFLIASKTFTTTAETITNANTAK  
ALKDVHFVSNIDGTHIAETLKVVDPETTLFLIASKTFTTTAETITNANTAK  
ALRSMHFVSNIDGTHLSEVLNLDLESTLFI IASKTFTTQETITNALSAR

15677249\_Neisseria  
15612158\_Helicobacter  
21673816\_Chlorobium  
15887753\_Agrobacterium  
15964236\_Sinorhizobium  
15806745\_Deinococcus  
37521261\_Gloeobacter  
29832844\_Streptomyces  
29828312\_Streptomyces  
15608086\_Mycobacterium  
17546438\_Ralstonia  
15676249\_Neisseria  
27573133\_Neisseria  
19173604\_Encephalitozoon  
16330760\_Synechocystis  
23125809\_Nostoc  
18030019\_Spironucleus  
13924524\_Giardia  
18030022\_Trichomonas  
15644137\_Thermotoga  
16080187\_Bacillus  
15669801\_Methanococcus  
23473348\_Desulfovibrio

ALKNMHFVSNVDGSQLRDVLSKVHPETTLFI IASKTFTTQETLTNALTAR  
ALKKMHFVSNVDGTQILDVLEKINPASTLFI VASKTFSTQETLTNALTAR  
ALKKVHFVSNVDGSHLVETLRGLNPETTLFI IASKTFTTQETLANAVSAR  
ALARAHFVSNIDGAHIADTLKLLDPETSLFI IASKTFTTITETMTNAATAR  
ALARLHFVSNVDGAHIADTLTLLDPETSLFI VASKTFTTITETMTNAATAR  
ALKTVRFVSNVDGDLTEKTRDLDPVETLFI VSSKFTTQETMTNARSAR  
ALKKVRFVSNVDGNSFAEVIHDLPEDETLFI VCSKFTTLETMTNAHSAR  
VLRTVRFVSNVDGADLHEAVRDLDPATLFI VASKTFTTITETITNATSAR  
ALRTVRFVSNVDGADLHEAVRDLDPATLFI VASKTFTTITETITNATSAR  
ALRSARFVSNVDPADLIATLADLDPATLFI VASKTFTLETLTNATAAR  
ALARVHFVSNVDGTDLAETLAGLNPDTTLAIVCSKFTTLETMTNAHSAR  
ALETVHFVSNVDPADLDAVLCRLNPETTVFCVASKSFKTPETLLNAQAVK  
ALESVHFVSNVDPADLDEVLCRLNPETTMFCVASKSFKTPETLLNAEAVK  
ALGETYFISNIDATDTIRVFEKIDPERALFI VVSKTFTTLETIKNAELAM  
ALAAIHFI DNSDPDGIDRVNLCLKAKSTLVVTTSKSGGTPPEPRNGLAETK  
ALAKIHFI DNNDPAGIDRVLNHLR-ASTLVVVISKSGGTPPEPRNGMIEVK  
AML-LHFLANTDSDSFSTLFDLLD-KKTLMTVISKSGSTAETATNTQCYL  
AKY-AHFVANTRDPASFAQVIELVTEERTLMVVISKSGGTAETATNCATYL  
AKYKIYFILQHRFRHVSTRSHPTS-MLPSWSTFPLAPHRQLVTHSTWM  
SLRRVVFVNDVDPDLMSVLDRIDPKTTLFNVISKSGSTAEMMATYSIAR  
ALNVIFIGNNSSSYMRDVMDDLEDVDFSNVISKSGTTEPAIAFRIFR  
AIS-AYFIDNSDPEKTLKILKVDLNESEIYI IISKSGTLETLVNYLIK  
AFFWLWIADNVADSLAQLSLPAEKTVVVVVISKSGGTIETIGQYFLIR

: . .

15595075\_Borrelia  
16752000\_Chlamydomyxa  
269\_185\_Cryptosporidium  
4835820\_Toxoplasma  
23509563\_Plasmodium  
15239045\_Arabidopsis  
1169797\_Oryza  
1169798\_Oryza  
1346073\_Zea  
37960111\_Phytophthora  
16273471\_Haemophilus  
15804618\_Escherichia  
23136626\_Cytophaga  
17510665\_Caenorhabditis  
17737445\_Drosophila  
18201905\_Homo  
Entamoeba  
Dictyostelium  
28923732\_Neurospora  
6319673\_Saccharomyces  
7437359\_Trypanosoma  
15677249\_Neisseria  
15612158\_Helicobacter  
21673816\_Chlorobium  
15887753\_Agrobacterium  
15964236\_Sinorhizobium  
15806745\_Deinococcus  
37521261\_Gloeobacter  
29832844\_Streptomyces  
29828312\_Streptomyces  
15608086\_Mycobacterium  
17546438\_Ralstonia  
15676249\_Neisseria  
27573133\_Neisseria  
19173604\_Encephalitozoon  
16330760\_Synechocystis  
23125809\_Nostoc  
18030019\_Spironucleus  
13924524\_Giardia  
18030022\_Trichomonas  
15644137\_Thermotoga  
16080187\_Bacillus  
15669801\_Methanococcus  
23473348\_Desulfovibrio

NKLKLNIGIYKQMVIIITLKDSMLAIEEKGYLEYFFMHDSIGGRFSPTSVA  
DYFAKGLFKDHFIAVTCESPMDDTGK-YLEVPHLWESIGGRFSSTSMV  
EWLHKALKVSKHLAAVSTNIKATSDFGIPVDHVFQFWDWVGGGRFVCSAV  
DWYLHHYKLGAFHCAVSTNLDGTSKFGIQSDRVFGFWDWVGGGRYSVTSVA  
KWLSLKIKLSKHMVAVSTNKLTDDEFGISRDNVFEFWDWVGGGRFVTSVA  
EWITAALGVAKHMVAVSTNLALEKFGIDPNNAFVAFWDWVGGGRYSVCSAV  
EWIVSSLGVAKHMIAVSTNLELVEKFGIDPKNAFAFWDWVGGGRYSVCSAV  
WSGLSLLLVAKHMIAVSTNLELVEKFGIDPKNAFAFWDWVGGGRYSVCSAV  
EWIVSSLGVAKHMIAVSTNLELVEKFGIDPKNAFAFWDWVGGGRYSVCSAV  
KWLVDGLTVAKHMVAASAVPLVQQFQIDRANIFGFWDWVGGGRYSVTSVA  
DWLLKAAKVAKHFVASTNNAKDVKEFGIDTNNMFEFWDWVGGGRYSLSVAI  
DWFLKAAGVAKHFVASTNNAKAVGEFGIDTANMFEFWDWVGGGRYSLSVAI  
AWFLQTAQVAKHFVASTNNAKVAAFVGFIDTENMFEFWDWVGGGRYSLSVAI  
EWFLAKAGVAKHFVASTNNAKVAEFGIDENMFEFWDWVGGGRYSLSVAI  
TWLLEHSAKVAKHFVASTNNAKVAEFGIDENMFEFWDWVGGGRYSLSVAI  
EWFLQAAKVAKHFVASTNNAKVAEFGIDENMFEFWDWVGGGRYSLSVAI  
KWFVEKTVGSKHFVAVSTNNAEVEKFGIDENMFEFWDWVGGGRYSLSVAI  
SWFLEKIGIGQHFVASTNNAEVEKFGIDENMFEFWDWVGGGRYSLSVAI  
SWFLEKTGITKHFVASTNNAEVEKFGIDENMFEFWDWVGGGRYSLSVAI  
NWFLSKTGIKHFVASTNNAEVEKFGIDENMFEFWDWVGGGRYSLSVAI  
NEFLKFLSVAKHFVASTNNAEVEKFGIDENMFEFWDWVGGGRYSLSVAI  
EWFLNHAGVAKHFVASTNNAEVEKFGIDENMFEFWDWVGGGRYSLSVAI  
KWFVERSGIAKHFVAVSTNNAEVEKFGIDENMFEFWDWVGGGRYSLSVAI  
AWFLVKAGVAKHFVAVSTNNAEVEKFGIDENMFEFWDWVGGGRYSLSVAI  
RFIAEKLVKHHFAAVSTALDKVAAFVGFIDENMFEFWDWVGGGRYSLSVAI  
AFIAGKLVGHHFAAVSTALDKVAFVGFIDENMFEFWDWVGGGRYSLSVAI  
KWLGLSLKVRHFVAVSTNNAEVEKFGIDENMFEFWDWVGGGRYSLSVAI  
QWCLAALGVIKHFVAVSTNNAEVEKFGIDENMFEFWDWVGGGRYSLSVAI  
DWLLTELKVAKHFVASTNNAEVEKFGIDENMFEFWDWVGGGRYSLSVAI  
SWLLAGLDVAKHFVASTNNAEVEKFGIDENMFEFWDWVGGGRYSLSVAI  
RWLTDALGVSRLFVAVSTNNAEVEKFGIDENMFEFWDWVGGGRYSLSVAI  
RWFIEHGLKQHFVAVSTNNAEVEKFGIDENMFEFWDWVGGGRYSLSVAI  
AWYRGAGFTACHFCVAVSADTAAAFVGFIAAERVFAMVDWVGGGRYSLSVAI  
AWYRGAGFTARHFCAVAVSADTAAAFVGFIAAERVFAMVDWVGGGRYSLSVAI  
KLLERKLGASKHFVAVSADTAAAFVGFIAAERVFAMVDWVGGGRYSLSVAI  
AVFEAQGLYAVAVTMPGSK-LSQQAQTEQWLQAFPMQDQWVGGGRYSLSVAI  
KAYAGHNLVAVAITSVDSN-LDKLAKDEGLARFPMYDQWVGGGRYSLSVAI  
NLLTKASLHCIAITIPGSN-LHKLAAS--WLQTFEMCEPTGGRTSICSAI  
TLLKDLRVSVFVAVTTEGSK-LDKQANEQQFRETFFHMNEATGGRTSICSAI  
KLCRDRNLHNAAVTIKDSL-LDKIAHENKFI RTWHMEIDTGGRTSICSAI  
GILEAYGLEHMLITDPEKGFRLKLVKEEGRSLEVPVGGGRYSLSVAI  
KLEEKYGARIVATTDKERGALKTLSNEEGFESFVIPPDDVGGGRYSLSVAI  
KRIEKLNSFKGLVFI TNGGKLRKREAKKNYDI FSI PENVPVGGGRYSLSVAI  
DWLRTARGNHVIAVTDERSGFLREEDTHGLASMPVDPDLGGGRYSLSVAI

. \* \* : :

15595075\_Borrelia  
16752000\_Chlamydomyxa

GLLTLCFTEKVAKEILKGANEDKKS LNKNVKNASLLAALISYERNVLN  
GVLGFAYGFVFLQLLQAGASMDQIALQPNARNLPLMSALISYERNVFLN

269\_185\_Cryptosporidium  
4835820\_Toxoplasma  
23509563\_Plasmodium  
15239045\_Arabidopsis  
1169797\_Oryza  
1169798\_Oryza  
1346073\_Zea  
37960111\_Phytophthora  
16273471\_Haemophilus  
15804618\_Escherichia  
23136626\_Cytophaga  
17510665\_Caenorhabditis  
17737445\_Drosophila  
18201905\_Homo  
Entamoeba  
Dictyostelium  
28923732\_Neurospora  
6319673\_Saccharomyces  
7437359\_Trypanosoma  
15677249\_Neisseria  
15612158\_Helicobacter  
21673816\_Chlorobium  
15887753\_Agrobacterium  
15964236\_Sinorhizobium  
15806745\_Deinococcus  
37521261\_Gloeobacter  
29832844\_Streptomyces  
29828312\_Streptomyces  
15608086\_Mycobacterium  
17546438\_Ralstonia  
15676249\_Neisseria  
27573133\_Neisseria  
19173604\_Encephalitozoon  
16330760\_Synechocystis  
23125809\_Nostoc  
18030019\_Spironucleus  
13924524\_Giardia  
18030022\_Trichomonas  
15644137\_Thermotoga  
16080187\_Bacillus  
15669801\_Methanococcus  
23473348\_Desulfovibrio

GPLSIHFGANLVQEFLDGCWMDQHYETAPVSNLPVLLGLLSVYNSTFMD  
GPLALQYGYDVAQEFNLNGAHAMDVHFKAELANLPLMLGLISVWNATFFF  
GPLSIAFGYKMNRFNLNGCHMDDEHFLHADLKNIPVLLALTSFYNSHFFD  
GPLSLQYGFMSVMEKFLKGASSIDQHFQSTPFENIPVLLGLLSVWNVSFLG  
GPLSLQYGFPIVQKFLGGAASIDKHFRRSSSFENIPVLLGLLSVWNVSFLG  
GPLSLQYGFPIVQKFLGGAASIDKHFRRSSSFENIPVLLGLLSVWNVSFLG  
GPLSLQYGFPIVQKFLGGAASIDNHFYSSSFENIPVLLGLLSVWNVSFLG  
GPLALQYGFDI TEKFLAGAHAMDKHLETPLRNLPVIMGLLGVWNVSSFLG  
GSIALSIGFENFEALLNGAHAMDKHFLSTPIENIPTTLLAVGLWNTNFFLG  
GSIVLSIGFDNFVELLSGAHAMDKHFSTTPAENLPVLLALIGIWIYNNFFG  
GSIALYIGYDNYIELLSGAHAMDKHFKEFPFANIPVLLALIGIWIYNNFYK  
GSIAVHIGFDNYEKLLDGAFSVDEHFVNTPLENIPVILAMIGVLYNNIYG  
GSICLSIGFENFEQLLDGAFHMDNHFKTTPFENAPVILALLGVWYNNFFK  
GSIALHVGFDNFQELLSGAHWMDQHFRTPPLENAPVLLALIGIWIYINCFG  
GTIMISIGEQQFLDLLSGAHAMDKHFRTTEFANIPVILAVLGVWYNNLYQ  
GSVALYVGMDFHESFLGAYHMDQHFLNTPLENLPVIMGLLGVWYNNFFG  
GSVALYVGYENFHFKFLAGAHAMDNHFRTPAPLKNIPVGGILSVWYNNFYQ  
GSVALYIGYDNFEAFLLKGAFAVDNHFTQTPLENIPLGGLLSVWYNNFFG  
GSVMISIGYDNFVELLTGAHIMDEHFINAPTENLPIILALVGIWIYNNFFG  
GPIMLYLGEENFIEMLNAGAHMDQHFINTPLENLPVILALIGIWIYINYYQ  
GSIMIYLGKKNFNALLKGAAYLMDHEFRNAPFENLPVLMGLIGVWYINFFQ  
GSIALYLGFDRFRELLAGAHAMDEHFLNAPLENMPILAMLGIWYNNFFG  
GPLMIAIGPDNFGKFLDGAHAMDRHFREAPIRNLPMLLGLIGFYNNRVLD  
GPLMIAIGKENFGRFLDGGHAIDEHFRSAPLRNIPLMLLGLIGFYNNRVLD  
GSLMVAVGPEHFREMLAGFHDMDHEHFRTPAENLPMLLGLLGVWYGDFFG  
GSTMIAVGPQFHFRLAGFHAMDEHFRTPAFENLPVLMALIGLWYNNFFG  
GSLMIAIGPERFREMLEDGFRIVDEHFEFSAENVPVLLGGLLGIWYGNFHD  
GSLMIAIGPERFRELLEDGFRIVDEHFRTPAENAPVLLGGLLGVWYGSFFG  
GSLMTVIGRDAFADFLAGFHIIDRFHATAPLENAPVLLGLLGLWYNNFFG  
GSIVLAI GPEQFETMLDGCARAMDRHFASAAFRNMPILGLLGSVWYRGGF  
GPVMVAVGGARFRELLAGAHAMDRHFSTPTRNIPVLMALIAVWYNNFQH  
GPVMVAVGGARFRELLAGAHAMDSHFHTPPRNIPLVLMALIAVWYNNFQH  
GSIALYVGFNNFMRLKGAFAVDEDFRRNRGRNAEMIHAIAELFY S-ENG  
GLLPAALQGDIDIQAML DGAKTMDAETRVREL RQNPAAALLALAWYAGDGG  
GLVPAALQGDIVRAMLDGAKEMDDA TRVPDVKNPAAALLALSWYFAGNGK  
AMVPCAFAGIKFDEFLLKGMAMQDCETRAEKN-PALEIAAMIDSLIKKFA  
GMVPAAFAGVDFESFLIGMSEMDKLRSHRENNPAIYATLLDWMKDKS  
GFVPIYAFARCDFGEFIKMSYMDTLTRAEGEN-PAALLATAIDANN-HKV  
GLLSAMAEGIDIDELHEGAKDAFEKSMKENIL-ENPAAMIALTHYLYLNK  
GLLPYAVSGVNIIDMMKALDASKDFATSELE-DNPAYQYAVVRNVLYNK  
GLAPLYSLGVQIDSKILEGAREMDKICQNEIDILKNPALLNGVIHYLYDKRG  
GLIPAAFTGMDWQALVRGAKTVGAPLCSPPDAGAHPAWHLAAWNHALMQA  
.  
:  
\*  
.

15595075\_Borrelia  
16752000\_Chlamydomyxa  
269\_185\_Cryptosporidium  
4835820\_Toxoplasma  
23509563\_Plasmodium  
15239045\_Arabidopsis  
1169797\_Oryza  
1169798\_Oryza  
1346073\_Zea  
37960111\_Phytophthora  
16273471\_Haemophilus  
15804618\_Escherichia  
23136626\_Cytophaga  
17510665\_Caenorhabditis  
17737445\_Drosophila  
18201905\_Homo  
Entamoeba  
Dictyostelium  
28923732\_Neurospora  
6319673\_Saccharomyces  
7437359\_Trypanosoma  
15677249\_Neisseria  
15612158\_Helicobacter  
21673816\_Chlorobium  
15887753\_Agrobacterium  
15964236\_Sinorhizobium  
15806745\_Deinococcus  
37521261\_Gloeobacter  
29832844\_Streptomyces

YSSNCIIAYSKAMENFYLHLQQLMESNGKSVNRFNETINTVRIIWGGIG  
YPTAVIPIYSSGLEFFPAHLQCCMESNGKSVIQDGRVGTSPVIWGEPE  
KSCVAVLPYQALCKFPAHVQQLMESNGKSSSIDGNMMLTGAIFFGEPG  
YSNVAVLPYAQALLRFAHIQQLTMESNGKRVMTDGTLDVGEIFFGEPG  
YKNVAIPLPYQNLKFSAHIQQLSMESNGKSVDRNQPPIHTCQVYFGEPE  
YPARAILPYSQALEKFAPHIQQVSMESNGKGVSIDGLPLPTGEIDFGEPG  
YPARAILPYSQALEKFAPHIQQLSMESNGKGVSIDGVQLSTGEIDFGEPG  
YPARAILPYSQALEKFAPHIQQVSMESNGKGVSIDGVQLSTGEIDFGEPG  
YPARAILPYSQALEKLAPHIQQLSMESNGKGVSIDGAQLSTGEIDFGEPG  
HSSRALPYSQALLRFAAHIQQVDMESNGKRVTVEGVDLPAGEVNFGEPE  
AQTEAILPYDQYLHRFAAYFQQGNMESNGKYVDRDGNVKTGPIIWGEPE  
AETEAILPYDQYLMHRFAAYFQQGNMESNGKYVDRDGNVVDGPIIWGEPE  
AESHAILPYDQYLMHRFAAYFQQGNMESNGKRVTKDGTAVTTGPIIWGEPE  
AETHALLPYDQYLMHRFAAYFQQGNMESNGKRVTRHGQRVDGPIIWGEPE  
AETHALLPYDQYLMHRFAAYFQQGNMESNGKRVSKSGKPVKTGPIIWGEPE  
CETHALLPYDQYLMHRFAAYFQQGNMESNGKYITKSGTRVDTGPIIWGEPE  
AQSHAILPYDQYLMHRFAAYFQQGNMESNGKRVTKDGVVNTGPIIWGEPE  
CQTQAILPYDQYLSRFPAYFQQGNMESNGKSVQRNGERVTTGPIIWGEPE  
AQTHLAPFDQYLMHRFPAYLQQLSMESNGKSVITSDGSAKTGPIVFGEPG  
AQTHLAPFDQYLMHRFPAYLQQLSMESNGKSVTRGNVFTDGTGPIVFGEPG  
SETQAILPYDQYLMHRFPAYLQQLSMESNGKSVTRGNVFTDGTGPIVFGEPG  
GGSHVAPYDQYLMHRFPAYLQQLSMESNGKSVTRGNVFTDGTGPIVFGEPG  
SKSHLAPYDQYLMHRFPAYLQQLSMESNGKRVTRGNVFTDGTGPIVFGEPG  
AHSQAIIPYDQYLMHRFPAYLQQLSMESNGKRVDRAGHEVDTGPIVFGEPG  
YPTAILPYDQYLMHRFPAYLQQLSMESNGKRVTRGNVFTDGTGPIVFGEPG  
YPSRAILPYDQYLMHRFPAYLQQLSMESNGKRVTRGNVFTDGTGPIVFGEPG  
AESLAVLPYDQYLMHRFPAYLQQLSMESNGKRVTRGNVFTDGTGPIVFGEPG  
AQTLAVLPYDQYLMHRFPAYLQQLSMESNGKRVTRGNVFTDGTGPIVFGEPG  
AQSHAVLPYSHYLSKFTAYLQQLSMESNGKSVQRDGNPVETGPIVFGEPG



16273471\_Haemophilus  
15804618\_Escherichia  
23136626\_Cytophaga  
17510665\_Caenorhabditis  
17737445\_Drosophila  
18201905\_Homo  
Entamoeba  
Dictyostelium  
28923732\_Neurospora  
6319673\_Saccharomyces  
7437359\_Trypanosoma  
15677249\_Neisseria  
15612158\_Helicobacter  
21673816\_Chlorobium  
15887753\_Agrobacterium  
15964236\_Sinorhizobium  
15806745\_Deinococcus  
37521261\_Gloeobacter  
29832844\_Streptomyces  
29828312\_Streptomyces  
15608086\_Mycobacterium  
17546438\_Ralstonia  
15676249\_Neisseria  
27573133\_Neisseria  
19173604\_Encephalitozoon  
16330760\_Synechocystis  
23125809\_Nostoc  
18030019\_Spirochete  
13924524\_Giardia  
18030022\_Trichomonas  
15644137\_Thermotoga  
16080187\_Bacillus  
15669801\_Methanococcus  
23473348\_Desulfovibrio

FTGNKPTNSILVQKITPFTLGLIAMYEHKIFVQGVIFNIFSFQDQWVEL  
FEGNRPNTNSILLREITPFFSLGALIALYEHKIFTQGVILNIFTDQWQWVEL  
FEGNRPNTNSIFFKKLTPFTLGLSLIAMYEHKIYTOQSVVWQINAFDQWQWVEL  
FEGNKPPTSIVLPPVTPFTLGLALIAFYEHKIFVQGI IWDICSYDQWQWVEL  
FVGNRPNTNSIVVKKVSPFTLGLALIALYEHKIFVQGI IWDINSFDQWQWVEL  
FEGNRPNTNSIVFTKLTPFMLGALVAMYEHKIFVQGI IWDINSFDQWQWVEL  
FPGNRPNTNSIVFKKMPYVVLGSLIAMYEHKIFVQGC IWNVNSFDQWQWVEL  
FEGNRPNTNSIFLHKLTPHSLGALIALYEHKIFVQGI IWNINSFDQWQWVEL  
FLGNRPNTNSILVGHIGPAELGALIVYEHFTTEGAIWDINSFDQWQWVEL  
FSGNRPNTNSILAQKITPATLGLALIAFYEHVFTTEGAIWNINSFDQWQWVEL  
FTGGRPSNSLLVKALTPRALGAI IAMYEHKVLVQGA IWGINSYDQWQWVEL  
FSGNRPNTNLILMDKVNPRNMGSLIAMYEHKTFVQGI IWGINSFDQWQWVEL  
FPGNRPNTNILLLEKISPSNIGALVALYEHKVVFVQGV IWDINSFDQWQWVEL  
FPGNRPNTNITVLDDELNPFNLGSLIALYEHKVVFVQGVVWNINSFDQWQWVEL  
FTGNRPSITFVYDKLTPFALGRLIALYEHRVFEVGVLF RINSFDQWQWVEL  
FTGNRPSLTIIVYDQLDFALGRLIALYEHRVFEVGA LFNINSFDQWQWVEL  
FEGNRPNTSTILADRLTPRTLGLALIALYEHKVVFVQ GAVWDINSFDQWQWVEL  
FEGNRPNTTLLAERLTPVVLGKLIALYEHRVFTQGV IWNLSFDQWQWVEL  
FKGNHPTTTTILAKELTPSVLQGLIALYEHRVVFVQ GAVWNIDSFDQWQWVEL  
FRGNHPTTTTILAAELTPSVLQGLIALYEHRVVFVQ GAIWNIDSFDQWQWVEL  
MPGNRPSTSIILASRLTPSVLQGLIALYEHQVFTEG VVWIDSFDQWQWVEL  
FEGNRPNTNITLQRLDAASLGLALIAEAHRTFVQGA VWNINSFDQWQWVEL  
FPGNRPNSNIIIDRLTPYNLGMMAAYEHKTFVQGA IWNVNPFDQWQWVEY  
FPGNRPNSNIIIDRLTPYNLGMMAAYEHKTFVQGA IWNVNPFDQWQWVEY  
FRGNRPITVVCYKLTPELGLAMIAHYEHKIFILGLY WGINSDQWQWVTL  
FENQRDSITVTIPEVDATSVGALIALYERAVSFYGS LVNWNAYHQPGVEA  
YENHRDSITVTIPOVNPRTVGLALIALYERAVGLYAS LVNWNAYHQPGVEA  
SQNGREFMTITVKKCEFCIGQLIALEERIVSTLGA FFWGINAYDQPGVQD  
YQNERAFVCITIDGFCERCIGQLIAFEERVVTTILSA FWNINAFDQPGVQD  
IQNKRSEFATSTFRERNEFTFGMMIALEERVVTF LASFFGINAYDQPGVQD  
RENNRPNMRVTFDGLTPYVNGQFFAYEAAATAFMG YLLEINPFDPQWQWVEL  
TDGNVNPNLIVNIPELNAYTFYGLVYFFEKACAMSG YLLGVNPNFDQWQWVEA  
TNNGVNPNVRIITLDEINEMAMGALLMYEMQVGF MGELYININAYNPQWAVEE  
HKTGTPLVEIGMAQPSEEQAGRLMALLEITTLTGL WLLDINPLDQWQWVEL

\* : . \* . : . \* . \*

15595075\_Borrelia  
16752000\_Chlamydomonas  
269\_185\_Cryptosporidium  
4835820\_Toxoplasma  
23509563 Plasmodium  
15239045\_Arabidopsis  
1169797\_Oryza  
1169798\_Oryza  
1346073\_Zea  
37960111\_Phytophthora  
16273471\_Haemophilus  
15804618\_Escherichia  
23136626\_Cytophaga  
17510665\_Caenorhabditis  
17737445\_Drosophila  
18201905\_Homo  
Entamoeba  
Dictyostelium  
28923732\_Neurospora  
6319673\_Saccharomyces  
7437359\_Trypanosoma  
15677249\_Neisseria  
15612158\_Helicobacter  
21673816\_Chlorobium  
15887753\_Agrobacterium  
15964236\_Sinorhizobium  
15806745\_Deinococcus  
37521261\_Gloeobacter  
29832844\_Streptomyces  
29828312\_Streptomyces  
15608086\_Mycobacterium  
17546438\_Ralstonia  
15676249\_Neisseria  
27573133\_Neisseria  
19173604\_Encephalitozoon  
16330760\_Synechocystis  
23125809\_Nostoc

GKIIANQILKNDN  
GKALANRVLELLE  
GKVLAKDIRNIIS  
GKVLAKGVRGILQ  
GKVLAKAEVRNYFN  
GKVLATQVRKQLH  
GKSLASQVRKSLH  
GKSLASQVRKSLH  
GKSLASQVRKQLH  
GKVLAKQVRAQLS  
GKQLANRILPELT  
GKQLANRILPELK  
GKQLANAAILPELK  
GKQLAKVIQPELA  
GKQLAKAIEPELD  
GKQLAKKIEPELD  
GKQLAKAVEKDLE  
GKQLAKSILPELS  
GKVLAKKILKEID  
GKVLAKVIGKELD  
GKLLAKSILLQLQ  
GKQLAKTILGELT  
GKELAVPILQELE  
GKQLAKAILPEFD  
GKELATGLLPVVE  
GKELATGLLPVIE  
GKVLAKKIDGELQ  
GKVLAGRIIPELE  
GKVLAKRVEPALT  
GKVLAKRVEPALT  
GKTQAKALLPVIT  
GKKLAKPILPELE  
GKQLAKTIIGELE  
GKQLAKTIIGELE  
GKKIATEVLETELE  
GKKAASILELQK  
GKKAASILDQLT

18030019_Spironucleus	GKKAATRVNELSM
13924524_Giardia	GKLACLVCNKLSK
18030022_Trichomonas	GKKAATGVNATSK
15644137_Thermotoga	GKKITFALMGR--
16080187_Bacillus	YKVNMFALLGK--
15669801_Methanococcus	EKKICWRLIKQ--
23473348_Desulfovibrio	GKRLANARLGAPG

\*

**Tryptophan synthetase  $\beta$ - chain (Fig. 4c)**

CLUSTAL X (1.81) multiple sequence alignment

```

34557127_Wolinella_succin      FGKYGGAYLPLLLIEHFKEIGEAYLKISQSFDFIQELKSIRKHYQGRPTP
222_2_Cryptosporidium_parvum  FGKYGGKYVPEVINNAMKEIEDAYNKISKSEDFINELKKIRKEFQGRPTP
22958844_Rhodobacter_sphaer    FGPGYGGAMLPPLPEPHFAAIREAYGRISKSADFAELRYIRKHFQGRPTP
28900440_Vibrio_paraha         FGGEYGGSFVPELEQIMRDINAAYECCQDPEFKDELARLYKHFVGRPSP
33456993_Cycloclasticus_sp     FGGEYGGQVIPPDLVAIMNDINRAYQEISKTSPFKDELNDLYTHYVGRPSP
15614226_Bacillus_halod        FGTFGGKYVPETLMSAIEEELALNDAMNDEAFIHDYKEHLREYAGRPTA
16079321_Bacillus_subti        YGDFGGKFVPETLMQPLDEIQTAFKQIKDDPAFREYYKLLKDYSGRPTA
14520675_Pyrococcus_abyssi     FGKFGGQYVPETLMEPLRELEKAYKRLKNDEEFNRQLDYLRWAGRPTP
11499193_Archaeoglobus_b2     FGFEFGRFVPEVLIPPLEELEKAYDRFKDDEEFKARLEYLKSAYAGRPTP
15606106_Aquifex_trpB1        FGPFGGKFVPETLMYALEELEEKYRELKSDPEFQKELDYLRAYAGRPTP
20808007_Thermoanaerobacter_te FGFRGGQYVPETVMNALIELEREFEAKKDEKFMEEYRYLKEYSGRPTP
15673445_Lactococcus_lactis    YGQFGGQYVPEILMPALELEEAAYEIVDSKKDPLFQAEKELLKDYVGRNP
32417954_Neurospora_crassa     FGFEFGGQYVPEALMDCLSELEEGFNKIKDDPAFWEEYRSYYP-WMGRPGQ
24215988_Leptospira_interr     FGFEFGRYSPEILHDALVELETTYKLLKKNKHFKELEYRKNYIGRPSN
2822116_Coprinopsis_cinerea    FGQFGGQYVPESLVDALAELEEAHKSIEDPAFWEVEVRSYLT-YSNRPSN
15642912_Thermotoga_maritima   FGPGYGGQYVPEILMPALELEEAAYEIMKDESFWKEFNLLRDYAGRPTP
29345943_Bacteroides          YGFEFGGAYVPEILHKVVEELKNTYLGVLESEDFKKEFDQLLRDYVGRPSP
34557707_Wolinella_succin      FGKFGGRFVPEITLMPALLELESAYNELRFDREFVSEVDYLLKEYVGRPSP
15963781_Sinorhizobium_melilot FGIFGGRFVAETLMPILLDLQDEWARAKNDPAFKAELNGLTHYIGRPSN
23502958_Brucella             FGIFGGRFVSETLMPILLDLEARYEHAKTDPDFWAEMDDLWKNYVGRPSP
22957700_Rhodobacter_sphaer    FGAFGGRYVAETLMPILDLAREYEAAKEDPAFKEELAYQRDYVGRPSP
464937_Pseudomonas_syringae    FGQFGGKYVPEITLMPALLELEKAYNHYKNDREFKEEFDYLRQYVGRPNP
23053570_Geobacter_b1         FGPGYGGTFVAETLSHALDELRLDAYARYQHDPEFIKEYEYELKHFVGRPSP
17546702_Ralstonia_solana     -----YANLRTDAEFAEAFREELKHFVGRPSP
Leishmania_sanger_major       FGRFGGKYVPEITLMPALTELESAPHALATDDEFQKELDGILKDYVGRESN
1174780_Zea_cp2               -GRFGGKYVPEITLMPALTELENAFHALATDDEFQKELDGILKDYVGRESN
1174778_Zea_cp1              FGTFQGGKYVGETLMPHALSELESAPFKLATDDDFQRELKDYVGRSLSP
7437014_Oryza_sativa          FGKFGGKYVPEITLMPHALSELETAFYSLATDDEFQRELAELKDYVGRESN
15236977_Arabidopsis_b2       FGKFGGKYVPEITLMPHALSELESAPYALATDDDFQRELKDYVGRESN
15239755_Arabidopsis_b1       FGKFGGKYVPEITLMPALTELESAPFRSLSGDQVQKELDGILKDYVGRESN
3915165_Campylotheba_cp_b2    FGRFGGKYVPEITLMPALAELETAYQYRNDPGFQAELOQLLRDYVGRATP
17231286_Nostoc              FGKFGGKYVPEITLMPALGELETAYQYRNDASQTELOQLLRDYVGRPSP
17227906_Nostoc_sp           FGRFGGQYVPEITLIPALTELEQAATAWKDESFTKELTQLLKYVGRPTP
33239640_Prochlorococcus      YGQFGGQYVPEITLIPALEQLEKDYNEAIDPAFKAEMEAAILKDYVGRPTP
7437011_Chlamydomonas         FGRFGGQYVPEITLMSCLQELEASYHRIKEDPEFLQELGGLLRDYVGRPSP
4325276_Gracilaria_cp        FGPGYGGRYVPEITLIPALELEEAAYREAKKDPALFLEELDHYLRFAGRPTP
136282_Thermus_thermo        YGRFGGRYVPEITLIPALDELEQAYEAAKQDPEFLDELDRLLREFVGRPNS
15805965_Deinococcus_radio    FGAFGGKFIPEALVAAVDEVAVEYDKAKADPEFARELDLLVHYTGRPSA
29832719_Streptomyces_avermi  FGTFGGKFIPETLVKNAADEEEYKAKNDPEFHQTLNLLRHVYVGRPTP
21673356_Chlorobium_tepid     FGPGYGGRYVPEITLMPAVSALEEAATAAIDPAFWAELNRLHRTYGRPTP
22974734_Chloroflexus_aurant  FGGEYGGQYVPEILVPALKQLEQAFIDAQEDPEFRSEFMTLLQYAGRPTA
28898734_Vibrio_paraha       FGFEFGGMYVPQILMPALRQLEEAQVSAQKDPEFQAQFNLLKDYAGRPTA
924784_E_coli                 FGFEFGGMYVPEILVPVVKQLEQAFVQAQNDPTFQAEFADLLKDYAGRPTA
16273336_Haemophilus_influ    FGFEFGGSFVSELLVPALRELEQAFDTCKLDEEFQKEYFRLLKDFVGRPSP
15612264_Helicobacter_pylori  FGDFGGQYVPEITLMPPLKELNALETVVPSPPEFQNEFHMLLREYVGRPTP
23473438_Desulfovibrio_desulf -----MPDSLKPILDELAATFERYRNDPGFLEEFNYLTRYSGRETP
23473141_Desulfovibrio_desulf FGFEFGGSYVPEVVQKALDKLEEAANYKDDDEEFLKEYHHYLLKDYSGRETP
19703662_Fusobacterium_nuclea PYPFGGQFMPEILMAPVQDLSNSWKSQNHSDFNKNELDVLLKDYAGRPTP
29840324_Chlamydomonada_caviae LETYGGQYVPEILVPLEDLDRSYEQKKDPQFRETFFHILKDYAGRPTP
29840318_Chlamydomonada_caviae DPETEPEIDIEKLRIFAELVKQEISRERYIEIPGELRKLKDYAGRPTP
14520795_Pyrococcus_abyssi    DPETEPEIDIEKLRIFAELVKQEISRERYIEIPGELRKLKDYAGRPTP
14520795_Pyrococcus_b2       DPDEGE-SRIGLLSKILPSALIDQEFATERWVSIPEEVREAYRRVGRPTP
18313363_Pyrobaculum         DPETMPEVVKPEKLLAIFPEPLVEQEVSDKEWIDIPPEVLDIY-SLWRPTP
15606594_Aquifex_trpB2       HPATQEPVKPEDLEPIFPKGLIQQEMSGERWIRIPEDVREIY-RIWRPTP
11498839_Archaeoglobus_b1     DPETKQIPISPEKLSVIFPMSLIEQEVSEERFIEIPEPVKEY-AVYRPTP
15643305_Thermotoga_maritima  NPRTLQPVTPDILLPIFPMALIEQEVSGERWIRIPEEVREIY-RLWRPSP
23053561_Geobacter_b2        G-LDNPQIGVDALAKVFPNMNLEIEQEVSTERWIDIPPEILGIL-KLWRPSP
21673033_Chlorobium_tepid    HPGTQKPLKAEDLYPIFAKELCHQELNQDAWIEIPEDVREMY-KYRSTP
29350072_Bacteroides

```







23053570\_Geobacter\_b1  
17546702\_Ralstonia\_solana  
Leishmania\_sanger\_major  
1174780\_Zea\_cp2  
1174778\_Zea\_cp1  
7437014\_Oryza\_sativa  
15236977\_Arabidopsis\_b2  
15239755\_Arabidopsis\_b1  
3915165\_Camptotheca\_cp\_b2  
17231286\_Nostoc  
17227906\_Nostoc\_sp  
33239640\_Prochlorococcus  
7437011\_Chlamydomonas  
4325276\_Gracilaria\_cp  
136282\_Thermus\_thermo  
15805965\_Deinococcus\_radio  
29832719\_Streptomyces\_avermi  
21673356\_Chlorobium\_tepid  
22974734\_Chloroflexus\_aurant  
28898734\_Vibrio\_paraha  
924784\_E\_coli  
16273336\_Haemophilus\_influ  
15612264\_Helicobacter\_pylori  
23473438\_Desulfovibrio\_desulf  
23473141\_Desulfovibrio\_desulf  
19703662\_Fusobacterium\_nuclea  
29840324\_Chlamydomonada\_caviae  
29840318\_Chlamydomonada\_caviae  
14520795\_Pyrococcus\_abyssi  
14520795\_Pyrococcus\_b2  
18313363\_Pyrobaculum  
15606594\_Aquifex\_trpB2  
11498839\_Archaeoglobus\_b1  
15643305\_Thermotoga\_maritima  
23053561\_Geobacter\_b2  
21673033\_Chlorobium\_tepid  
29350072\_Bacteroides

TAGTATLKDAMNEAMRNWVYIHNFTFYVIGTVAGPHYPVMMVRDFQSVIG  
ESGSRTLKDALNEAMRDWVTNVDTFYIIGTVAGPHYPVMMVRDFQAVIG  
ESGSRTLKDALNEAMRDWVTNVENTFYIIGTVAG-----  
HSGTATLKDATSEAIRDWTNVENTHYILGVSAGPHYPVMMVREFHKVIG  
HSGTATLKDATSEAIRDWTNVENTHYILGVSAGPHYPVMMVREFHKVIG  
HSGTATLKDATSEAIRDWTNVENTHYILGVSAGPHYPVMMVREFHKVIG  
HSGTATLKDATSEAIRDWTNVENTHYILGVSAGPHYPVMMVRDFHAVIG  
HSGTATLKDATSEAIRDWTNVENTHYILGVSAGPHYPVMMVREFHAVIG  
AAGTGTLLKDATSEAIRDWTNVENTHYILGVSAGPHYPVMMVRDFHAVIG  
EAGTGTLLKDATSEAIRDWTNVENTHYILGVSAGPHYPVMMVRDFHAIIG  
TAGTATLKDATSEAIRDWTNVENTHYILGVSAGPHYPVMMVRDFHAVIG  
NSGTATLKDATSEAIRDWTNVENTHYILGSAAGPHYPVMMVREFQSVIG  
TSGSATLKNALSEAIRDWTNVENTHYIVGVSAGPHYPVMMVRDFQACIG  
AAGSRTLKDATNEAIRDWTNVENTFYILGVSAGPHYPVMMVRDFQSVIG  
TSGTATLKDATNEAIREWVTNVDTFYILGVSAGPHYPVMMVRDFQSVIG  
KSGSRTLKDAINEAFRDWVANVDRTHYLFGTVAGPHYPVMMVRDFHRVIG  
TAGSRTLKDATSEAIRDWTNVENTFYIVGVSAGPHYPVMMVRDFQSVIG  
SSGSRTLKDAINEAMRDWVTNP-DSYLLGSLGPHYPVMMVRDFQSVIG  
HSGSATLKDACNEALRDWGSYEDAHYLLGTAAGPHYPVMMVRDFQSVIG  
HSGSATLKDACNEALRDWGSYEDAHYMLGTAAGPHYPVMMVRDFQSVIG  
QKGSCLKDACCEAMRDWSEANYETHYLLGTAAGPHYPVMMVRDFQSVIG  
NSGSATLKDAVNEALRDWASSYKDTHYLLGTSAGPHYPVMMVRDFQSVIG  
QSGRTLKDAINEAMRFWIANIATHHYCFGTAAGPHYPVMMVRDFQSVIG  
TSGQTLKDAVDEALAAWIREADTFYLLGSAVAGPHYPVMMVRDFQSVIG  
EKGRTLKEAVDAAFKAWININDTFYVLSAGVAGPHYPVMMVRDFQSVIG  
NOGAQTLKDAVNEALRDWSKNYESTHYCLGSLGSPYPMVRFQSVIS  
KRGNSGLKEAVDAIEDFIFKHDTHTFCIGSALGPYPKIVRDFQSVIS  
PSGKRFLKEDPNHPGSLGIAISEAIEDVLKARYSLGSLVNLHVLHQTVIG  
PSGKRFLKEDPNHPGSLGIAISEAIEDVLKARYSLGSLVNLHVLHQTVIG  
PSGRRHYR--PDHPGSLGIAISEAVEYVLSRKYLPGSVMFVLLHQTVIG  
PSGRKYLEENPEHPGSLGIAISEAIEEAASKYSLGSLVNLHVLHQTVIG  
PSGRKYLEENPEHPGSLGIAISEAIEEAASKYSLGSLVNLHVLHQTVIG  
PSGRKILEEDPNHPGSLGIAISEALEVAVSTKYSLGSLVNLHVLHQTVIG  
PSGRAILLADHPDSNGSLGIAISEAIEQAVETRYALGSLVNLHVLHQTVIG  
PSGRRILEEDPDTGSLGIAISEAIEQAVETRYALGSLVNLHVLHQTVIG  
PSGKIDILTAHPTYQSLGTAISEAIELAQMCKYTLGSLVNLHVLHQTVIG

\*

34557127\_Wolinella\_succin  
222\_2\_Cryptosporidium\_parvum  
22958844\_Rhodobacter\_sphaer  
28900440\_Vibrio\_paraha  
33456993\_Cycloclasticus\_sp  
15614226\_Bacillus\_halod  
16079321\_Bacillus\_subti  
14520675\_Pyrococcus\_abyssi  
11499193\_Archaeoglobus\_b2  
15606106\_Aquifex\_trpB1  
20808007\_Thermoanaerobacter\_te  
15673445\_Lactococcus\_lactis  
32417954\_Neurospora\_crassa  
24215988\_Leptospira\_interr  
2822116\_Coprinopsis\_cinerea  
15642912\_Thermotoga\_maritima  
29345943\_Bacteroides  
34557707\_Wolinella\_succin  
15963781\_Sinorhizobium\_melilot  
23502958\_Brucella  
22957700\_Rhodobacter\_sphaer  
464937\_Pseudomonas\_syringae  
23053570\_Geobacter\_b1  
17546702\_Ralstonia\_solana  
Leishmania\_sanger\_major  
1174780\_Zea\_cp2  
1174778\_Zea\_cp1  
7437014\_Oryza\_sativa  
15236977\_Arabidopsis\_b2  
15239755\_Arabidopsis\_b1  
3915165\_Camptotheca\_cp\_b2  
17231286\_Nostoc  
17227906\_Nostoc\_sp  
33239640\_Prochlorococcus

AEAKEQFLEMLPDHIVACVGGGSNAMGIFSAFIVELGWVEPLGK-GKSGE  
IESRQFLEMLPDIVTACVGGGSNAMGIFSGFIVELICVEPLGK-GNKGE  
VEAREQFLEMLPDMVAACVGGGSNAMGIFSGFLVALYGIIEPCGT-SSKGD  
NEARVQFKEMLPNNLVACVGGGSNAMGLFSAFLVAIHGVEPAGRSLDKGE  
KEARAQFLEELPDVITACVGGGSNAIGLFTAFLVDIVGVEPAGRSLDKGE  
DEARAQFQELLPTEVIACVGGGSNAIGTFYFVFLVKLIGVEAAGKGVETSQ  
EEAKQDLKRIMPDKVVACVGGGSNAMGMFQAFVLELIGAEAGKIDTPL  
REARQILEALPDVIVACVGGGSNAMGIFYPFVRLIGVEAAGKGVETSQ  
KEARRQILEAMPDAIACVGGGSNAMGIFHPFVRLIGVEAAGKGVETSQ  
RETRQILEALPDVIVACVGGGSNAMGIFYPFVRLIGVEAAGKGVETSQ  
DETKEQILEKLPDYVACVGGGSNAMGIFYPFVRLIGVEAAGKGVETSQ  
QEARAQFLEKLPDALVACVGGGSNAMGLFYPFVVMYGVAAAGLGDITPH  
NETKQMLKLPDAVACVGGGSNAVGMFYPFVVKLLGVEAAGKGVETSQ  
IESRQFLEMLPDVIVACVGGGSNAIGMFGYFVVKLIGVEAAGKGVETSQ  
EETKQILEKLPDYVACVGGGSNAAGIFYPFVVKLIGVEAAGKGVETSQ  
EETKQILEKLPDYVACVGGGSNAAGIFYPFVVKLIGVEAAGKGVETSQ  
EETKQILEKLPDYVACVGGGSNAAGIFYPFVVKLIGVEAAGKGVETSQ  
FEAKQILEKLPDYVACVGGGSNAAGIFYPFVVKLIGVEAAGKGVETSQ  
EEAKQILEALPDVIVAAVGGGSNAIGLFPFVVKLIGVEAAGKGVETSQ  
TEARQILEALPDVIVAAVGGGSNAIGLFPFVVKLIGVEAAGKGVETSQ  
REVRVQLEALPDVIVAAVGGGSNAIGLFPFVVKLIGVEAAGKGVETSQ  
KETRDQLQALPDSLVACVGGGSNAMGLFHPFVVKLIGVEAAGKGVETSQ  
REAKAQHKKLIPDLVACVGGGSNALGLFYPFVVMYGVAAAGHGIASGA  
EECKVQMPLEQPDVIVACVGGGSNAMGIFYPFVVKLIGVEAAGKGVETSQ  
-----  
KETRRQAMDKKPDVIVACVGGGSNAMGLFHEFVVRVVLGVEAAGKGVETSQ  
KETRRQAMHKKPDVIVACVGGGSNAMGLFHEFVVRVVLGVEAAGKGVETSQ  
KETRRQALEKKPDVIVACVGGGSNAMGLFHEFVVRMIGVEAAGKGVETSQ  
KETRRQALEKKPDVIVACVGGGSNAMGLFHEFVVRMIGVEAAGKGVETSQ  
KETRRQALEKKPDVIVACVGGGSNAMGLFHEFVVRMIGVEAAGKGVETSQ  
KETRRQALEKKPDVIVACVGGGSNAMGLFHEFVVRMIGVEAAGKGVETSQ  
QETRAQLEKLPDILLACVGGGSNAMGLFHEFVVKLIGVEAAGKGVETSQ  
KETRRQALEKLPDILLACVGGGSNAIGLFPFVVKLIGVEAAGKGVETSQ  
EETKQILEALPDVIVACVGGGSNAMGLFHEFVVRMIGVEAAGKGVETSQ



23473141\_Desulfovibrio\_desulf  
19703662\_Fusobacterium\_nuclea  
29840324\_Chlamydomonada\_caviae  
29840318\_Chlamydomonada\_caviae  
14520795\_Pyrococcus\_abyssi  
14520795\_Pyrococcus\_b2  
18313363\_Pyrobaculum  
15606594\_Aquifex\_trpB2  
11498839\_Archaeoglobus\_b1  
15643305\_Thermotoga\_maritima  
23053561\_Geobacter\_b2  
21673033\_Chlorobium\_tepid  
29350072\_Bacteroides

HAATLCLGEPGSYMLKDQAGEPAVYSISAGLDYPGVGPESHKLDLGRA  
HAATLTLGTVDTYALFNEDGSVKPVYSISPGLDYPGIGPEHAFLRDSKRA  
HAARFATGRPGSYLLQDNEGOVLPHTHSISAGLDYPVAVGPDHVLYESGRA  
HAARFATGKPGSYVLQDEEDGNCADTYSISAGLDYVSVGPTAEMHESGRA  
YTYDYGDSAGLKLKMHMTLGHRYVPPPIHAGGLRYHGLAPTLVSLINHGIV  
YTYDYGDSAGLKLKMHMTLGHRYVPPPIHAGGLRYHGLAPTLVSLINHGIV  
YKYDFPDVAVGVMLKMYTLGHDFVPPAIHAAGLRYHGAAPSLSLRKLKGLV  
YKYDFGDSVGLLIKMYTLGHDFVPPAIHAGGLRYHGDAPLVCKLYNLGYI  
YKYDFGDVAGLLKMYTLGHDFIPPPAIHAGGLRYHGDAPTLVCKLYNLGYI  
YDYDFGDTAGLLKMYTLGHDFIPPKIHAAGGLRYHGSAPIIARLVKEGLV  
YAFDYGDTAKMIAMMYTLGHDFMPPGIHAGGLRYHGESALVSQLHAGLI  
YAYDSGDVAMLLPMHSLGHTFIPPAIHAAGGLRYHGMAPLVSHTKQLGLI  
FOYDFGDEAGYLLPMFTLGHNFAPAIHAGGLRYHGAGVIVSQLKNDLM

34557127\_Wolinella\_succin  
222\_2\_Cryptosporidium\_parvum  
22958844\_Rhodobacter\_sphaer  
28900440\_Vibrio\_paraha  
33456993\_Cycloclasticus\_sp  
15614226\_Bacillus\_halod  
16079321\_Bacillus\_subtilis  
14520675\_Pyrococcus\_abyssi  
11499193\_Archaeoglobus\_b2  
15606106\_Aquifex\_trpB1  
20808007\_Thermoanaerobacter\_te  
15673445\_Lactococcus\_lactis  
32417954\_Neurospora\_crassa  
24215988\_Leptospira\_interr  
2822116\_Coprinopsis\_cinerea  
15642912\_Thermotoga\_maritima  
29345943\_Bacteroides  
34557707\_Wolinella\_succin  
15963781\_Sinorhizobium\_melilot  
23502958\_Brucella  
22957700\_Rhodobacter\_sphaer  
464937\_Pseudomonas\_syringae  
23053570\_Geobacter\_b1  
17546702\_Ralstonia\_solana  
Leishmania\_sanger\_major  
1174780\_Zea\_cp2  
1174778\_Zea\_cp1  
7437014\_Oryza\_sativa  
15236977\_Arabidopsis\_b2  
15239755\_Arabidopsis\_b1  
3915165\_Camptotheca\_cp\_b2  
17231286\_Nostoc  
17227906\_Nostoc\_sp  
33239640\_Prochlorococcus  
7437011\_Chlamydomonada  
4325276\_Gracilaria\_cp  
136282\_Thermus\_thermo  
15805965\_Deinococcus\_radio  
29832719\_Streptomyces\_avermi  
21673356\_Chlorobium\_tepid  
22974734\_Chloroflexus\_aurant  
28898734\_Vibrio\_paraha  
924784\_E\_coli  
16273336\_Haemophilus\_influ  
15612264\_Helicobacter\_pylori  
23473438\_Desulfovibrio\_desulf  
23473141\_Desulfovibrio\_desulf  
19703662\_Fusobacterium\_nuclea  
29840324\_Chlamydomonada\_caviae  
29840318\_Chlamydomonada\_caviae  
14520795\_Pyrococcus\_abyssi  
14520795\_Pyrococcus\_b2  
18313363\_Pyrobaculum  
15606594\_Aquifex\_trpB2  
11498839\_Archaeoglobus\_b1  
15643305\_Thermotoga\_maritima  
23053561\_Geobacter\_b2  
21673033\_Chlorobium\_tepid

HSVGVSDDEEAIKNFFALSRLGII PAIESAHAIAYGMKLAKERLKGKILIN  
KTVCITDQEAINGFFELSRTEGII PAIESSHAIGYVLKIAKEMKKGKILIN  
TYTGADDREALAFAFALSRLHEGII PAIESAHAVAFAMREAKHPGSILVN  
NYGSINDDEAIDAFELSRHEGII PAIESSHAVAYAIKLAQQGESSILVN  
KYVAIDDQECIDAFMTLSRVEGII PAIESAHAVAYAMKIAKMSSTILVN  
KYDAVTDKEALDALKQLSVHEGII PAVESAHALAKAFERAKQLSQSILVC  
TYDSTDEEALADALAFKLLSEKELPAIESAHALAKAFKLAQMDRIILVC  
EYVTVTDEEALRAFHELSTRTEGII PAIESAHAVAYAIKLAREMSRVIIVN  
EYVTVNDEEALRAFKTLKLEGI PAIESAHAIAYAMKMAEEMQRVLVNV  
KYTTATDEEALGFKLLARTEGII PAIESAHAVIKAVEVARELKGKIVVN  
EYVYATDEEALAAAFMDLSRTEGII PAIESAHALAYAMKLAQKDEIIVVN  
KYVDITDEEALGFKILSRTEGII PAIESSHAIAYAVKLAKELGASMIVC  
KFVAATDAQAFEGFRMSQLEGI PAIESSHGIWGALELAKTMKPDVVIC  
EYVNVDDDEGALDAFIEVCRIEGII PALETAHAFRYAKDLARTMSKNILIC  
DYVVTDEEDALRGFRMLTQKEGI PAIESSHAIWEGVKIAKSLPKDIVIC  
LYDAVTDDEALDAFIELSRLEGI PAIESSHALAYLKKINIKG--VVVVN  
TVLAVNDEEAIEAYELTKLEGI PAIESAHALGALKKLFKPE--VVVLT  
EYDHVTDAAEALAFVWLSQKEGI PAFESAHAIAYLKKARERFKDVIIVS  
EYVPTDAAEALAFQTLTRLEGI PALEPSHALAEVIKRAPKMGKIIILMN  
DYVPTLDNEALDAFQLCTRTEGII PAIESAHAIQAQVMAPTMGKVMIVN  
EYVSTDAEALAFQLCCALEGI PALEPSHALAHVIKIAPTLPRIIVMN  
QYTSVTDDEALDAFHKCCRLEGI PAIESAHALAEVFKRAPTLPKMLVNV  
EYVSVTDDEALAFQTLTREGEI PAIESSHAIHVVKLAPTLPRITIVAC  
QYVGTIDKEALQAFHDLCRMEGII PAIESSHALAYACKLAPTLPKILLVN  
-----  
EYDSVTDQEAALDAFKRVSRLGII PALETSHALAYLEKLCPTLADRNVVN  
EYDSVTDQEAALDAFKRVSRLGII PALETSHALAYLEKLCPTLPDRVVVN  
EYDSATDEEALAGFRKVSRLGII PALETSHALAYLEKLCPTLPDRVVVN  
EYFVSTDEEALAFKRVSRLEGI PALETSHALAHLEKLCPTLPDRVVVN  
EYYSITDEEALAFKRVSRLEGI PALETSHALAYLEKLCPTLPDRVVVN  
EYYCCTDEEALAFKRLSRLEGI PALETSHALAFLEKLCPTLPNKVVVN  
EYYSVTDDEEALAAFRQLSRLEGI PALETAHAIAYLETLCQQLDGRIVIN  
EYYSVTDKQALDGFQQLSRLEGI PALETSHAIAYLETLCQQLNGRIVIN  
EYAAVTDQEAALSALQLLSEKLEGI PALETAHALAFLEKLCPTLSSEIVIN  
EYAVTDAEALGFKLLSRLEGI PALETSHAIAYLEKLIPTLKRVRVIN  
-----  
EYASVTDDEEALGFKLLARLEGI PAIESAHAIAYAAKVVPEMDKVVVN  
EYVPTDAAQALEGLQLLTRLEGI PAIESAHAIYAVQLARQMKPIVVVN  
EYRAVTDAAQALRLLSRTEGII PAIESAHALAGALEIGKELGKLILVN  
CYTSTTDKEALDALAKTEGII CALESASHAVHYAMKRAAEMPKIIVVN  
TYTAASDDEEALAAQQLLARLEGI PAIESAHAVAEAIRLAPTLMPRIILVN  
EYDNVTDDEEALAFQELARSEGI PAIESSHALAHALRMARENPELLVNV  
DYVSTDDDEEALAFKTLCLHEGI PAIESSHALAHALKMMRENPDLLVNV  
DYVSTDEEALNAFQELAKHEGI PAIESSHALAYALKLIKQNPPELLVNV  
VYESASDIEEALAFRLLCQKEGI PAIESSHALAYALKLAQKCAIIVVN  
RYGMVNDQAALHAFKTLRMEGI PAIESSHAVAVWLENRDGIPRDRVVVN  
EYVNASDAEALSAPFDLSRTEGII PAIESAHALAYAMRLAPTLSSIIIVN  
EYVSATDDEAVNALLLTKKEGI PAIESSHALAEVIKRAPKLNKIIIVN  
SYTVATDKAALAAFFLLSRLEGI PAIESSHALAELEINIAPTLPKVVVVN  
QYTYATDDEEALAFRLSKTEGII PAIESSHALAHMIKIAPSLDKIAIVN  
KPIAYHQTEVFEEAVLFAKAEIVPAPESAHAVKAVIDKALEARREMIIL  
KPIAYHQTEVFEEAVLFAKAEIVPAPESAHAVKAVIDKALEARREMIIL  
EAVAYPQEEVMEARALLFARTEGIVPAPESAHAIKAAVDLAKLPR--VIA  
DAVAYQTEVFEEAVTFARTEGIVPAPESAHAIKAAIDEALKCKEEKVIL  
KARAVKQDPTFEAGLLFARTEGII PAPETNHAVRAAIDEAIKAREEKVIL  
EAQAFDQDETFEAAKIFAKLEGI PAPERSAHAIAGAIAREAKKAREKIV  
EAKSYRQACFEAAHLFARSEGI PAPERSSHAVRAAIDEAVLAKEEKVIL  
EATALPQTECYEAAALLFAHTGFI PAPETSHAIQTIAREAKQAKKEKVVIL

29350072_Bacteroides	EAVDIQQLESFEAGCLFAQAEGII PAPESSHAIAAAIREAKACKEEKVIL
34557127_Wolinella_succin	LSGRGDKDID
222_2_Cryptosporidium_parvum	LSGRGDKDLD
22958844_Rhodobacter_sphaer	LSGRGDKDID
28900440_Vibrio_paraha	LSGRGDKDID
33456993_Cycloclasticus_sp	LSGRGDKDAD
15614226_Bacillus_halod	LSGRGDKDVH
16079321_Bacillus_subti	LSGRGDKDVN
14520675_Pyrococcus_abyssi	LSGRGDKDLD
11499193_Archaeoglobus_b2	LSGRGDKDMD
15606106_Aquifex_trpB1	LSGRGDKDMA
20808007_Thermoanaerobacter_te	LSGRGDKDVN
15673445_Lactococcus_lactis	LSGRGDKDVV
32417954_Neurospora_crassa	LSGRGDKDVQ
24215988_Leptospira_interr	LSGRGDKDVA
2822116_Coprinopsis_cinerea	LSGRGDKDVE
15642912_Thermotoga_maritima	LSGRGDKDLE
29345943_Bacteroides	VSGRGDKDIE
34557707_Wolinella_succin	LSGRGDKDMI
15963781_Sinorhizobium_melilot	LSGRGDKDIF
23502958_Brucella	LSGRGDKDVH
22957700_Rhodobacter_sphaer	MCGRGDKDIF
464937_Pseudomonas_syringae	LSGRGDKDMQ
23053570_Geobacter_b1	LSGRGDKDIH
17546702_Ralstonia_solana	LSGRGDKDMH
Leishmania_sanger_major	-----
1174780_Zea_cp2	CSGRGDKDVH
1174778_Zea_cp1	CSGRGDKDVH
7437014_Oryza_sativa	FSGRGDKDVD
15236977_Arabidopsis_b2	FSGRGDKDVQ
15239755_Arabidopsis_b1	FSGRGDKDVQ
3915165_Camptotheca_cp_b2	CSGRGDKDVH
17231286_Nostoc	CSGRGDKDVQ
17227906_Nostoc_sp	CSGRGDKDVQ
33239640_Prochlorococcus	CSGRGDKDVN
7437011_Chlamydomonas	CSGRGDKDVN
4325276_Gracilaria_cp	-----
136282_Thermus_thermo	LSGRGDKDVT
15805965_Deinococcus_radio	LSGRGDKDVT
29832719_Streptomyces_avermi	LSGRGDKDMD
21673356_Chlorobium_tepid	LSGRGDKDMG
22974734_Chloroflexus_aurant	LSGRVTKNIL
28898734_Vibrio_paraha	LSGRGDKDIF
924784_E_coli	LSGRGDKDIF
16273336_Haemophilus_influ	LSGRGDKDIF
15612264_Helicobacter_pylori	LSGRGDKDLS
23473438_Desulfovibrio_desulf	LSGRGDKDMG
23473141_Desulfovibrio_desulf	LSGRGDKDVA
19703662_Fusobacterium_nuclea	ISGRGDKDVA
29840324_Chlamydomonada_caviae	LSGRGDKDLE
29840318_Chlamydomonada_caviae	LSGRGDKDLS
14520795_Pyrococcus_abyssi	FNLSGHGLLD
14520795_Pyrococcus_b2	FNLSGHGLLD
18313363_Pyrobaculum	FNLSGHGLLD
15606594_Aquifex_trpB2	FNLSGHGYFD
11498839_Archaeoglobus_b1	FGFSGHGLLD
15643305_Thermotoga_maritima	FNLSGHGLLD
23053561_Geobacter_b2	FCLSGHGQLD
21673033_Chlorobium_tepid	MNWSGHGLMD
29350072_Bacteroides	FNLSGHGLID

**9) 1,4-alpha-glucan branching enzyme (Figure 5)**

CLUSTAL X (1.81) multiple sequence alignment

The *Toxoplasma gondii* sequences were obtained from <http://ToxoDB.org> release 3.0.



15606119\_Aquifex\_aeolic  
23021493\_Clostridium\_thermo  
17228208\_Nostoc\_sp.  
45509261\_Anabaena\_variab  
23126496\_Nostoc\_punctif  
15618386\_Chlamydomophila\_pneumo  
18309045\_Clostridium\_perfri  
16080150\_Bacillus\_subtil  
19704191\_Fusobacterium\_nuclea  
32398905\_Cryptosporidium\_parvu  
T\_gondii\_TgTwinscan\_5121  
23135626\_Cytophaga\_hutchi  
28829250\_Dictyostelium\_discoi  
17554896\_Caenorhabditis\_elegan  
32564391\_Caenorhabditis\_elegan  
16416077\_Neurospora\_crassa  
171569\_Saccharomyces\_cerev  
28573410\_Drosophila\_melano  
4557619\_Homo\_sapient  
30680140\_Arabidopsis\_thalia  
15227650\_Arabidopsis\_thalia  
1169912\_Solanum\_tuber  
2801805\_Gracilaria\_gracil  
29346181\_Bacteroides\_thetai  
29245217\_Giardia\_lambli  
32470978\_Pirellula\_sp.  
45916029\_Mesorhizobium\_sp.  
45548449\_Rubrobacter\_xylano  
22968918\_Rhodospirillum\_rubrum  
22986678\_Burkholderia\_fungor  
23471845\_Pseudomonas\_syring  
22971470\_Chloroflexus\_aurant  
23124761\_Nostoc\_punctif  
23475793\_Desulfovibrio\_desulf  
22968087\_Rhodospirillum\_rubrum  
46135186\_Anabaena\_variab  
23130645\_Nostoc\_punctif  
32398951\_Cryptosporidium\_parvu  
T\_gondii\_TgTwinscan4106  
23123689\_Nostoc\_punctif  
23050680\_Methanosarcina\_barker

YQITGYFAPTSRYGTPQDFMYLIDKLHQEGIGVILDWVPSHFPTDRKRWH  
YQVTGYAATSRYGQPKDFMYFVDKCHQNGIGVIIDWVPAHFPPKDKQGEH  
YQVTGYAAPTSTRFGTPEDFMYFVDQCHQNNIGVLVDWVPGHFPKDRKGEH  
YQVTGYAAPTSTRFGTPEDFMYFVDQCHQNNIGVIVDWPVGHFPKDRKGEH  
YQVTGYAAPTSTRFGTPEDFMYFVDKCHQNDIGVIVDWPVGHFPKDRKGEH  
YQVTGYAAPTSTRYGTLOEQFYFVDYLHKENIGIILDWVPGHFPVDSQALH  
YQVTGYYSITSRYGDIKGLKRLINALHKEDIGVILDWVPGHFPKDKWADN  
YQGTGYYSPTSRFGPPHDLMKFVDECHQONIGVILDWVPGHFPKDRDREN  
YQGTGYYSVTSRYGTPDDDFMYFVNYFHKNLGVILDWVPGHFPKDKSLGEN  
YQVTGQYAPYSRLGTPDDFKYLVDKLHQSGIGVFLDFVPAHFCKDREGEH  
YQVTGLYAPYSRLGSPDDFKYLVDKLHQSGIGVFLDFVPAHFCKDREGEH  
YHVSNNFAPSSRFGTPEDLKYLVDRAHKFGLAVIMDVVHSHAVKNDRGYH  
YQVTNFFAISSRFGTPEELKEMIDKAHEMGLLVFLDVVHSHASKNDRGNH  
YQVSNFFAVSSRCGTPEDLKYLVDKAHSLGIFMLLDVVHSHASKNARGYH  
YQVSNFFAVSSRCGTPEDLKYLVDKAHSLGIFMLLDVVHSHASKNARGYH  
YQVTNFFAASSRFGTPEELKELIDTAHSMGILVLLDVVHSHASKNDRGEH  
YQVTSFYAASSRYGNPEQLKRMIDVAHSHGLFVLLDVVHSHASKNARGEH  
YQITSFFAASSRYGTPEELQELVDTAHSMGIIVLLDVVHSHASKNPRGTH  
YHVTNFFAPSSRCGTPEELKSLIDRAHELGLVLLMDIVHSHASKNPRGYH  
YHVTNFFAPSSRFGTPEELKSLIDKAHELGLVLLMDIVHSHASKNDRGYH  
YHVTNFFAASSRYGTPEELKYLIDKAHSLGLQVLVDVVHSHASKNDRGYH  
YHVTNFFAISSRCGIPEDLKYLIDKAHQGLYVFMDDVVHSHASSNERGRH  
YHVSFFAASSRFGTPEELKALIDAAHEMGIIVMDIVHSHAVKNPREH  
YQVTSFFAPSSRFGTPEELKYLVDKAHEAGIAVLLDLVHSHASKN-KEDH  
YNPAHIFAVEQAYGGPDALKAFVDAHEAGFAVIDVYVNHFGPDSHRSD  
YNPAHIFAIESSYGGPAAFKAFINAHKQGIIVLDVVYVNHFGPDSWRSD  
YDGVLPYAPDSSYGTPEELKELVQAAHGRGIMVLLDVVYVNHFGPETERHA  
YDGVLPYAPDSSYGTPEELKALIDAAHGLGLMVFLDVVYVNHFGPETQAHQ  
YDGVLPYAPDSSYGRPEELKALIDAAHGLGLMVFLDVVYVNHFGPDREGTH  
YDGVLYHAPQSSYGTPEELKHLIDTAHGLGLMVFLDVVYVNHFGPDRSDLH  
YDGVLYYAPHTAYGGVTGLKRLVDAHAHAGIAVLLDVVYVNHFGPETNRYQ  
YDGVYVYAVQNSYGSYPAELKQLVNACHQNNIAVLLDVVYVNHFGPETKTYK  
YDGVYVYIHHRAGPDALRRFVDTAHGLGLSVYGDVVVNLHGLPESRRRH  
YDGVNLFAPTRLYGQPDMMRFVDRALHGLAVILDVVYVNLHGLPDTDRYK  
YNPRYFFATESSYGSTADLKKLVDECHQIRIIMDGIYNHSEASPRDPD  
YKVRHFFATESSYGSTEDLKRIDECHGRGIRVFMDDGIYNHTDEEPEDPA  
YNPRLLLSINSHLGSVDDFIYLVQKVHEKDLFIVVDLVLHHSAGEKGGGD  
YNPRLLLATHEGKYGTPDQLRKLVDRECHRGKGLAVIFDLVNLHGSAGEKGGGE  
YETGSHIFALEYSAGGRDQLKHFIRECHYGIIVLDVVYVNHFGPDSGDYF  
FAPIGYFGVDERFGRNKDMQELIDKAHQKDIIVDSVYAHSTHSLGPFPAE  
:  
.  
\*  
:  
::  
\*  
.  
:  
\*  
:  
\*  
:

45548447\_Rubrobacter\_xylano  
45916434\_Mesorhizobium\_sp.  
22968855\_Rhodospirillum\_rubrum  
22986676\_Burkholderia\_fungor  
23468472\_Pseudomonas\_syring  
29829347\_Streptomyces\_avermi  
15806848\_Deinococcus\_radiod  
22999765\_Magnetococcus\_sp.  
23474508\_Desulfovibrio\_desulf  
32472075\_Pirellula\_sp.  
22971471\_Chloroflexus\_aurant  
15606119\_Aquifex\_aeolic  
23021493\_Clostridium\_thermo  
17228208\_Nostoc\_sp.  
45509261\_Anabaena\_variab  
23126496\_Nostoc\_punctif  
15618386\_Chlamydomophila\_pneumo  
18309045\_Clostridium\_perfri  
16080150\_Bacillus\_subtil  
19704191\_Fusobacterium\_nuclea  
32398905\_Cryptosporidium\_parvu  
T\_gondii\_TgTwinscan\_5121

PDWGTLIYNYGRNEVRNFLISNALFWLDEYHIDGLRVDVAVASMLYAFITVA  
RDWNTLIYNYGRNEVRNFLANALFWLDRYHVDGLRVDVAVASMLYTTTIA  
MDWGTYIFNYGRNEVRNFLANALFWLEQFHIDGLRVDVAVASMLYAVTIA  
QDWNMTIYNLGRNEVSAFLIASGLAWLKRYHVDGLRVDVAVASMLYAITIA  
QDWNMTIYNLGRTEVHGFMLASALHHLKHYHIDGLRVDVAVASMLYTMVIA  
PDWGTLEFDYGRNEVRNFLVANAVYWCEEFHIDGLRVDVAVASMLYVVVIA  
YDWNTRYIFDYGRNEVVMFLIGSALKWVQDYHVDGLRVDVAVASMLYCLMIA  
QDWDTLIFNFRLEVKSFLLISSALHMLTRYHLDGLRVDVAVASMLYVLMIA  
PDWGTYVFNRYGRNEVRNFLANALYWFKEFHVDGLRIDAVASMLYVMMIA  
PDWGTMIYFNFRNEVRNFLANALFWLDKYHIDGLRVDVAVASMLYVITAA  
PDWGSYIYNYGRNEVRSFLLISSALCWLKDFHIDGIRVDVAVASMLYVQVIA  
PDWNSFVFDYGKPEVRSFLLSSAHFWLDKYHADGLRVDVAVASMLYVQVIA  
PDWGTTHIFNYGRNEVRNFLANAMFWFDKYHIDGLRVDVAVASMLYVMMIA  
KEWGTLVFNYSRNEVRNFLVANALFWFDKYHIDGIRVDVAVASMLYVLSIA  
KEWGTLVFNYSRNEVRNFLVANALFWFDKYHIDGIRVDVAVASMLYVLSIA  
KGWGTLVFNYSRNEVRNFLANALFWFDKYHIDGIRVDVAVASMLYVLSIA  
PHWNTFTFDYSRHEVTNFFLGSALFWLDKMHIDGLRVDVAVASMLYVLTFA  
KLGWGTFFNFDLGPKEVKSFLISNAFYFINEFHIDGLRVDVAVASMLYVITIA  
WLWGTANFNLGKPEVHSFLLISNALYWAEFYHIDGIRVDVAVANILYVMMIA  
E-WGSANFNVSREVRNFLSNLYFWIKFPHIDGIRVDVAVANILYVMMIA  
KQWGTAVFNFRNEVRNFLGAAHYHLRYYHIDGLRIDAVSSMLYVFTMA  
KMWGTLVFNFRSEVRNFLGAAHYHLRYYHIDGLRIDAVSSMLYVFTMA

23135626\_Cytophaga\_hutchi  
28829250\_Dictyostelium\_discoi  
17554896\_Caenorhabditis\_elegan  
32564391\_Caenorhabditis\_elegan  
16416077\_Neurospora\_crassa  
171569\_Saccharomyces\_cerev  
28573410\_Drosophila\_melano  
4557619\_Homo\_sapien  
30680140\_Arabidopsis\_thalia  
15227650\_Arabidopsis\_thalia  
1169912\_Solanum\_tuber  
2801805\_Gracilaria\_gracil  
29346181\_Bacteroides\_thetai  
29245217\_Giardia\_lambli  
32470978\_Pirellula\_sp.  
45916029\_Mesorhizobium\_sp.  
45548449\_Rubrobacter\_xylano  
22968918\_Rhodospirillum\_rubrum  
22986678\_Burkholderia\_fungor  
22971845\_Pseudomonas\_syring  
23471470\_Chloroflexus\_aurant  
23124761\_Nostoc\_punctif  
23475793\_Desulfovibrio\_desulf  
22968087\_Rhodospirillum\_rubrum  
46135186\_Anabaena\_variab  
23130645\_Nostoc\_punctif  
32398951\_Cryptosporidium\_parvu  
T\_gondii\_TgTwinscan4106  
23123689\_Nostoc\_punctif  
23050680\_Methanosarcina\_barker

TGWDSKLFYDGKLEVKRFLLSNLRWLEYYHFDGFRFDGVTSMYSITIA  
ELWDSRLFNFGNVEVMRFLLSNLRFYVDEYHFDGFRFDGVTSMIYIVTIA  
NLWDSRLFDYDTQTEFLRFLLSNVRWVVEYGFDFGFRFDGVTSMIYMITIA  
NLWDSRLFDYDTQTE-LRFLLSNVRWVVEYGFDFGFRFDGVTSMIYMITIA  
DLWDSRLFNFGHHEVMRFLLSNLRWVDEYAFDFGFRFDGVTSMIYVITVA  
PLWDSRLFNFGKFEVQRFLLANLAFYVDVYQDFGFRFDGVTSMIYAVTVA  
SLWDSRLFNFGYVEVLRFLLSNLRWVHDEYNFDFGFRFDGVTSMIYIITIA  
DLWDSRLFAYSSWEVLRFLLSNLRWVLEYYRFDGFRFDGVTSMIYITIA  
WMWDSRLFNFGSWEVLRFLLSNLRWVLEYYRFDGFRFDGVTSMIYAITVG  
15227650\_Arabidopsis\_thalia  
KLWDSRLFNFGYANWEVLRFLLSNLRWVLEYYRFDGFRFDGVTSMIYATVIA  
SLWDSRLFNFGHWEVLRFLLSNLRWVMEYYHFDGFRFDGVTSMIYAVTIA  
PAWDSLCFDYGNKNEVIFLFLSNCKYWLEYYRFDGFRFDGVTSMIYAIISIA  
PLWDSKIFNYKNPETLRFLLQNVRWLQEFKIDGFRFDGVTSMIYILITIA  
TPWGDTRPDYGRGEVRQFIHDNAMMMREYHADGLRYDMTAYIRTCYLIA  
TPWGDTRPVYGRGEVRQFIRDNALFWLEYYHLDGLRFDGVTSMIYIRNRITIA  
TPWGAANFDDAGSVREFFVHNALYWIEEYGFDFGLRLDAVHAIRDVHVLV  
TPWGAAIDFERP--VRDFFVANALQWLEYYRFDGLRFDVHAIRDVHVLV  
TPWGPDAIDFERS--VSDFFTDNAVYWINEYRIDGLRFDVHAIRDVHVLV  
TPWGDAIDFRP--VRDFFVDNALMWLLEYYRFDGLRMDAVHAIRDVWLN  
TPWGAAINFDGPEVSRWLIENALEWLREYHIDGLRLDATHAIFDAYLFA  
TPWGEALNFDDAHSVRNYFIENALYWLREFHIDALRLDQAIYDRHLIA  
TPWGACPSFDGASAVRRYFLDCCRLMAHFHMDGLRLDQAIYDRHLIA  
NDWGVAINFDPGPGSVRAFYLANAGYWIEEFHLDGLRLDQAIYDRHLIA  
NNWGPPEFNHYEYDEARKFIGDTRVRYWVGEYHLDGIRYDAAQIANFYNVA  
NYWGPPEFNHYDYDKAWYIIGDVRVFWIQEYHIDGIRYDVAQRANFNINVA  
TGWGAKFNHFKK--VQDMLYEACRVLLGEYGVDFIRFDSVHNMNPNRFLIA  
SGWGRKRFVHKKR--VRDMILAAALCFIEEYAGDGLRLDSVHSPMPWKILIA  
YPDGRRFPEGGYLDVRKMFISSAVTLMEEFHVDFGFRVDQTTSIHSAILIA  
NAFGISTDFTYKFT-RNFFLTVNNHWLDSYHIDGFRYDYVPGYWLLIQCA

\* . \* \*

45548447\_Rubrobacter\_xylano  
45916434\_Mesorhizobium\_sp.  
22968855\_Rhodospirillum\_rubrum  
22986676\_Burkholderia\_fungor  
23468472\_Pseudomonas\_syring  
29829347\_Streptomyces\_avermi  
15806848\_Deinococcus\_radiod  
22999765\_Magnetococcus\_sp.  
23474508\_Desulfovibrio\_desulf  
32472075\_Pirellula\_sp.  
22971471\_Chloroflexus\_aurant  
15606119\_Aquifex\_aeolic  
23021493\_Clostridium\_thermo  
17228208\_Nostoc\_sp.  
45509261\_Anabaena\_variab  
23126496\_Nostoc\_punctif  
15618386\_Chlamydomyxa\_pneumo  
18309045\_Clostridium\_perfri  
16080150\_Bacillus\_subtil  
19704191\_Fusobacterium\_nuclea  
32398905\_Cryptosporidium\_parvu  
T\_gondii\_TgTwinscan\_5121  
23135626\_Cytophaga\_hutchi  
28829250\_Dictyostelium\_discoi  
17554896\_Caenorhabditis\_elegan  
32564391\_Caenorhabditis\_elegan  
16416077\_Neurospora\_crassa  
171569\_Saccharomyces\_cerev  
28573410\_Drosophila\_melano  
4557619\_Homo\_sapien  
30680140\_Arabidopsis\_thalia  
15227650\_Arabidopsis\_thalia  
1169912\_Solanum\_tuber

EESTAWPMVSRMGGGLGFGYKWNMGWMDTLQYMKEDPVHR  
EESTAFPGVSRAGGLGFGFKWNMGWMDTLQYMKEDPVHR  
EESTAWPMVSRMGGGLGFGYKWNMGWMDTLQYMKEDPVHR  
EESTAWPGVTADGGLGFGFKWNMGWMDTLQYMKEDPVHR  
EESTAWPGVSEQGGGLGFGYKWNMGWMDTLQYMKEDPVHR  
EESTAWEGVTRDNGLGFGFKWNMGWMDTLQYMKEDPVHR  
EESTAFPGVTTTEG-LGFDYKWKAMGWMDTLQYMKEDPVHR  
EESTAWPQVTRLGGLGFGFKWNMGWMDTLQYMKEDPVHR  
EESTSWPGVSRGTGGLGFGFKWNMGWMDTLQYMKEDPVHR  
EESTAWPGVSRDGGGLGFGFKWNMGWMDTLQYMKEDPVHR  
EESTAWPMVSRVGGGLGFGFKWNMGWMDTLQYMKEDPVHR  
EESTAWPMVSRVGGGLGFGFKWNMGWMDTLQYMKEDPVHR  
EESTAWALVTKTGGGLGFGFKWNMGWMDTLQYMKEDPVHR  
EESTDWPMSVWTGGLGFGFKWNMGWMDTLQYMKEDPVHR  
EESTDWPMSVWTGGLGFGFKWNMGWMDTLQYMKEDPVHR  
EESTSWPMVSRGTGGLGFGFKWNMGWMDTLQYMKEDPVHR  
EESTAFPGVTKQGGGLGFGFKWNMGWMDTLQYMKEDPVHR  
EESTSWPNVTNYGGLGFGFKWNMGWMDTLQYMKEDPVHR  
EDSTEWPQVTGEGGLGFGFKWNMGWMDTLQYMKEDPVHR  
EDSSAWPLVTKDGGGLGFGFKWNMGWMDTLQYMKEDPVHR  
EESTAWQGVTH--GLGFDKAWDLGWMMDTLQYMKEDPVHR  
EESTSWQGVTDGGLGFGFKWNMGWMDTLQYMKEDPVHR  
EDVSGMPGLSRDGGVGFDFYRLAMGIPDKWIKLKLQDEQ  
EEVTGLATLCREGGGDFDYRLAMGIPDKWIKLKLQDEQ  
EEVSGMPGICREGGQGFDFYRLAMALPDMWIKILKHTSDED  
EEVSGMPGICREGGQGFDFYRLAMALPDMWIKILKHTSDED  
EDVSGMPALCLLGGVGFDFYRLAMALPDMWIKILKHTSDED  
EDVSGYPTLCLIGGTGFDFYRLAMALPDMWIKILKHTSDED  
EDVSGMPTLCREGGIGDFYRLAMALPDMWIKILKHTSDED  
EDVSGMPALCSQGGGDFDFYRLAMALPDMWIKILKHTSDED  
EDVSGMPTFCIDGGVGFDFYRLAMALPDMWIKILKHTSDED  
EDVSGMPAFVCGGVGFDFYRLAMALPDMWIKILKHTSDED  
EDVSGMPGLGREGGIGDFYRLAMALPDMWIKILKHTSDED

2801805_Gracilaria_gracil	EDVSGMPTLCVRGGLGFDYRLAMAIPDMWIEVLEKEKDEN
29346181_Bacteroides_thetai	EEVSGMPGLAADGGYGFYRMAAMNIPDYWIKTIKEKIDED
29245217_Giardia_lambli	EDVSGYPCMATDGGIGFDYRFQMAVPDLWINMMKHGFDTG
32470978_Pirellula_sp.	EDLQTNWLTQGGAGFTTQWDAAFVHPIRSVQEIIDAH
45916029_Mesorhizobium_sp.	EDLRNKAEITGEGGAGFGAQWSEFVHPIRETLIAPDDES
45548449_Rubrobacter_xylano	ENEENEASLLREGRSLYDAQWDDLLHHALHVALTGEDAAY
22968918_Rhodospirillum_rubrum	ENDANEARRLIAVRGRYDAQWDDIHAFHVVTATGETGGY
22986678_Burkholderia_fungor	ENEHNSANLLE---THFDAQWDDAHNTLHVLLTGETEGY
23471845_Pseudomonas_syring	ENEFNQASLLE---KGFDAQWDDGHNTLHVLLTGETDAY
22971470_Chloroflexusaurant	EHPLNDPRFARLGGYGLSGVWSDDFHHALHTLLTGEQRGY
23124761_Nostoc_punctif	ESDLNPNQIIRLGGYGLDAQWSDDFHSLHALLTGDRQGY
23475793_Desulfovibrio_desulf	ESTRNEIHLTTAGGYGMAAQWSDDFHHAHVHARLTGESCXY
22968087_Rhodospirillum_rubrum	ENEPQEVRLIEAQGYGLDALWDDLLHHAAMVALTGRREAY
46135186_Anabaena_variab	EHIPETTSITN-LDGPMDGCWHDSEFYHTIKAHICGDT---
23130645_Nostoc_punctif	EHIPDTNTVVK-PDGPLDACWHESFRYFLVPYICGKS---
32398951_Cryptosporidium_parvu	EVPENPNYLK--ECGFDSQWHSYDYDIISQFRNQK---
T_gondii_TgTwinscan4106	EVTPENPQICN--SAGFDSQWHSAYDAIKVTRGQD---
23123689_Nostoc_punctif	EDHSDWEPVTEVGGLGFDAAWYANFYHHLIGDARQTEYA
23050680_Methanosarcina_barker	EQLQKPVIEILNCTWQNRRTLNTARGIYGNCEKLTGSLFG
	*