

#NEXUS
[This alignment contains the confidently alignable positions of UK, UK-UPRT and UPRT. This file is a single alignment of all three genes. UK sequences are the first sequences in the alignment, the fused UK-UPRT sequences are in the middle and the UPRT only sequences are the last sequences in the alignment.]

Begin data;
Dimensions ntax=67 nchar=487;
Format datatype=protein interleave symbols="1234" gap=-;

Matrix
18699734_Homo NGGEPFLIGVSGGTASGK-SSVCAKIVQLLGQNEVDY-RQK-QVVILSQDSFYR--VLTS
13899253_Homo HQR-PFLIGVSGGTASGK-STVCEKIMELLGQNEVEQ-RQR-KVVILSQDRFYK--VLTA
24649624_Drosophila EVKSPFLIGVAGGTASGK-STVCKKIMEQLGQAEMDH-TQR-QVVSISQDSFYR--ELTP
17538198_Caenorhabditis TLKLP LLIGVAGGTSCGK-STIVERI IENLNANAKQSGRQI-DIVHLSLHSFYR--ELSA
28828393_Dictyostelium HSKHPFIIGVTGGTASGK-TTVCDEIMKRLE-----NK-RIAIICLDSFYR--PLS-
Entamoeba_sangerUK -NNPTVLI AVAGGTASGK-TTFCQEIAN TLKGE-----KFVVISQDSFYR----PL
Paramecium1UK YEPQRILVGV TGGSAAGK-TTLCETIKREIQYDS-----DV-DMTILSCDSFYK--GVD-
15641051_Vibrio ENNQCVIVGIAGASASGK-SLIASTIYNELR----AKVGDH-QIGVITEDCYK--DQS-
26248441_Escherichia QSHQCVIIGIAGASASGK-SLIASTLYREL R----EQVGDE-HIGVIPEDCYK--DQS-
16272099_Haemophilus SNPSCIIIAITGASASGK-SSIASTVHKELC----NELGCQ-EIGIITEDSYK--DQS-
15613838_Bacillus MGKRPIIIGVAGGTGSGK-T----TVAKEIF----YQFNEK-SIVLIEQDAYYK--DQS-
15594361_Borrelia -----IIGISGGSGSGK-TTVVSKISEFIP-----EFVLISQDNYYK--SVG-
20455345_Chlamydomphila M----MIIGITGGSGAGK-TTLTQNIKEIFG-----E-DVSVICQDNYYK--DRS-
20455343_Deinococcus P----FVIGVAGGSGSGK-TTVTRRVIETVG-----RE-GVAVLNQDNYYR--DQS-
1351377_Mycoplasma --EKGILVAISGGSCSGK-TTVAEMIYQLLS-----KCLKVAIICQDNYYK--SYK-
20455342_Ureaplasma S--P-ILILIAGASGSGK-TTFANEI IARIP-----KNTTSVVICQDSYII--SNS-
20455333_Listeria P----IVVGV TGGSGSGK-TSVTKAICDHFS-----GH-SILMIAQDVYYH--DQA-
17229842_Nostoc P----IILGIVGDSAAGK-TTLTRGIAQVLG-----PE-NVTLIC TDDYHR--YDR-
16331050_Synechocystis L-DRVVLIGVAGDSGCGK-STFLRRLTDLFG-----EEFMTVICLDDYHS--LDR-
15790282_Halobacterium S----FAIGIAGGTGAGK-TTVAREITDNVG-----E-SATLIPLDNYE--DLS-
6320223_Saccharomyces VEEKRIVISIGGGHATGVG-AIALDLQNTFK-----S-----LYN--SIN-
1730776_Saccharomyces K---VILVALSGCSSSGK-TTIAKLTASLFT-----KATLIHEDDFYKH-DNE-
6324339_Saccharomyces PWTTPYIIGIGGASGSGK-TSVAAKIVSSIN-----VP-WTVLISLDFNYYN--PLGP
22026891_Drosophila -----MDCFYK--ILNE
8923487_Homo QSKEAFAIGLGGGSASGK-TTVARMII EALD-----VP-WVLLSMDSFYK--VLTE
TC909826_Mus QSKEAFAIGLGGGSASGK-TTVARMII EALD-----VP-WVLLSMDSFYK--VLTE
15232108_Arabidopsis APHQPFVIGVTGGTASGK-TTVCDMIIQQLH-----DH-RIVLVNQDSFYR--GLT-
15222778_Arabidopsis EHGQPFVIGVAGGAASGK-TTVCDMIMQQLH-----DQ-RAVVVNQDSFYH--NVN-
TC162174_Glycine -----
18416764_Arabidopsis -----MIIQQLH-----DQ-RVVLINLDSFYH--NLT-
TC193979_Zea -----Q-RVAVTQESFYI--GLT-
15232164_Arabidopsis PLKQPFVIGVAGGTASGK-TTVCNMIMSOLH-----DQ-RVVLVNQDSFYH--SLT-
Dictyostelium_sanger ---NPLVIGVCGGSASGK-TTVCDKIIANLN-----VR-WVLLSMDSFYKN-----
chlamyUKUPRT -----
CryptoUKUPRT SNSNVFVI AVAGGSASGK-TSVCTRIFSELG-----DK-RVTVIETDSFYK--TPV-
137110_Saccharomyces PWTTPYIIGIGGASGSGK-TSVAAKIVSSIN-----VP-WTVLISLDFNYYNPLGPE-
17539892_Caenorhabditis SLKHPFVIGVCGGSASGK-TTVAEKIVERLG-----IP-WVTILSMDSFYK--VLTP
28923467_Neurospora ---DVSIIIGIAGSSGSGK-STLSQAI VKKLN-----LP-WVILSIDSFYK--SLDA
2499951_Toxoplasma -----
21358379_Drosophila -----
21450816_Homo -----
Paramecium4UPRT -----
18312741_Pyrobaculum -----
29348200_Bacteroides -----
17546907_Ralstonia -----
22968125_Rhodospirillum -----
16080742_Bacillus -----
15835447_Chlamydia -----

15803021_Escherichia -----
 23475563_Desulfovibrio -----
 22970181_Chloroflexus -----
 15643484_Thermotoga -----
 19551914_Corynebacterium -----
 23126288_Nostoc -----
 6647900_Nicotiana -----
 15607101_Aquifex -----
 28917005_Neurospora -----
 29841457_Schistosoma -----
 17506069_Caenorhabditis -----
 18203426_Deinococcus -----
 19703818_Fusobacterium -----
 22654174_Streptococcus -----
 120609_Saccharomyces -----
 22654176_Mesorhizobium -----
 Leishmania_sanger -----
 Tbrucei_sanger -----
 Crypto_uprt -----

18699734_Homo EQKAKALKGQFNFDHPDAFDNELILKTLKEITEGKT-----VQIPVYDFVSHS-R
 13899253_Homo EQKAKALKGQYNFDHPDAFDNDLMHRTLKNIVEGKT-----VEVPTYDFVTHS-R
 24649624_Drosophila AEKAKAQKGLFNFDHPDAFNEELMYSTLQNILKGHK-----VEIPSYDYRTNS-L
 17538198_Caenorhabditis EEKILAREGKFNFDHPDQINFDLLAETLQNMIDGKT-----VEIPKYDMITSS--
 28828393_Dictyostelium -KEDRETVASYNFDHPDAFDWLSVENALRDLKDGKD-----IEIPTCYFKSHSRL
 Entamoeba_sangerUK TKEEHDNVAEYNFDSPSSFDWDLIIDTLKKIKAKKN-----VSLPVYDYVTHSRK
 ParameciumlUK --KSFYDISQYNFDHPDSDLDFDNAFEVISTLLSGRP-----ASVPIYCFVNHK-R
 15641051_Vibrio -HLSMEERVKTNYDHPDSDLDHLLCQHLQMLARGEA-----VEVPEYSYTEHT-R
 26248441_Escherichia -HLSMEERVKTNYDHPDSDHSLLEHLQALKRGS-----IDLPVYSYVEHT-R
 16272099_Haemophilus -HEMTERVKTNYDHPNSMDRDLIQLHLKLNKNGSA-----VDVPVYSYVEHT-R
 15613838_Bacillus -QLSLEERLQTNFDHPLAFDNDLLIEHLHSLLNQQA-----IEKPVYDYKLHT-R
 15594361_Borrelia -DYEHE-FSKVNFDPDAFDNLFYEHLKLNKKNKNSP-----IDMPYDFINHK--
 20455345_Chlamydomonada -HYTPEERANLIWDHPDAFDNDLLISDIKRLKNNIEI-----VQAPYDFVVLGN--
 20455343_Deinococcus -DIPFESRLHTNYDHPDAFDWALLREQLDALLAGVP-----IEMPEYDFVQHT--
 1351377_Mycoplasma -NKPLLKRKTINFDHPDAFDWKLRLSHIEDLLNGSI-----VNVPLYDYINYT--
 20455342_Ureaplasma -QLNKKERRLINYDHPSSFEWELMREQLSDIKKRKK-----IKVPIYDYKTEI--
 20455333_Listeria -DISFEERLKVNYDHPDPLAFDLDLISHIAALRRYET-----IEKPIYDYEKYT--
 17229842_Nostoc -QORA--EIGITALHPDCNHLDIMQOHLSSLRTGQP-----ILKPVYSHKTGS--
 16331050_Synechocystis -QGRK--AAGVTALDPRANFDLMEQIKTLKSGQS-----IMKPIYNHETGL--
 15790282_Halobacterium -DRPFEEERANANYDHPDFAFEWELLRTHIDALLSGQS-----IEMPQYDFERHV--
 6320223_Saccharomyces -IRVINLDNMIENIKSYNNNDYDFDNILNLVYEKH-----AVTSQNDMIQHD--
 1730776_Saccharomyces -VPVDAKYNIQNWDSPREALDFKFLGKELDVIKQTGKIATKLIHNNNVDDPFTKFHIDR--
 6324339_Saccharomyces EDRAFAFKNEYDFDEPNAINLDLAYKCILNLKEGKR-----TNIPVYSFVHHN-R
 22026891_Drosophila KQHEQALINEYNFDHPDAFDIELLLDVLTKLKEGKR-----VEVPVYNFVTHG-R
 8923487_Homo QQEQAAAHNNFNFDHPDAFDLDLIIISTLKKLKQGS-----VKVPIYDFTTHS-R
 TC909826_Mus QQEQAAACNNFNFDHPDAFDLDLIIISTLKKLKQGRS-----VQVPIYDFTTHS-R
 15232108_Arabidopsis -SEELHVVQEYNFDHPDAFDTEQLLHCVDILKSGQP-----YQIPIYDFKTHQ-R
 15222778_Arabidopsis -EVELVRVHDYNFDHPDAFDTEQLLSSMEKLRKQQA-----VDIPNYDFKSYK-N
 TC162174_Glycine -----DYNFDHPEAFDTEQLLRVMDKLNKHSQA-----VDIPKYDFKGYK-N
 18416764_Arabidopsis -EELARVHEYNFDHPDAFDTEHLLSCMEKLRQQA-----VDIPKYDFKTYR-S
 TC193979_Zea -DEELVHVHNYNFDHPDAFHTELLSCMQLKKGKA-----VDIPNYNFKTYK--
 15232164_Arabidopsis -KEKLNKVHEYNFDHPDAFNTEVLLSCMEKLRSGQP-----VNIPSYDFKIHQ--
 Dictyostelium_sanger -LSKDNDPSKYNFDHPDAFDYDLMVKTISELRAGKK-----VNIPKYCFKTHS-R
 chlamyUKUPRT -----VQEGKS-----AEVPIYDFAKHS-R
 CryptoUKUPRT -LEEGQTMADYNFDHPNSVDFELLYNVLLSLKNGEG-----VHIPNYCFKQHK--
 137110_Saccharomyces -DRARAFKNEYDFDEPNAINLDLAYKCILNLKEGKR-----TNIPVYSFVHHN--
 17539892_Caenorhabditis EEIKAAHESRYNFDGPNAFDLDLLEVLKRLREGKS-----VDVPVYDFNTHS-R
 28923467_Neurospora ESSRKAFMNEYDFDSPEALDFDVLVDRQLDLKAGKR-----AEIPVYSFEKHA-R

2499951_Toxoplasma -----
 21358379_Drosophila -----
 21450816_Homo -----
 Paramecium4UPRT -----
 18312741_Pyrobaculum -----
 29348200_Bacteroides -----
 17546907_Ralstonia -----
 22968125_Rhodospirillum -----
 16080742_Bacillus -----
 15835447_Chlamydia -----
 15803021_Escherichia -----
 23475563_Desulfovibrio -----
 22970181_Chloroflexus -----
 15643484_Thermotoga -----
 19551914_Corynebacterium -----
 23126288_Nostoc -----
 6647900_Nicotiana -----
 15607101_Aquifex -----
 28917005_Neurospora -----
 29841457_Schistosoma -----
 17506069_Caenorhabditis -----
 18203426_Deinococcus -----
 19703818_Fusobacterium -----
 22654174_Streptococcus -----
 120609_Saccharomyces -----
 22654176_Mesorhizobium -----
 Leishmania_sanger -----
 Tbrucei_sanger -----
 Crypto_uprt -----

18699734_Homo KEE-----TQTVYPADVVLFEGLAFYS---QEVRLDFQMKLFVDTD----ADTRL
 13899253_Homo LPE-----TTVVYPADVVLFEGLVFYS---QEIRDMFHLRLFVDTD----SDVRL
 24649624_Drosophila DFE-----NVLVIYPADVVLFEGLVFYF---PKIRELFHMKLFVDTD----SDTRL
 17538198_Caenorhabditis -MN-----GTVTVEPAKVIIIEGILLLYD---ERVKLLSTKLFVEKN---AESRL
 28828393_Dictyostelium EET-----VGLGDVDVILFEGLISFYT---QSLREQMDIKIFVDTD----SDTRL
 Entamoeba_sangerUK PD-----WVSVETGDVVIFEGLYTFYQMKKEYENYFDMFDLKIIFIESD----NDTRL
 Paramecium1UK LDQ-----CSIMLP-AQVIVFEGLMALYD---ERIRDLMSYKIFIHCD----DDIRL
 15641051_Vibrio TEQ-----TTTMTPKKVIIEGILLLTD---PRLRDLMHATVFMDDTP---LDICL
 26248441_Escherichia MKE-----TQTVVEPKKVIIEGILLLTD---ARLRDELNFSIFVDDTP---LDICL
 16272099_Haemophilus TNE-----THTFTPKRIVILEGILLLTD---ERVRLQADISVFVDDTP---LDICF
 15613838_Bacillus SNE-----VILVEPKDVIIEGILLLED---ERLRELMDIKLFVDTD----ADIRI
 15594361_Borrelia RQL-----KTVLVVPVTPVVIVEGIMIFVE---ERVRNLIDLKIYIDTP---NDIRF
 20455345_Chlamydophila RSK-----TEIETIYPSKVIIVFEGLVFEN---QELRDLMDIRIFVDTD----ADERI
 20455343_Deinococcus RAA-----HTTRVLPGRVVVLEGGFFALYD---EELRSRMGLKVFVDAD---ADVRF
 1351377_Mycoplasma RAK-----KTAKIGPIDVVILEGLMPWFD---EKLSRLSKLKIIFIETN---GEERL
 20455342_Ureaplasma RLD-----KTIDISDVDVIVFEGLIYAIYD---DVINQIADLKIIFIETP---KDECL
 20455333_Listeria RKQ-----EVEIQEPREVIIEGILILED---KRLRDLMDIKVYVDDTP---DDIRF
 17229842_Nostoc FE-----PPQYIKPNKFVIIIEGLLGYST---RAARDCYDVKVYLAPP---ENLRA
 16331050_Synechocystis LD-----PPEKVEPNKVVVIEGLHPLYD---ERVRELVDVFGVYLDIS---EEVKI
 15790282_Halobacterium RKA-----DRVVVEPTDVIVLEGILALSD---ETVNDMLDLHIYVETD---ADVRI
 6320223_Saccharomyces YE-----D---PIDLIIVCGCYALYD---KRINEISQLKVFLLSD---ADKRL
 1730776_Saccharomyces QVWDELKAKYDSINDDKYEVVIVDGMIFNN---TGISKKFDLKIIVRAPYEVLLKRRRA
 6324339_Saccharomyces VPD-----KNIVYIGASVVVIEGIYALYD---RRLLDLMDLKIYVDAD---LDVCL
 22026891_Drosophila ESQ-----TKTMYG-ANVIFEGILTFHS---PEVLKLLDMKIFVDTD---PDIRL
 8923487_Homo KKD-----WKTLYG-ANVIFEGIMAFAD---KTLLELLDMKIFVDTD---SDIRL
 TC909826_Mus KKD-----WKTLYG-ANVIFEGIMAFAD---KTLLELLDMKIFVDTD---SDIRL
 15232108_Arabidopsis KVD-----AFRQVNACDVIIEGILVFHD---SRVRDLNMNKIFVDTD---ADVRL
 15222778_Arabidopsis NVF-----PPRRVNPSDVIIEGILIFHD---PRVRDLNMNKIFVDAD---ADVRL

TC162174_Glycine	DVF-----PARRVNPADVIIILEGILVFHD----PRVRALMNMKIFVDTD----ADVRL
18416764_Arabidopsis	SVF-----RRVNPTDVIILEGILLFHD----PRVRKLMNMKIFVCTD----ADVRL
TC193979_Zea	SVP-----NARKVNPSDVIILEGILVFHD----SRLRDLNMNMKIFVDTD----ADVRL
15232164_Arabidopsis	SIE-----SSSPVNPADVIIILEGILVLDND----PRVRDLNMNMKIFVDTD----ADVRL
Dictyostelium_sanger	LVH-----QDTVYGADVIIILEGILTLYS----KELRDLMDIKIFIDTD----DDVRL
chlamyUKUPRT	SEE-----THTTGPADVIVVEGILVLM----QEVREMCNMKIFVDTD----DDLRL
CryptoUKUPRT	RLE-----TGRKVSPASIIIVEGIFILFH----PKIRHLINMSIFVDTD----DDIRL
175110_Saccharomyces	RVP-----DKNIVIIYGASVVVIEGIYALYD----RRLDLMDLKIYVDAD----LDVCL
17539892_Caenorhabditis	DPN-----SKMMYG-ADVLIFEGILAFHD----ERIKNLMDMKVFVDTD----GDLRL
28923467_Neurospora	LDR-----TTAIYS-PHVLILEGIFALYD----PRVLELLDMKIFCEAD----ADTCL
2499951_Toxoplasma	-----
21358379_Drosophila	-----
21450816_Homo	-----
Paramecium4UPRT	-----
18312741_Pyrobaculum	-----
29348200_Bacteroides	-----
17546907_Ralstonia	-----
22968125_Rhodospirillum	-----
16080742_Bacillus	-----
15835447_Chlamydia	-----
15803021_Escherichia	-----
23475563_Desulfovibrio	-----
22970181_Chloroflexus	-----
15643484_Thermotoga	-----
19551914_Corynebacterium	-----
23126288_Nostoc	-----
6647900_Nicotiana	-----
15607101_Aquifex	-----
28917005_Neurospora	-----
29841457_Schistosoma	-----
17506069_Caenorhabditis	-----
18203426_Deinococcus	-----
19703818_Fusobacterium	-----
22654174_Streptococcus	-----
120609_Saccharomyces	-----
22654176_Mesorhizobium	-----
Leishmania_sanger	-----
Tbrucei_sanger	-----
Crypto_uprt	-----

18699734_Homo	SRRVLRDISERGRDLEQILSQYITFVKPAFEEFCLPTKKYADVIIIPRGA--D-----NL
13899253_Homo	SRRVLRDVR-RGRDLEQILTQYTTFVKPAFEEFCLPTKKYADVIIIPRGV--D-----NM
24649624_Drosophila	ARRVPRDINERGRDLDAVLTQYMTFVKPAFEEFCSPTKKFADVIIPRGA--D-----NT
17538198_Caenorhabditis	RNRLATYIRDYHRAPLSIIROYTEFVKPAFEEFCRPTKKYADVIIIPRGA--D-----NH
28828393_Dictyostelium	SRRVMRDIAERGRSLEGLVLYQYEFVKPAFDDYILPTKKHADVIIPRGA--D-----NV
Entamoeba_sangerUk	ARRILRDINRGRRTLDSVLFQYKKFVKPAYDKWVYPQRKRADIIVP-----
ParameciumlUK	CRRILRDVKERQRTVESVLFQYTKFVKRAFENFIEPLMSQADLIIPGHR--S-----NH
15641051_Vibrio	LRRVKRDVQERGRTMESVLKQYQETVVRPMMQFIEPSKQYADIIVPRGG--K-----NR
26248441_Escherichia	MRRIKRDVNERGRSMDSVMAQYQKTVVRPMMFLQFIEPSKQYADIIVPRGG--K-----NR
16272099_Haemophilus	IRRLQDMEERGRSLQSVLDQYRATVVRPMMFLQFIEPSKQYADIVIPRGG--K-----NR
15613838_Bacillus	IRRMVRDIRERGRTLESVIEQYTKVVRPMMHMQFIEPTKRYADVIIPEGG--Q-----NR
15594361_Borrelia	IRRLRRDISKRGRTVESVIDQYLNTRRWGYRFFIEPTKEYADLIIPEGG-----HND
20455345_Chlamydophila	LRRMVRDVQEQGDSVDCIMSRYLSMVKPMHEKFFIEPTRKYADIIVHGNY-----RQN
20455343_Deinococcus	IRLLRDTQERGRTPESVIEQYLGVRPMMHLSFVEPTKRYADVIIIPGG-----MNE
1351377_Mycoplasma	IRRIERDQW-RGRNIDSIIKQWREIVAPMYEIFVEKMKRNADLILPWSQ-----RRE
20455342_Ureaplasma	IRRIIRDVNERNRSFESVITQWRSTVSPMYDQFVEPSKKNANVSVLWNE-----HNR
20455333_Listeria	IRLLLRDMKERGRTMDSVIDQYLSVVKPMHNEFIEPTKKFADIIIPGG-----ENH
17229842_Nostoc	KWKVKRDTQKRGYTFEQVLAELE-KREPDSEQFIRPQRQWSDIIVVSFYPPHEDGDETNGH

16331050_Synechocystis	NWKIQRDMAERGHTYEDILASIN-ARKPDFTAYIEPQKQYADVVIQVLP-----TRL
15790282_Halobacterium	LRRIERDVVERGRELEGVMDQYLSTVKPMHEQFIEPTKRHADII IPEG-----ANS
6320223_Saccharomyces	ISLIKKKNVGSNEQLAQLITEYMDHLRPEMQQYIEPTRTFADLIIPSTNEN-----LGRA
1730776_Saccharomyces	SRKGYQTLDSFWVDPYFYDFEFVYESYRANHAQLFVNGDVEGLLDPRKS-----KNI
6324339_Saccharomyces	ARRLSRDIVSRGRDLGCIQQWEKFKVKNPNAVKFKVPTMKNADAIIPSMS--D-----NA
22026891_Drosophila	ARRLRRDISQRGRDLKGVKQYLNVMVKPSYCNYIAPTMAHADIIIVPRGG--D-----NK
8923487_Homo	VRRLRRDISERGRDIEGVIKQYNKFKVPSFDQYIQPTMRLADIVVPRGS--G-----NT
TC909826_Mus	VRRLRRDISERGRDIEGVIKQYNKFKVPAFDQYIQPTMRLADIVVPRGS--G-----NT
15232108_Arabidopsis	ARRIRRDIVERGRDVSLEQYAKFVKPAFDDFVLPSSKKYADVIIIPRGG--D-----NH
15222778_Arabidopsis	ARRIKRDTVEKGRDIATVLDQYSKFKVPAFEDFILPTKKYADIIIPRGG--D-----NH
TC162174_Glycine	ARRIKRDTADNAREIGAVLDQYSKFKVPAFDDFILPTKKYADIIIPRGG--D-----NH
18416764_Arabidopsis	ARRIKRDTVENGRDIGTVLDQYSKFKVPAFDDFILPTKKYADIIIPRGG--D-----NH
TC193979_Zea	TRRIHRDTIDKGRDLKAVLEQFSKFKVPAFEDFILPTKKYADVIIIPRGG--D-----ND
15232164_Arabidopsis	SRRIQRDIVERGRNIQNVLEQYTKFKVPSFDEYIQPSMKYADIIIPRGG--D-----ND
Dictyostelium_sanger	ARLKRDI AERGR TLESVLHQYNTFVKPSFDYII-----PRGFDPE-----KT
chlamyUKUPRT	ARRIQRDVAHRGRDVASVLKQYTEHVKPMFDMYVAPSRRHADVIIIPWGNRDN-----NL
CryptoUKUPRT	VRRIRRD TIERGRQIDDI LNQYEKTVKPSYDEFIYPTRRYADIVIPHY-----NE
137110_Saccharomyces	ARRLSRDIVSRGRDLGCIQQWEKFKVKNPNAVKFKVPTMKNADAIIPSMS-----DNA
17539892_Caenorhabditis	ARRIVRDVTDGRGRDIDGIMEQYFTFVKPAFDKYIAPCMDSADLIVPRGG--E-----ND
28923467_Neurospora	SRRI LR DVKDRGRDVGIIKQWFNYVKPNFEKFKVDPQRKVADIIIVPRGV--E-----NH
2499951_Toxoplasma	-----MAQVPAS-----GKLLVDPRY--S-----TN
21358379_Drosophila	----SSGSQSEEGSSSDHQPPQ-----EAQEQQQLHTPT--H-----AH
21450816_Homo	----SMPCHNQVNSASTPSPEQLRPGDLILDHAGGNRASRAKVILLTG--Y-----AH
Paramecium4UPRT	-----
18312741_Pyrobaculum	-----
29348200_Bacteroides	-----
17546907_Ralstonia	-----
22968125_Rhodospirillum	-----
16080742_Bacillus	-----
15835447_Chlamydia	-----
15803021_Escherichia	-----
23475563_Desulfovibrio	-----
22970181_Chloroflexus	-----
15643484_Thermotoga	-----
19551914_Corynebacterium	-----
23126288_Nostoc	-----
6647900_Nicotiana	-----
15607101_Aquifex	-----
28917005_Neurospora	-----
29841457_Schistosoma	-----
17506069_Caenorhabditis	-----
18203426_Deinococcus	-----
19703818_Fusobacterium	-----
22654174_Streptococcus	-----
120609_Saccharomyces	-----
22654176_Mesorhizobium	-----
Leishmania_sanger	-----
Tbrucei_sanger	-----
Crypto_uprt	-----
18699734_Homo	VAINL-IVQHIQD--ILN-----
13899253_Homo	VAINL-IVQHIQD--ILN-----
24649624_Drosophila	VAIDL-IVHHIGE--ILA-----
17538198_Caenorhabditis	VATDL-IAKNLQV--RFK-----
28828393_Dictyostelium	VAIDL-IVQHISS--KLS-----
Entamoeba_sangerUk	-----
Paramecium1UK	VSVAF-IVNHLKN--MAK-----
15641051_Vibrio	IAIDV-LKAHIK--LLK-----

26248441_Escherichia	IAIDI-LKAKISQ--FFE-----
16272099_Haemophilus	IAINM-LKAQILH--LLN-----
15613838_Bacillus	VAIDL-MVTKIRA--IIE-----
15594361_Borrelia	KALYV-LSTFLKS--LSK-----
20455345_Chlamydomphila	VVTNI-LSQKIKN--HLE-----
20455343_Deinococcus	PALDM-LAARIR-----
1351377_Mycoplasma	VSTSV-LDVAIEH--LFH-----
20455342_Ureaplasma	VALHL-IHKWINN--IH-----
20455333_Listeria	VAIDL-MTKIES--ILQ-----
17229842_Nostoc	LNVRLVLRPTIPHPDFTS-----
16331050_Synechocystis	IEDK--ESKLLRV--RLV-----
15790282_Halobacterium	VAVNL-LEEKVQA--ETS-----
6320223_Saccharomyces	VLVDG-IVKAIEDTKSQI-----
1730776_Saccharomyces	KEFINDDDTPIAKPLSWV-----
6324339_Saccharomyces	TAVNL-IINHKS--KLEIHELPTN-QVLSLHTMLLNK--NLNCADFVIFYDRLATILL
22026891_Drosophila	VAIHL-IVQHVHT--QLQLHLLHPTP-QIKGLHTFIR--CRNTSRDEFIFYSKRLIRLVI
8923487_Homo	VAINL-IVQHVHS--QLELSVLKSTP-QVRGMHTIIR--DKETSrDEFIFYSKRLMRLLI
TC909826_Mus	VAIDL-IVQHVHS--QLELSVLKSTP-QVRGMHTIIR--DKETSrDEFIFYSKRLMRLLI
15232108_Arabidopsis	VAVDL-IVQHIHT--KLGVFVIETTF-QIRGMHTLIR--EKDISKHDFVIFYSDRLIRLVV
15222778_Arabidopsis	VAIDL-IVQHIHT--KLGlyVIQSTF-QIRGMHTLIR--DSKTTKHDFIFYSDRLIRLVV
TC162174_Glycine	VAIDL-IVQHIRT--KLGlyVIQSTF-QIRGMHTLIR--DAKTTKHDFIFYSDRLIRLVV
18416764_Arabidopsis	VAIDL-IVQHIRT--KLGlyVIHSTF-QIRGMHTLIR--DSQTTKHDFVIFYSDRLIRLVV
TC193979_Zea	VAIDL-IVQHIRT--KLGlyVIQTTY-QIRGMHTIIR--DAATATHDFIFYADRLIRLVV
15232164_Arabidopsis	VAIDL-IVQHIRT--KLCIFIISSTF-QIKGMHTLIR--DINTTKHDFVIFYADRLIRLVV
Dictyostelium_sanger	AQLDL-----IHVIKETN-QIKAMLSILR--NKDTKVGFVIFYSDRLCSLII
chlamyUKUPRT	VAIDL-IVEHIRT--KLOFQIIPSNF-QIQGMHTIIR--DRETSKEDFVIFYADRLNRLVV
CryptoUKUPRT	VAVDL-VVQHLRY--KLKLIIPSNF-QIRHMHSIIR--NKDTSVVDVFWSDRLIRLVV
137110_Saccharomyces	TAVNLIINHISKLELKSQSSPSQDV-LNRNIIHELP--PTNQVLSLHTMLLNKNLNCAD
17539892_Caenorhabditis	VAIDM-IVQNVMA--QLVLFILKETP-QVKGLVTFVR--DRETSrDNHIFYSDRLMRLII
28923467_Neurospora	VAMTM-VTQFIQQ--KLLVVQLDQTP-QLRGMNTLIH--DIDTSKEDFIFYFNRF'TLLI
2499951_Toxoplasma	DQEES-ILQDI-----VVLKQTA-QLRAMMTIIR--DKETPKKEFVIFYADRLIRLLI
21358379_Drosophila	AAVPA-ATSE-----LKLLECNs-QVAELLTILR--DKNTTRSDFKIFYADRLIRLVI
21450816_Homo	SSLPA-ELDSGAC--GGSLKLLPMND-QIRELQTIIR--DKTASRGDFMFsADRLIRLVV
Paramecium4UPRT	-----MFKQILWY--QQQVGFDFRKNADRLIRILI
18312741_Pyrobaculum	-----VRIID-HV-YAQYLLTKLR--DKNTGSIEFRKGLVRLGRIG
29348200_Bacteroides	-----IDFSQTNs-ILNQYISEIRNVEVQNDRLRFRNRIERGEIMA
17546907_Ralstonia	-----LFILNHP-LIQHKLTHMR--DKDTSTRTFRELLREITLLMG
22968125_Rhodospirillum	-----LFVLDHP-LIQHKLTHMR--RRETSTRTFRMLLREIALLLMG
16080742_Bacillus	-----VYVFDHP-LIQHKLYIR--NENTGTKDFRELVEVATLMA
15835447_Chlamydia	-----AVYVLSHP-LIQHKASLLR--NKNTKSKIFRECLKEISLVC
15803021_Escherichia	-----IVEVKHP-LVKHKLGLMR--EQDISTKRFRLEASEVGSLLT
23475563_Desulfovibrio	-----VYAVDHP-LVKHKLGRRLR--QSDVPVAEFRALSNEVCRLLT
22970181_Chloroflexus	-----VFVSRHP-LVQHKLALLR--SKWTEPKKFRRELVEIAQLLF
15643484_Thermotoga	-----LVVVDHP-LIKHKLTIMR--DKNTGPKFRRELLREITLLLA
19551914_Corynebacterium	-----ITIVNHP-LVASRLTLR--DERSDNAAFRAAANDLGAMLI
23126288_Nostoc	-----RVYVPPHP-LIKHWLAVAR--DAGTPSVLFRSAMTELGRWLT
6647900_Nicotiana	-----LVFVPPHP-LIKHWVSVLR--NEQTPCVIFRNAMSELGRLLM
15607101_Aquifex	-----MIVELSHP-LIKHKVNTAR--IQDTSAEKLRKTLKELGFMLV
28917005_Neurospora	-----VTVLSQTP-QLIALLSIIR--DKRTDRGDFIFYSNRIIRLLV
29841457_Schistosoma	-----IILLQSD-HVRVLQTVIR--NRETPrNEFLFNADCLIRLVV
17506069_Caenorhabditis	-----FVLLKKTp-QILELQTIK--DRSTNHSDVFNADRLMRLVI
18203426_Deinococcus	-----VTLV-EHP-LVQHKLsLMR--DERTGVKEFRRELAGEsLLLA
19703818_Fusobacterium	-----VIEI-NHP-LIEHKMTILR--SVDTDTKsFRLENLIEIAKLMT
22654174_Streptococcus	-----IEVI-NHP-LIQHKLSILR--RTDTSTKAFRELVEIAMLLMG
120609_Saccharomyces	-----VYLL-PQTNQLLGLYTIIR--NKNTTRPDFIFYSDRIIRLLV
22654176_Mesorhizobium	-----VTVV-DHP-LVQHKLTIMR--KKETSTAGFRLLREISLLLG
Leishmania_sanger	-----HLHLLPQTPQLHFLFTVIR--DVETQRTDFIFYSERIIRLIF
Tbrucei_sanger	-----TFHLLPQTNQLHFLFTIIR--QRETGRTDVIFYSERIIRLIV

Crypto_uprt -----SLHICTQTPQLKGVMTILR--DANTSKEDFVFYTDTRICRIVL

18699734_Homo -----
13899253_Homo -----
24649624_Drosophila -----
17538198_Caenorhabditis -----
28828393_Dictyostelium -----
Entamoeba_sangerUk -----
ParameciumlUK -----
15641051_Vibrio -----
26248441_Escherichia -----
16272099_Haemophilus -----
15613838_Bacillus -----
15594361_Borrelia -----
20455345_Chlamydomydia -----
20455343_Deinococcus -----
1351377_Mycoplasma -----
20455342_Ureaplasma -----
20455333_Listeria -----
17229842_Nostoc -----
16331050_Synechocystis -----
15790282_Halobacterium -----
6320223_Saccharomyces -----
1730776_Saccharomyces -----
6324339_Saccharomyces SWALDDIP-VAHTNIITPGEH-TMENVIAC-QFDQVTAVNIIRSGDCFMKSLRKTIPNIT
22026891_Drosophila EYALSFLP-FKKTIVETPQGV-LYEGKRME-S-RKICGVSILRAGETMEQAVCDVCKDIR
8923487_Homo EHALSFLP-FQDCVVQTPQGO-DYAGKCYA-G-KQITGVSILRAGETMEPALRAVCKDVR
TC909826_Mus EHALSFLP-FQDCTVQTPQGO-DYVVKCYA-G-KQITGVSILRAGETMEPALRAVCKDVR
15232108_Arabidopsis EHGLGHLF-FTEKQVVPTPTGS-VYSGVDFC---KKLCGVSVIRSGESMENALRACCKGIK
15222778_Arabidopsis EHGLGHLF-FTEKQVVPTPTGS-VYSGVDFC---KKLCGVSVIRSGESMENALRACCKGIK
TC162174_Glycine EHGLGHLF-FTEKQVITPTGS-VYTGVDFC---KRLCGVSVIRSGESMENALRACCKGIK
18416764_Arabidopsis EHGLGHLF-FTEKQVITPTGC-VYSGVDFC---KRLCGVSVIRSGESMENALRACCKGIK
TC193979_Zea EHGLGHLF-FQEKQVITPTGS-VYTGVDFC---KRLCGVSVIRSGESMENALRACCKGIK
15232164_Arabidopsis EHGLGHLF-FTEKQVITPTGS-VYSGVDFC---KRLCGVSVIRSGESMENALRACCKGIK
Dictyostelium_sanger EEALTYLP-FTEKIVTPTPTGS-LYHGEELN---SRICALVVLVLRAGGCMEQPLRSICKGIR
chlamyUKUPRT EAGLGFLLP-FTERSIVTPTGN-AYSGVGFA---RKLKCGVSVIRSGEAMEAALRACCKGIK
CryptoUKUPRT ENALSHLS-FTGQTIETPIGE-LYDGVQFN-YKDKLCAVSVIRGGESMEIGLSAVCKDIP
137110_Saccharomyces FVFYFDRL-ATILLSWALDD-IPVAHTNII-TPGEHTMENVIACQFDQVTAVNIIRSGDC
17539892_Caenorhabditis EECMNHMP-YKDEIEMAGGR-KTIGKRKD---AQICGLPIMRAGECMETALRSIVKDCV
28923467_Neurospora EKAMDNIP-FASTTITTPLN-AYAGLRPR---GQVSAVVVLRGGALEAGLKRVLDPCK
2499951_Toxoplasma EEALNELP-FEKKEVTTPLDV-SYHGVSFY---SKICGVSIVRAGESMESGLRAVCRGCR
21358379_Drosophila EESLNQLP-YTHCDVETPTGA-IYEGLYK---SGNCGVSVIRSGEAMEQGLRDCCRSIR
21450816_Homo EEGLNQLP-YKECMVTTPTGY-KYEGVKFE---KGNCGVSVIRSGEAMEQGLRDCCRSIR
Paramecium4UPRT EQAISQIE-KKKHIKQSPLG-----YY-DAHEICFVSVILRSGNAFLIEALKVMTGAS
18312741_Pyrobaculum YELAKTFP-VKYVEIETPLG---KAVGVDIIGL-DKVVIVQVLRRAAMPLVEGLVKAFDQAR
29348200_Bacteroides YEMSKEFR-YSVKNIRTPLG---IAPVSTPD--NQLVISTILRAGLPHFQGLSYFDGAE
17546907_Ralstonia YEITRNLP-LTTRHIETPMG---PMDAPVIA-G-RKLAVVPVLRAGVGMDSGLLELIP SAR
22968125_Rhodospirillum YEVTRSLP-LTTESIETPLE--AMEAPVIG-G-KKLVVVPILRAGQGMADG LLELIP SAR
16080742_Bacillus FEITRDLP-LEEVDINTPVQ---AAKSKVIS-G-KKLGVPVILRAGLGMVDGILKLIPAAK
15835447_Chlamydia YEATRDLA-LKNISIQTPLM--QAECPLT-G-TKIVVIPVLRAGLGMVDGFLALV PNAK
15803021_Escherichia YEATADLE-TEKVTIEGWNG--PVEIDQIK-G-KKITVVPILRAGLGMVDGLENV P SAR
23475563_Desulfovibrio YEATKDLA-TEKTQVEGWAG--PVEVEQIK-G-KKITVVPILRAGLGMVDG LLD MIPGAK
22970181_Chloroflexus YEATQDLA-LAPLTVETPLA--TCAGYEVA-E-R-IGIIPILRAGLGMVAEIVEILPTVH
15643484_Thermotoga YEATRHLK-CEEVEVETPIT--KTIGYRIN-D-KDIVVVPILRAGLVMADGILELLPNAS
19551914_Corynebacterium YEASRDLE-VEHFDTKTPVAMAEG--TRL--K-QPPIIVPIIRAGLGMIDPALS MIPDAQ
23126288_Nostoc YEAARDWLPTQETAVQTP LDT--CPATVIDPQ-IPVAVVPILRAGLGLLEGAQTLLPLAS
6647900_Nicotiana YEASRDWLPIITGEIQSPMGV--ASVEFVDP-EPVAIVPILRAGLALAEHASSILPATK
15607101_Aquifex YEALKDIL-LEEKEVRTWIGN--KRFNYLN-E-EEIVFVPILRAGLSFLEGALQVVPNAK

28917005_Neurospora EEGLNHLP-TVEHPVTTTPVGR-IYDGLAFQ---GKICGVSIMRAGEAMEQGLRDCCRSVR
29841457_Schistosoma EEGLNQLP-YENVCVTTPTGN-LYHGKIFL---RGNCGVSIMRSGEAMERGLRDCCRSMR
17506069_Caenorhabditis EECLNHLP-FTEHTVTTPTGF-KYEGIQFN---RGNCGVSLCRSGEAMEVSLRQCCRCIR
18203426_Deinococcus YEAMRDLE-VEPVRFATPIE-E-GDFPMLS--GKKLALVAILRAGLVMTDAIVTLVPAAK
19703818_Fusobacterium YEATKNLK-LETTEVTTPLM-K-TQAYTLQ--DK-VALVPILRAGLGMVDGILDLIPTAK
22654174_Streptococcus YEVLRLDP-LEDVEIETPIT-K-TVQKQLA--GKKLAIVPILRAGIGMVDGLLNLVPAAK
120609_Saccharomyces EEGLNHLP-VQKQIVETDTNEN-FEGVSFM--GK-ICGVSIVRAGESMEQGLRDCCRSVR
22654176_Mesorhizobium YEVTRNLE-LTTTTIETPIE-T-MEAPTL--GKKLVFASVLRAGNGLLEGLLDLVPAAAR
Leishmania_sanger EAALCLIP-VKPFNVITPVGAVYRGRVDD---RGIIGVSIVRAGESMERVLREMCPCGVR
Tbrucei_sanger EAALNLIP-TIPYDVITPTGSKYQGCMTDG---QGIIGISILRAGEAMERVLRETCRQVR
Crypto_uprt EHALNLLP-YDYKEIKTPNGIEVKGI AFN----TPICGVSLIGSGEAMENALRFVCRGCR

18699734_Homo -----
13899253_Homo -----
24649624_Drosophila -----
17538198_Caenorhabditis -----
28828393_Dictyostelium -----
Entamoeba_sangerUK -----
Paramecium1UK -----
15641051_Vibrio -----
26248441_Escherichia -----
16272099_Haemophilus -----
15613838_Bacillus -----
15594361_Borrelia -----
20455345_Chlamydomyxa -----
20455343_Deinococcus -----
1351377_Mycoplasma -----
20455342_Ureaplasma -----
20455333_Listeria -----
17229842_Nostoc -----
16331050_Synechocystis -----
15790282_Halobacterium -----
6320223_Saccharomyces -----
1730776_Saccharomyces -----

6324339_Saccharomyces IGKLLIQ-SDSQT-----GEPQLHCEFLPPNIEKFGKVFLMEGQIISGAAMIMAIQ-V
22026891_Drosophila IGKILIQ-TNLKT-----GEPELYLRLPKDIKDYK-VILMDATVATGAAAMMAIR-V
8923487_Homo IGTTILIQ-TNQLT-----GEPHYLRLPKDISDDH-VILMDCTVSTGAAAMMAVR-V
TC909826_Mus IGTTILIQ-TNQLT-----GEPHYLRLPKDISDDH-VILMDCTVSTGAAAMMAVR-V
15232108_Arabidopsis IGKILIH-RDGD-----NGMQLIYEKLPDISERH-VLLLDPVLGTGNSANQAI E-L
15222778_Arabidopsis IGKILIH-REGD-----NGQQLIYEKLPDISERH-VLLLDPILGTGNSAVQAIR-L
TC162174_Glycine IGKILIH-REGD-----NGQQLIYEKLPNDISDRH-VLLLDPILGTGNSAVQAI S-L
18416764_Arabidopsis IGKILIH-REGD-----NGQQLVYEKLPNDISERH-VLLLDPILGTGNSAVEAIN-L
TC193979_Zea IGKILIH-REGD-----NGQQLIYQNLPKDIANRH-VLLLDPILGTGNSAVQAI S-L
15232164_Arabidopsis IGKILIH-RENN-----DGRQLIYEKLPKDISSRH-VFLLDPVLASGYSAVKAIT-L
Dictyostelium_sanger TGKVLIQ-SDEMK-----KPHLFYEKLP-NVTD SH-VLVLDPTIATGASSEM AIR-V
chlamyUKUPRT LGKILVHRHRDSS-----DVIYEKLP SDIADRH-VMLLDPLLTGGTAVKAIQ-I
CryptoUKUPRT IGKILLE-FQNPKTELDAQFDKPKIIYCKLPDDIASRN-VFILDPI LGNGFGVFS AIK-Y
137110_Saccharomyces FMKSLRK-TIPNITIGKLLIQSDSQTGEPQLHCEFLPP-NIEKFGKVFLMEGQIISGA-A
17539892_Caenorhabditis IGKILIQ-TNETT-----FDPELHYIRLPPHITRYK-VIIMDATVTTGSAAMMAIR-V
28923467_Neurospora TGRVLIQ-SNMRT-----GEPHYLALPKDIDQHEAVLLLD AQMSSGGSALMAVQ-V
2499951_Toxoplasma IGKILIQ-RDETT-----AEPKLIYEKLPADIRDRW-VMLLDPMC ATAGSVCKAIE-V
21358379_Drosophila IGKILVE-SDANT-----HEARVVYARFPDDIGSRQ-VLLMYPIMSTGNTV LQAVN-V
21450816_Homo IGKILIQ-SDEET-----QRAKVVYAKFPDDIYRRK-VLLMYPILSTGNTV IEAVK-V
Paramecium4UPRT IGQILIQRNEETS-----QPSYFFQKLPQNIK DQQ-VILVDPMLATGGSASMAIN-I
18312741_Pyrobaculum LGVVAAK-RREE-----GGVVDVDFYSKVPKIAEDDV-VIVADPMLATGITMSRVIE-E
29348200_Bacteroides NAFVSAY-RKYKD-----TLKFDIHIEYIASPRIDDK-T-LIITDPMLATGGSMELSYQ-A
17546907_Ralstonia VGHIGVY-RDEQH-----R-PVEYLVR LP-DLEERT-FILCDPMVATGHS AVHAVD-V
22968125_Rhodospirillum VGHIGLY-RDEVT-----RQPVEYLVKLP-ETTERT-FILVDPMLATGHS GAHAAT-V

16080742_Bacillus VGHVGLY-RDPET-----LKPVEYYVKLPSDVEERE-FIVVDPMLATGGSAVEAIH-S
15835447_Chlamydia VGLIGMS-RNHET-----FQPSSYCCKLPEDIADCH-VFIVDPMLATGGASATIQL-L
15803021_Escherichia ISVVGMY-RNEET-----LEPVYPYFQKLVSNIDERM-ALIVDPMLATGGSVIATID-L
23475563_Desulfovibrio VSVVGMF-RNEET-----LEPVTTYTKLANNIEERQ-AIIIDPMLATGGTLVATID-L
22970181_Chloroflexus VVHLGLY-RDHET-----LQPVTYYNKLPSKPDIDL-TIIVDPMLATGGSAVAVD-I
15643484_Thermotoga VGHIGIY-RDPET-----LQAVEYYAKLPPLNDDKE-VFLLDPMLATGVSSIKAIE-I
19551914_Corynebacterium VGFIGLA-RDEET-----HEPVPLYEALPQDLSNQP-VFLVDPMLATGGSLHAIR-L
23126288_Nostoc VYHLGLA-RDEET-----LQPHCYLNKLPKDFPQTRVLITDPMLATGGSIMATMA-E
6647900_Nicotiana TYHLGIS-RNEET-----LQPSVYLNKLPDKFPEGSRVIVVDPMLATGGTIVAVID-L
15607101_Aquifex VGFLGIK-RNEET-----LESHIYYSRLPELK--GKIVVILDPMLATGGTLEVALR-E
28917005_Neurospora IGKILIQ-RDEET-----AQPCLFYDKLPEDIADRW-VLLLDPMFATGGSAIMAVE-V
29841457_Schistosoma IGKILIQ-KAEEN-----KIIDAKVYYAKFPNIEHRK-VLLMYPILGWN-----S
17506069_Caenorhabditis IGKILIG--DEQ-----KVLARLLPDITSRR-VLLLYPTIGSGTTVCKAIE-V
18203426_Deinococcus VGHIGMY-RDPQS-----LAPVAYYSKLPADIAERR-VFLTDPMMLATGGSANAAIQ-N
19703818_Fusobacterium VGHIGVY-RNEET-----LEPVYYYCKLPTDIASRK-VILVDPMLATGGSAVYVID-Y
22654174_Streptococcus VGHIGMY-RDEET-----LQPVYELVKLPEDIDQRQ-IFVVDPMMLATGGSAIIVD-S
120609_Saccharomyces IGKILIQ-RDEET-----ALPKLFYEKLPEDISERY-VFLLDPMLATGGSAIMATE-V
22654176_Mesorhizobium VAHIGLY-RDHET-----LEAVEYFFKAPSDLDLADRL-VIVVDPMLATANSAAIAID-K
Leishmania_sanger IGKILVQRDETST----DKTPDARFTYSKLPDVSARR-VLLLDPMCATGGSVIKATEIL
Tbrucei_sanger IGKILVQ-RDEKS---LEKAPDERFNYSKIPSDVSGRH-VLLLDPMIATGGSVIKATEIL
Crypto_uprt IGKVLLNNSSENE-----FDNAISAAVVKLPEDV SERV-VIVMSPVLGTGNSLCLIEVL

18699734_Homo -----
13899253_Homo -----
24649624_Drosophila -----
17538198_Caenorhabditis -----
28828393_Dictyostelium -----
Entamoeba_sangerUK -----
Paramecium1UK -----
15641051_Vibrio -----
26248441_Escherichia -----
16272099_Haemophilus -----
15613838_Bacillus -----
15594361_Borrelia -----
20455345_Chlamydomyxa -----
20455343_Deinococcus -----
1351377_Mycoplasma -----
20455342_Ureaplasma -----
20455333_Listeria -----
17229842_Nostoc -----
16331050_Synechocystis -----
15790282_Halobacterium -----
6320223_Saccharomyces -----
1730776_Saccharomyces -----
6324339_Saccharomyces LLDHG-IDLEKISVVVYLATEVGIIRILNAFDNKVNI FAGMIISREKLQ--NHQYKWALT
22026891_Drosophila LLDHD-VPEDNII LASLLMAEIGVHSIAYAFPVKVIVTSALDPEINSKF--YVIPGIGNF
8923487_Homo LLDHD-VPEDKIFLLSLLMAEMGVH SVAYAFPRVRIITTAVDKRVNDLF--RIIPGIGNF
TC909826_Mus LLDHD-VPEDKIFLLSLLMAEMGVH SVAYAFPRVRIITTAVDKRVNDLF--RIIPGIGNF
15232108_Arabidopsis LIQKG-VPEAHII FLNLI SAPEGIHCVCKRFPKLKI VTSEIDQCLNEEF--RVIPGLGEF
15222778_Arabidopsis LISKG-VPESNII FLNLI SAPEGVNVVCKKFPRIKI VTSEIELGLNDEF--RVVPGMGFEF
TC162174_Glycine LIKKG-VPESNII FLNLI SAPKGVHVVCKSFPRIKI VTSEIEI GLNDEF--RVIPGMGEF
18416764_Arabidopsis LISKG-VPEGNII FLNLI SAPQGVHVVCKKFPRIKI VTSEIDNGLNEEF--RVIPGMGEF
TC193979_Zea LLKKG-VQEANII FLNLI SAPQGVHVVSKRFPRVKI VTSEIEI FGLNDDF--RVVPGMGFEF
15232164_Arabidopsis LLSKG-VPESHII FLNLI AAPQGIHALCKKFPMLKI VTSEIDSSLNEDS--RVIPGLGEF
Dictyostelium_sanger LLDHG-VPENKII FVSVIASLKGILYLNRYFPDVQFVVS AIDKELSDEG--FILPGCGFY
chlamyUKUPRT LKDRG-VPEDRILFLTI IAAPEGVHRVCSFPGVKLLTSE-----
CryptoUKUPRT LLSKG-VLQRKII VLSLIVAHNAIHRICKEFPEVTLITTEIDRDVNSDG--FVIPGLGCF
137110_Saccharomyces MIMAI-QVLLDHGIDLEKISVVVYLATEVGIIRILNAFDNKVNI FAGMI--ISREKLQNH

17539892_Caenorhabditis LLDHD-VKEEDIFVASLLMGQOGAHALAYAFPVKVLLITTAMDHQMTENC--YLIPGMGNF
28923467_Neurospora LVDHG-VKMDRIVLATYSAGRLGLHRLTTVFPEVTVVVGNVIQDVEERW--VEK-----
2499951_Toxoplasma LLRLG-VKEERII FVNILAAPOGIERVFKEYPKVRMVTA AVDICLNSRY--YIVPGLGDF
21358379_Drosophila LREHG-VPESCI ILSNLFCTPIAARTVVNAFPKLIKILTSELHP-----VAP--NHF
21450816_Homo LIEHG-VQPSVI ILLSLFSTPHGAKSIIQEFPEITILTTEVHP-----VAP--THF
Paramecium4UPRT LKHHG-VKEENITFLTLVSC EQGLSKLFSEYPKIKIITAQVDPILLKDIN-YLAPGLGDF
18312741_Pyrobaculum VYRAG-NPGR-LIVVSVIATPMGI ERVLSRWPSAEIYAVAVDPTLNDKA--FIVPGLGDA
29348200_Bacteroides MLTKG-HPAE-IHVASIIASQHAIDHIKSVFPEDKTTIWCAAIDPEINEHSYIVPGLGDA
17546907_Ralstonia MKQRG-VPDEHILFLALVAAPEGVEVFQQAHPGVKLYVASLDSHLNEHA--YIIPGLGDA
22968125_Rhodospirillum LLRHG-VPAERIRFMALVAAPEGVRTFHAAHPDVPVYVAALDRQLNDQA--FICPGLGDA
16080742_Bacillus LKKRG-AKN--IRFMCLVAAPEGVEELQKHSDVDIYIAALDEKLNEKG--YIVPGLGDA
15835447_Chlamydia VKEHG-AKN--ITLLNLVAVPEGIERIQKDHPDVTIYVASLDEKLNEKA--YILPGLGDA
15803021_Escherichia LKKAG-CSS--IKVLVLAAPEGIAALEKAHPDVELYASIDQGLNEHG--YIIPGLGDA
23475563_Desulfovibrio LKKAG-CKQ--IRGLFLVAAPEGIKRVQDAHPDVDIYVA AVDDCLNENG--YILPGLGDA
22970181_Chloroflexus LKQWG-AQR--IKFLGLIAAPEGVRALSEAHDPVAIHAAIDSHLNERG--YIVPGLGDA
15643484_Thermotoga LKENG-AKK--ITLVALIAAPEGVEAVEKKYEDVKIYVAALDERLNDHG--YIIPGLGDA
19551914_Corynebacterium LADRG-ATD--ITAICMVSAQPGVDALAESGLPVRVLTATIDPGLDENA--YIVPGLGDA
23126288_Nostoc LTQRG-ADPSLTRIVCVVAAPPALQKLSEAYPGLIVYTATIDEKLNSKG--FIVPGLGDA
6647900_Nicotiana IKERG-VDNSQIKVICAVGAPPALQKLSEKFPGLHVVYAGILDPTVNDKG--FIIPGLGDA
15607101_Aquifex ILKHS-PLK--VKSVAIAAPEGKRIEEKFKEVEIFVGNVDERLNDKG--YIIPGLGDI
28917005_Neurospora LKSRG-VPEERII FLNILASPEGISNFASKFPKLRVTAFVDQGLDDKN--YIIPGLGDF
29841457_Schistosoma FKSFG-CFANL-----
17506069_Caenorhabditis LKEAR-VPDENIYLVSLFISPTGLKNI TRKYPYITVVASDITS-----LYP--NHF
18203426_Deinococcus LKDA--GAQ-SIKLMTILAAPEGIHAVQQAHPDVDIVTASVDSRLNDHG--YIVPGLGDA
19703818_Fusobacterium LKEQ--GVT-DIIFMCLVAAPDGIARLLNKHPDVPIY TAKIDQGLNENG--YIYPGLGDC
22654174_Streptococcus LKKR--GAS-NIKFVCLVSAPEGVKALQEAHPDVEIFTAALDERLNEHG--YIVPGLGDA
120609_Saccharomyces LIKRGVKPE-RIYFLNLICSKEGIEKYHAAFPEVRIVTGALDRGLDENK--YIVPGLGDF
22654176_Mesorhizobium LKGRGA--T-NIRFLCLLAAPEGIERFTKAHPDVPVFTASIDRQLNEKG--YIMPGLGDA
Leishmania_sanger INEYG-VLEEDII FLNLISAPAGIRKYLGRFPKIQIVTAAIDDDL DENR--YIVPGLGDF
Tbrucei_sanger INEYN-VPEENII FLNLISCEPLRRYLTAFFPHVNVVSAALDPGLDSRK--YIVPGLGDF
Crypto_uprt IKNN--VKEKNIICMALLSSKLA IERVFSQFPEIRLVISSIDYNLDENM--QVVPVGNF

18699734_Homo -----
13899253_Homo -----
24649624_Drosophila -----
17538198_Caenorhabditis -----
28828393_Dictyostelium -----
Entamoeba_sangerUK -----
Paramecium1UK -----
15641051_Vibrio -----
26248441_Escherichia -----
16272099_Haemophilus -----
15613838_Bacillus -----
15594361_Borrelia -----
20455345_Chlamydophila -----
20455343_Deinococcus -----
1351377_Mycoplasma -----
20455342_Ureaplasma -----
20455333_Listeria -----
17229842_Nostoc -----
16331050_Synechocystis -----
15790282_Halobacterium -----
6320223_Saccharomyces -----
1730776_Saccharomyces -----
6324339_Saccharomyces RFFDSKY
22026891_Drosophila GDRYFGT
8923487_Homo GDRYFGT
TC909826_Mus GDRYFGT

15232108_Arabidopsis	GDRYFGT
15222778_Arabidopsis	GDRYFGT
TC162174_Glycine	GDRYFGT
18416764_Arabidopsis	GDRYFGT
TC193979_Zea	GDRYFGT
15232164_Arabidopsis	ADRYFGT
Dictyostelium_sanger	SN-----
chlamyUKUPRT	-----
CryptoUKUPRT	ADRYFGT
137110_Saccharomyces	QYKWALT
17539892_Caenorhabditis	GDRYYGT
28923467_Neurospora	--RYFRC
2499951_Toxoplasma	GDRYFGT
21358379_Drosophila	GQKYFGT
21450816_Homo	GQKYFGT
Paramecium4UPRT	GDRYFGT
18312741_Pyrobaculum	GDRAFAT
29348200_Bacteroides	GDLAYGE
17546907_Ralstonia	GDRLFMT
22968125_Rhodospirillum	GDRLFMT
16080742_Bacillus	GDRMFGT
15835447_Chlamydia	GDRLCGT
15803021_Escherichia	GDKIFGT
23475563_Desulfovibrio	GDKIFGT
22970181_Chloroflexus	GDRQFGT
15643484_Thermotoga	GDRLFRT
19551914_Corynebacterium	GDRLYGP
23126288_Nostoc	GDRTFGT
6647900_Nicotiana	GDRSFGT
15607101_Aquifex	GDRLYAV
28917005_Neurospora	GDRYYSM
29841457_Schistosoma	-----
17506069_Caenorhabditis	STSYFGA
18203426_Deinococcus	GDRIYGT
19703818_Fusobacterium	GDRIFGT
22654174_Streptococcus	GDRLFMT
120609_Saccharomyces	GDRYYCV
22654176_Mesorhizobium	GDRMYGT
Leishmania_sanger	GDRYFGT
Tbrucei_sanger	GDRYFGT
Crypto_uprt	GDRYFGI

;

End;

[MacClade 4.03 registered to Jessica C. Kissinger, University of Georgia]

#NEXUS

{TK Alignment - only unambiguous positions are included.}

Begin data;

Dimensions ntax=21 nchar=213;

Format datatype=protein interleave symbols="1234" gap=-;

Matrix

```
Halobacterium_sp      -WVEVITGSMFSGKTEELLRRLRRAEIAQDVAAVTPAVDDRYGEA----
Deinococcus_radiodura -HLEVVGPMFSGKSEELIRRLTRSILIARQRVAVFKPAIDNRYHAS----
Bacillus_subtil       -WLELICGSMFSGKSEELIRRVKRATYAKQEVRFVKPVIDNRYSEA----
Mycoplasma_pneumoniae -WIEVICGPMFSGKTEELLRRIKRWKLAIPVIIIFKPKIDTR-QQH----
HUMAN                 GQIQVILGPMFSGKSTELMRRVRRFQIAQYKCLVIKYAKDTRYSS-----
Vaccinia_virus        GHIQLIIGPMFSGKSTELIRRVRRYQIAQYKCVTIKYSNDNRYGT-----
DICTYOSTELIUM_discoide -GKIQVIFGPMFSGKTEELIRRIKRFNFANKKCLLIKYSKDTRYNDNIDKS
Entamoeba_hist        -----MFSGKTTELIRLIKRFYRSKTTTVVIKYSKDTRYGS---ED
rypanosoma_brucei     GRIELIIGPMFAGKTTMELMRRVQRHKAQRSCYIIKYAGDTRYSEG----
Oryza_sativa          GEIHVIVGPMFAGKTTALLRRVQVEAGTGRNVALIKSDKDNRYGLD----
Thermotoga_maritima   -KLTVITGPMYSGKTTTELLSFVEIYKLGKKKVAVFKPKIDSRYHST----
Escherichia_coli      AQLYFYYSAMNAGKSTALLQSSYNYQERGMRTVVYTAEIDDRFGAG----
Haemophilus_influenzae -AKLYFYYSMTNAGKSTLLQSSYNYRERDMNTLVYTAIDDRFGVG----
xanthomonas          AKLYFYYSAMNAGKTTLLQSAHNYRERGMRTSILTPKLDHRAGSG----
Cparvum              AKLYFYYSAMNAGKSTVLLQSSFNYQERGMKTILFTISKDCRFEEG----
Mesorhizobium_loti   AKLYFNATMNAGKTTMLLQASNYRERGMTTMLFVAGHYRKGDSG----
Agrobacterium_tumefac -AKLYFNATMNAGKSTMLLQASNYRERGMRTLIFTAADFDRAGVG----
Streptococcus_pneumon -AQLYRYGTMNSGKTIIEILKVAYNYEEQGGVIMTSALDTRDVG----
Listeria_monocytogene -AQLFFRYGSMNSGKTIIEILKVAHNYEEQNKTVAIFTSIGIDDRDQVG----
Enterococcus_faecium -AQLFYRYGTMNSGKSIIEILKVAHNYEEQGRVIMTSAIDTRSGFG----
Streptomyces_avermiti -PELVFFSGTMDCGKSTLALQIGHNRSARGLQGVIFTR--DDRAGEG----

Halobacterium_sp      TLGSHAGRSWAATVVEPTAE-GVASIP-----TLLNGEQVVAIDEAN
Deinococcus_radiodura -EVASHAGRTLEAVAVPDAQA-IRAQLSGQGE-LLSAAPEGVDVIGIDEAQ
Bacillus_subtil       -AVVSHNGTSMTSYAISSAAD-IWDHIS-----ESTDVVAVDEVQ
Mycoplasma_pneumoniae -LVKSRNGHSDEAIEINSPLE-IYDYLTK-----DRFDVVAIDEAQ
HUMAN                 SFCTHD--RNTMEA-LPACL-LRDVA-----QEALGVAVIGIDEGQ
Vaccinia_virus        GLWTHD--KNNFEA-LEATK-LCDVL-----ESITDFSVIGIDEGQ
DICTYOSTELIUM_discoide -FLVTHD--KONYQA-FPCSI-LEDVK-----EQAQNYDVIGIDEGQ
Entamoeba_hist        EAISHD--KESWKA-IPMK-LMPVL-----ETALNYEVIGIDEGQ
rypanosoma_brucei     AITSHD--QRALTANVSNSN-LHDVG-----DEWRKYDVIAVDEGQ
Oryza_sativa          SVVTHDGTMKPCWALPELSS-FQDKLGT-----EAYDKVDVIGIDEAQ
Thermotoga_maritima   MIVSHSGNGVEAHVIERPEE-MRKYIE-----EDTRGVFIDEVQ
Escherichia_coli      KVSSRIGLSSPAKLFNQSS-LFDEIRA----EHEQQ--AIHCVLVEDEQ
Haemophilus_influenzae -KVTSRIGISQDAFLRSETN-LFDEINE----HLKKE--KVHCVLVEDEQ
xanthomonas          VVASRIGLRADGQTFDRQTELLQLIERD----IAAHG--PLHCVLVEDEQ
Cparvum              SICSRIGLSEKAHTFTPDLK-LLDIINQ----ENNKE--KIDCVLVEDEQ
Mesorhizobium_loti   LISSRIGLETEAEMFRDGDD-LFARVAE----HHDHT--TVHCVFVDEAQ
Agrobacterium_tumefac -RVASRIGLSEARTFDENTD-LFAEVAA----LHAQA--PVACVFVDEAN
Streptococcus_pneumon -YVSSRIGMKRPALAIETTD-IFGYIRD----LPE---KPYCVLVEDEAQ
Listeria_monocytogene -FISSRIGLKREATPIFSDTN-IFEIVVN----IKP---KPNCVLLDESO
Enterococcus_faecium -KVESRIGVSKPAIAITSEMD-VYGELVE----IIDQYGVDCILVDESO
Streptomyces_avermiti -KLSSRLGLVTEAVEAAPGMD-LHAYLVD----QMSKG-GKVDYLIVDEAQ

Halobacterium_sp      FFP---AELVDVCQELAA-DGRRVVLSGTDQTFRGEFPFVPPQLMAIAEY
Deinococcus_radiodura -FFG---PELVPLVLELAD-AGVRVVLAGLDLDFRAEPPFGSMPELLARAES
Bacillus_subtil       -FFD---QEIVEVLSLAD-KGYRVIAAGLDMDFRGEPPFGVVPNIMAIAS
Mycoplasma_pneumoniae -FFS---SEIVEVVKSLND-LGINVIVSGLDTFRAEPPFGSIPQLLAIADK
HUMAN                 FFP----DIVEFCEAMAN-AGKTIVIVAALDGTFFQRPFGAILNLVPLAES
Vaccinia_virus        FFP----DIVEFCERMAN-EGKIVIVAALDGTFFQRPFNINLNLIPSEM
DICTYOSTELIUM_discoide -FFP----DVVQFSEDLAN-QGKTVIAALDGTFFQRPQSVIDLVSKAEY
Entamoeba_hist        FFP----DLIEFSEACAS-YGRLVIAALDGTFFQRPFGQITDLIPLCES
rypanosoma_brucei     FFP----GVAAFCSKAAD-SGKVIVSALDADYLQEPFEEICLLVSRADS
```

Oryza_sativa FFD---DLHDFCCKAADRDGKIVVAGLDGDYKRNKFGSVLDIIPLADS
 Thermotoga_maritima FFN---PSLFEVVKDLLD-RGIDVFCAGLDLTHKQNPFFETTALLLSLADT
 Escherichia_coli FLT---RQQVYELSEVVDQLDIPVLCYGLRTDFRGELFIGSQYLLAWSKD
 Haemophilus_infulenzae FLS---KQQVYQLSDVVDKLRIPVLCYGLRTDFQAELEFEGSKYLLAWADQ
 xanthomonas FLS---SAQVWQLSEVVDRLRIPVLCYGLRTDFRGELFEGSQYLLAWADE
 Cparvum FLT---KFQVRELICIVVDKLNIPVLCYGLRTDFKGNLFEGSKYLLAWADK
 Mesorhizobium_loti FLE---EEQVWQLARIADRLNIPVMCYGLRTDFQKLFSGSRALLAIADD
 Agrobacterium_tumefac FLS---EHHVWQLANIADRLNIPVMTYGLRTDFQKLFASKELLAIADAE
 Streptococcus_pneumon FLK---RHHVYDLARVDELDIPVMAFGLKNDFRNELFEGSKYLLLLADK
 Listeria_monocytogene FLE---KEHVFQAKIVDELNIPVIAVYGLKNDFRNELFEGSKYLLLYADK
 Enterococcus_faecium FLS---RENIIDLSLIVMNEEVPVICYGLKNDFQNNLFKGSSEALLTFADK
 Streptomyces_avermiti FLA---PEQIDQLARVDDLDLDFVAFGITTDFRTKLFPQSQRLIEELADR

Halobacterium_sp VD--KMRAICMQCGE-PATRNORLIEGEPAHYDDPTVMVGAEEF-----
 Deinococcus_radiodura VD--KLTAICTVCGA-PARTORLIGGQPARFDDPVVLVGAQES-----
 Bacillus_subtil VT--KLQAVCSVCGS-PASRTQRLIDGKPASYDDPVILVGAES-----
 Mycoplasma_pneumoniae IC--KLDAVCNVCGQ-LAQRTRQIVS-----KSNETVLIGDIEA-----
 HUMAN VV--KLTAVCMCEFR-EAAATKRLGT-----EKEVEVIGGADK-----
 Vaccinia_virus VV--KLTAVCMCKCFK-EASF SKRLGE-----ETEIEIIGGNDM-----
 DICTYOSTELIUM_discoide IT--KLTAVCMVCYN-EAAFSKRIVE-----SDDIELIGGIDK-----
 Entamoeba_hist VK--KLSAVCVNCGK-KAAFSRLTSS-----EESIEVIGGVDK-----
 rypanosoma_brucei VV--KLSAVCMECHNRKASFTYRTVK-----SDERKLVGGSDM-----
 Oryza_sativa VT--KLTAARCELGR-RAFFTLRKRTR-----ETKTELIGGADV-----
 Thermotoga_maritima VI--KKKAVCHRCEYNAATLTLKVAG-----GEEIDVGGQEK-----
 Escherichia_coli LV--ELKTICF-CGR-KASMLRLDQAGRPYNEGEQVVIIGGNER-----
 Haemophilus_infulenzae LE--ELKTICY-CGR-KANFVLRRLNDQGEVIKEGAQIQIGGNS-----
 xanthomonas LE--EIKTICH-SGS-KATMTVRVDAQGHAVQDGPQVEIIGGNER-----
 Cparvum LT--EIKTICR-CGK-KATMTIRLNSNGEPVFSGEQILIGDNSI-----
 Mesorhizobium_loti LR--EVRTICR-CGR-KATMVVRLGADGKVARQGEQVAIG-KDV-----
 Agrobacterium_tumefac LR--EIRTICR-CGR-KATMVARFDSEGKVVTEGAQIEVGGNGK-----
 Streptococcus_pneumon ID--EIKTICQYCKK-KATMVLRTQD-GLPVYDGEQIQIGGNET-----
 Listeria_monocytogene LE--EMKTICWFCAK-KATMVLRVDDKGPVYTGEQIMIGGNDH-----
 Enterococcus_faecium IE--EIKTICHFCNK-KAVMNARIVD-GIPVYSGEQIVVGGNED-----
 Streptomyces_avermiti IETLQVEAMCW-CGA-RATHNARTVG-GEMVVEGEQVVVGDVNRPAAEVGG

Halobacterium_sp YEACRNCHVVR
 Deinococcus_radiodura YEALPRAPHRAQG
 Bacillus_subtil YEACRHHHEVPG
 Mycoplasma_pneumoniae YEPRCKLHQPSAG
 HUMAN YHSVCRLCYFKKA
 Vaccinia_virus YQSVCRKCYIDS-
 DICTYOSTELIUM_discoide YISVCRGCYNSDQ
 Entamoeba_hist YCAVCRK-----
 rypanosoma_brucei YMSVCRSCYETKR
 Oryza_sativa YMPVCRQHYLDGQ
 Thermotoga_maritima YIAVCRDCYNTLK
 Escherichia_coli YVSVCRKHYKEAL
 Haemophilus_infulenzae YLSVCRLHYKEKC
 xanthomonas YVSVSRAEFKKIM
 Cparvum YTSVCRKHHIINC
 Mesorhizobium_loti YVSLCRRHWEEEM
 Agrobacterium_tumefac YVSFCRRHWVETF
 Streptococcus_pneumon YISVCRKHYFAPM
 Listeria_monocytogene YYPVCRKCHANPP
 Enterococcus_faecium YLSVCRNHYYNED
 Streptomyces_avermiti YEVLCRRHHRRM

;
 End;

Thymidine kinase (TK) accession numbers are: *Agrobacterium tumefaciens*, NP_355239; *Bacillus subtilis*, NP_391587; *Cryptosporidium parvum*, AY466379; *Deinococcus radiodurans*, NP_295707; *Dictyostelium discoideum*, AAB03673; *Entamoeba histolytica*, AY466383; *Enterococcus faecium*, ZP_00035626; *Escherichia coli*, NP_309767; *Haemophilus influenzae*, P44309; *Halobacterium sp.*, NP_280328; *Homo sapiens*, P04183; *Listeria monocytogenes*, NP_466067; *Mesorhizobium loti*, NP_104016; *Mycoplasma pneumoniae*, NP_109732; *Oryza sativa*, AAO17013; *Streptococcus pneumoniae*, NP_345493; *Streptomyces avermitilis*, BAC70125; *Thermotoga maritime*, NP_228211; *Trypanosoma brucei*, AAM95606; *Vaccinia virus*, NP_063745; *Xanthomonas campestris*, AAM43364.

GenBank accession numbers for additional *Cryptosporidium parvum* genes identified in this study are: adenosine kinase, AY466390; adenosine monophosphate dehydrogenase, AY466391; guanosine monophosphate synthase, AY466392; adenosine transporter, AY466393; uracil phosphoribosyltransferase, AY466394.