

Supplementary material: Amino acid sequences of algal proteins belonging to the SUF machinery or Fe-S proteins discussed in this study.

Most sequences presented in Tables 2 and 4 have been retrieved by blastp search on the jgi phycocosm portal or on NCBI GenBank (mostly plastid genome-encoded proteins). When sequences were missing only in one or two organisms but present elsewhere, we have performed a tblastn analysis against the genomic sequences to assess whether the absence was not due to a wrong annotation or to a genome assembly problem. If there was no significant hit exhibiting the typical motifs representative of the protein families, the sequence was considered as being absent. Only *Chlamydomonas reinhardtii* sequences have been manually curated and reannotated parts appear in red characters. Additional observations appear in red in the define.

NFS2/NIFS2 orthologs

>Cre07.g322000

MAISCLLRHRDDAASARSLHAAANESTSHESSQAVPQVPTVGLGQALPHLTRTGCIYLDYNATTPIFPEVANEMAPFVFEHFG
NPSSGHVYGRACKSALDTARQVRVAALIAAASPSEVHFTSCGTESDNWAVYGAVMAARAAAAGAAGAAGYVPHVVASAVEHPAV
LAHLTHLQEQGLLSYTLVPVDGEGLVSPADVAAAVTPATCLVTVMHSNNEVGAVQPIAHIAAAARAHAARLTAGGGGATSPSP
SLSPSSSTGGSSSGTSGSSWRPRLLVHSDAAQSTGKVDLDVGALDLDLMTLVGHKFGAPKGVAAALYVRHGTALPNFYGGGQE
GGRRAGTENVMQVVLGAAAAITVRERRQLQTHMGDMAQRLLAAVRAQVAPAEQDKLRLNGPADPTRRLPNTLSLSIRGLNSG
AALAALSGRLAASAGAACHSSGAASVSAVLRAMKVPTDYATGTLRLSTGRHTTTAEVDEAAGLIVAEARRQGVLAH*

>jgi|Cyapar1|22355|g21891.t1

MEKPIYLDNFNGSTPLSREVVDAMKSVLDSGLGTGNPSSSHFYGAESKKAIEAARKQVAAALHCAEDELIFTSGGTESNNYAIK
GYFAANKGRGNHIISSSTVEHPAVTQVCDFLAETAGARVSYCPVDGTGRDLAALAEALITPATILISVMHANNEVGTLQPIREI
VEIAKRGGRNIAVHTDASQSVGKVPVRVDELGVDMTLVAGHKLCAPKGI GAVYTGRGTAVRQMHGGDHELKRRAGTENVLYA
VALGTAAAAAARDLEKNMAHCRRTDRLLARFKEAGLDIRVNGHPEHCLPNTLSVGFGRIEAPTLLSELAASVSASAGAACHS
DHKMSGVLKAMKVPVAVYGMGTVRFTSTGATTTEEEVDRAAALVVAVRRLTPGEEAGAGDGE GEGGPEGLPALADVRMTRYTK
GMGCACKLKPQLLEGVLRGLPRATDPRVLVSTETSDDAAVYKISDELAHVHTLDFFTPVCDDPYEF GAVAAANAISDVYAMGA
TPATALSIVAFPSLRLPITALQRILQGAADKCAEAGVSI VGGHSIDDEPKFGLSVTYGHPERILRNS SARAGDVVVLTKPI
GTGVLTTATKRGVASAEAAAACAMA VGPVHAATDVTGFGLAGRRARLGG RALLPGAFAHAAAGIVPGGT VSNMEYAGRWRV
WGDGVGETTRALLCDAQTS GGLLLAESAEALLAALRERKTLAAA AVARLTAPGPGIITVR*

>jgi|ChlNC64A_1|144931|IGS.gm_8_00259

MPRLGEPLPHIQASGCVYLDYNATTPIFPEAADEMRFPLTAFGNPSSSAHAFGRPCKAAVDAAARVAAMVGADPDEIFFTSCG
SESDNWAIWGAVMEARRQQRAAPDTS GAAAF LPHVVSRIEHPAVIECLAMLA AQGLLEYTLVPVSC EGLVRVADVEAAMTPH
TCLLTFMHSNNEVGSIQPVAQLAALARRAGALMHCDAAQSLGKVDVDVQQLGVDMTLTVVAHKFGGPKGVAAALYVRRGVRLERL
LCGGGQEGGRRAGTENVVLVAGLGKAAELVSREL PATAAHMAAMRDSLQQQLLAGLP PSTALIHGPADDALRLPNTLSIGIRG
IAAARLLAELSEQLAASAGAACHSGGGHGVSAVLQAMAVPTEHAVGTLRLSTGRHTTQQD VDRAAALILDYVQRHGSS*

>jgi|Chlpril1|4911|rna-gnl|IITBIO|A3770_07p49110_mRNA_A3770_07p49110

MAKEVASSRRCIYLDNNGTSP IYREVAEEMQPF LDFHFGNPSSSHAYGRVPKQALDLARERVANLLGAKASQVYFTGSGTESN
NWAIRGALEYGSQKRPKRSDFLPHVVT SNIEHPAVAEALAKYKREGLCDYTEVAVDTKSGVLSVEAVEAAFR TGETVLC TVMH
SNNEIGSLQPIAEIAKIAHAHAGALMHTDGAQSVGKVDVDVEALGVDLCTVVGHKFGASKGVAAALYCSLDSREFPSLLHGGGQE
NGMRPGTENVLLISIGIKASEIALLEGKEIRRNMLKTRNLLRGR LVEALGSHGIDYRINGPAEEANSLPNTLSISIPAVQAQE
ILADLSESVAASAGAACHSGGFHVSGLVKAIGLDDAHARGTLRLSTGRLTTEEEIERASDLIVEAVLRATSSSG*

>jgi|MicpuC3v2|5257|wlab.205949.1

MPCAIVAAPAPSPPLASSVARRRRRRRGPPSHVASAKPSSSSSSSSSTDLAAASASVSPAAPPKPPALLSAPSTPGLVAGLV
FGGLP LLAALRRKEESAEEADADRSSRRRPASPLPAATTASAMTTTTTPPSTPLPKGEGAIGSKLP GRAHETCAYLDYNA
TTP IYPEVAAAMEPFLWEHFGNPSSGHAFAPCRDAIATARNRVATVLGCAPDEIVFTSCGSESDNHAIASAVEHARRSAGGR
GGRVPHVVTTAVEHPAILEYLVSEEAKGTLAYDAVRVNAEGIVDPGDVAAAVREETCLVTVMHANNEVGAIQPVAAACAAAAA
KNPRVLVHDAAQSLGKIPVKVDLLGVDMCTIVGHKIGAPKGVAAALYVKKTAPFSKLFHGGGQESGRRAGTENVMHVVLGAA
CALVTKEEDALPSYMATLRDAMQRLVAELGGDDGAGVRVNGPANDADRLPNTLSVGVKGVSAVLLQTLSSSVAASAGAACH
TGAAAASISSVLRAMEIPEAYAVGTLRLSVGRHTTTS DVEVGVKRIADAARAQISEIDAMPEGERPAWCVRNVFK*

>jgi|Chlat1|251|Chrsp1S03050

MGGVGGSEVQEVQSDGEAKPIYLDYNATTPIDKEAADAMLPFLYSEFGNPSSSHAYGQRAKAVVESARAELAALIGCHPSEIY
FTSGGTESNNWAILGTALHAREARGGVQGVHIVTSAIEHPAVLEPCRYLERHHGFDMTVVGVDSQGLVDVDVAVREAVRDDTM
LITVMHANNEVGSIQPLAEVAAVAVWERGIPLHTDASQSVGKVPVDVGT LGVDMTLTIAGHKFYGPKGVGALYIKSSHRLCKFMH
GAGHERGM RAGTENITLIAGLGAAARIARRGLQQNMEHVMRMREALVRELT PPLDGVRYLINGPQDALKRLPNTLSISIEGVS
ASDVLHDIETDVAASAGSACHAGTSTVSSVLKAMNVPVEFALGTLRLSTGKPTTQAEVTEAARI IAQSVRKYTTSSKQQSTSN
*

>jgi|Mesovir1|3004|Mesvi192S03833
MSACYANSNLASVAKIPASFPLLGRGNYSGKNTRKRVRVYAKAAMQTSQGPANMPHTVPSDRGAIYLDTAATTPPIYEEVAQAM
LPYLLKHFGNPGSRHAFSSPTRSAVLHSREQVAVLLNVDDKDDVFFTSCTESNNWAI FGTVARAERFLASHDSPAADTAAQGT
PRLPHVISSVIEHPAILKCLQYLNACGRIQYDLVPVDAQGLVDPAAVAAAVTPTTTTLVTIMHANNETGAIQPIAEIVKHVRER
ASSVLLPGLQLLIHTDAAQSVSKVDVRPQDWGVDLCTLVGHKMGAPKGIAALYIRDGAQIDNFLHGGGQEGAGHRGGTENVLLI
VGMGEAAAIAVKEAGVTPKNMRKLKGLFLLSLLREGLGAENVRVNGPPVDSMPALPNVLSVGIRGIEASAFDLTDLANDVAASAG
AACHAGEASISSVLVAMNVPQEFAGKTLRFSFGRLTTEADVMTAAELVIAARK*

>jgi|Klenit1|2855|rna-KFL_000530350
MDASTGSSSLEHPQRHPLGWEKLEGAPIYLDYNATTPISPEVSAAML PFLQGFGWGNPSSSTHVFGGEQAKRAVEYARRQVADLI
GCAEDEVFFTSGGTESNNWAIKGAVEAALAHLPGRRRSGPAFRLVTSAVEHPAVTEVMEHLRDSRGFDLRVAPVDSQGLVDAD
AVIAAAFGGAESGDLGRLESVTSADDSFQHSEASKVTNGIKEEGQAPESVEVAKAIMTDRSAVEETSTTGVETEGAEKAGPVK
VLNGVFDADKVDGKVDGVGQNGQASQTLIVSIMHSNNEVGAIQPIAEIGRALRKAGVVFHTDASQSIGKVPIKWDELEVLLT
IAGHKLYAPKGVGALIVRRGTRLAKLMHGAGHERGFRAGTENIMLIAGLGKACEVSRDRMARNVAHVTDVRDRLLSRLQQNLA
SGFQGSKRLALKVNGPIDAAKRLPSTLNISMGSVDGPRLLADLAPLVACSAGSACHSGGHGSPVLKAMGVPEFAACTLRLSV
GRETEEEVDAAASCISEAIKKQHEFELPTIAVPDSFVSLFRQLSYSMSGSPSPKKG*

>jgi|Chabra1|325782|rna-CBR_g8345
MAFLFLALLCCLFTPSDSEVAEAILPFLYEEFGNPPSSSHFYGKKAKLAVENARREVALLIGCKADEVVFTSGGTESNNWAI RC
GAQANSRMKGKHLVTSSIEHPAVSEVINHL CQNEGFCKTVVPVDLEGRVDAEAVVKVATADLETSLVSIMHANNEVGS LQPIA
EIAASLRGKGLVHTDASQSVGKIPINWRTLGVDDLTIAGHKLYGPKGVGALIIRRGVILPKFMLGAGHESGRAGTENVALI
VGLGKACEVARKGLDRNMRRSRAMRDLLEKLEDALPGLRCRGDDIELSRIDFRVNGPADPEKQLPNTLSVSFKGVGDGVKLMA
AISP FVACSSGSACHAGHVA VSPVLAAMGVPAEYAQCTLRLSMGRETRTRTDVDTAAEVIMSHIKSLFP SGVTCT*

>jgi|Guith1|109959|au.42_g11061
MVYLDYNATTPLDKVPQAMKPYIEGRFGNPPSSSHWYGVKEKEAVAKAREEVADLIGCSSEEVFTSGGTESINWALKGTAER
LRKTKRHIVTSAAEHVAVLECKYLEEVHGM EITHVGVNEYGEVDPEQVRAAIREDTALVTIMLANNETGTINDIRSI AQIAR
SKGVLVHTDASQAIGKIPVDVNDLGVDFLT VAGEHAKHV VANTLTDSGTWPGHKLYAPAGVGALYIRHGTELPNFMHGAGHER
GRRAGTENTMHLVALGAACKAIPGIELNGHPERRLPNTLNVLNLPFCVSRSHGDLQIIAQ TADRLAFSAGSACHAGQTGHVSH
VLKAMNISDKAFRCCHLRIRNH*

>jgi|Guith1|139922|fgenes2_pg.41_#_133
MFGKGDPYLDYGGTSPVDPSVFD SMKEFFVSNWGNPSSSHVYGQLTTGRGLKEVSKKREKTWLI ALDALLMKWGTE SNNWVL
KDSPDMNQGA VSLARKRLQDHFTSNMGG SALAWSSMYVDQLRPHIVTTAIEHPAILEPLKWLETEQQCDVTIIQPGKDGSKDY
DIAGTSPREREPGVVNQDVVDALRTQTILV SIMHANNEVGAVQPIREISDVNDLGVDDL SMAGHKLYAPKGIGALYVRGGAK
GLDGDGDFRLQRI MHGGGQELGLRAGTENVG LAGVLMRQASLIRSLHVQLQAKFIGQIHLNGPELPASPINGAWPRLPNTLSI
GFRGLKAQEILGKVQDVVSASAGAACHSSNIPKISSVLAAMQVPEDAALGTIRFTVGRYTTED E INRAEAVADAANELFKQS
V*

>jgi|Emihul|424605|estExtDG_fgenes2_newKs_pm.C_3090007
MSVREQKLATTPVWPEVAAAMTPYLTHHWGNPSSSGHAFGRPCGA AIVQARA AVALIGAE PSEICFTGCGSES DNFAIVGALE
MEEARRRGAGGAALPRPHVVASNIEHPAVELCIEALCAAGRLDATYVPVDEEGLVSAEAVAAATTEQTFLVSV MHSNNEVGSV
QPVRQIADAVRALRPSVLVHTDAAQSIGKVDVDTALRGVDDLTLVGHKFGAPKGIAALYIRSGALPRMIHGGGQESGRAGT
ECTPMVVALGTAAAIARA EAAPLREHMRRTD ALESALVERLPAGSTRVNGPADASRRLPNTLSIGLRDVRAAE LLETLSERL
AASAGAACHANAASVSAVLRALRVPPEFAVGT LRLSTGRHTTLEEVALAADLIAEEARRQWGGDGAAGKVSDSGPS*

>jgi|Bigna1|92143|estExt_fgenes1_pm.C_40030
MGATCKPAKGPSVTYLDYNATTPIDPRVQAVMQPLLSSGWGNPSSSHVYGKVANAALKAARQKVAAAVGCADSNEIVFTSGGT
ESINHAIRGA AFASKRIRYGGNHIISSIEHVAVSATCRYLQDHGFKVTTL PVDGSGRV SPEALKKAITRKTVVVSIMMANN
EVGTIQPIAELVKVAKAASKKIIIFHTDASQAVGKIDVNVQELGVDDLTIAGHKLYAPKGIGALYIRNELPVTLEPLIHGAGHE
MGRRAGTENILLAAGLGEACQIAKQDLEERMHRMKLCRDHLQRKIMDEFKGALVRINGDEKARLPNTLSISFHGLASHEILHD
IQDLVACSAGSACHSNEVRSISIRISHVLAAMKVPLDFAKGT LRLSTGINTTIEADKAAKFVCETNIPLLV*

>jgi|Phatr2|2966|gw1.19.75.1
IYLDYNGTTPVYPQILQAMPLPYLTTFGNPSSGHLFHRQPRQAVDRARKQLLTLLGQPEADLSSIWFTSCGTESDNLAIQIAL
QSSSKLISTRFGAHTLPHIVSCNVEHPAISGYLDALVEEKVCEVTYVPVQSDGMVSADAMIAAIQPQTILMTLMLANNESGAL
QPVQKVAQHCSKANILFHTDAAQAVGKVSVDLDDLGHDPDMISIVGHKMGAPKGIACLYVRPGCCEQHGRALHNRGILLIGGGQ
EHGRRGGTENVPIYVGFGEAASLATKDWKSNSIKMEGLRTRLLFNLENWLKGDMVRTNGPSNPAHRLPNTLSVGLKGIHSGAL
LAAIGDQVAASAGATCHSASGVSSVLKAMGVPETFARGTLRLSLGTHTEQEVD

>jgi|Ectsil1|32276|rna4872
MHLLSMFSVAAVAAGMCVGSVTAEVTAAGIETPAT TGNLMRFS DAGFWYIVAMTFFVLVIVDATNKKFS PRACLNTDGMSSK
TGIAVALPYRGTRRCTYLDYNATSPIFPEVSREMFPFLGECFGNPSSGHALGRACKKAVVLARARVASLLGCSPEEVVFVASG
SEADNHAILASLALSSKAATAPPSAASSSLPHVVSIAIEHPAVTKCLNLHLVAEGKVEVTFVGVDEEGRVSVEGVVGA FRPETA
LVTIMHSNNEVGSLQPI TEIAEACRQRGIMCHTDAAQSIGKVPVKVADLGVDM LTLVGHKFGAPKGVAALFVRRLGALPPMIH
GGGQEGARRAGTENVLLISGMGKAAEIVDVELERLSKHLAAMRERLELLMAADLGDGVKVHGPSPSLRPLPNTLSIGLPGVE
ARVLLERVGA*

>jgi|Ochro1393_1_4|849870|fgenes1_kg.4_#_1171_#_TRINITY_DN23625_c1_g3_i7
MLKRNRIITKKPAAGKSQDSEKIIIPKNLSEQTDPSNVRTSYITGNSSGKQHEHGKENTSQVGIPHPYTKRQEYSGCVYLDY
NATTPIFPEVTSAMLPFLSDFFGNPSSSHVFSKPCRAALDQARINVGRVLVNAANSAKEIYFTSCGTESDNRAIDIAIHHRQH
KKKLLAVRGVASANIVSVPHIITCSIEHPAVLVVYRSLEAQGEIRLSIVPVNTEGIVDIAMLKRELSPTALVTIMHSNNEVG
SIQPIREIANCVREFNQRLSHGEACVLMHSDAAQSLGKVSVDVQGLGVDLLTIVGHKFGAPKGVAALYIRSNVQTTTMLVGGG
QERGVRRGTENVGLVCGLGEASRLAYTEASELLHFLVLKLQLVTSIVEGFSAQQVAECIRFNGPQRSCDISSELSSDLKLLRV
MLKSAPSMNIEQIANQNNGNNSRTAAAASTSSGSSRAVGAIIDLLEQLPNTVSVSFKNIHSHLLIHHLSSRVACSAGSACHA
VTPATAVDNTRSDTDKSGTNSVSGANNSSSTNSSNSGAQTQSSAKAAGMNSRGKVSVDVLLAMKVPLDFAQGTLLRSLFGRHT
TATDIEAAATHIIAAVKSQWATNGYSG*

>jgi|Auran1|11983
VYLDYAATCPIIPEVADAMVPFLYDHWGNPSSGHAYGAPCRSAVADARGSIARLLGCASEEVAFCSCGSEADNMAILASLGK
RRHVVS SAVEHPAILACVDALEREGRCDVSRVGVDPAAVADAVKPGETALVSVMLANNEVGSMVDVAAISAAVKARD
AGILVHTDAAQAVGKIDVNAKRLGVDYLTIVGHKFGAPKGVAALYHRRGAPLRSMFLGGGQEGGLRAGTEAVPSVVALGAAAA
IWTREGAAIAAHSAMRDLRRRLAALGPERCAVNGPLGAGGEALPNVLSFAVRGARAARVLGDVGDVRAASASAACHSGSD
AVSAVLLAVGVETELALGTLRLSVGRHTTAREVDAAADVL

SUFS1 orthologs

>jgi|Galsul1|1124|XM_005707326.1
MTHCAMNRCGGWKTSCSFVTQCGFKEENVFGKRWCSFSPSVKRLRSLRPQRLLTCAILLSSPYSDTLNNLQCPVKREDFPILQ
QTIENTKLLVYLDAAATSQKPLVLEALQKYLLHNNANVHRGAHTLSSRATEAYERARQILQHFIGAESREQIIFTRNATEAIN
LVAYAWGLHNLKPGDELILAIISNHHSNIVPWQLIASRTGAILKYVVLNDNQEQDSLESFKMLLESKGKRLVACSHVSNVLGFIN
DIKPIITRWAHSAGALVLVDACQSVPHMPVNVVDLDCDFLVASGHKMCPTGTIGFLYGKESLLENMSPFLGGGEMIADVYEDYS
TFAVLPHKLEAGTPAIAEAVGLGAAVEYLSIGMDVIHSYERCLSEYLYRHLIDIPNVTIYGPKDVKRRVALCSFNIQGIHAS
DLASILDLLEGVAIRSGHHCAQLLHRSLGISGSARASLYMYNTEKEIDSFIGSIYRSIDMLKR*

>jgi|Caulen1|89248|g4590.t1, **bad annotation**
MVQSPSISGSSSLADTVRPDPFILTQEVNGKPLIYFDNAATSQKPSVIEAMDRIYITITNSNVHRGVHALSAEATSLFENARSK
VANLVRADKRSEIVFTRNATEAINLVAYGWAMSNLKMGEILISVAEHHSNMVPWQIVAKKTGAQLKFISLTPETEELNMEHL
RELISDRTKLISTFHASNVLGAISPVKEISRIAKKYGSKLLLDCCQSVPNMSVDVGEGLGADWIVASGHKMCPTGCGFLWGKY
EVLEDMEPFMGGGEMIEDVFLDHTTYADPPARFEAGTPAIAEAIGLGAACDYLTSSIGMENIRSYYYYIGTYLYNKIVGLNAVK
VYGPPPSQGRGPLCAFHLDGVHPTDLAMVLDQYGDANINKGV*

>jgi|Cyapar1|25252|g24670.t1
MDRQAFALSAPARGSFSPASCPTGLASQNTCDRAPRRAHRSVLVAGASLNGASSSEASRRARFVLPSEFVGRSPASRSAGSDQ
RGVSACPPPAPPPTCAAVETRSLGDEVVRKDFPILDQEQINGQPLVYLDNAATSQKPNVLDVLRYYNEDNSNVHRGVHTLSAR
ATDAYEGARRIIANFINAEDPACVIYTRNATEAINLVAYSWGTLTKPGDEIVLSVMEHHSNLIWPQIVAQRTGAVIRYGLA
EGGVLDVESVEGLINERTKLVS LAHVS NVLGC FNPVKRIGAAARAVGARMLVDACQSVHPVPDVQDLACDWLVFSSHKMCGP
TGIGALYGRREVLEAMPFFMGGGEMIQDVSLEHFTCNTLPHKFEAGTPAIAEAIGLGAACEYMSGIGMGRITHEYEMELTRYLW
KRMGELPFEVELYGPHPDAAGPEGRAALCAFNVKGVSNDLCTILDQDGVAIRAGHHCTQPLHKELGVSHSARASLHFYNTRE
VDDRFIKALADAVDFFTSV*

>jgi|ChlNC64A_1|34562|estExt_Genewise1Plus.C_50023
MPAPQVQDAPAAAAAADLGTATRADFPILHQSVNERPLMYLDNAATSQKPRQVLQTMDEYYGEGGYNSNVHRGVHALSARAT
AAYEAAREKIAGFINASSPQEIYVTRNATEAINLVANTWGAAQLREGDEVVLSVAEHHSNIVPWQLLAQRRGLVLKFAELTGS
EEVDLEKLAALITPRTRLVSLVHVS NMLGCVLP TQ RVAEMA HVGAKLLLDCCQSVPNMPVDVQTLGADWIVASSHKMCGPTG
IGFLWARYSLLEQMPFWMGGGEMIQEVRLEGSTYAEPPSRFEAGTPAIAEAIGLGAACDYLSGLGMERVAHAHERELGAHLYWQ
LRSIDRVRIYGPSPEARLGRAALATFNVEGIHPTDISTILDSTGVAVRSGHLCTQPVHRHLGISSSVRASPHYIYNKAEVDAF
VDALKDAIQFFT*

>jgi|Chlpril|3067|rna-gnl|IITBIO|A3770_04p30670_mRNA_A3770_04p30670
MTTTTTTTMVQCSRSRWGCPRGGFRGASSVAARSAAQGTSPRLTAEDVAGIRREFPILEREVHGDKRLVYLDNAATSQKPRRV
LGAMEDYYADYNSNVHRGVHALSAKATERYEEARDKVARFVGAESSREIVFTRNASEAINLVAYTWGEENLGEGDEVVLTVAE
HHSNIVPWQILSRKKFTLKYVRLDEASQTLDEGLGEAIGPRTKLTVGHVSNVLGCVPVGDIVALARSGGHPGCKVLLDA
CQSVPHMPVNFSELQVDFAVCSGHKMCAPTGTIGFLWGKMEVLEAMPPFMMGGGEMIQDVFEHSHTFARPPGKFEAGTPAICEAI
GLGAACDYLSDIGMHRLEEFELMGTYMYEELEKNIEGIEIYGPSPKQAPRAALAAFNVEGLHATDISMILDQYGIIVRSRGGH
CTQPLHHHLGINASARASLYMYNTEKEIDQFVAKLKDITDFFGQAS*

>jgi|Pico_ML_1|55742|g1387.t1
MRATIPTIATPASRRRNARHAKTSAQATGVVGN SQDLAHRVRADFPILERQVYESFPLVYLDAAATSQKPRCVIEAMSDYYMQ
YNSNVHRGVHALSSQATEEYETARSKVADFVNAETDREIVFTRNASEAVNLVAYSWGDLNLKEGDQVVSVMEHHSNLPWQI
ICKRTGAKLCHVPMTEDEELDNLRLIEITKSGQTKLIALAHVSNLTGCITPVEEVVQIAKNCGAKTATKDDRIVQEDLITGRP
DNNVTEHIVEKLGAQLHLRENHPLSILRSAIEEWFAQEHGDKFRCFNDLPPLVTTKANFDDVLPADHVSRSMDNTYYVDANT
VLRCHTSAHQAEMLRQGERKFLVTGDVYRRD TVDATHYPVFHQMEGVRI FEPEEWERAGTDGTKLAEAE LKKTLEGLARHLFG
DVEMRWIDEYFPFTDPSFELEIFFNNEWMEVLGCGVMEQAILLEEAGLGNKKAWAFGLGLERLAMVLFEPDIRLFWSTDERFL

SQFKSGDMSAKFKPYSKFPPCFKDVSEFWIPETFTENNLCCEVTRGIAGDLVEEVKLIDEFHNPKKGKTSHCYRIAYRSMDRSLT
DEEINQLQENVRSSLSVEQLGVELR*

>jgi|MicpuC3v2|917|wlab.201976.1

MATMATLTLAASASAPRASASAASRDGAPPPRGASPASHGSRARVAAASPPNPPRLPGTARTRASVSTAAASTSVRSTLADDEL
RGDFPILHQTLPLDSGKPLVYLDASAASSQKPNVVIDAHAAYYREINANVHRGVHYLSGKATDAYELARVKVADFIGAETDREIV
FTRNASEAINLVAYTWGVANLKPGEV I I SELEHHSNIVPWQLVCAQTGATLRHVGLAPDNGIDMDALEASAREGKTKLIAT
AHVSNVLGSEVDVPRVLKLAKEVGLLDACQSVPHMPIDVKTGLVDWIVASGHKMCAPTGIGFLWGRADVLETMPWMMGGG
EMIQDVFMDRSTYAPPPARFEAGTPAIGEAIALGAACDYLSAIGMDRVHEHEVDVGTRLYERLAAIDGTVYGPPTPAKGRASL
AAFNVEGLHANDVCTLLDASGVATRSQHCAQPLHAALNVPASARASLYLYNTREEVDARTLEETIAFFREINAGM*

>jgi|Chlat1|1843|Chrsp14S02228

MVAVAVAGSSGSSAAVAGFRVGNRVAAGVVDGEGVQVARASLPGRSGRRRTSSRGVAVRLPAAQRGVSAAASVSASQVAEKQ
SQASTASPEVSGKLSPEALRQDFPILHQENDFGVPLVYLDASAATSQKPKAVLDALTAYYEGFNSNVHRGVHTLSAKATTAYEE
ARVKIAKFAVNAASDREIVFTRNASEAINLVAYSWSGLKNLKPGEI I I LSVMEHHSNIVPWQILSQQTGAVVKFVGLTPEETLNM
EEFRSLSDRTKLVSLSVWVSNTLGCNFPVEEVVQLAHAKGARVLVDACQTVPHMPVDVQAIDCDWLVASGHKMCAPTGIGFLY
GKLDVLRGMPFFMGGGEMIADVFLDHSTYAEPPSRFEAGTPAIGEAIALGAACDYLSLGMMDRVHALEELGLYLYEQLSQVP
KVRIFGPKPAADGTGRAALTAFAFNVGLHPTDISTFLDLSFIAIRSGHCTQPLHRYLNISSSARASLYFYNTKEEIDTFVREL
RNTINMLTGAGDDGFQQA*

>jgi|Mesovir1|5648|Mesvi433S09449

MAAYYEGMNSNVHRGVHYLGLTATEAHNLARTKVAKFVNAARDEEIVFTRNASEAINLVAYSWSGMNNSLSPGDEI I ISTVMEHHS
NLVPWQLVAQKTGAKIVHVGLTKDQTFDMDHFRSLSDRTKLVAIGHVSNALGCTNPVREVCQLAHARGAKVLIDACQSVPHM
RVDVQSIGCDWLVASGHKMCAPTGIGFLYGYDVLKAMPYMGGEMIVDVFLDHSTYADPPMRFEAGTPAIAEAIALGAADV
YLTGIGMDRVHDYEVELSTYLYHEMEKVPGITLYGRPEIGRAALCSFNVEGLHGVDVATLLDNKGIAIRSGHCTQPLHRLH
GIAGSARASLYIYNTKEEVDVFVAALTDVVKFFRELGM*

>jgi|Klenit1|6882|rna-KFL_001880270

MPPACKRAAVRAIAAPERAAATVSAPAKSLGDETRPDFAILHQEVEGNPLVYLDNAATSQKPRQVLQALENYYKGYNSNVHRG
THKLSAMATDEYEKARIKVAKFVNSPTDRQI I FTRGATEAINLVAYSWSGLNNLKGDEIVLSQMEHHSNIVPWQLVAERTGAL
IRFVGLTEHQDLMEQFRNALSNRTKLVAVHVHSNTLGCNPAEEI VRLSHSVGAKVLLDACQSVPHMPVDVQALDADWLVAS
GHKMCAPTGIGFLYAKWDLLLSMPPWQGGGEMIRDVYFDHSTYQEPSPRFEAGTPAIGEAIGLGAACDYLSVGMERIHQYEV
EVGRYLYEKLSDIPEIRIYGPPLSLADRASLCAFTIEGVHATDLASHLNDFRGVAVRSQHCTQPLHRLVGNATARASLYFY
NRKEEIDVLVDAIREGIDFFTAPLEPPPGF*

>jgi|Mesen1|2876|ME001742S01985

MLTIKNLHAGIEGKKILKGINLEIKPGEVHAIMGPNNGSGKSTLASVLAGRKEYEVSTGSVQFLGKDLLDSPEERAGEGLFLG
FQYVEIPGLSTTNFIKTALNEVRKYRGQEPLDAVAFKLKMEKMAIMNVQDQSLLSRSLNEGFSGGEKKRNEIFQMAMLPQKL
AILEDTSGLDIDAIRIVANGVNQLRNKDNAVLAIHYQRLLDYIIPDVIHVLYNGRIVKSGPKELALELEERGYDFIKTNGY
SDIVSAHFGHSSGYIKDGIHALNTAFTEGGVFIHLRKTKEKQPVYIYHLTTAEHEPGFTQPRSLVYLGEQAEMQLVETSMCL
GQQDSLNSQVIEIVLEEAALVLEYKIQHDHDNASQVNTTHIRQIGKYSHTVTLSLNGNMVRNNLVNVEASHSEAHLYGLYF
PKGKSHIDNHTIVDNRVPHCESNELYKGILDEQGTGVFNGKIFVRQPAQKTNAFQSNKNVLLSDTAINTKPLEIFADDVVC
SHGCTVGRSLSEGLFLYLSRGIPEKTAQSMLLHGFAEDVLEKIKLEPIRRYADALIRIRYCMITTATIPVFDVYALRKQFPVL
EREVKGKPLIYLDNAATSQKPAVVIDALVQYYTGYNANIHRGIHTLAEETAFAFEATRETAKHFIHAAGREEI I FTRGTTEGI
NLVAYTWGRANVGKDEI I I SAMEHHSNIVPWQILCEEKGAQLKVI PVDDAGELIMEAYEKLLSDKTRLVAI THVSNALGTVN
PVKQMIRSAHKAGALVLVDGAQSTVHLDIDVQDMDCDFVCSAHKLYGPTGVGVLYGKKHVLEAMPVFQGGGEMIKEVTFEKT
TYNELPYKYEAGTPNIADTIVFKTALDFIEQTGKEQIRQHEHELLQYATAALETVPGLKI I GRAKEKTSLSVSFVIDKVHPQDI
GILLDNKGIAVRTGHCAQPLMQRYGIPGTIRASFAAYNTKEEIDELLGLLEKSIKLLV*

>jgi|Mesen1|4618|ME000236S03869

ARATDEFERAREKVARFVNARSSREIVFTRNASEAINLVAYSWGAAANLQKGDEIVVTVAEHHSNIVPWQLLCQRTGAVLRVV
ELTPGEQRVDVEQLRQLVGARTLVAIHHVSNTLGAESPVAATAAARAHAAGVLLDACQSVPHMPVDVQRLDCDFLVASGHK
MCGPTGIGFLYKSEVLQAMPPFMGGGEMISDVFMDSHTYAEPPSRFEAGTPPIAEAI GLGAADVLTIDGMDRIHAYEMELG
RYLHEQLQAVPGVTTYGPPPSDNGEGRSALCAFNVEGVHATDLSTFLDQQHGAIRSGHCTQPLHRYLGTNASARASLYFYNT
TKDEIDTFIAGLKDITDFFTSFSDA*

>jgi|Chabra1|349042|rna-CBR_g45931

MASTATAPSLAGASMLPSPSSPRLASGKVVAAPPITMQWVQRTAQPRAMPDRSTNSFSPCRFAAARSCASSSSSPSSSSSLSS
LSRSASCRSHSGEVVHRVILVPAAAAKERKSAPRRRKEARADPLPSFDLTVTALTAGRCRRRYRSDAVGKERASNRCYPPTGC
MAPALRAAEDMTAHGRGRRRRWCRPSCRSNNVEGTAGTAKGERRDRVGAKGSRFRLPTAAQRVTKGPGFAQGASIPDDVGQDN
AAAAAAAAYCYSKTPYSGFGGVAPPRSRRKVKVINTLIRCAAAISSVDDASSSSPSSPSSSAAAAAAAVSVGTTTSLGEITVK
DFPILSQEVNRRIRIVYLDNAATSQKPRQVIDAVQRYRYQDYNNSNVHRGIHTLSVKATEAYEDARGKIAKFINAERDSEI I YTRN
ATEAINLVANTWGVANLTPGDEVVLSVMEHHSNLVPWQIIAERTGAVLRFVSLSQCESLMDHLRLSVGNKTKLVALQHVSN
LGCVNDAMVIGEIAHKGAKVLLDGCQSVPHMPVDVQALDCDFVSGSGHKMCGPTGIGFLYKMEVLKAMPFFLGGGEMIQDV
YLDHSTYAEPPNRFEAGTPAIGEAIGLGAACDYLMGIGMGKIHDYEVELSGYLYKKMAELEGMKIYGPKPTDPGGRAALCAFN
VPQIHPTDLMSVLDQSGFAIRSGHCAQPLHRRFGINASARASLYFYNTKEEIDMFIEGLKEAIEFFSVAGGV*

>jgi|Guith1|121029|au.210_g22131 C-terminal missing

MAGTAPALVMALLVLGGWTTQTGEGATNMCGNCSFTVSDAGLLQRTGSCLAGTGPCVLKAAGIAEIERGTFDGLADMTYLDLS
HNQLTNISQGLFDGLSSQLLDLRYNQLSS IAPDAFSNCTQLSS IDVGFNQLTCYYASWPLLAMALDSTVLARLCQNVTTDVT
PCYPHIRVELDSTSLPSDMAGEYCEAFNCAHGEGISPRDDSSHCFRKNGDEAWKILNTGCGWTIGHIEQNGERLYPSRFLQW
QEYARSYIGDCSLIPPSQLDVRLANTSNFYDGGGRIVVGIHTVLVLP PHHSSDADPSADIFNATSAVNSSSASTPLPLVDT
LLSFPECVEQRDWAAMLELLRNGTAATGGGLGMMPEPPLESWVCIREHPCLQRIDGAQPSDMMCFVYHEVDLVFLPWSTNSLLK
LVGSRMSRARVAWAWRMLLL VAGLVGRVQVCTFRAPHLTSSMGYRGIHHQLKVSMRASSTSLSLGKEVRKDFPILKQKVHGGK
DLIYLDASAATSQKPVQVLEAMKKFYEESNSNVHRGSHALSVKATDMYEGAREIVAKFVNARREEIVFTSGATEAINLNSQQG
FDMDFHRSILSDKTKIVAVVHVS NMLGCINPVQEI VAAAHAVGAKILLDACQSAPHMPLDVRLDCDFLVASGHKLCGPSGSG
FLYGKKEILESMPWK

>jgi|Bigna1|147067|aug1.128_g21775
MSMGEALQGEFPILMQKVHEGKPLLYLDSAATSQKPRKVTEVLREASKNATTFLESDIQSGINKARSKVASLINAEPPEEVF
TRGATEALNLVARGWDLHVSEGDEIILSVMEHHSNLVPWQNLAEKGA ILRFVPLSNDQAFDLEAFKSMNLNPGTKIVAVNH
VSNVLGCVNPVNEIARLTHEAGAVLSLDACQSVPHMPVDVRELGADFSASGHKMLGPTGIGFLWGRKELLE TMEPLNTGAQI
FTEVSLDPSAPPPSTAFLPVPWRFEAGGVPEAVALGEACRYLQSLDMGRVHSYEVELGRYLWESLSEMEGVRLYGPSPDSA
AGEGRGALVAFNIPGVHANDLAFFLDQEGVAIRSGHCTQPLHRS LGIAGSARASLYVYNTRKDVDRFCKKLQKTLKMFDRMG
FLVEEKSPCPLVSTADPTLKDPAINRPCTDLWPRVEEEEMGEESAAS*

>jgi|Phatr2|19482|estExt_gwp_gw1.C_chr_50292
MRPSNTILSFVFTLLLLATCLGASHGWT SRVAVGSSRRRAVVTFVSVHDGSTENASIGDHTDRADFP HGYIATRVRPDFKILS
RTVTDNRNPLIYLDASAATSHKPPVLDKLDEYYQYLSNVHRGAHQLSREATAAYEASRDKVQAFVKAPSRNEIVFTSGATDA
INLVVQSYGRAFLQAGDEVLLTATEHHSNLVPWQMLAEKGLVLKFVPVDPETGGLDWNAFESLLSPRTRFVAFQHVS NVLAC
INPVSDMVAMIRNKSPDAKILLDACQSVPHMPVNVQQLGVDFLVASGHKMGCGPTGIGFLWGKEDLLNSMPWKG GEMIDQVT
LEGSTWQPSPARFEAGTPAIAQAVGLGAAIDYLN EIGMDNIEAYEHEIGAYLYAQLSTVPGVTVLGPSVNRVALSAFAHTTVH
PSDLSTFLDVEGVAIRAGHHCCQPLHQALGYSHSARASLYFYNTKDDVDDFIRHLKTTIAFFENLESPSEEQVDDDFVPFV*

>jgi|Ectsill|23950|rna8870
MVSAAAVDGVTSMKEEGDKDGLGLSLGERVRADFPILDQEAYPGKPLVYLDASAATSQKPKVMDVLRDYYERDNANVHRGAHQ
LSIRATEAYEAAREKVGA FVNAETSREIVFTRGATEAINLVAQTWGLDNLAAGDEIVTTVMEHHSNMVPWQLLAERTGAVLKF
AQIREDMSLDLDHMKSLITDKTKLVAVVHASNALGGVNPVSEIAEAAHAVGAKVLVDGCQSVPNMPVDVQTLGVDWLVASGHK
MCGPTGIGFLWGRMSVLETMRPWQGGGEMIDQVYFDHSTFAEPPARFEAGTPAIAQAVGLGAACDYLS SIGMENGAAEHEMS
KYLWERLSEVEGLDFYGP PPNASGDNRNPLLAFNSRDVHAHDL SFFMDQEGVAIRAGHHCTQALHRQLGAAGSLRASLYIYNT
KDDVDQFIEALRSTLDMFKAMDGDGGGAGGGLVGGEPSIF*

>jgi|Nanocel779_2|581086|fgenes h1_kg.5_#_1192_#_TRINITY_DN7801_c17_g5_i1
MSEYYEHSNANVHRGAHALSIRATEAYEATRDKVQNF IHA STREEIVFTRGATEAINLVANTWGIQNLKAGDEI LLTVMEHHS
NLVPWQLIAKKTGAVIKFVGLDEEGLDLPQFHSLLS SKTKLVAFPHVSNNTLGCILPVAEMAAAHA VGALVQLVDACQSI PHM
KVDVQALDVDVFLAASSHKMCGPTGAGFLYKGLGV LKQMPWPWHGGGEMIDEVFF EYSTFAEPPARFEAGTPPIAEVVGMAACD
YLMEVGMEKVEEHEHQLGKLYERLSEVEGVRIYGP PPKGEKGKRAALASFN TDQVHASDLAFFLDHEGV AIRSGHHCTQPLH
RMLHAKGSLRASLYVYNTKGDIDVFIARLKETMEFFGNFQSGGAGAA TIL*

>jgi|Ochro1393_1_4|872714|fgenes h1_kg.13_#_1149_#_TRINITY_DN19211_c0_g1_i6
MNYGFILISAF LILQGWSPYVLGFGQKSSPIVKFSRLSRTLAATVAVDEMNAIEREDFPILNQESH PGKPLIYLDASAASSQK
PTFVLQMDAYYKTS HSNVHRGAHALAVRATDAYE AARQKVQTFIN AQHREEVIFAKGATEAINTVALSWGQRLQAGDEI ILT
VMEHHANLV PWQMLAQR TGAVLKFVQMTPSQQFDVDHFHSL LSPKTKMVAIAHASNV LGSINPVEEVIQSAHEVGAVLLDAC
QSVPHMPVDV VHLDCDFLVASSHKMCGPTGIGFLYK KALLESMPPIVGGGEMIDVTLEGATYAMP PSRFEAGTPAIAEAVG
LGAACDYLSRIGMAR IHEHESKLGAYLYDQLAKIDGLTLLGPSTTEPGAQRTGLVAFNSNTI HATDLSFFLDQEGVAVRTGHH
CAQPLHKQLGIAGSLRASLYFYNDKEDIDNFIVK LKETIHM FENMESIF*

>jgi|Auran1|11607
AEALRAEFPALHQRTNGRD LIYFDSGATSQKPARVLGALEK FYARDNANVHRGAHALAT RATDAYEAARDKVAA FVGARRDEI
VWTRGATEAINLVAQAWGPSNLGAGDEIVLTELEHHSNLVPWQLLARRTGAKIRYARLDADGAGVDEAHL LAQINGRTKLVAL
VHVS NVLGSVAPVA AVVAAARTRGAPGCAVLLDACQSAPHAPLDVGALGVDFLAASGHKMGCGPTGIGFLYGRREVLAAMD PWQ
GGGEMIADVFLDDGETTFADPPARFEAGTPPIAEAVGLGA AVDFLADVGMEHVAAHEAALGARLYDGLAAFGDRLELYGPTDA
ADRGAALVAFNARGVHAADLATFLDLEGVAVRAGHHCTQPLHARL GAPGGSVRASLALYNTADEVDAFLAALARVLD DFDGAG

>jgi|Vitbras1|1502|Vbra_6995.t1
MTVMQTTVVVPTSTALFCGLLIS AAGAA FVGLVAAAPTSTTLGSP LRSDFPILDERIDGKPLVYFDNAASSHMPKPVLEA IER
YYQHQS NVHRGAHTLALRATEAFESARSSVAAFLNARREEIVFTSGATEGINLVAQTWGEANIGEGDEI ILSVMEHHSNIV
PWQLLAKRKGATIKYIQMDENECLDMS SFESLLSPRTKMVGIVAVSNTLGC SNPIAEIVRQSHAVGAKVLVDACQAVPHKKID
VNALDCDWLVASGHKMMGPTGIGVLYGKYEVLES MPWPMMGGGEMIEDVYFDHSTYLDPPSRFEAGTPPIAQAVALGA AVDYLN
KLGM DNIAEY EKDVSSYLYRRLG DIPDLRLYGPDP IAGHSRAALVAFNHDNIHASDLSTFLDQEGIAVRAGHHCTQPLHRL
GAAGSIRASLYVYNTRQEIDCFIDALHRVIQMFSDIPEAPGIDLGFAALGMEGLSLLTAGEMGP NDEGEGEVME

>jgi|Symmic1|19232|rna14008
MALLAQGLTSPPPGLPLARRSNPRKTELASARGVQPTHAAAATTALVAGALTNRRTPGVGARVVS KAAKPGKSSWSAAEAEG
LRSDFPALQQEVKPGVPLVYLDNAATSQKPKQVLD ELNRFYSQDNSNVHRGMHTLSMRSTEA FETARQVAKFVNAPDSHEIV
FTRNATEAINLVARTWGAANIAGDEIVLTVMEHHANLV PWQLLAESTGAVLRFGRRLRDDGTL DVEQFESLIGDRTKLVALCH

VSNVLGCVNPVERVAKLAHSVGAKVLVDACQSVPHMPVDVQTLGADWLVASGHKMCGPBTGIGFLWGRGEILRGSPFFLGGM
IDQVTLLEGSTFNDIIPHKFEAGTPAFAEAVALGAACEQYLTDLGMAAVSQRRRAGGEHELAKHLWSRLEALPGIHLYGPSPETQ
PDRASLVCFNVDGCHANDLATLVDQDRGVAMRSGHHCTQPLHGELGVTASARLSAYFYNTIEEIDVAVDALEVAIGLIRTGPA
GDEPELDPAMAAALDA*

>jgi|Symmic1|35087|rna24877

MNSISSRNVLCKVSFFRAGRSGWLAGGIERHRHGKALCLRFLRQLTACRSGKRAKAEAKAACPSHCLPNMACHQRTGSPSFLSL
CVLPLRPGQLRPSGGGRPKDGQAKPSFQTAVLVAAATAAVQKSRRLRVAAPLEAPPAPSAEHQGVADRSLASKVRGDFPALST
EAYPGVPLIYLDGATSQKPTCVLKALTEYYEHSANVHRGAYALAEATEAFEHARSSVATLIGAKGSQEIIVFTSGATDAINL
VAESWGSRNLPDDEVLTITVMEHHSNIVPWQLATQRTGATLKHVGVTEEGILDMEEFHRLLSPRTKMVAFVHISNTLGCINPV
KEIAEAAHAVGAKVLDDACQSVPHLKIDVQDLGIDFLVASSHKMYGPTGVGFLWGRPEILEAMPPWKGGGEMIKVHMDSEFY
ADVPARFEAGTPPIAQAVGLGVACDYLLELGMDQVEAYEQFLARYLWESLSAIPGLRLYGPPASAGPRAALVAFNDTEPDIYP
QDVAGVLDVDGVAIRAGHHCAQPLHRALEATYGSARVSCAFYNTTEEIDKFVKALSEALETVRAGEGCVFDPDDPESCNCSSG
RMRS*

SUFE1 orthologs

>Cre06.g309717

MAMISQRQCGLRVASSLSSRSSPAVMPCPMRRQSLGRAMTGVCRAQKVSDLPASLKKIVGAFQMVDPDMARYKQLLFYATKLA
PMPDEDHIPANKVEGCVSQVWVPEMRADGLIYWRADSDSQLTKGLAALLVTGLSGCTPAEILTVQPEFIEMGLGKQSLTPSR
NNGFLNMFRLMQRTLELVAASGASASAAAAAAPAGESVPPQAATAAATAATSSSTNGAHASTPAPAPEAPAASAASSTRTPL
QDGMRRKLTEALKPSTLTII DESAQHAGHAAMMVAKPGKAGGPGETHFKVEVVSEAFEGLTQVKRQRMVYQLLADEFNMGLHA
LALVTRTPAEASK

>jgi|Galsul1|5938|XM_005703315.1

MRMGHFDRLMISHLDCLVWTSPEANFYRLAYTRQGREGPNCFFHNKKRLHFPKSTFRLQIRCQVDEKKLGL
TPELARLVKSFAAAPDPKLRVQQLLYLAQTLEPLPFQYKTNENKVPGLSTVHVVGDCDNEKIFFKGDS
AQLTKGLLALLIKGLNGCTVEEIERISPEFVTVAGLSVSLTPGRNNGFLNMLQTMKNKAKEAASKLNGVS
REGNYSRDPVKQTQASPSTAETASGPIYSAIVEKLQKLKPSRLEVHDDSFQHAGHVGAQLRSSETHFSV
YVVSDAFVGLSLVRRHQLVYTLLGQELKEGLHALRIQAKTPSEVSDE*

>jgi|Cyapar1|2397|g2322.t1

MFQVALSQPLSVRSAAAPSARGAICTASKPQRPKRAAAAARSWLGSQAQLSWSACVAAHGPPVRSVAPARV
PRVLAAGEVVAAGGMNLTPQLETIVESFQMVDPDKMYRQQLLFFAKELAGCPPEEVKKEENKVI GCVSQVW
IVGSLDAEGNVSYVGSDSELTKGLAALLIRGLSGNKPADILRVSPFVKELGLNQSLTPSRNNGFLNML
KRMQRIAVELAAGDGGSSSGGAGAAAAEPAPQAAQEEQDDGTVYSAIRKRLTAALKPAKLVIHDESAQ
HAGHAGARGYNGESHFRVEVVSAEFEGKNAVARQRMVYQVLADEMRERIHALALSTKTPAEAA

>jgi|Caulen1|93255|g7875.t1

MQFRVSRSLSSLYFPPTFRKRHLIFHSLHSSRRTTIRPQAQQSVMDLPSSSLKTIVTAFQTVPDPMARYQ
LLFFASKLDPLPSEFHTTEENRVQGCVSQVWVQATMTDGLVQFRADSDSQLTKGLAALLVQGLSGCSPQEI
LTIQPDFIEMGLGKQSLTPSRNNGFLNMFKLMQKRTLELAAIDLKQESSGSDSPSSNGLQEDGGDQSSSQ
TPVEDAIRYKLTQGLKPKELVI INESSQHAGHMGNPNNDPETHFKVEI ISEQFTGANSVKRHRMVYALLK
DELAGPIHALTLSTKTPAEASE*

>jgi|Caulen1|90246|g5535.t1

MAFSMSKRISLWNCTRSFTGAQSI SRSFKSTLLTRRTTFKLDARESAVELPTSLKKIVTAFQMVDPDMA
RYKQLLYFASKLDPLVSEFHTKKNRVQGCVSQVWVTANLIDGLVQFRADSDSQLTKGLAALLVQGLSGCT
PREILSIQPDFIEMGLKQSLTPSRNNGFLNMFKLMQKRTLELATADIRSQQSTGGDSSSNKGVSSNAGQ
PPSKTPIEDAIREKLTLEGLNPDQLLI INESSKHAGHMGNPNNDPETHFKLEI ISSEFEGANSVKRHRMVY
ALLQEELAGPIHALTLNTKTPAEVK*

>jgi|Tetstr1|426704|TSEL_016974.t1

MPAAARPVVGNARCAAGTPAACQQLSQPHRIALSACAARTHLPLQHSVTRPLARCRRSFAMARPVRAAS
RRNSQIVCAQAGAEPLPNLKKIVTSFQMVDPKARYKQLLFFAAKLQPMAAEDHTDDNKVTGCVSQVWVK
PELREGKVFWAADSDSQLTKGLAALLVQGLSDCTPEEIVALDADFIEALGLKQSLTPSRNNGFLNMLRLM
QKKSLQLMASA*

>jgi|Chlat1|2784|Chrsp187S02951

MEVVSMSMPAAALVGTSSVARPPPLPGSHRHACSSSSSPSPHPSCSLPSSSSASLCCFPLTSSLPSRRL
RRLTRPTRPAHSTTTTASQSVNELPAALQKIVRGFQMVDPDMQRYKQLLFYATKCLKPLPAEDHNDGNKVT
GCVSQVWVVRLEDGKVYFSADSDSQLTKGLAALLVEGLSGASPEEVLRTPDFITDLGLKQSLTPSRNN
GFLNMLKLMQKKTLELSMQVGASS*

>jgi|Mesovir1|1777|Mesvil412S09175

MMTQEVEDLPEALQAIVRGFQMVDPDMERYKQLLFYASKLPPMPDEDHVEDNKVKGCVSQVWVKPRMEDG
KVYFTADSDSQLTKGLAALLVKGFSGAKPEEIVRVTPDFIQMLGLKQSLSPSRNNGFLNMLKLMQKKTVE
LYMATQKDA*

>jgi|Pracol|14963|PRCOL_00000686-RA
MQRYSQQLFFATKLAFLDDAARTDDNKVKGCVSQVWVVARLDAATGTMLYEADSDSQLTKGLAALLVEGL
SGATPDEVAAVQPDFVEELGLKQSLTPSRNNGFLNMLALIQAATAALAAGGATGAAAGTDAEPAASAATD
DTDAARPVYAAIVSKVSAALAPSRLELDDESASHSGHAGVNGLRGEETHFRMEVVSASFEGLNVRKQRK
VYELLADEFKGSLLHALALTCRTPEEDAAAAAN*

>jgi|MicpuC3v2|823|wlab.202019.1
MSSATATAAASALASRAAPHRGDHRRRAVIATRNNAAAASRATTAAAAAETEKLPANLKKIVGAFQMVDP
PMQRYKQLLFFAAKLKGFDEKDRVEDNKVQGCVSQVWVPRMGEDGLVYFTADSDSQLTKGLAALLCEGL
SGSAPKEIMAVEPDPFVLLGLGQSLTPSRNTNGFMNMLRLMQKKTLECYMEA EKAAKDA*

>jgi|Ostta4221_3|72643|OT_ostta02g01250T0
MSARLRTPWFRTDAGVARRRDVQERSRWRTLRTPRSIILVRAETGEISPKSLASVVSFQAVPDPMQRYR
QLLFMAKTLTPVQAERLCMDYKVPGCVSQVWIIPSLKDGLVYDAESDALLTKGLAALLIKALSGNSPEE
ISSVTPNFIIADLGLKSALTPSRNTNGLNMLSLMQNQARSFL*

>jgi|Mesen1|4166|ME000219S03292
MATLVCQQSLAAITVQATRRLDYKEGFRSACFGSFPAAKCSRLAGSLLSSASVSNNNTSYLGLVYTTSN
ARKTHELSICCEQAIEELPEKLQKIVKSQFQAVQDPRARYQQLLFYAKKLKPLAKEHQVSENKVTGCVSQV
WVLPISIDVEGRVHFQAESDSALTKGLAALLVEGLSGATPTEIVRITPDFVQMLGLKQSLTPSRNNGFLNM
LRLMQKKTLELFMQWQASQKSSASTNDTQDIVGEAETTLITVGSEKEGPAAGVPQDGSAGISTEPSEGSE
STSSSGSADDVVRDPKPIYSIGILFKLEKALEPVELVDDVSHQHAGHAGVHRGATETHFNVKVVSAPFD
GLSTIKRHLIYITLLDEELKAGLHALSLITKTPSELERV*

>jgi|Guith1|66468|e_gw1.13.229.1
MSVNEVKLTPELQKIVQQFSLVPDPKLRYYQQLLFFAAKLGAMEEVHKVEENKVKGCQSTVYVHATKDEEG
KIWYTGDSDSLTKGLCAMLVRGLSGNTVQDILEVSPFVKEAGLSVSLTPSRNNGFLNMLNTMKAQAKA
LA*

>jgi|Emihu1|58720|gw1.152.59.1
LEKVVRGFKSVDPQKLRYYQQLLFLAQKLAPMDASLAVDANKVPGCLSTVYVHGSGLDDDTINFGESDSQ
LTKGIAALLINGLSCGTNEEIQAVDPSFIIQAAGIAQSLTPGRNNGFLNMLKMMKGQAAKLA

>jgi|Pavlov2436_1|1033487|fgenesh1_kg.256_#_63_#_Locus3407v1rpk48.27
MQLSLSWVLAVGWAGPQASQVLPVAVSCARGPAHRVWVGQNPENPEI GLTPYLEDVVETFKMVDPQKLR
YQQLFLAQKAKPMDSLKVPDNKVPGLSTVVFVHARPEAGRIFYDGDSDAMLTGLVLLTEGLSGHTV
EEILAVDPKFIAEAGIASSLTPGRNNGLLNMLALMKKKAQQIASGSAAAEFASVVAQQSVVAGPVVEG
GSAPVADSIIRKVCEKLKPTRVKLVDNSAQHAGHKSREGLADGETHFALEVVSEVFAPLSMVQRHQLVYA
LLDAEMKGSVHALQIVARTPSE*

>jgi|Bigna1|92600|estExt_fgenesh1_pm.C_380005
MSTPYHTGASLLLLVVVASSLAALSPTRHLSTGLKPSMVLQRPQQTGVLQQVTTTRKLEVSRLNRAGGFS
RAMEQLSRRSSRQREMHAFAATSSSLNLTNPNEKEVVRKLSAPDNTLRHQQLLYAKEAQPLPSQFQKAE
NKVPGCLSTVYVVAVKDEDGKVFFRGDS DALITKGLVNLLIRGLSGYSVEEIVAVKPEFIQEAGITQSLT
PGRNNGFLNMLNTMKKKAVQV TRE*

>jgi|Thaoce1|72048|rna32474
MKTMIAGIPVAGIALYSLASTTGGYGGVEAWSPSRQSSSTSRTASSSTVTRPRRGDTTATALHSSTEID
TDALNLTPELEMMTGAFSSIADEKTRHKQLLYMANQLPTVDDSI RIPENKVPGLSTVHVDCTTEEKDGD
TLVNYVGDS DGLLTKGLLALLIRGLSGCTAEQIAAVDPQFIQAAKISQTLTPGRNNGFLNMLAVMKKKAS
DAVEGGDESPASGDAAGAENSEGVVTSFDEVEGKPMYNAILSTLVPIPKPVSIELVDNSSQHAGHGGAKE
WEESGESHFALDVVADVDFGLPLVKRHQLIYMLLGEVMPKIHAIERAKSPGEI*

>jgi|Phatr2|7639|gw1.5.347.1
DDKLRYKQLLYMANQLEPIPESSQIPENKVPGLSTVVFVDGTAKYSDNGQDVLIYFRGDS DGLLTKGLVA
LLVRGLSGNTAESIQKIDPAFIQKAGIAASLTPGRNNGFLNMLATMKRKATQLA

>jgi|Thaps3|36605|e_gw1.11a.172.1
LTMMTNAFASIPDEKTRHKQLLYMASKLPDVGDEVVRKENKVPGLSTVVFVDCVLEKNDEDEGEYVYNYFG
DSDGLLTKGLLALLIRGLSGCTPQQINEVDAQFIQAAKISQTLTPGRNNGFLNMLAVMKRKA

>jgi|Nanoc1779_2|577053|fgenesh1_kg.3_#_1240_#_TRINITY_DN5402_c0_g1_i1
MRPLSPVPLQLGLVLLII VALARVTQGFVVVVPRQQVGRPSCRQTVVMSSTEIDESSLGLTPRLLEVS
R AFQLMDERTRVQHLLAMAGALKPFKDFRKL PENKVPGLSTVYVHAYLKEGKVYFEGSDAQLTKGLVAL
LIEGMSGCESHEIQAVKPEFIKFAGIAASLTPGRNNGFLNMLAMLKRKAVEAEAKGEGTGQEEPEIEFEE
IITRPVYSSTMKKLINFLKPKQLIITDDSAQHAGHAGAKGLDGESHFSLKIVADCFEGLSLVQRHKL IYT
VLTAE MQQIHALSIDAKTPAEAGL*

>jgi|Ochro1393_1_4|727987|e_gw1.239.22.1
MSVKEEEADDFNLTPVLKTYANGLRNVDDKLRYQQLFLAAKCPEMDASLKLDSNKPVGCLSVVHVHAD
LDSEGKISYKGDSDSLTKGLVTLTVKGLSGHTPEETEAVDPKFIQYAGVAKSLTPGRNNGFLNMLNLMK
VKAKQLADGDVDGDANDSPGSIATPTVTASETTQGPVYSSIQTKLSMLKPSLLEVEDESYKHAGHAGIAG
VTAGSGSGETHFNVKIVAECFDGLSRVQRHQMVYALLAQEMKNGVHALSINAKTPSEV*

>jgi|Pelago2097_1|314629|CE314628_4120
MLSAFVALVLQAGLCAGFSPGSHARRRPSRTGAAPRAAAASLDALPAPLRAIVESLASLPDDKYRYKQLL
FMAGDCDPIADALKTPENKVPGLSTVHVAATLDDGKVYYTGDSDGQLTKGLLTMLVKGLSGATPAQIGE
VKPDFIRDAGIAASLTPGRNNGFVNMLAVMQKKAQALLDAPAADAGGQKLAAMLEILACLKPEALEL
VDESGGAETDFKLTIIVTASFDFEVEMRQVVIETLLADLIIPRCESFRILALTPEER*

>jgi|Vitbras1|825|Vbra_4677.t1
MVISRLAVAVAWQLPSGQHQPRLPALPSSLVHRQQPSFRRSSTVLRESVGDGGEAAGDLPPKLRTLIATF
RRMPDDRFRYQQLLVFAKGLKPMNNAYKTKENKVKGCLSVVHVHAEGREDDSGDMRIYFQGDSDGLLTGK
LVNFLVQGLSGCTPRQIESVPAAAVMEARVTNSLTPGRNNGFLNMLETMKAKARDIAAVQGRA

>jgi|Symmic1|13793|rna10871
MNCTFARCLYCEKLAPTHPHPPPGIFALSPAMLVVSATPKAPPAPTLRSRSDSRCPETSWASGVRAGA
GLVSLLPAMPaelKLPENKVPGLSTVHVSARLEEDGTVSFQGDSDALISKGLCALLVLCLSGCTPEETIA
GVKPEFIKASGISAAALTPGRNNGFFNMPRPKSVEYRLSQVESALLSN*

SUFE3 orthologs

>Cre06.g251450 SUFE3
MRLFANTRGCASATHERSRCILVGSTRFIGGGAPRVAHQSSLEQSKPCLISRRGRPGFQLAAVQASAAGDTAQVATPSTGAPG
AEWLPASLEAQLAPITSAPDAKEALRRLLDAGSGPLAPLAAARTASNRMVGCTAQVWLAAETDAAGRMFAQGWSDSEVSRGLV
ALLVRGLSGCTPEEVMQVSASQVQQLSRVLGRSVLPGRANGLGNMLESARKRAALAAAAAAGRRLDVFPSSLITADALTPQ
GAFAEQAARYLAPDAAVSQLVRVCRDKHIGVVAHFYMDPQVQGVLSAAAEWPHIAISDSLVMADTAVRMAEAGCTTICVLG
VDFMSENVRALIDEAGHSAVQVYRLAESDIGCSLAEAAESDSYRYLQQAHTPNSVHVYINTSLRTRKARAHALVPTITCTS
SNVVQTVLAAAFADVPATVWYGPDYMGANLAQLFADLASGAASDDVRLHPAHTVDSIRSLPLRLRYFTDGTCIVHHIFGG
EVTTELVAAGYDAYLAAHFEVPGEMFRLAMQAKRSRGMGVVGSTSNILDFIADKLREALSAPHPERLQFVLGTEAGMITSIVR
KVQGLLRQSGRTDVEVEVFPVAPSAVATPQQRPEGAAPLTLPGLALVPGPASGEGCSLEGGAACPYMKMNTLAALVSVC
ERVGSPAGEASLERYRPRTYGGETVGGRSLLAAGCVPILHMRNFQRSQGRRLGPDLLQDIASRHTAR*

>XP_005850636.1 hypothetical protein CHLNCDRAFT_19680, **partial** [Chlorella variabilis]
LERLATELLAAAGPQERSRLLGLYARRLAPLPDAARTDANRMVGCTAQVWVSAELDGEGRRLMADSDSELTRGLAALLVEGL
SGLTLEELLQVDSAVLGLGLGAAVLTRSRANGLNMLESLKRRARMLLGDLPRFPSSLIGAERTSAQGAFAEAQNAFLRPDG
AVVDRLVEQLAAKKVGVVAHFYMDPEVQGVLSAAERWPHINISDSLVMADGAVKMAEAGCTAIAVLGVDFMSENVRALIDEA
GHADVAVYRMSADSIGCSLAEAAESPAYDAYLAEAGDTPNSLHVYINTSLKTKALAHSVVPTITCTSSNVVQTVLTAFAPQVP
DVHVWYGPDTYMGRLAQLFQSLANLSDEEVRELHPAHTQASIHALLPRLRYFEQGTICIVHHLFGGEVCELVKEGYRDAYLTA
HFEVPGEMFSLAMDARQRGMGVVGSTQNILDFIAAKLGAALQFPFNRLQFVLGTESGMITSIVRKVQGMRAAGRDDVEVE
VFPVSPPEAITTDRQQQVVRAGLPTGLSVVPGPAGGEGCSLQGGCASCYPYMKMNSLQALMTVCQRVGSPAGEALLEAFKPRPY
TELVDGKTMAQAGCVPILHMRGFQKGGKLPEALVADITGRHSA

>QDZ26067.1 quinolinate synthase [Chloropicon primus]
MSQRLGPLSRSSGSGRRKRRTGFVSAVYKAEAPVEEAGGTGFVENLHVELIADLGVKDDSDTGALLKRLIELGSSSPFVLP
QHELTEEKRVPGCASRVWIDCEVREEDSLVHFRGYSDAISKGMCNLLCRNFSGKLTPEFSLIDESFLKEVTRGLGPVLSSS
HRRHGLFSIFDSMRKRVHQLVSGGIDEPFPLVITKEDLIPQGYAESQATYLPKPDQSKVDDLVRSLKETKCGVVAHFYMDPE
VQGVLAQAQKDWPHIHISDSLVMADRAVDMAAEGGCENICVLGVDFMSENVRATLDHQGYNDVNVMYRMDVLDIGCTLAEEAES
ERYERYLEEAARATENALHVYINTSLKTKAVAQSKVPTITCTSSNVVQTVLTQSSSLDGDLSVFGPDAYMGANIKEMLEQL
CGTSDEEVRLKIHKDHDISSIKNLLSRFRYYQDGCILVHDMFGSEVAQTIRESYADAFCTAHFEVPGEMFKLALHASHNGDRGV
IGSTKNILDFILGKVAEANEHNLESERLQFILGTESGMITSVNVNAVQESLREAGSNVEVEIVFPVNSSAITSASQANAPAMPF
GLDIVPGPAGGEGCSMEGGCASCYPYMKINSGLDMFVLSKLEEDSDPILQDYAPRNYEESIDGKSFAQVGTVPILHMRHFTQ
HKCMSDDLVDADIKARGLQ

>XP_003064062.1 predicted protein [Micromonas pusilla CCMP1545]
MSSAPIAGLRARVAKRVARRNVDRRAIVSTERSAANNLGNELKRSVAGAVAAEKDVSEKLKALIARGGDLPALPESDRAPS
NRVMGCTSQAWIDVSLDADGKVLLRGHSDAQITTGAGVLAAGLGGLSPEDVLGVDDDVVKDLGLGASALPRSRANGLFRNMLE
SVKKQTRLLRDGRDAMPFPLLVTDAGIEARGPFAEAQANYLEPDAAEVAALVKELKAKKIGVVAHFYMDPQVQGVLMAAAKE
YEHVFISDSLVMADAAMVMEAGCDVAVCVLGVDFMSENVRALIDDAGHADAKVYRMAAEDIGCSLAEAAQSDAYYEYLAAADG
VENAVHVIYINTALDTKANADAIVPTVTCTSSNVVQTVLQAAQVPGVNIIFYGPDYMGNLAEFTMMSTWTDEEVAHVPA
HSRATIKALLPRLRYFQDGTICMVHMHFGGDVCDTVRFYFGDAYQTAHFEVPGEMFKLAMEAKRERDMGVVGSTKNILDFVVD
VDDAIARKHPNGERLSFVLGTETGMVTSIVRAAQARLKKAKADGVDPVECEIVFPVSADAVTATGETQAPSIGGLAVVPGPAG
GEGCSSEGGCASCYPYMKMNSLDALMRVCARVGTGGAATLAPMEPKKYAAAKEGEKSVAARGCVPILHMRDFTDRKKFSDAFV
NDIYGLNLIIITYHITRARASRRAPPPPSRASLAHALAAMLRSKSIIVSVAFFGGIST

>jgi|Klenit1|14322|rna-KFL_008730030

MEVATRSAASFPHLLPDGQVLHGQGGARGKPARPAALASSPSPLKIPRRGPLFCRCGKELPRSDFFGTSTW
HPDASSQRQPQSTRPAFARAEGQAGREGGRKQRGSIGATARERRQVRAAVSVVQDVVTETGGWGLEEVRRR
FEALADGQSKYRLLQYAAGLPRLEGDKTADNRVMGCTAQVYVTVCLDDAGLVQLGGDSSELSTGLVA
VLVAALGGRPPAEVLAVQEDALAGLQLGSAVTSRSRINGFYNNMLQSIQRRTRAAVAAAAGEAPPAPFPSSL
LITAGGEAVEAQGEFARAQAEYLSPDPAAVTELVNLLREKKIGVVAHFYMDPQVQGVLA AAAARQWPHIHI
SDSLVMADRAVSMAAGCERIAVLGVDFMSENVRAILSRGGYPEVPVYRMAAEHIGCSLAEAAESPAYVD
FLEEAAGTPGSLHVIYINTSLGTAKAEAHARVPTITCTSSNVVQTVLQAFQVPECSVYYGPD TYMGNNLG
TLRLQLATQSDDEDVAKVHPAHTQASIRALLPRLHYFQDGACMVHDMFQQAVTQVRVASYGDAYQAAHFV
PGEMFDLAEAAAARGMGIVGSTKNILDFIAERTQAALAAGYAQRLQFVLGTEAGMVTAINNKVKILKEA
RAAGNDTPVEVEIVFPVSADAVTATGENSGGSEALAGLTIVPGVAAGEGCTAGGCASC PYMKMNSLDA
LLMVCRRIGTPEGELLDLAFQPRMVDRTVAGKAAVDLGSDSLHMRHFQTNGKLSDELVADIRTRNQHGRRQ
GPL*

>jgi|Chabral|329227|rna-CBR_g20372

MEAQVVAGAVRMESAWATGIVVGQEGEGRGRNGCCCSAQAASSSLSMSLCHSSRLQRRSGSLRFSCHGS
SDSGCSHRWESAHSVYGRSRQSSPLSRSDPWGCGSRWVSKSQICGCFSLRVRRKMEGKVREGVVRCLAD
AVVLRHVGEVILQKIAEEFSRCEDNKDKMRLLLHYAERLPRLDKEDCIPENQVMGCTARVWVTARLDDE
GWVWFSGDSDAELSRGLCAVLVQGLSGLTPEAILDVREETLTSMHLSPIATRSRINGFYNNMLKMMQKRAG
ELALLRDGLTPVAPFPSSILISANGELQPRGDFAQSQARFLSPDPHAVDTLVSLISSKKIGVVAHFYMDPE
VQGVLA AAARSRWPHLLISDSLVMADGAVKMAEAGCTTIAVLGVDFMSENVRAILDQAGF SHVGVYRMSNL
EIGCSLATAAQGPSYVQFLDEASKTPNSLHVIYINTSLETKAEAQ LKVPTITCTSSNVVQTVLQASAQVP
EVSVMYGPDTYMGANLAFVLRREMANLSDEEIAAIIHPAHNRESIGALLTRFHYFQDGACIVHDI FGSEVVK
RVKAGYSDAYQTAHEFVPGEMFALAMEAKRRQMGVVGSTSNILSFIADKTQEAIARGHDQRLQFVLGTES
GMTSIVGKVIKLSNMEAAAGARGPPPQVDVEIVFPVSSDAITASSEATGGNLSSPSLSAPAAGQGLDQ
LKGAVKPASGVLGELGVVPGVQGGEGCSISGGCASC PYMKMNSL NALLRVCQWIGTDSEVMLSAYKPRLV
RDQVGGRSVLVDLGCTPILHMRHFQKTGSFADDFVDDIRARKYLSS*

>CBJ28535.1 conserved unknown protein [Ectocarpus siliculosus]

MFGTKAAVLRGARLIGRRSLHRRQPPDAPAAAAAVCAGPGVSQRERGAHYSRVAGLSTAAGLKVATAARQVQHGGERFP
SL
LITADDITAEGSFAETQATFMTDPDSLVS ELDHLLPTKNMGVVAHYMDAELQGALAALKWEHVS VADSLAMGDAAVRMAEAG
VSSVACLGVDFMAESVRATLDSRGFQGVVPVYRLSESEIGCSLAEAAERKAYEAWLRKASATANS MHVVIYINTSLVTKALSHSL
VPTITCTSSNVLHTVLQAFQVDPDLTIWYGPDTYMGENLHNMTEL AGMTDGGQIREVHPKHDQASVRSLLPRFHYFRQNCIV
HHMFDDVVERVRSNHADAFHTAHLEVP GEMFRLAMEAQNKGRGVVGSTSNILGFIKDQTRGAVARAGGAGGAGEEAAAGEKLS
FVLGTESGMTSIVRAVQEE LRAHNNKTQVEIVFPVASEAMTATGESGDGSMQVVPVGVSGGEGCSTAGGCATCPFMKMNDLHS
LLDVVG DANESGAVPDLKGF LPKRRAFTIDGQDKTELGVVPIHMRDLMDKGALSDDLVS DVQSRG

>jgi|Pico_ML_1|52398|g3107.t1

MRAQVRASISRDRGIGTPAKGKVETQEQLQKLLARGARAIAWPEERRKAENRVLGCTTRTWLEARRCRGG
KGVEFEVDSDSLQTRGVAAIACEALRGKDAQQVARIDVQEFGEIGLSQAFLAPSHANGFQNLVNAMKKR
AAVLQAQDEADGADVPEAVRFPSLRITADDVEAGSF AAAQAQFLRPDLQVKELVHVIQVKRIGVVAHF
YMDPEVQGV LSEALDWP HICISDSLMLADKAVAMVEEGGCKSICVLGVDFMSENVRAILDEAGHKDVQV
YRMASDAIGCSLAEAAESLSYETYLKEASKTPNSVHVVIYINTSLRTKARAHGLVPTITCTSSNVVQTVLT
AYAQIPGVTVVYGPDTYMGANIEKLTLDS EMDDAATQELHPDHNRETIKSILPRFKYFKDGSCIVHDLF
GGEVCDTVRTRYRD TYLTAHFEVPGEMFILAMEAKQRGMGVVGSTQNILDFICHRLDEAIHDGETDET LR
FVLGTESGMTSIVRHVQAKLKPVASHPGLAVEVIFPVSTDAISRPEEQPAGLGAVPGTLLRAGAGARA
VRRTKQTTDRDACTKNEEGTGNV IIPKKDYSEHILKIQRKKTP*

>jgi|Caulen1|90988|g6234.t1

MMTRYNFASRPIASKSPCVQTQFNQRALRSPQSLHVS RPRTRYVSNVVCQVTINNP AVASVLT ELDSDS
SATEKLKVL LHRGRQLIEFPESNQTIENRVMGCTTKTYLTVELVDGLVEFAGTSDSELTKGFIWILTQLL
NARPVDEVLQFDVQSISEILSMSEIMTPSRTNGFLNLIETLKRKTRMLVRQFPKFPSLIIRSDKLEPVGA
FAEAQAQYLNPNAGAVEELADLLREKKVGVAHFYMDPEVQGLLT MASDLWPNIVISDSVAMSEHAVKMV
REGGCEHIAVLGVDFMSENVRIVLDRAGFDQIPVYTMSKDHI GCSLAEAAADSIAYEDYLRNSMKLTSGPK
LHLCYINTSLKTKAISQSIVPTISCTSGNVMMKTL LQSYAQIEDLT VLYGPDTYMGENIRDMFTSIADMSD
EIKQIHPGHDKESIRSLLRKFEFFQEGTCVVHHLFGGEVCKIVREAYGDAYIAAHFEVPGEMFALAMEA
RRRGMGAIGSTRDIHDFISRIVEAQVNERTTDRQLF ILGTESGMTSIVRSVQSIIRKNSTLEDLEVEIV
FPVASSAITTQDQQTEKVISLPDEMVLVPGAAAGEGCSAEGGCANC PYMKMNTLTALKQIVERIGQGGGL
EGFY PQEMDSVVQVNGREESI HDLGF LPVYEMRNFQENGVLSD ELVSSVKATT*

>jgi|Tetstr1|432101|TSEL_021572.t1

MAALTRAPLRPAAAGLHRHVAAGSVPIPPAAGRSTQPRRLREHFRSNLVAAAASASSVQVAQLPTAALQE
IAAELDAAADPKQRIKTLVAYGAQLEGLPEEAKVMNRVMGCTAQVWLTAKLAADGTVQLAADSDSDVTR
GLCAVLVKGHLHKFQELAEVSTDALLALPLGPVMA PSRTNGFLNMLETARKLARGLMGEMETFPSSLL
KANSISAQGPFAESQAQYLSPNPEAVDRVAELLSSKKVGVAHFYMDPQVQGV LSSAAERWPHIHIISDSL
VMADTAVRQPARPVWAPTISTTILTSIITTTTTTTT VITITITTTVANTITTLIIITITATVATTIM

>jgi|Pracol|16967|PRCOL_00006555-RA

MARCAPARPAAAAAAMPARGRARPRRAAARAAGGGGGSVAERWEAVAARIAAAGDAKARLALLLERGAAA
PRMADAARTAAANVPGCAARTWVAHPLDGDASGVREFADSDSALTRGLAAYFADALEGLPPREVAALDA
GAALRALKLGDAAVANSRGNLRLTLEAFARRAKELDAEERERAGAGATAARARAAPFPSLVLRAGAPA
APRGAFALAQAEYLRPDAARARAMASVLSAKRIGVVVHYMDPEVQGVLSAARDEWPFVHISDSLMLMADQ
AVVMVEEGGCESIAVLGVDFMAENVRAVLDRAGHTSVPVYRMDPGDIGCSLADAARGPEYDEYLGA AAAQ
GRSVHVYINTSLATKAAANATVPTITCTSSNVVRTVLQAFAQVEGVHVWYGPDSYMGANLVELLESLL
LPEEEVARVHPAHTHASVRDALSRHYFEDGMCVVHDMFGEDVVRTLEEHYVGERDTFLTAHFEVPGGLF
GLALKERAAGDSVVGSTAEILAFIKAKTADAVEANEDRALRFVLGTEVGMVTSIVLAVQDMLDGARERLG
GDAADVSVEIVFPVSSDAVQGVDAAGAITKSKAGSSPSLLGLEVPGAAGGEGCSAEGGCASCYPYMKMN
SLDALERVVGAIGTDAEAMLEAHKPRVMFEDASEMDRGVEPIIYMREFTKTQRFPPDLAERIEGSSVGAVA
ANAT*

>jgi|Chlat1|7489|Chrsp60S00549
MAAAAMETAALSGAVRGTVWVGRSVEIAADEAKQRKKSISVGGKKKPPPPQQRVPAVAVASSSGRGGSRRAA
TVSVATVKEGTRIPKDA AAAVAEVALGRSTPAWSQERAEAWRVAGCAARTWVRCSVDRATGGIDVRAD
SDSDIIRGVGTLLSKACQGRTPVEVASMSAPELLSLLGLPFPQVLSPPSRSGGARALADAILRQARAI
GGDVGTRFPVSLVIEADKLSPVGEFAEAQSRYLSPDPAQVQQLASLLASKRTGVVAHFYMGPEVQGVLSA
REVYPHIHVSDSLVMADSALRMVEEGGCQAICVLGVDFMAENVRAILDRAGHADIKVYRMSNDVIACSLA
DAARTNNYIEWLQHAKDNVSGPHLVVYINTALDTKAHAHSIMPTITCTSSNVVQTILTAFAKQIPNLHVW
YGPDSYMGANIKDLLTRLASESGVEGEKYDLLGHDATSLKSLLERYDYYKEGVCIVHDMFGEEVVKRVS
AYGDAYQAAHFEVPGGMFSLAMEARARGMGVVGSTNILDITFITEKTRDALERDFDDKLTFLGTESGMLT
AIVRRVQSLQQAALSHPSPSPIQVEIVFPVNSDSITTTSATSPQLSFQGLSIIPGPAGGEGCSSEGGCA
SCPYMKMNTLEALIKVVGRIGTPGEGMLSGYVPEVYGSVGEGEGGGSVADKGCPIVHMRGFQKSGKLTE
ELVEDILARATP*

>jgi|Mesvir1|12745|Mv20420-RA.1
MTSGLAATYVANAVLPSFVQKVGGCTGARGATVKGEVARKSGFAGRHDTCRLLKSAHQASRQSSAYQTL
ASPSGAELSRADEQMAMPSSMAKIGKEFQAVEDMQARMKMLIERGSSCPPLDHERVNNENRMVGC SARAW
VAASLDSQGRHLHYRADSDSAITRGMCAVLCEALSGMTCEEVIRTPATVRTLNIPTGVVGLSRTAGLHNI
FESMRRAARKLKEGAAGLAPFPSLLVSAHSIKAQGTFAEAQRAYLQPNPEAVTRLAQLLEKKKISVVAHF
YMDPQVQGVLSAAELWPHILISDSLMLADRAVMVADSGCQRVVVLGVDFMSENVRAILDKSGFADVPV
YRMSADEIGCSLADAAISDSYFKYLQQAALTPRSLHVYINTSLETKARAHALVPTITCTSSNVVQTVLQ
AFAQIPDCTVWYGPDTYMGGNLCMFKRLAVLSDEEIQAIHPAHNRSTLKALLPRLHYQEGMCVVHDLF
GEEVVKTTRRELYGDAFQAAHFEVPGEMFALAMEAKSRNMGVVGSTANILDITARTTEAIERDLPDERLQ
FVLGTESGMVTSIVRAVQALLREAEKAREGKPCLEVEIVFPVSSDAISRSPDSSAGGSGKGTVTAFFGQ
LAVVPGVQQGGEGCSISGGCASCYPYMKMNSLEALMKVCERIGTPGESILLSSYEPRKYQEGGLGSRSLADVG
VDPILHMRVYFQQNKFSDELVDVLTTRYTA*

>jgi|Mesen1|9282|ME000060S08723
MFLENGKAFAQVPDLHVWYGPDTYMGGNLVDLLNRLAEQSDDEEVQAIHPGHTAASIAALLPRLHYFQEGT
CIVHDLFGSEVVERVRKGYNDAYLTAHFEVPGEMFALAMDARVHGMGTVGSTQTILDFILARTREALARG
FDERLKFVLGTESGMTSIVNAVRKELAAARQAGSTAEVEVEVFPVSSSAIATTNQEGIASSSLGGGAG
AALSELPIVPGVQGGEGCSVSGGCASCYPYMKMNTLDALVRVAKLAGSGGEPLLNAYAPRAFSHQVAGRTV
ADIGCDPILHMRHFQIRSSIEFAEEQVGMVSAQVVENTLIEKTRSSRYLSAVSVEDKGTGKVEGGVSGAK
GQQRLEASLYFDGYALLKETSSPTALQ TALQ TALQ TALQSEQNSFSATGSGSSSGSGSGGRGDADSWKE
GVMGVLEQHLGSCSSEGEWELEGGMVEQAPGHGCSDEASCPSPWWSASWHVEGFLFREDGGVLLPQQALT
VVAASNLLGASSGGLAVYPRSHWLVASWLQAALKGEAATSMPPPSALPASAPLQVAMAAGDVLLMHPWLA
SAPAPNASVHPAWRLHARLRPNRDSGCGDDSSASSALLPKAASADDVVRAMWRHRAWAAAQAPRGPPV
RGDQVAALQALQDEAAKAQAGNWA VPSLPPLPFSSSSAAAAAATAEGEELKLPFPCAYASPADVRA
WSLSVEARRFLAEDSWQEALVVLRLQLAGLRPLSFSIFYEAGRC SVRQAPRGQLAGRAGRKCLLEDGERFL
TRAIELAPGWPPAYAEALALTLAMGRGREVEVGPLIASLLQVPRLPCLCRQHPEAMAAMRQLLQCAQEVA
AQRDDDFQFFEWDLLEGLAGTGEHPSPDLT*

>jgi|Phatr2|14735|e_gw1.16.187.1
GSFAVAQATYLPDAEAVQDLVDLLERTQSGIVAHYMDVELQGVLAAGARLPGRVGIADSLKMGDMAV
DMAANHNTAIACLGVD FMSES VQAILQQNGHGHIPVLRATHKPIGCSLAESAERESYRVWLEQESVGQA
ALHVYINTSLETKAVSQSIVPTITCTSSNVLQTLQASAQMGPELKILYGPDTYMGENLVTLSTIEAR
WNDEQIASLHASHSRDTI

SUF8 orthologs

>jgi|Galsul1|3981|XM_005704982.1 **SufBC fusion**
MNQSLDATTLLKLEQIINKSYKYGFSTKIQNEDFPKGINEEIVTLISKKKNEPKYMLDFRLKAYKKWKQMYVPKWQYLYLKEQIN
YQNILYYSVPKQKQLQSLNEVDPELLKTFEKLGIPLNEQKRLTNVAVDFDSVSGTTFKQELADQGIIFCSISEAIQKYP
SLIQKYLGSVVPIDGNYFAALNSAVFTDGSFCYIPPDITCPLLESTYFRINNEESGQFERTLI IADRNSYVSYLEGCTAPQYK
KNQLHAAVVEIIAENSEVKYSTVQNWYSGDENGNGGIYNFVTKRGLCAERNISKISWTQVETGSAITWKYPSCILVGNYSKGE
FYSVALTNHYQQADTGSKMIHIGKNSKSKIISKGISTGKSINSYRGVKI SSNASKARNYSQCDSLLIGNSSEANTFPYIEVH
NRSSIIIEHEASISKINEEKLFFYMQRGISIEEAI SLIVSGFCKEVFTELPLEFALEADKLLGLKLEGTNTEEYI LNGINLNVK

QGEIHAIMGPNSGEII FQEQNLLDYTIEDRANLGIFLAFQYPLEISGVNNIDFLRLAYNSKLFKNQTSTVDPLKFLEIVYPK
LKLVLGDESFLHRKVNEGFSGGEKKNEILQMALLDAKLAILDETDSGLDIDALKDISNAIKSILKISQFKQSIIIIITHYQRI
LNYIQPDYIHVMYKGKIIKTGDASLASELESKGYEWITSE*

>Cre15.g643600 SUFB/NAP1

MLPCGTRTCSGASTSTRTORAAAPGVPIAARSLRRGPTTCRTATVEEAQISDEKRDIQKILDRPYKYGFKTIIESDTFPKGLN
EDVVRAISAKKGEPEWMLLEFRLKAFRKWLTMEEPKWSDNAYPEIDYQDVSYSEPKFKEKLES LDQVDPPELLKTFEKLGIPLN
EQKRLSNVAVDAVFDVSISIATTFKADLLKHGVI FCSI SEALKDYPEMVRKYMGSVVPVGDNYFAALNSAVFSDGSFVFPKGV
KCPMELSTYFRINASETGQFERTLIVAEAGAYVSYLEGCTAPAYDNNQLHAAVVELYCDKDAEIKYSTVQNWYAGDVNGKGGI
YNFVTKRGLCSGSHSKISWTQVETGSSITWKYPSVVLADGNSVGEFYSVALTNNRQQADTGTKMIHVGRNTRSRI VSKGISAG
NSRNCYRGLVQVQPSARGARNFSQCDSMLIGDNAAAANTYPYIQVREPSAVVEHEASTSKISEDQLFYFQQRGIDPEKAVGAI I
SGFCREVFNELPLEFAAEVNELMSLKLEGTVG*

>jgi|Caulen1|89441|g4767.t1

MWNWNNIENLCSTGHFAWKGECDGSGSVIGLDLQELSLQGSITGKLALLDSLVTLRLRGNLFTGQVPDLTELSYKILDLS
NELKGEFFPSAHAFWATPAGVSLLEEIYVQGNKFSGYFVDDDAVSI EESSIRIISASQNHYHGKLPESWFRVLVHLEQMDLSNNR
GLCGEIPSFQNGVSI PEDWLSSQICEIPKSSAQEEATLLLSLKNELAPQSSVEQVLKNWRPEKSEELCESWERVECEEGRVFL
TSTSIGVPCDENLEVSWMATEVLPNTNDQVTQITTSVLVPEVAQAPQEF LNVSQDSEDSMPDAMQSSPAYTETENSAIVPSPE
PELEISGFDGRKLPLTVSTTDSSNGVDSGEDSFEEKKGLSQHAI IWTIVSFLLGILLVILIFLCIHHRNQDPNKDSKESRTW
KMYSQFMSSSLGLIANTTKTKSGSGKEGSSNGSKPKEDVESPTTSSVLASDHDSETNKNVSELGKLASIDSFAPLRMSFS
FSRKNKAVQLQNTTADRVQDGTDMALFGIQTDDSRSLTLDTNDSQONTYDNPLSSNDEPMATNLNDNEVFNDCGVMFKNR
AFSEQTLNSEPSVTHTGLQRNYNDFHSIDE EPIEGAKSNHPPERLKRQPDFVVMGRCLLTQCSRGYFGDLGTLTQTTSDTM
GLRPHSTDATELNLNSKSTTWSITEDLGGRTRTITSEPI MITRELTGSGNDVEDTKSIQIVSGRRIRPFHTVCQSSTGVADSI
GDKSSAEKADLQKLLNKPYPKWFKTI IESETFPGKLSDEVVRAISVKKEEPEWLTDFRLKAYKRWLTMKEPDWSNTYPRINY
QDLSYSEPKQKEKKQSLDEVDPPELLATFDKLGIPLNEQKRLTNVAVDAVFDVSISIATTFREELSKVG VVFCFSFEAVKEYPE
LVRKYLGSVVPVNDNYFAALNSAVFSDGSFVYIPKG VHCPEMELSTYFRINASETGQFERTLIVAEEGSYVSYLEGCTAPAYDT
NQLHAAVVELFACEDSEIKYSTVQNWYAGDEEGKGGIYNFVTKRGLCHGNNSKISWTQVETGSAITWKYPSCVLRGDNVGEF
YSVALTNNAQQADTGTKMVHVGNTRSRI VSKGISAGRSRNAYRGLVKIAPTAKNARNYSQCDSMLLGDNAGANTYPYIQVRD
PSAQVEHEASTSKIGEDQLFYFQQRGIDLEEAVGMIISGFCREVFNELPLEFAAEVNQLMSLKLEGSVG*

>jgi|ChlNC64A_1|29265|estExt_Genewise1.C_10416

MLAKPYKYGFKTI IETE QFPKGLSEDEVVRAISAKKEEPEWMLDFRLKAYRKWLTMEEPNWSDNRYPRIDYQDLSYYSAPKVVD
KKTSLDEVDPPELLATFDKLGIPLNEQKRLANVAVDAVFDVSISIATTFKEELGKAGVIFCSI SEAVKEYPDLVRKYLGSVVPVA
DNYFAALNAAVFDSDGSFVYIPKGV RSPMELSTYFRINASETGQFERTLIVAEAGAYVSYLEGCTAPAYDENQLHAAVVELSAA
KDAEIKYSTVQNWGYAGDAEGRGGIYNFVTKRGICL GERSKISWTQVETGSAITWKYP SVVLKGDHSVGEFYSVALTNNRQQ
ADTGTKMIHVKGKTRSRI VSKGISAGHSVNAYRGLVQVQPTAAGARNYSQCDSMLIGDQAGANTYPYIQAGGGAAAGSGQGAR
VEHEASTSKIGEDQLFYFQLRGIDAEDAVGMIISGFCRDVFNELPLEFAAEVNALMSLKLEGSVG*

>jgi|Tetstr1|442129|TSEL_030283.t1

MAAATASACSSASLRGAPGRPTARFGRSGVLQLPVRRATGRAGRRAAVVRATATPQESAELSEKLAKLENIDSLESEEKREIR
DILKRPYKYGFTSAVESVAIPKGLDEDTVRLISAKKEEPEWLLDFRLRAYRKWLTMKPEWSDNHYPPI NYQDVVYYTEPKQK
KKLNSLDEVDPELLAFAFEKLGIPIGE QKRLTNVAVDAVFDVSISIATTFREELAKAGVIFCSI SEAVKEYPEIVKKHMGSVVPI
ADNYFAALNSAVFSDGSFCYIPKGVTSPMELSTYFRINSEESGQFERTLIVAEEGSYVSYLEGCTAPAFDTNQLHAAVVELSC
AKDAEIKYSTVQNWYAGDEDGKGGIYNFVTKRGLCAGDNSKISWTQVETGSAITWKYP SVVLADGNSIGEFYSVALTNNKQQA
DTGTKMIHVGNTRSRI VSKGISAGESLNCYRGLVQVQPGAKNARNYSQCDSMLIGDKAGANTYPYINVRDPSARVEHEASTS
KIGEDQLFYCQQRGIDMEAAVAMIISGFCQEVFNELPMEFAAEVNQLMSLKLEGSVG*

>jgi|Tetstr1|450808|TSEL_037844.t1

MSDYDIPTLKENIDRDTVDQVLALDVDDKYKYGFETIIESDKSPKGLNEDTVRFISAKKNEPEWMLLEWRLEAFRRWLTLEEPTW
ARVSYPEIDFQDQYYYSAPKSTTGPKSLADVDPPELLRTYEKLGIPLSEQKILAGVQGEDATQGT PFGSNVAVDAVFDVS VVT
TFREELAKHGIIFCSI SEAVREHPVLVKYLGSVVPVTDNFYATLNSAVFTDGSFVYIPKGV RCPMELSTYFRINERN TGQFE
RTLIIAEEGSYVSYLEGCTAPMRDENQLHAAVVELVALDEAEIKYSTVQNWYPGDKDGKGGIYNFVTKRGDCRGKNSKISWTQ
VETGSAITWKYPSCILRGDGRGEFYSAISNGKQQVDSG TKMIHLGKNTSSRIISKGIAAGRSNTYRGQVSSHRRATGARN
FTQCDSLLIGDRCAHTVPYIESRNSSTVFEHEATTSKISDDQMFYCLQRGLSEEEAVALIVNGFVRDVLQHLPMEFMVETQK
LIGISLEGSVG*

>jgi|Tetstr1|447492|TSEL_034873.t1 (NIFS fusion in Nter)

MAQGRAIYLDHNAGAKLRPAVADAMRELIGVVGNASSVHAPGRAARAQIEAARSABAALAGVEAANVFTSGATEANVTALSP
VWVDGRGSRSFDRLLVGACEHPSRLGELTADGRSVLVSMAANNETGVIQPLEEIGQAVHEVGATLHVDVQAAGRLPLDLER
WRADSVLSAHHKIGGPQGAGALVLRNAGLHPVSLLTGGGQERRKRAGTENVLA IAGFGVAARIVLEEIADWDRLRLVRDGLER
DLLHIKRETRVFGQNAQRLPNTSCFAVPGMPAETVLI ALDLEGIAVSSGSACSSGKVSSSHVLAAMGVDRDLARCALRVSLVQ
ETIDRVRSIDVDQYKYGFSTDI ESKAPKGLNEDIIRLDLGEKRANARVDAGLASTGRSNAGRQMDPEWAMPVSKGAEGSSP
RKARKVAEAVFDSVSVGTTFQKELKEAGVIFCSI SEAIQEHPELVKKYLGSVVPITDNFFATLNSAVFSDGSFVYVPPGVRC
PMELSTYFRINAENTGQFERTLI IADKGSYVSYLEGCTAPGRDTHQLHAAVVELVALDEAEIKYSTVQNWYPGDKDGKGGIYN
FVTKRGDCRGDRSKVMWTQVETGSAITWKYPSCILRGDDSGGEFYSAIANNYQQADTGTKMVHLGKNTKSRI VSKGISAGKS
QNTYRGLVSAHRKASNARNFTQCDSLLIGQDCGAHTVPYIDSKNATAVFEHEATTSKISDDQMFYCQQRGLSDEEAVALIVNG
FVKDVIQQLPMEFAVEAQKLISISLEGSVG*

>jgi|Chlpril1|6983|rna-gnl|IITBIO|A3770_13p69830_mRNA_A3770_13p69830
MKAKVVTRGGLGLGKGRKPAACAASCSYEEAASSLRRCIVARSASTLAKGEDGWVLGGGRRRERGALTRVGCSVEGVVEVD
AVELEKKEIKDILAKPYKFGFVSDIESEAI PKGLSEETVRMISKKKNEPEWMLLEFRLKAYRKWLTMEEPDWSNRYPKIDFQD
IVYYSEPKQKEKKQSLDEVDPPELLRTFDKLGIPLGEQKRLANVAVDVFDSDSVSIATTFREDLLKAGVIFCSISEAVHEYPDIV
KKYMGSSVSEGDNYTALNSAVFSDGSFCYIPKGVKSPMELSTYFRINASDTGQFERTLIVAEEGSYVSYLEGCTAPSYDTNQ
LHAADVVELYCAKDAEIKYSTVQNWYAGDETGKGGIYNFVTKRGLCAGENAKISWTQVETGSAITWKYPSVVLKGDNSVGEFYS
VALTNNYQQADTGTKMIHLGKNTRSRIVSKGISAGKSQNCYRGLVQMGGNAENARNYSQCDSMLIGDTAHANTYPYIEVKEPS
SRVEHEASTSKISEDQLFYFQQRGIDAEAEAVGMIISGFCREVFQELPMEFAAEVNQLMSLKLEGSVG*

>jgi|Pico_ML_1|52773|g3433.t1
MATLRRPCAAPPTRSRTRSAPLCHSKVHRRVVEARAATDAPAPATETEETTEIRRLLSQPYKWGFKTDIDAVTI PRGLSEDTV
RLISAKKNEPEWMLDFRLKAFRKWLTMEEPDWSNHYPPIDYQNI IYYSEPKMKEQKQSLDEVDPDLLATFDKLGIPLGEQKRL
ANVAVDVFDSDSVSIATTFREELAKAGVIFCSISEAIKEYPDIVKYNLGSVVPVGDNYFTALNSAVFSDGSFCYIPKGVKSPM
ELSTYFRINAQETGQFERTLIVAEEGSYVSYLEGCTAPAYDTNQLHAADVVELYAAKDAEIKYSTVQNWYAGDEQGRGGIYNFV
TKRGLCDGDHISKISWTQVETGSAITWKYPSVVLKGDNSIGANTYPYIQVKDPSAKVEHEASTSKIGEDQLFYFQQRGIDQEQA
VGMIISGFCREVFNELPMEFSAE*

>jgi|MicpuC3v2|6075|wlab.223738.1
MTSTSLCAAANAARVGAKSRRAGRAATLSSRGITSPGDVAASVSGRARGGKRVGWRRADATPRAEAI SDPAGVVEEEKDAKQE
EVDKIQQMLSKPYKPGFKTDIETDTPKGLSEDTVRIISAKKNEPEWMLLEFRLKAFRKWLTMEEPDWSNRYPDIDYQAVSYY
SAPKVMKEKKSLDEVDPPELLKTFDKLGIPLSEQKRM TNVAVDVFDSDSVSIATTFREDLAKVGVI FCSISEAIREYPELIKQHL
GSVVPVGDNYFSALNAAVFDSDGSFCYIPKGVKSPMELSTYFRINAESSGQFERTLIVAEEGSYVSYLEGCTAPAYDENQLHAA
VVEISAADAEVKYSTVQNWYAGNEDGVGGIYNFVTKRGLCKGDNSKISWTQVETGSSITWKYPSVVLGADNSVGEFFSVALT
NNKQQADTGTKMIHVGNTRSRIVSKGISAGKSRCYRGLVQVSPSAVGARNYSQCDSMLIGDKAGANTYPYISVREPSARVE
HEASTSKIGEDQLFYFQQRGIDPETAVGIIISGFCNEVFNELPLEFAAEVNALMSLKLEGSVG*

>jgi|Pracol1|14672|PRCOL_00000395-RA
QFSSPARVRGERRRRATARHGARRVGGGGEVARSSAGRRRWGGGGEARRARGDGGGVGGGGAPLRGAAAGARARGRCRRAP
PRAAVPDSGAAAAAPAEELTETERLNKLLAKPYKHGWKTEIESETIDKGLTEDTVRLISAKKKEPEWMLLESRLKAFKVVWQ
SMEEPIWSDNYHDYKIDYQDISYYSAPKTKKASMDVDPVLVTFDKLGIPLNEQKRLANVAVDVFDSDSVSIATTFKEDLAK
AGVIFCSISEAVKDYPELVKKYLGTVVPPRDNYFAALNSAVFTDGSFVFIPKGVRCMELSTYFRINEMETGQFERTLIVAE
GSHVSYLEGCTAPAYDENQLHAAIIVELVCEKDAEIKYSTVQNWYAGDAEGRGGIYNFVTKRGLCAGENSKISWTQVETGSAVT
WKYPSCVLKGDNSVGEFYSVALTNNMQQADTGTKMVHIGKNTRSRIVSKGISAGKSQNAYRGLVQVQPTAVGARNYSQCDSML
IGDNAAANTYPYISVRDPASRVEHEASTSKIGEEQLFYFQQRGIDLEAEAVGVIISGFCQDVFRELPLEFAAEVSQLMSLKLEN
AVG*

>jgi|Chlat1|2051|Chrslp17S02525
MAAAMTAAAAVPAERLLFRAPATRPESRTKTELSAKRSSAFGSPLSLPQSVSTRQQPVSSSTRAEVIQEPAPMPVLPKKSASD
DTISKFLEREYKYGFVNIESETIPKGLNEETVRMISAKKEEPPDWMLLEFRLKAFRKWQTMKEPTWSDNKYPKIDFQAISSYSA
PKNKPTLDSDQVDPEMLKYFEKLEIPLGEQKRLANVRQVAVDVFDSDSVSIATTFKKELAEAGVIFCSISEAVKEHPDLIKKY
LGSVVPIDGNYAALNAAVFDSDGSFCYIPKGVTCPMDLSTYFRINASETGQFERTLIVAEEGSYVSYLEGCTAPSYDENQLHA
AVVELYCAANAEIKYSTVQNWYAGDKNGVGGIYNFVTKRGLCAGKSKISWTQVETGSAITWKYPSVVLGADNSIGEFYSVAL
TNNHQQADTGSKMVHVGNTRSRIVSKGISAGQSVNCRGLVSI GPNAAGARNYSQCDSMLIGDNAGANTYPYIQVREPSAKV
EHEASTSKIGEDQLFYMQQRGIDTEQAVAMIIISGFCRDVFNELPLEFAAEVNGLMSLKLEGSVG*

>jgi|Mesovir1|3765|Mesvi232S04464
MAAVHLNATAQLAMAAVCAAPAASQVFRRRGEVLSKGGFRGASLASRPRASGRIVYKTSRKC VATSLPEVVVEESED MKTIS
KFLDREYKFGFVSDIESV SIGKGLTEDTVRLISAKKNEPEWMLDYRIKAYRKWLTMEEPQWSDNKYPKINFQDIIYYSEPKFK
PSVDSDLDQVDPELLRTFDKLGIPLGEQKRLANVKPVAVDVFDSDSVSIATTFKKDLAKAGVIFCSISEAVKEHPDLIKKYLGSV
VPVADNYAALNAAVFDSDGSFCYIPKDTVCPMELSTYFRINASETGQFERTLIVAEENSHVSYLEGCTAPSYDENQLHAAVVE
LYAASGAEIKYSTVQNWYAGDKDGLGGIYNFVTKRGLCAGARSKISWTQVETGSAITWKYPSVVLGADNSVGEFYSVALTNNH
QQADTGTKMVHVGNTRSRIVSKGISAGNSVNCYRGLVQVNPATGARNYSQCDSMLIGDKAGANTYPYIQVRDPSAKVEHEA
STSKIGEDQLFYFQQRGIATEQAVGAIISGFCREVFNELPMEFAAEVNQLMSLKLEGSVG*

>jgi|Klenit1|10541|rna-KFL_003990110
MEAVASPALHTQAAAAGFGQRTGHVQTQVVRTQAALLPRQQRHPKLSHFGSAKSIAGQKVCAESEGFRRDRKGRGHARAAGT
QSPTRERPSVKEAAEKIVGGGLGGGLDEVKLDEVAGQVAKETGSEKAQISEFLKREYKAGFVTDIQSERIPRGLSEETVRTISA
KKNEPQWMLDYRLRAYRQWLKMEEPWSDNKYPRIDYQDIVYYSEPKKKDVKASLDEVDPDLLATFEKLGIPLSEQKRLSNVA
VDVFDSDSVSIATTFRKELAQAGVIFCSISEAVREYPDILKEYLGSVVPVPGDNFFAALNAAVFDSDGSFCYIPKNVECPMELSTY
FRINDSDTGQFERTLIVAEEGSKVSYLEGCTAPSFDTNQLHAADVVELYAHKNSEIKYSTVQNWYAGDEEGVGGIFNFVTKRGL
CEGKNAKISWTQVETGSAITWKYPSVVLKGDNSIGEFYSVALTNNRQQADTGTKMIHVGNTRSRIVSKGISAGNSVNCYRGL
VQVAPTAHGARNHSQCDSMLIGDQAGANTYPYIQVKDPTARVEHEASTSKIGEDQLFYFQQRGIEAEKAVGMIISGFCRDVFN
ELPLEFAAEVNQLMSLKLEGSVG*

>jgi|Mesen1|10608|ME000086S10145
MGRKDRLQRGNWRCEAAEGGVLAPERTSGREKTSSDETIDNFLKRDYKYGFVSNIESVAIPKGLSEETVRIISAKKKEPEWML
EFLNAYRQWLKMTPEKWSDNKYPAIDFQVLVLVLVLVLLCLCLWLAVDQGVYGGGFDEVNAGGKTRVSQCCPSTSYTSDWGW
AGGGGGGGGDS AAGGGIQDVVYYSEPKMKEKKASLDEVADLLDAFEKLGISLSEQKRLANVAVDVFDSDSVSIATTFRKDLAK

AGVIFCSISEAVREYPELIKKYLGVVPGVDNYAALNAAVFSDGSFCYIPKDVTCPELSTYFRINAMETGQFERTLIVAED
NSYVSYLEVKDPSARVEHEASTSKIGEDQLFYFQQRGIDSEKAVGMIISGFCRDVFNELPLEFASEVNQLMSLKLEGSVG*

>jgi|Chabral|342956|rna-CBR_g37549

MAAMAAMAAFAATAGAPKTPAPALVAYSARGVGGEGASSVSAPVALIARKPGALSAREYEGSACSLCSSSSKSSSFLGGSASF
LPNKDKSKSKSSSSSFTFPSSSSSSSSSSSSSSSSIPRAAAQPQSLIPAESSTPSSSSVLDKPEAEDDVIQQFLKKDYKWGFV
TDIESDSIPKGLSEETVRVISAKKKEPQWMLLEFRLNAYRQWLKMRPTWSDNRYPPIDLQNVCIYSEPRQKEKKASLDEVDP
LLRTFEKLGIPLGEQKRLANVAVDAVFDVSISIATTFREELAKAGVIFCSISEAVREYPELIKKYLGRVVPVADNYAALNAAV
FSDGSFCYIPENTISPELSTYFRINASETGQFERTLIVAEGSYVSYLEGCTAPSYDKNQLHAAVVELYCAEGAEIKYSTVQ
NWYAGDEEGKGGIYNFVTKRGLCEGKSGKISWTQVETGSAITWKYPSVVLKGDNSVGEFYSVALTNNRQQADTGTKMIHVGN
TRSRIVSKGISAGKSVNRYGLVQVGPHTAHNARNYSQCDSMLIGDTAGANTYPYIQVKDPTARVEHEASTSKIGEDQLFYFQQ
RGIEMEKAVGMIISGFCRDVFNELPLEFAAEVNQLMSLKLEGSVG*

>NP_050730.1 putative ABC transporter (chloroplast) [Guillardia theta]
MSDDLKSRSLRELVSQPYKYGFHTDIENEEFPKGLDEDIKEISKLKCEPSYMLDFRLKSYILWKKMSLP
EWACLTLYLNINYQDIVYYSAPKNSTKLDSELDVDDKILETFDKLGIPLNEQKRLANVAVDAIFDSVSVGT
TFKQELSNVGVLFCEPLSEATNKYSTLVEKYLGSVPIGDNFYAALNSAVFSEGSFCYIPPNVKCPLELST
YFRINNENSQGQFERTLIADFNYSVSYLEGCTAPMYDKNQLHAAVVELIALENAEIRYSTVQNWYSGDTN
GKGGIYNFVTKRGLCAGKSSKISWTQVETGSAITWKYPSVILVGEDSVGEFYSVALTNNYQQADTGTKMI
HVGGRGSKSRIISKGISAGYSKNTYRGQVKININALGSINNSQCDSMLIGPYSQANTYPYIQVSNAMSRVE
HEASTSKIEEQFLFYFLQRGISVEQAISLLISGFCRDVFNELPLEFAEADKLLSVKLEGTVG

>YP_277324.1 putative ABC transporter (chloroplast) [Emiliana huxleyi]
MEKQTNDSKVLLEIVVTPYKYGFKTEIENENFPKGLNQEIVDQISRKDEPLFMSQFRAQAFSQWQKMP
SPDWAYLGIPIPIDYQNIQYYSVPKTKKKLGLSDEVDPPELLKTFDKLGVSLNEQKRLTNVAVDAVFDVSIS
GSTFKVELQRAVIFCSFAEAVERYPELIKKYLGSVPIGDNFFAALNSAVFSDGSFCYVPKDVKCPMEL
STYFRINNEEAGQFERTLIVVEDRGEVGYLEGCTAPQFSTNQLHAAVVELIAFEEAVIKYSTVQNWYAGD
ENGIGGVYNFVTKRGLCSGRKARIAWTQVETGSAITWKYPSVILAGDSSIGEFYSVALTNNKQEADTGK
MIHLGNSKSRISKGISTGNSKNSYRGLVKVSPRAINTQNYSCDSLLIGKDSSANTFPYLEIQNSETK
IEHEASTSKIAEEQLFYLLQRGISMEEAVALMVNGFCKEVFNELPFEEFASEADRLNLKLEGSVG

>XP_001712812.1 SufB involved in iron sulfur protein biogenesis (nucleomorph)
[Bigelowiella natans]

MITKTNYKAKLTITYTIKKNCKIKKISFRTFQQTESYKYKLKSFSECETIRKSLNSNTLKSISLKNLEPLWI
IKFRDLALKSLYNRDPKWFSEFNYSYFNIDNNSIYYSALIKKDTALVSSNNETSIISSSSKKKTAIDLVD
SISIDTSYQKELGKHGVIFTNLNSLKNYSIDLISNFMGSVVSLRDNYTSLNSAVFSEGSFYIYPANLVC
PLDLSTYFRINGNVIGQFERTLIVTEKNASVTYLEGCTAPYFTENQLHAAVVELCAEEGSEIKYSTVQNW
YPGDGSMGKGGIFNFVTKRGLCNGSRKISWTQVETGSSLTWKYPSVILKGDNSVGEFYSIALTKLRQQAD
TGSKMIHIGRKTFSKISKGISLGNKNCYRGLVLHNKSLKSKNTTQCDSIIFGDSSSANTYPYVTVRD
SSAWLEHEASARIQDDQVFYLLQQRGLGFEEAINLIVGGFCKEVIEELPFEEFLTEVNELLNSSMIKKIG

>YP_874419.1 iron-sulfur cluster formation ABC transporter (chloroplast)
[Phaeodactylum tricornutum]

MVNKSNKVLNKNITNLVNQTYQYGFSTTIEKDIIIEKGLNEKTIHLISQKKKETKFLNFRKLKAYKKWKQM
PEPEWAYIKFPQIDYQDVIIYYSAPKSQKKLKDSEVDPELLKTFEKLGISLTEQKRLANVAVDVFDSVS
IGTTFKEELNKGVSIFSSISEAVTEYPELIEKYLGSVPIGDNFYAALNSAVFTDGSFCYIPEDTICPLD
LSTYFRINDEKSGQFERTLIIEKNSQVNYLEGCTAPQYDSNQLHAAVVELIALENANIKYSTVQNWYAG
NEFGKGGVYNFVTKRGLCAGSSSKISWTQVETGSSITWKYPSVILGDNSSQGEFYSVALTNNYQQADTG
KMMHIGRNTSRRIISKGISAGKSKNSYRGLVNVTSKALGARNYSQCDSLLIGDLSNANTFPFISVQNSTT
KIEHEASTSKIGEEQIFYFLQRGICLEKAVELMISGFCREIFTELPLEFAAEADKLLTLKLEGSVG

>YP_003289159.1 Chloroplast conserved hypothetical protein (chloroplast) [Ectocarpus
siliculosus]

MMTETNNATLQQLVNQPYKYGFSTDIETLPPGLNEKIVRNIIKKNNEPDYMFMHFRIGALKKWRQMKSP
TWAKLDFLEMDYQKIVYYSAPKTKKTLANLDEVDPKIDTFDKLGISLNEQKRLANVAVDAVFDVSISIAT
TFKEELEKYGVIFCPISEAIKNYPHLVKQFLGSVVPFGDNFFAALNSAVFTDGSFCYVPKNLRCPLELST
YFRINNKEGSGQFERTLIIEEGSYISYLEGCTAPQYDTNQLHAAIVELIALKNAEIKYSTVQNWYAGDKN
GIGGIYNFVTKRGLCIGENSKISWTQVETGSAITWKYPSVILIGNESQGEFYSVALTNMQQADTGTKMI
HVGSNTKSQIISKGISSGYSKNSYRGLVKIGTKALDSRNFSQCDSLLFGADAQANTFPYIQVQNSTSKIE
HEASTSKVGEEQFYLLQRGINAEDAVTLILTGFCKIVFDELPIEFASEVEHLLRIKLEGSVG

>YP_008520066.1 cysteine desulfurase activator subunit sufB (chloroplast)
[Nannochloropsis oceanica]

MTTFLDHKQIDLVQRQGRFTERCVKPLRDNTYKYGFATVIESDEIKSGLDLNTIKLISAKKDESGFIIDF
RSRAFKKQSIKDPAAWELNYPIDYQAIKYSSPKVQEKLSLDEVDPKIDTFERLGIPLNEQKRLAN
VAVDAVFDVSISITTFKRELLKLGIFCSISEAINTYPELVKKYLGSVVPITDNYFVALNSAVFTDGSFA
YIPKNLQSPIELSTYFRINNEQSGQFERTLIIEAGSMISYLEGCTAPKYDSNQLHAAVVELVSHGAHI
RYSTVQNWYAGDKKGGVYNFVTKRGLCLGNESKISWTQVETGSAITWKYPSCLLMGNASQGEFYSVAL

TTNYQQADTGTKMIHIGKNTRSRILSKGISGGHNSKNTYRGLIQFGRKALQAVSYSQCDSLLLLAHQSETNT
FPYINVRNNSARVEHEASISRIGEEQIFYFHQRGIPEEDAIAKLIVNGFCQEVFRKLPMFALEAEKLLDL
KVEKSLT

>AIM52780.1 iron-sulfur cluster formation ABC transporter (chloroplast) [Ochromonas
sp. CCMP1393]

MKFLFFNNKKFMENSKNQTNFEINQVINQPYKYGFTEVETEFKSPGLNTEVIKLIKKEEPFFLTEFR
LKAYKKWLKMKFPQWANLTINPIDFNAITYYSIPKKKKQLNSLDEVDPILRTFEKLGISLEEQKTLNSV
AVDAVFDVSISIGTTYKEKLAKEGVIFCSISEAIIQYYPKLVQKFLGTVVSPGDNFFASLNSAVFTDGSFCY
IPQNVTCPLELSTYFRINDKDSGQFERTLIIAEKNSSVNYLEGCTAPQYSENQLHAAIVELIAFENANIK
YSTVQNWYSGDEKGIGGIYNFVTKRGLCIGKNSNISWTQVETGSSITWKYPSCITLIGKSSGEFYISALT
NNYQQADTGTKMIHLGEETTSKIIISKGISCGSSKNSYRGLVKIASNANCSRNFSCDSFLLGSSSLACTY
PYLDIWNSTSIIVEHEATISKVSEEQLFYLTQRGISVEQAVGLLINGFCKEVVTMLPMEFATEADKLLSLK
LEGTVG

>ACS36807.1 ABC transporter (chloroplast) [Aureococcus anophagefferens]
MSDVKNKGVSTIDQIVSQPYKYGFTEIEKERIPVGLDENTVRLLSSKKGEPEFMCQSRLQALKIWKNL
EPNWAEHLHPEINYQDITYYSAPKKKKKLDNLSEVDPPELLKTFDKLGISLNEQKRLSNVAVDAVFDVS
IATTFKKELAKAGVIFCPLSEAVQQYPELIKAYLGKVVPAAGDNYFAALNCAVFDGSDGFCYVKNVCP
MELSTYFRINDEESGQFERTLIIAEQNSVVSYLEGCTAPQRDNQLHAAIVELVALENAEIKYSTVQNWY
AGNEKGGGIYNFVTKRGLCSGKNSKISWTQIETGSAITWKYPSCILVGDSSVGEFYVSALTNNYQQAD
TGTMVHIGNNTKSTIIISKGISAGNSKNSYRGLVKISPAKAGARNFSCDSLLIGDKSNANTFPYIEVKN
STKVEHEASVSKIGEEQLFYFSQRGIEVEQAISLMISGFCHDVFLKLPMEFSVEADKLLSLKLEGSIG

>jgi|Symmic1|40445|rna28602

MAGPGIFSFLDLPPPEEPAPQNESRDSSDDSPAIRDAPTEKAANSFFPTISTAGLLHSDSDSDGGLYLQHPGQVPQDV
LAVACEPPESFPGLPTHVVLKVLISIEGLPLRKEERNVPGSGKGLGYFLEGKDTIDRSGASAAILPEPQSRPPGRH
RCAEPRIFAELSDDHWQAVWDRPGEVRVKLKDQDSVVVGIVLQGVEVVAVTAELDLTGTMRRFFQPPQQLYRPAFSG
VALGLIEGARAKLALIELYPGSQLPGRPVAEPEKPSMVRPPKDVPISDCRFCDDGYGRRSCDGGGHGVLVVCSTCDG
MPALPCPVCRGRGTLHDNLEGIGGPRSRRAQDGAVLSTVKGRRQCNCWGAPLTCKECFGAAALRCNVCNGAGWSAGS
VELKQKQEEAVALRDGQNAELGEVLDQEYSAGFYTDIESESLPKGLNEDIVRKIWEIKEEPEWMLFEFLNAFRKWSMKMPEWAQ
LQMEETDFQDIVYYSRPKTKKKKQSLDEVDPPELLDTFEKLGIPLTEQKRLTNVAVDAVFDSESIGTTFQKELEE
AGVIFCSINEAMKEHPPELVKKYLGSVVPVADNYFAALNSAVFSDGSDGFCYIPKGTTCPEIISTYFRINNKE
SGQFERTLLIADSSVVSYLEGCTAPMFDSKQLHAAVELHAAESAEIKYSTVQNWYAGDKNGKGGILNLVTKRGM
CEGKSKISWTQVETGSAVTWKYPSCILKGDDSIGEFYSVALTNNYQQAVSSVILTEARMVTVLSCLVPGSDSLAT
DPAMLSFDDGLLTQVYEAGAEDDEEAESAEAEYENRGASRRREVEVASPAQVRRREPEPDYPSPRELQLPHRY
PVWSAGAASSARLKSSSSAPALPRVRSARQGEESLLSPIDGPGHPAAQRQAHHWKPVRRKLPQQGHLRPLSHLR
RLDTPEDPKKWMPPKVPVGGGAWGGILQSQHQRLAGVLAKQHQEMKKMQQLEVLQQLNRLGSGYVGGDARVDGA
VHSHYVMEVMAKPVGRKAPSWKPLSPIRRSALPVLPEDEEARASLPGSRGQSSEAFVEEAEEVLADPALEAQQVSE
DEQGETDAFEQKLEAEGAPETEAEGVPAAEACATGQGSLEGEIEVSVQVTLPSDEDEEASLQGLQPSVEAQLASG
DVLVPEEVGSHTDGDAVEQLRKEVWPPALMLCICQLA*

SUFC orthologs

>jgi|Galsull1|3981|XM_005704982.1 **SufBC fusion**

MNQSLDATTCLKLEQIINKSYKYGFSTKIQNEDFPKGINEEIVTLISKKKNEPKYMLDFRLKAYKKWKQMYVPKWQY
LKFEQINQONILYYSVPKQKQLQSLNEVDPELLKTFEKLGIPLNEQKRLTNVAVDAVFDVSISIGTTFKQELADQGI
IFCSISEAIQKYP SLIQYLGSVVPIGDNYFAALNSAVFTDGSFCYIPDPTICPLELSTYFRINNEESGQFERTLII
ADRNSVVSYLEGCTAPQYKKNQLHAAVVEIIALENSEVKYSTVQNWYSGDENGNGGIYNFVTKRGLCAERN
SKISWTQVETGSAITWKYPSCILVGNYSKGEFYVSALTNNYQQADTGSKMIHIGKNSKSKIIISKGISTGKSIN
SYRGVKV KISSNASKARNYSQCDSLLIGNSSEANTFPYIEVHNRSSIIIEHEASISKINEEKLFFYFMQRGISIE
EASISLIVSGFCKEVFTELPLEFALEADKLLGLKLEGTNTEEYILNGINLVNKKQGEIHAIMGPNNGSGEIIIFQ
EQNLDDYTIEDRANLGIFFLAFQYPLEISGVNNIDFLRLAYNSKLKFNQSTVDPLKFLEIVYPKLKLVGLDES
FLHRKVNEGFSGGEKKNEILQMALLDKAILDETDSGLDIDALDKDISNAIKSILKISQFKQSIIITHYQRI
LNYIQPDYIHYVMYKGIKIKTGDAASLASELESKYEWITSE*

>AIU44589.1 sulfate ABC transporter protein (chloroplast) [Cyanophora paradoxa]

MSTEKTKILEVKNLKAQVDGTEILKGVNLTIIYSGEIHAIMGPNNGSGKSTFSKILAGHPAYQVTDGEIIFK
NKNLLELEPEERARAGVFLAFQYPIEITAGVSNIDFLRLAYNTRRKEEGLTELDPLTFYSIVKEKLN
NVKMDPHFLNRNVNEGFSGGEKKRNEILQMALLNPSLAILDETDSGLDIDALRIVAEQVNLNKN
DNSIILITHYQRLLDYIVPDYIHYVMQNGRILKTGGAELAKELEIKGYDWLNELEMIKK

>YP_009051123.1 sulfate ABC transporter protein (chloroplast) [Galdieria sulphuraria]

MTHNSLLQIKNLHVKLANTEEYILNGINLVNKKQGEIHAIMGPNNGSGKSTLSKVIAGHSYKIVKGEIIFQ
EQNLDDYTIEDRANLGIFFLAFQYPLEISGVNNIDFLRLAYNSKLKFNQSTVDPLKFLEIVYPK
LKLVLDESFLHRKVNEGFSGGEKKNEILQMALLDKAILDETDSGLDIDALDKDISNAIKSILKISQFKQSIIITHYQRI
LNYIQPDYIHYVMYKGIKIKTGDAASLASELESKYEWITSE

>jgi|Chlre5_6|16639|Cre07.g339700.t1.2

MAHQILRSSRIASSTRVAQRSSRVHAVSVRCAAAPADVMLEVKDLTASIAGTHTPIILKGVNLTIRNGEVHAIMGKNGSGKSTL
SKVLVGHDPDYEVTTGGTAVFKGKNLFEMEPEERSHAGLFLSFQTPIEVPGVSNVDFLRMACNSRRKALGHPELDPLEFYAYIMP
KLEMLNMDPTFLNRNVNEGFSGGEKKRNEILQLAVLEADMAILDEIDSGLDIDALRDVAKAVNQLRSDDTGVLMTVHYKRLLD
YIKPDFVHIMQAGEIVKTGDMSLVDQLEAGGYATL*

>jgi|Caulen1|97221|g3371.t1
MKSMRLVQGSSLRDSGTPVRNVIPSLKRRIRCSAASAKESTTELFKVEGLKAEIAASGESVLNGVNLTVNYGEIHALMGKN
GSGKSTLAKVLVGHDPDYNVTEGSAIFKGQNLFELEPEERSHGLFLSFQSPPIAIPGVNNVDFLRACNARRKAKGETELDPLE
FYAYLTPKLEALKMDPTFLNRNVNEGFSGGERKRNEILQLAVLESMAILDEIDSGLDIDALRDVANAVNGLRTEKTGILLVT
HYKRLLEYLRPDYVHVMQDGGKITKTGDMQLADQLEEKGYAPLKN*

>jgi|ChlNC64A_1|37406|estExt_GenewiselPlus.C_290103
MGKNGSGKSTLSKVLVGHDPDYEVTTGGSAAAYKGQDLFELEPERRSHMGLFLSFQSPVEIPGVSNIDFLRIATNARRSANGEPEL
DPLEFYAHVMPKLESINMDPAFLNRNVNEGFSGGEKKRNEILQLACLEADMAILDEIDSGLDIDALRDVSKAVNGLKAQRPDM
GILMTVHYKRLLDYIRPDQVHIMQDGRIVTTGDMGLVDKLELEGYSVLRS*

>jgi|Tetstr1|432597|TSEL_021967.t1
MNTCKATRGGQALRPAPSSAGRLQARRLPRLSRVAPRRRAVRVLAAGDVLVLEVKDLRANIAGTDQAILKGVNLTVRAGEIHAI
MGKNGSGKSTLSKVLVGHDPDIYVTTGGSVMYKGEDLFALEPEERSHKGFLFLSFQSPIEIPGVSNVDFLRSCNSRRKAQKDE
DPLEFYSYIMPKLEALNMDPTFLNRNVNEGFSGGEKKRNEILQMSVLESEMSILDEIDSGLDVDALRDVSKAVNGLRTEDNAV
LMVTHYKRLLDYVVPQVHVMRDGEI IKTGGIEIVDTLEAEGYAALSTA*

>jgi|Chlpri1|7803|rna-gnl|IITBIO|A3770_16p78030_mRNA_A3770_16p78030
MASTARGMRPSMARTPRATVGTSTRRGIGIGIPAFGMKRGVTAPMQEAGGGCGSCCSRCSEERRRRRCRVAATTEEAGGANG
CVLEVSGLEASIDDKSILKGLNLRINRGEVHAIMGQNGCGKSTLSKVLVGHDPDYKVDGGSVTYKGGNLELEPEERSHAGLFL
SFQYPVEVPGVNNVDFLRACNARRAALGETELDPLEFYGHVTPILQELDIDPKFLSRNVNEGFSGGERKRNEVLQLAVMDS
LAIMDEVDSGLDIDALKNVADALNKLRRKKKDMSVILITHYQRLNLYIKPDYVHIMKDGRCIKTGGAEALALELESEGYASF
FDGV*

>jgi|MicpuC3v2|90|wlab.202455.1
MSAATISSRTLTPASLTRARGASSRAARAASRAVGPVVRAAKGDVLVLEVKDLTAKVAETGETILNGVTLTIKEGETHAIMG
KNGSGKSTFTTKVLVGHPSYEVTTGGTAVFKGKDLLAMPEDRAREGLFLSFQSPPIEVPGVSNVDFLRMMCNERRKARGEELDP
LEFYGFLTPKLDQLNMDPSFLARNVNEGFSGGEKKRNEILQLAVMESEMSILDEIDSGLDVDALRDVAAAVNVLKEGDKGLL
ITHYQRLLDIAHPDFVHVMKKGKITQTGDKSIASKLEEGLFASLA*

>jgi|Pracol|20197|PRCOL_00003471-RA
MACLSARGLCAEVEATGKKILKGVDLTVREGEVHAIMGKNGCGKSTLSKVLVGHDPAYKVVTAGTATYRGDDLFEMEPEERSHAG
LFLSFQAPIEIPGVSNVDFLRMMYNARRKALGMPELEPLEFYGYLTPKLEQLNMDPSFLSRNVNEGFSGGEKKRNEILQLAVL
EELSILDEIDSGLDVDALRDVAEAVNGLKSDENAVVMITHYRRLDYIEPDFVHIMERGKIKQTGGIELAEVLEEKGYAGLS
S*

>jgi|Chlat1|3274|Chrsp22S03442
MSAAAALAGGALSWSAAGRAESSSPSASPSLLRPATGLRASRLRCRAVPSSRSISRPAVSAAAGDVILEVKGLCAKVVDSELD
ILRGVDLTVREGEIHAVMGKNGSGKSTLSKVLVGHDPQYEVTTGGTAMYGQDLFELEPEERSHNGFLFLSFQSPVEIPGVSNVDF
LRMANARRKALGKPEMPDFEFYGLWTPKLEQLNMDISYLQNRNVNEGFSGGEKKRNEILQLAVLEADLSILDEIDSGLDVDAL
RDVATAVNALRSPDRASILITHYQRLLEYIQPDFVHIMESGRIVRTGDKSLAKQLESGGYAAIK*

>jgi|Klenit1|1526|rna-KFL_000240090
MATSAAAASIHNLTCGTSRQSRYETNARCPLLPQSPLHKADVHSHGLKCGASRKAVSRVGKTVVFRQPTCVKAAVSTSEAP
TSTSTEPGKVLLEVEGLEAEVADGGRQILKGLNITIREGEVHAIMGKNGSGKSTLSKVLVGHDPDYDVTGGSAGFKGKDLFELE
PEERSHAGLFLSFQSPVEVPGVSNVDFLRMAHNARRQAQGLKELDPLEFYAFLTPKLDQLKMDASFLSRNVNEGFSGGEKKRN
EILQLAVLEADLAILDEIDSGLDIDALRDVAEAVNGLRKSNNAVLLITHYQRLLDYITPDFVHIMQSGRIVKTGGKELAKELE
QGGYGALGPM*

>jgi|Mesen1|8851|ME000053S08269
MALCISNAAIVPVSSEL SHNKSSSRTSQKRQALASCRVLRGANS AHLKLNETSLRSDSLRTARKAPSVGRCRRGTSVTSAA
AEAEAAAFAPVAEGDII LEVKGLCACIAESGKEILKGVDLVIRAGEVHAIMGKNGSGKSTLSKVLVGHDPDYDVTGGTAVYKGRD
LFELEPEERSHAGLFLSFQSPVEVPGVSNVDFLRMACNARRAALGLPELDPLEVLEAEMAILDEIDSGLDVDALRDVSAAVNG
LRQPTRAVLMITHYQRLLDYIKPQFVHIMESGKIVKTGDSSLALELEAGGYAAVRG*

>jgi|Chabra1|335080|rna-CBR_g26048
MATALRAAAGAATAAAASCCCAPSRLHSEVTSTASCRSRNGPQSWVGAETACAPRALPLAQQRGVCAPQPQQQLVAPRRGKAQ
CSAPLGGGLLKLDRPHSMAVAAMASAAAAASLADQSTVAEGGEEGGRKVLLEVKGLTAAIADGGAEILKGVHAIMGKNGSG
KSTLSKVLVGHDPDYEVTTGSAHYKGENLFEMEPEERSHAGLFLSFQSPVEIPGVSNVDFLRMACNARRAARGLPMDPLEFYA
FLTPKLES LKMDISYLQNRNVNEGFSGGEKKRNEILQLAVLEADMAILDEIDSGLDIDALRDVANAVKGLQGPQNAVLMITHYQ
RLLDYIEPDFVHIMEAGKIVRTGDKTLANELEVGGYAAALKN*

>SUFC G. theta, **reconstituted sequence**

MKKKILEVTNLHAAVNEIKIVKGLNLVVNAGEIHAIMGKNGSGKSTFAKIIAGHPDYTTITNGDITYQHTSILELTPEDRAKRG
IFLSFYQYPIEIPGVTNADFLRLACNARRIYQGLPEMEPLEFFEYINSKLPVLVDLKPSFLTRDVNEGFSGGEKKRNEILQMSIL
DTKLAVLDETDGLDIDALRTVANGIKSLANDGNAILITHYQRLLDYIKPDFVHVMQEGKIIKTGSASIALDLEKYGYDWLK
NELK-

>jgi|Emihul|443440|estExtDG_fgenes_h_newKgs_kg.C_2400030
MLLVASSSGAAVALGSRPAPSSAARRGGALQMEAGETVLDVSDLCAAVGETPILKGVNLKVRKGEVHAIMGPNNGSGKSTLSK
VIVGHPAYEVTGGSALFEGEDLLELEPEERAQQGVFLAFQYPVEIPGVANNDLRLAVNKRRAQGLDEYDPIEFFGVLESEKM
SQVSMSPDFLGRDVNSGFSGGEKKRNEILQLALLEPSLAILDDETDGLDIDALKTVAGVNAYRGEENGIVLITHYQRLLDYI
KPDFVHVMSDGEIVRSGGPPELALALESDGYAFLDAGARSDA*

>SUFC *Bigelowiella natans*, **reconstituted sequence**
IYKDCRKLNLTVNPGIEHVLNMGNGSGKSTLSKTIIVGHPSEYVIDGSITYKNIDITQIDASLRALNGIFMCFQSPVEIQGVKN
LDLFLRKISNSRRKFLSQPIFDPLDFYSFISDKIESVGLDNTFLTRNVNEGFSGGEQKRNEMLQLITVQADFCIFDEIDSGLDI
DSIKNLVDIISTLRNRGVGMIIITHYNKLIDLVRPDYVHILKNGKIIKSGDMELSKIIEREGFDAIN*

>sulfate ABC transporter protein (chloroplast) [Phaeodactylum tricornutum]
MTLNSPLLEIKNLQVSINENQILKNLNLTKNGEIHAIMGPNNGSGKSTFSKVLAGHPAYSILGGDILFKGSSILDLEPEERSH
LGIFLAFQYPVEIPGVSNEDFLRLAYNSKQKFYNKVEVDPIEFLLTIITQKLQVLNMSARFLGRNVNEGFSGGEKKRNEILQMI
LLDSELSILDETDGLDIDALKIISNGINNFMGPEKSIILITHYQRLLDYVNPVHVMQNGKIIKTGSALAKELEAKGYEW
LET

>D1J705, ycf16, *Ectocarpus siliculosus* (chloroplast)
MTNKTIPLLLEIRNLHANINNNKILRGVNLISFEGEIHAIMGTNGSGKSTLSKVIAGHPSYDVIEGEILFQGNICDLDPDLRS
HLGLFLAFQYPVEIAGVDNQEFLEAIYNAKQVSMNLPKANPLFFYELLREKLKMNLDSEFSIRNVNEGFSGGEKKRNEILQF
ALLDSILGILDETDGLDIDALKSISKTINILMENKNKSIILITHYKRLLEYIKPDFIHMVQDQGVKTDATLANLLEEKGY
DWLKPISN

>AGI99421, FeS cluster assembly protein (chloroplast) [Nannochloropsis oceanica]
MLEIKNLTAATSIKEGKRNVASILKGVNLNLKAGEIQIVMGNSNGSGKSTLAKLLAGHPAYNMLHGSIRFRNEKLIRCLPETRA
RLGLFLGFQYPMEVGVSNQDFLRLAYHSRYKKKDDILNTSLRFANRILNYVKILDMDPDFLPRPLNQGFSGGEKKKNEILQM
AITGCELAVALDEIDSGLDVDALKTSLKTIILYQDLSLNSRFLQTTKSLILITHYQRLLEYIRPDKIQVMKEGKIVCSGKSNL
GNIVDRKGYDWLSGTGLGSLIENINKKLQTLKNSVF

>AIM52779.1 iron-sulfur cluster formation ABC transporter ATP-binding subunit
(chloroplast) [Ochromonas sp. CCMP1393]
MNKTLLKLENLEASIQNTQIIKDLNLEVKENEIHVIMGPNNGSGKSTLSKILAGHPSEYIKKGSVNFCEKN
LLDMAPEERSHEGLFLAFQYPIEISGVTNYDFLRRIAYNEKRKYLQEEELDPIEFMQLTQRLNLKLMKNE
FLNRNLNEGFSGGEKKRNEILQMLLLDPKLVILDEIDSGLDVDAIKIICEGINNTLSQNSSLIVITHYPR
ILNLYKPTYVHIMIDGKIIKTGNLELVESLEKDGQYQAVI

>jgi|Vitbras1|19345|Vbra_4547.t1
MTSLVACAFSLIISLVALPRPCTAFRMPSPVHRTPRTRVRSRQPSLLRMSASEDASDTIMRIDNVHAEVGESEEEKVKILKLDG
FSLEVKKGEVHAIMGPNNGSGKSTLSKVIITGHPDYRVIEGSITFKGEDVLDMDVDERALAGIFLAFQYPVEIAGVSNSDFLRAA
VNIRRKREGLDDDLDAIEFSMKMFEEELDKLGLDPSFLDRGINAGFSGGEKKRNEVLQMVMLNPDVLILDETDGLDVIDSLKTTA
EAVNNFKASHPKDTFLIITHYQRLDLIKADRVSVMKKGRVVLTTGGPDATTIEQWGYGDWLEKEAAGGREAVLA

>jgi|Symmic1|43348|rna30659
MASMSTSVARQLQSMHNFYGPALKSWGLANLVRVRRQPLLYAVATAVQQRLESSESQSSSTKIAWCSSVVRAPPPVELWVPERQGSF
DACIDEEMLAHFGKNLIPRLMLCYGACGPRPSLNTQADSSRWLQRGIDLGMSTAVEDLTWICLLQHTSRPGPCQEVIRKYSC
GRELPWFVFRVQDGPWSKSAKYMQRFSNLEKFAQPASGNHWEKLGAVIDFVEHRLSAAAPQVPGTVNEMLRLLDTFAQRREHW
LKVAGGAKAEVALECGTFIGSRITYAGVDPVQACIARHLVDLAGASFVLSGGCHWTGEVRDIVPKIVELFGATSVGMVFM DYK
EDTGAIGINSRLLADNVALPGAPLLWNLAFFSPSWELTAFAMTEFHEPNTEDWMALGRITCAAEPASVDLQKETCLERTRTVD
RAARSKKWCILTQDVLPTTMMKPLVAEIGKVARAMRQLAGRLALAAVALASEDYMARVDPSSLHTWTRTVRVAEDAFLLLEMHQG
NLKALLNGEAEVDLPQEQQEQLLSAFPAAEALTVEARVPQIQSQRWAAKILRHAVGARVDPDCAVNTENHLEKEILAVFRRLRN
PPIGSVLTFGGLYGGVHGEVDPDPAVELMLELQQRGAGKITDSRAGVAALRTLNI PDPLRWADLQLPAEFAGELDLLRLD
GLRHGVSCVLRQVRPRLVALLVLSQVPPPFQYFPLGVPGYNSPPALMSCSLSGAIEVLAKHDLFLIRLTGPYALFVRQSEWP
EALPINEFDCYRRASVWGLKDIPLSFVREWLREEVDQVLPRLWRNLTHLHSDGGPFLLVLFCKSCRVEWIRKQGGAMDAMEVA
TMAAPLVSSSTSSVHPRWRRTTSLRSPATSFQGRSADSASSTWHTGWLASTAIVGLAASRPTRRRRLQSRQLRQTRRAGDSP
VVLEVEVEAKSTDEDEKQILKGLNLITRAGEVHAVMGPNNGSGKSTLAKVLIGDSAYEVTNGSAKLGETDLLEMEPHERALEG
LFLAFQSPPAVGGVSNLDFLRAAYNAQSIFAAMRMSSGIRRRKVPGESEADEGLSNSGALETLRKFDVYNKVHEDYMQKRQLGG
AVTLVTCAILAVLVYCEVCEFFSVEVLHSITVDTNIDRKLPISLDITFPHLRCEVSVDTVDSAGDTQVDAHGGGLDMHNLDA
GKISTGDPVASKGDCWPCLEAEDAKHKCCNSCQELKDAYNDKELPYFHVLDTAMQCKSSIGCRKIGKVVNVKVSNGNIHVALGK
SVRRDGKLVHEFNIEDIGDGFNTSHHHSITFGEYVYGLQSPLEGIRKIAGAGSWMYHYVVKLVPTVYISRWGTTTYYTQYQSV
TDSARNVQVREGELSGLPVFLVYDFSPFLMQQTEQVQKPSYVFTSMCAIVGGAFSVATLVEMALSSAREEPELDVIIFYGLV
TQKLQDLKINPDFLNRNVNEGFSGGERKRNEMLQMAVLQPKLAILDEIDSGLDIDALKDVAAEIRSVREQDPDRAMLVVTHFE
RFLRYVEADHVHVMYQGRILKSGGKELADKLDEEGYDWVLKEAK*

SUFD orthologs

>SUFD galderia sulphuraria 074W

MKSEEMCWREICFVPCLSQGNEKLLVQRETRFLSSFPYGYSLHPRKCTEFNKRKSPLVTRKRPTTLTTF
MSSSTIQESAKQTEDKTLWLSILDCRESTANHEDTAVLNSLHREAEEELKTRLRIPGRKDELWRFNTLNRN
LFLTRFQTSSHRVDNQLLESLEFQDVQSQRVVLVDGVDFPFLSNISNIPDRIFIGSVTELDVERQRWIYE
LTKKGETGIQETNIFGAINLASFRDITIVWIPQDVSLQDYIHLCCYSSNCLNSSFYGISHPRIVVVVEGG
SHVKILQHHFGASGGYLDNFATSSISIGDSATLEYIYVNECPKDALWLGLSLNAQLEQYCTFQFRFLSSGSK
VGRLTMTTDLVGRESNTSLHGLAIAANEERTLDFHSFVDHRVPSCHSEQLQRNLVADRSQCVCFRGLVKIRR
EAPHTCAHQLCRSMLLSNRAQTKTMPMLEIANDEVECSHGATVSEIEEDELFFYLLSRGIPewDARVLLVR
SFSLELLKEFPFPIVARMESLIAQVNNNSKEEKVIEADTVEEGQWDWTT

>AIU44588.1 putative ABC transporter (chloroplast) [Cyanophora paradoxa]

MVNTQSPKNSGLENLVNQPYKYGFTTDIEIETISKGLTEDTVRLISAKKNEPEFMLEFRLQAFRKWLEMK
NEPEWAHLNYPKINYQDMVYYSAPKQKQKLLQSLDEVDPTLLETFEKLGIPLTEQKRLANVAVD AIFDSVS
VATTFKEELAKEGVIFCPISEAIIQKYPDLIKKYLGSVVSTSDNYFSCLNAAVSDGSFCYIPKNVRCPL
LSTYFRINNGESGQFERTLIVADEGSYVSYLEGCTAPQFDTNQLHAAVVELVALKDAEIKYSTVQNWYAG
DDNGKGGIYNFVTKRGLCAGENSKI SWTQVETGSAITWKYPSCVLLGDNSIGEFYSVALTNRYQQADTGT
KMIHIGKNTRSRIISKGISAGHSQNSYRGLVKISPKALGARNYSQCDSLLIGDNSQANTFPHLQIKNPTA
KVEHEASTSKIGEEQIFYFLQRGINAEEAISLIISGFCREVFNNLPMFALEADKLLGLKLEGSVG

>Cre12.g513950 SUFD/NAP6

MQASTSLGSACRTQLAGKRVVSGHRYPNARIALKPAKSGAVASAVVAEDKWLARTVKEVATPSSLAALREQSKSALATLRMPT
TRNEDYRFTDIAPLLRTNVQPAAGAAVPSSELLAAHELPGAAARLVVVDGVLRLPELSSGLSGLPAGTYVGPLAGAPEAVKQKL
GSLNSNRGGPFVAVLNGSLVPEVLVAVPAERQLEGLPLFVLHLASGATGADEVAANAPRLLHAGTGSSAELVEEYVALPGGRR
YLTAVAGELFLEAAASVRHSYVQREGESFHFKSTLVHQAERSQYTLAEASVGGAIARHDLVIQQGGPETHTQMSHFLLCGPS
QTHDLHSRLTLDHPHGTANQLHKCIVSHASGRGVFDGNVKNRLAQKTDAGQLSRNLLLVPLATVNVKPNLQIIADDVKCTHG
CAVSDLRDDELFFYFRARGISAESARQALVFSFGAEVIRMQMGPVLQKRVQDDVSRITLLSAEPFAASA*

>jgi|MicpuC3v2|6075|wlab.223738.1

MTSTSLCAAANAARVGAKSRRAGRAATLSSRGITSPGDVAASVSGRARGGKRVGWRRADATPRAEAI SDPAGVVEEEKDAKQE
EVDKIQQMLSKPYKPGFKTDIETDIPKGLSEDTVRIISAKKNEPEWMLLEFRLKAFRKWLTMEEPDWSDNRYPDIDYQAVSYY
SAPKVMKKKSLDEVDPELLKTFDKLGIPLSEQKRMNTNVAVD AVFDSVSIATTFREDLAKGVIFCSISEAIREYPELIKQHL
GSVVPVGDNYFSAALNAAVSDGSFCYIPKGVKSPMELSTYFRINAESSGQFERTLIVAEEGSYVSYLEGCTAPAYDENQLHAA
VVEISAAKDAEVKYSTVQNWYAGNEDGVGGIYNFVTKRGLCKGDNSKISWTQVETGSSITWKYPSVVLGADNSVGEFFSVALT
NNKQQADTGTMIHVGKNTRSRIIVSKGISAGKSRNCYRGLVQVSPSAVGARNYSQCDSMLIGDKAGANTYPYISVREPSARVE
HEASTSKIGEDQLFYFQQRGIDPETAVGIIISGFCNEVFNELPLEFAAEVNALMSLKLEGSVG*

>jgi|Pico_ML_1|52773|g3433.t1

MATLRRPCAAFPTRSRTSRAPLCHSKVHRRVVEARAATDAPAPATETEETTEIRRLLSQPYKWGFKTDIDAVTIPRGLSE
DTVRLISAKKNEPEWMLDFRLKAFRKWLTMEEPDWSDNHYPPIDYQNI IYYSEPKMKQKQSLDEVDPELLATFDKLGIP
LGEQKRLANVAVD AVFDSVSIATTFREELAKAGVIFCSISEAIIKEYPDLVKNYLGSVVPVGDNYFTALNSAVFSDGSFCY
IPKGVKSPMELSTYFRINAQETGQFERTLIVAEEGSYVSYLEGCTAPAYDTNQLHAAVVELYAAKDAEIKYSTVQNWYAG
DEQGRGGIYNFVTKRGLCDGDHDKISWTQVETGSAITWKYPSVVLKGDNSIGANTYPYIQVKDPSAKVEHEASTSKIGED
QLFYFQQRGIDQEQAQVGMIIISGFCREVFNELPMFEFSAE*

>jgi|Praco1|14672|PRCOL_00000395-RA

QFSSPARVRGERRRRATARHGARRVGGGGEVARSAGRRRWGGGGEARRARGGDGGVGGGGAPLRGAAAGARARGCR
RAPPRAAVAPDSGAAAAPAAEEELTETERLNKLLAKPYKHGWKTEIESETIDKGLTEDTVRLISAKKGEPEWLLSRLK
AFKVVQSMEEPISVDNHYDKIDYQDISYYSAPKTKEKKASMDVDELPVRTFDKLGIPLNEQKRLANVAVDVVDVSDV
TTFKEDLAKAGVIFCSISEAVKDYPELVKKYLGTVVPPRDNYFAALNSAVFTDGSFVFIPKGVRCMELSTYFRINEMET
GQFERTLIVAEEGSHVSYLEGCTAPAYDENQLHAAIVELVCEKDAEIKYSTVQNWYAGDAEGRGGIYNFVTKRGLCAGEN
SKISWTQVETGSAITWKYPSVVLKGDNSVGEFYSVALTNMQQADTGTGMVHIGKNTRSRIIVSKGISAGKSQNAYRGLVQ
VQPTAVGARNYSQCDSMLIGDNAAANTYPYISVRDPASRVEHEASTSKIGEEQLFYFQQRGIDLEEAVGVIIISGFCQDVF
RELPLEFAAEVSQLMSLKLENAV*

>jgi|Chlpril1|6983|rna-gnl|IITBIO|A3770_13p69830_mRNA_A3770_13p69830

MKAKVVTRGGLGLGKGRKPAACAAKASCSYEEAASSLRRCIVARSASTLAKGEDGWVLGGGRRRERAGLTRVGCSEGVVEVD
AVELEKKEIKDILAKPYKFGFVSDIESEAI PKGLSEETVRMISKKKNEPEWMLLEFRLKAYRKWLTMEEPDWSDNRYPKIDFQD
IVYYSEPKQKEKKQSLDEVDPELLRTFDKLGIPLGEQKRLANVAVD AVFDSVSIATTFREDLLKAGVIFCSISEAVHEYPDIV
KKYMGSSVSEGDNYTALNSAVFSDGSFCYIPKGVKSPMELSTYFRINASDTGQFERTLIVAEEGSYVSYLEGCTAPS YDTNQ
LHAAVVELYCAKDAEIKYSTVQNWYAGDETGKGGIYNFVTKRGLCAGENAKISWTQVETGSAITWKYPSVVLKGDNSVGEFYS
VALTNYYQQADTGTMIHLGKNTRSRIIVSKGISAGKSQNCYRGLVQMGGNAENARNYSQCDSMLIGDTAHANTYPYIEVKEPS
SRVEHEASTSKISEDQLFYFQQRGIDAEAEVGMIIISGFCREVFQELPMFEFAAEVNQLMSLKLEGSVG*

>jgi|ChlNC64A_1|29265|estExt_Genewise1.C_10416

MLAKPYKYGFKTIIETE QFPKGLSEDVVRAISAKKEEPEWMLDFRLKAYRKWLTMEEPNWSDNRYPRIDYQDLSYYSAPKVVD
KKTSLDEVDPELLATFDKLGIPLNEQKRLANVAVD AVFDSVSIATTFKEELGKAGVIFCSISEAVKEYPDLVRKYLGSVVPVA

DNYFAALNAAVFS DGSFVYIPKGVRSPEMELSTYFRINASETGQFERTLIVAEEGAYVSYLEGCTAPAYDENQLHAAVVELSAA
KDAEIKYSTVQNNWGYAGDAEGRGGIYNFVTKRGICLGRSISKISWTQVETGSAITWKYPSVVLKGDHVSVEFYVSALTNNRQQ
ADTGTKMIHVKGKTRSRIVSKGISAGHSVNAYRGLVQVQPTAAGARNYSQCDSMLIGDQAGANTYPYIQAGGGAAAGSGQGAR
VEHEASTSKIGEDQLFYFQLRGIDAEDAVGMIISGFCRDVFNELPLEFAAEVNALMSLKLEGSVG*

>jgi|Tetstr1|438113|TSEL_026736.t1
MRVSI SPIGARSSKPLATARRATPAARVGRARRVPAVSAAAVQSAVATKDEWLSSTLQALEQSADGASAPLAEMKGAANDKL
LRSSVPTTRDEEYRFTDLAKLLSSKMCAPAAGGVDTSKLDAMAFSEASGSR LAIDGKLDLSASNLSSIPSNVYVGNLADAPD
WAVSKLNKQASSSRGNVSTFNSATATDVAVVAVPAGVKMECPHIMTLASSAPNEGELSASAQRLLVCLGEDAAVEVVEEFQ
TTQSGGAFTTSVAEIIYLDEGAALKHGVQLQLSDSGSGVHMKTTLVQQATNSCYKVVESCLGGNITRHDNLIVQLGPDNTTMR
HFLLAGKQDLHDLHKLILDHPRGVADQLHKCIAVSPTSRGVFDGNVKVNRAAQQTADAGQLSRNLLAPRATVNVKPNLQIVA
DDVKCTHGCTVSDLSEEEIFYFASRGIDKETARSCLVYSFGAEVLREHGFEGESSRLQAEMTSM LAAV*

>jgi|Tetstr1|447494|TSEL_034875.t1
MNMEPRTLETKAEALAEAYETMRDALPGGAGVARLRTEAFDAFRATGLPHRRVEEWKYSDLRRLKDVPAFAARRSEADAKA
VLASAGEAYASVDRFRLVIVDGHFYPELSDVAGLAGEGVTVRLFAEALAADGEALLDLPVVALADAAVGLNAIFCADGLDLAI
ADNTKLTKPVEIVHVS GAEAASTAERVLIRAGAKSEAVILQTFVGGVEGTFSNTLTDIRAGKGAGLTVARLQAEAAGTTHIAT
ATLLTAAEEAQAKLICAGIGSGFARHQSFVAFDGENARADLLGITMVDRQHV DQTLVVDHAVPHCESREIFKTVVDDGAKGVF
QGKIIVRPHAQKT DGMQMTQSLLLSENADMSLKPELEIFADDVQCAHGATSGQIDEDLLFYLRARGIPEKEARTLLVLAF LSE
AVEEINDEIKPQAVIDAVTRAYSHEYANVHRGLHYLSNTATENFEAARETVRRFLNAPSVDEVIFTRSTTEAINLVAESYGAQ
HIGEGDEIVLTIMEHHSNIVPWHFLRERKGAVLKWVYVREDGSFDLDAFEQAIGPRTKMVAITHMSNVLTGVVPIKQVCEIAH
AHGVPMVDGSGSAVHMPVDVQDLGCDFYCFTGHKVYGPSGIGVLWGKMELLEKMAPFNNGGEMILDVTEGDVTYNAPPHRFE
AGTPPIVQAIGLGAALDYMDAIGRDRIAAHEARLRDIAQEQLKKLNSIRIFGEAPGKGAIISFEMKGVHAHDVSMVIDREGVA
VRAGTHCAQPLLARYGVTSTCRASFGLYNTVEEVDLSLVAALQKAQDFFA*

>jgi|Tetstr1|450807|TSEL_037843.t1
MSTPLLEIKNLHVRVEDNEILKGVNLT LNRGEVHAIMGRNGSGKSTLSYVLAGKEDYEVT EGEILLDGENLLEMEADERAAAG
VFLAFQYPIEIPGVATMTFLKAAAMNAQRKARGEGELSTPDFMRAVKDAAGLLEIKQDMLKRLNVGFSGGEKKRAEILQ MALL
APKLCILDETD SGLDIDALKVVSEGVNKLRAPE RSM LVI THYQRLNLYIVPDVTPVRLSGAEESLVAQLESAGATDAAERLRV
IGLPTRRVESYHYTDLRTLLGQIPRAKAASDSSAPAVDIPGAYRILIANGVVQSTSTAPAGVIVGKAAGSVLTTRDDVLVRL
NAALVSES LNLELAGSVDPV IHI DHMEGPAGHVQSGAKIAFGPGAKATVIETVSGSDAAHMGNVGSYVSGVDGAEVTHILVD
LSARQTHFATVEYRIGAESSFKSVVINAGAHLARTNVFADFVGE THGDFFGLNVVDTDEHRDITIDITHGVPNTTSAELYK
QIARGRGKAVFQGINVAIDAQKTDAKMMTQGLMLSDEAEILSKPELEIFADDVVC GHGATCGDLDETS LFYLMSRGISRADA
ETILIRAFLEEITGAVEDEYVSEALAGVVERWLK KDVLDRMEHAYRHEYANVHRGLHTLANRATEAFEGAREKVRAFLNAERV
EEIIFTRSTTEAINLVASSFAGPRIGEGDEIVISIAEHHSNIVPWHFHRERKGA VIKWVDVADGSGFDLDAFTAALTDR TKIV
AITHMSNVLTGINPAKQIVEIAHARGIPVLIDGSGAAVHTKVDVRDIGADFYVFTGHKLYGPTGVGVLYGKYDLLAAMQPFLG
GGEMIEEVSQDAVTYNAPPHRFEAGTPPIVQAIGLGAIDYVEAIGRDAIAAHEHEVAVYAGEQLSRINSLRMFGIAPSKGGI
FSFMLENAHAHDVSTILDRYGVATEAEAKSFVEGGNALPTEDVERITADLIAALKTVYDPEIPVDIYELGLIYKVDLDDDRNL
TIDMTLTAPGCPVAGEMP GWVENAARSVEGIQDVEVKMVFDPWPWGPDRMSEEAQVALNWW*

>jgi|Caulen1|89441|g4767.t1
MWNWNNIENLCSTGHFAWKGV ECDGSGSVIGLDLQELSLOGSITGKLALLDSLVT LRLRGNLFTGQVPDLT ELSYLKILDLS
NELKGEFFPSAHAFWATPAGVSL EEEIYVQGNKFSGYFVDDDAVSI EESSIRIISASQNHYHGKLPESWFRLVHLEQMDLSNNR
GLCGEIPSFQNGVSIPEDLWSSQICEIPKSSAQEEATLLLSLKNELAPQSSVEQVLKNWRPEKSEELCESWERVECEEGRVFL
TSTSIGVPCDENLEVSWMATEVLP TNN DQVTQITTSLVVPEVAQAPQEFLNVSQDSEDS PMDAMQSSPAYTETENSAIVPSPE
PELEISGFDGRKLPLTVSTTDSSNGVDSGEDSFEEKKGLSQHAI IWTIVSFLLGILLVILIFLCI HHRNQ RDPNKDSKESRTW
KMRYSQFMSSLSGLIANTTKTSGSGKEGSSNGSKPKEDVESPTTSSV LASDHDSETNKNVSELGKLASIDSFAPLRMSFS
FSRKNKAVQLQNTTADRVQDGT TDMALFGIQTDDSRSLRLTD TND SQQNTYDNPLSSNDEPMATNLDLNDEVFNDCGVMFKNR
AFSEQTLNSEPSVTHTGLQRNYNDFHSIDE EPIEGAKSNHPPERLKTROPDFVVMGRCLLTQCSRGYFGDLGTLTQTTS DTM
GLRPHSTDATELNLNSKSTTWSITEDLG GTRTTISEPIMITRELTGSGNDVEDTKSIQIVSGRRIRRP FHTVVCSSSTGVADSI
GDKSSAEKADLQKLLNKPKYKWFKTIIESETFPKGLSE DVVRAISVKKEEPEWLTDFRLKAYKRWLT MKEPDWSDNTYPRINY
QDLSYYSEPKQKEKKQLSDEVDPELLATFDKLG IPLNEQKRLTNVAVDVAFDSVSIATTFREELSKVGVVFCSFSEAVKEYPE
LVRKYLGSVVPVNDNYFAALNSAVFS DGSFVYIPKGVHCPMELSTYFRINASETGQFERTLIVAEEGSYVSYLEGCTAPAYDT
NQLHAAVVELFACEDSEIKYSTVQNNWYAGDEEGKGGIYNFVTKRGLCHGNNSKISWTQVETGSAITWKYPSCVLRGDNVSVEF
YSVALTNNAQQADTGTKMVHVGNTRSRIVSKGISAGRSRNAYRGLVKIAPTAKNARNYSQCDSMLLGDNAGANTYPYIQVRD
PSAQVEHEASTSKIGEDQLFYFQQRGIDLEEAVGMIISGFCREVFNELPLEFAAEVNLQIMSLKLEGSVG*

>jgi|Chlat1|2051|Chrsp17S02525
MAAAMTAAAAVPAERLLFRAPATRPESRTKTELSAKRSSAFGSPLSLPQSVSTRQQPVSSSTRAEVIQEPAPMPVLPKKS
ASDDTISKFLEREYKYG FVGNI ESETIPKGLNEETV RMI SAKKEEPDWMLEFRLKAFRKWQTMKEPTWS DNKYKPIDFQA
ISYYSAPKNKPTLDSLDQVDPEMLKYFEKLEIPLGEQKRLANVRQVAVDAVFD SVSIATTFKKELAEAGVIFCSISEAVK
EHPDLIKKYLGSVVP IGDNYAALNAAVFS DGSFCYIPKGVTC PMDLSTYFRINASETGQFERTLIVAEEGSYVSYLEGCT
TAPSYDENQLHAAVVELYCAANA EIKYSTVQNNWYAGDKNGVGGIYNFVTKRGLCAGKKSKISWTQVETGSAITWKYPSV
LAGDNSIGEFYSVALTNNHQADTGSKMVHVGNTRSRIVSKGISAGQSVNCRYGLVSI GPNAAGARNYSQCDSMLIGDN
AGANTYPYIQVREPSAKVEHEASTSKIGEDQLFYMQQRGIDTEQAVAMIISGFCRDVFNELPLEFAAEVNLGMSLKLEGS
VG*

>jgi|Mesovir1|3765|Mesvi232S04464

MAAVHLNATAQLAMAASVCAAPAASQVFRRRGEVLSKGGFRGASLASRPRASGRIVYKTSRKC VATSLPEVV EEESEDMK
TISKFLDREYKFGFVSDIESV SIGKGLTEDTVRLISAKKNEPEWMLDYRIKAYRKWLTMD EPQWSDNKYPKINFQDIYY
SEPKFKPSVDSLDQVDPPELLRTFDKLGIP LGEQKRLANVKPVAVD AVFDSVSIATTFKKDLAKAGVIFCSISEAVKEHPE
LIKYYLGSVVPVADNYAALNAAVFDG SFCYIPKDTVCPMELSTYFRINASETGQFERTLIVAEENSHVSYLEGCTAPS
YDKNQLHAAVVELYAA SGAEIKYSTVQNWYAGDKDGLGGIYNFVTKRGLCAGARSKISWTQVETGSAITWKYPSVVLGAD
NSVGEFYSVALTNNHQADTGTKMVHVGNTRSRIVSKGISAGNSVNCYRGLVQVNP GATGARNYSQCDSMLIGDKAGAN
TYPYIQVRDPSAKVEHEASTSKIGEDQLFYFQQRG IATEQAVGAIISGF CREVFNELPMEFAAEVNQLMSLKLEGSVG*

>jgi|Klenit1|10541|rna-KFL_003990110

MEAVASPALHTQAAAAGFGQRTHGTHVQTQVRTQAALLPRQQRHPKLHSFSGSAKSIAGQKVCAESEGEFRDRKGRGHARA
AGTQSPTRERPSVKEAAEKIVGGGLGGGLDEVKLD DEVAGQVAKETGSEKAQISEFLKREYKAGFVTDIQSERIPRGLSEET
VRTISAKKNEPQWMLDYRLRAYRQWLKMEEP RWSNDNKYPRIDYQDIVVYSEPKKKDVKASLDEVD PDLATFEKLGIPLS
EQKRLSNVAVDAVFD SVSIATTFRKELAQAGVIFCSISEAVREY PDLIKEYLGSVVP PGDNFFAALNAAVFDG SFCYIP
KNVECPMELSTYFRINDSDTGQFERTLIVAEEGSKVSYLEGCTAP SFDTNQLHAAVVELYAHKNSEIKYSTVQNWYAGDE
EGVGGIFNFVTKRGLCEGKNAKISWTQVETGSAITWKYPSVVLKGDN SIGEFYSVALTNNRQQADTGTKMIHVGNTRSR
IVSKGISAGNSVNCYRGLVQVAPT AHGARNHSQCDSMLIGDQAGANTYPYIQVKDPTARVEHEASTSKIGEDQLFYFQQR
GIEAEKAVGMIISGF CRDVFNELPLEFAAEVNQLMSLKLEGSVG*

>jgi|Mesen1|10608|ME000086S10145

MGRKDRLQRGNWRCEAAEGGVLAPERTSGREKTSSETIDNFLKRDKYKYG FVSNIESVAIPKGLSEETVRIISAKKKEPE
WMLEFRLNAYRQWLKMT EPKWSDNKYPAIDFQVLVLVLVLVLLCLCLWLAVDQGVYGGGFDEVNAGGKTRVSQCCPSTSY
TSDWGWAGGGGGGGGSDAAGGGIQDVVYVYSEPKMKEKKASLDEVDADLLDAFEKLGISLSEQKRLANVAVDAVFD SVSIA
TTFRKDLAKAGVIFCSISEAVREYPELIKYYLGVVPGDNYYAALNAAVFDG SFCYIPKDVTCMELSTYFRINAMET
GQFERTLIVAEEDNSVSYLEVKDPSARVEHEASTSKIGEDQLFYFQQRGIDSEKAVGMIISGF CRDVFNELPLEFASEVN
QLMSLKLEGSVG*

>jgi|Chabral|342956|rna-CBR_g37549

MAAMAAMAAFAATAGAPKTPAPALVAYSARGVGGE GASSVSAPVALIARKPGALS VARYEGSACSLC SSSKSSSFLGGS
AVFLPNKDKSKSKSSSL SFTPPSSSSSSSSSLSSSFSSI PRAAAQPQSLIPAESSTPSSSSVLDKPEAEDDVIQQFLKKD
YKWGFVTDIESDSIPKGLSEETVRVISAKKKEPQWMLFRLNAYRQWLK MREPTWSDNRYPPIDLQNVCIYSEPRQKEK
ASLDEVDPELLRTFEKLGIP LGEQKRLANVAVDAVFD SVSIATTFREELAKAGVIFCSISEAVREYPELIKYYLGRVVPV
ADNYAALNAAVFDG SFCYIPENTISPMELSTYFRINASETGQFERTLIVAEKGSVSYLEGCTAPSYDKNQLHAAVVE
LYCAEGAEIKYSTVQNWYAGDEEGKGGIYNFVTKRGLCEGKGSKISWTQVETGSAITWKYPSVVLKGDN SVGEFYSVALT
NNRQQADTGTKMIHVGNTRSRIVSKGISAGKSVNCYRGLVQVGPTAHNARNYSQCDSMLIGDTAGANTYPYIQVKDPTA
RVEHEASTSKIGEDQLFYFQQRG IEMEKAVGMIISGF CRDVFNELPLEFAAEVNQLMSLKLEGSVG*

>ALG63564.1 putative ABC transporter (chloroplast) [Guillardia theta]

MEEIQRSANTRTEESDFNVKKLIAQPYKYGFTTNVENE EFPRGINADIVRLISEKKNEPSFMLDFRLRAY
AQWCKMIAPEWACVKHNSIDYQDIIYYSAPKYKKKLD SLDEVDS DILDTFEKLGIPLTEQKRLANVAVDA
IFDSV SIGSTFREELSNVGVIFCPISEALAKYPDLVKYLGTVVPIGDNYFAALNSAVFSEGSFCYVPED
VVCPLLELSTYFRINNQD SGQFERTLI IADKNSKVSYLEGCTAPQFDKNQLHAAVELIALD NAEIKYSTV
QNWYAGDTKGLGGIYNFVTKRGLCAGKKS KISWTQVETGSSITWKYPSCVLVGEESKGEFYSVALTNNYQ
QADTGTKMIHVGGKTKSKIISKGISAGNSRNTYRGQVKISGKAENSINYSQCDSMLIGDKSQANTFPYIQ
VSNATSKVEHEASTSKIEEQLFYFLQRGIDLES AISLLISGFCKDVFSELPMEFALEADRLLSLKLEGS
VG

>jgi|Emihul|454654|estExtDG_Genemark1.C_1180073

MLTACIASVGAAYLPSQRPAPGAPLCGSARGRAAVRCAEAGTLTQSAVDDWLGS AVAARGSVASGLLDEV RSEATDVVPP
EVREAVLESRLREIPEREADPRTEL GCHTFAALNQAALSDVACVYVPSGVAVDTLRLILLSTPSYPDSGAEAGASPQLAAS
HPNLESLWLDGASL LLDQYAGRGAYFCNGLSRLRLGENASLTHAYLQE QSDGAVHVD SVLATLAAGARYDANIVQSGGR
LSRVNLAVRLEGRHSHASLNGIALGSESQ LLDLHSAVRHVSPDCSSEQEQRNALAGRARVVFRGAVQVPHGSDNTTANQL
CRSLLLSDSARVDISP NLEILTDDVCTHGATVADLDDEMVFYLQARGLGRAEARVLLLEGWARS LMG EAMSSI*

>CBN77201.1 FeS assembly protein SufD [Ectocarpus siliculosus]

MERTAQRSRNGGSESRRVSTLLAVAAGLLSLAPTAD SFLSAPMKGAGGSGAAQGLRRRSTGEARSAQAGRQGDGVLSATG
VATEDNQGAVAGSLLETQ QKGVSNNEMKPFVRVPRDRPMWFDAIREGAVPRPDSLVELSKAGEAELAEALPTRKQEPY
RYTDLES LYRTDFTSAAPTTTAAETA AAIKPHLLEASQGGQMVFN VGVFSD ELSDV SALGGVEGLVAGHIGAIEGAPLDQ
AKEMLSYLPEKDADFR TTQGS LAFASLNQACFADAGVIMVADGVTVEKPIQVIF FSEGEEGPTVSHPRLVVKAGADSHVT
LTQSYLSRGGVCLANGFTRVLVGD RATVTHDYAEEMSASDRLVDTVSV DQTTNSTYSVNQVLSGAFDARANFQIDLLETG
SHCNIFTAALT TNNQRQDVHSTIRHKAEGTTCLQEQRNVVSDTADCVFKGRIVIDQVAQKTDANQLCRTLLVSDKARVTV
MPSMEI IANDVKCTHGATICDLEDEELFYLM SRGISKIQMGFHI

>XP_002184318.1 FeS assembly protein suf [Phaeodactylum tricornutum CCAP 1055/1]

MKFTSVTLACFFFV ESSGVSAFTFTPTQRQRSIPLIVSSSRRTSALHVSIGLGPDPNETSGKRSETLVAGVDYEVNP
HDDYRLSRRSKLDQTCDEWFAGLLDETEDGV LGS LVT DARTVLTTPVPLVNDLENPDSSHDDWTPYV NTRLPWTPLPVAF
GLEQFGLPVPRNAETWRHFDVASMVAESYSSAVKHDH EAKVDKIELFREQLAVAGGWLD DNACEARLVY INGRFVPQLS
KSNVDVARNLDSTDGLSKQSVQYLSRLTDGFTDE LASVPNNDI LETSYARLSRPNHNLGNATTQLAVNSQQGTACFAALN
TIKTGAVAYVHAPKGYNHIVADEQGDENDEEGPTRPNLKPVLIVNAVTKSGDAADHAKGVACHPRTLIVAEESLLSVVQ

SCVDLDDNSAHTPKLYNGYTQCFVHQKANVTHSYLEESGGIVTAGVERSDDEFVADETKPREIESRRKDRLRDTHLEAIDV
HIMGADAAAYEGTIMSLGGSGRVRVALSVALLPQGAHAGVNGFSLAGGAQRVDIKTNIHHVAQGTTSRQVQKNMIGGRATG
AFRGRIRVEQSAQQTESEQLSRTVLLSDKARAWAVPSLEIIADDDVQCAHGATVSDLSEELFYLRSRGMQAMARNLLMY
AFSGDVSDCVDPVAVLGGFDGNAGLRQRVIRRLLENLVPQGERAIKGEFQSI

>jgi|Nanoc1779_2|575559|fgenes1_kg.2_#_2054_#_TRINITY_DN1698_c0_g1_i1
MNMQHRPQRRQWGASYLFATAALLGTVQAFLLPFNTPAQPPPSLAKDSGLWWQQRRQQRSTKLQAALAEPRSPSVTLTD
DSTALIEAETVALKPPSREEQARDWFNSVLSSREPPLFHAATELGVGRILDGETRAPTRKEEAWRFTDLSRVYSRKPGK
APAVLQPAQVALIKEKVKHEHTLEVCEGRALVFVDGVFHFPEFSPAGSVEGLKEEGVREGWKVGSIFDFTGEEQEELLREIE
WSPEVGMHRHTLAAGSLPFCSLNQACVKDAACLVVGERGEMAAPVQFLFISTSTEGEGGKEGGKEARVSHPRLVVIAKKRS
AGKFVQSYVDIGEGVGLANACTTVRVDRRAEVQHLHANEMGGKQLFDVSVTTKKLSAYRNVFINVGGDMGRINLQATM
NGTEGRVDTEGLTVSGEEQELDIHSYIHHLRKDGKSNMYQCNIVSGDSHCVFVKGRMKLEPRARGVTADQLIRSIMLTDS
TVDAMPTLEVCSGDVDCTHGATVADIDENEAFYLSRGIGRREARQVLMMGFAAQVYNQVPCERMKRRLSDRLYEVCPEV
GMAESYEYLSI*

>jgi|Ochro1393_1_4|710221|e_gw1.51.30.1
MPRIVLLSLIFLQLTFFISRVSGYVIRHSVTPFCNIRSNGRKQAAIAIRGKNSFRIDSTSLNTSETTDSGEEDTISNFFR
KSSSISSNNEMVNLNANVAGKELLSGMKRPYKDEAWRHTNLRKLFAYNYKRPEVDYAAKLSRSSVEHLIDDACKESCLVF
IDGAYASKLSSISKIPENVGASSLLHGISSSGGAVQVEMLKHVPDEKELPRDTFASDSLALNMANMVD SAYIHIPTN
TVVETPIQVLFNCNTVAGSATSNAEEAAEFGTYPRLMVRDLSGAELQLKQSFATIPVDSASQTPIEIEIEPGGGPGFVGS
FTRIVLAEGAQLTHYTVQEVQSASRLVEVLAASVSGNSSYDVAAVQSGAVIGRINAHIELTDSNANCSLSGVTLAGKRQS
LDIHSNILDAPAANSEQQQRNVIADRGEAIFKGRIQIPKHAQLTDSACLRSIMLGQRARI IAMPTEITADNVVCSHG
ASVQDLDENSMFYLAARGVDRQEARKLLLRGFVFDLLKENVMDARSSQRLIRKLESMYPKQYSDVVDPSGGSQAQKFMSI*

>XP_009032168.1 putative FeS assembly protein [Aureococcus anophagefferens]
MRNRLGAGCWCALNAACAGDGVVWAREDAAAAAAAAAARDDDGARVAAGGAAAAPARLHVHVRSAGATGVLHPRTVVHV
GDGARVALRESFVAVGPAAEKRSRLTNARTNCRVGAGVLTHTVLDSEADQVHVHHVQCGVRRDGEFRSRAISDGAAVGR
VAVEAEILGPHARFDFLGLQLGAGDQNLDRALTALRHTAGDAESVQDIRNVAGERAKLTFKGRIAPVPTGQKTADQICKS
LLDDGATVNAMPSSLEIVADDVKCTHGATIADLDEGLFYLNRMLESEKDARKLLLKAFICYDIVQRTDLLPPATVDRDLA
KLSQLTV

>jgi|Vitbras1|16540|Vbra_21980.t1
MVLLEALIGHAVSISDLSSAPASLRPAFIGTAGGWRSLGHTTGGRDLLLRARRRRNQHHQRLFSMSTAEADTAAKEKEQP
PAAKEGEGEEDGVDESSSYKLVKRLEQERVSGLINDMVDPNYRREDYADRWAKEDDPEYVDSESPNSWERLRWKEEGK
LRFEDWSQFDPKRLEDPEYSLPEDWSAEPVGGGRHNIFQPTLVTSKDLKAVEQLDGGVVKVDEAPVDESVPQYRGLMKE
VGFGFGYRNISMNKADIQRLMWWDKMLELQPLPANEDFEKLHENGTAVMKELYVPERKREPWRQKDLTIYKTD FVAY
GDLSSTEDISEYVMNDTAALLVIKDLGVDRRLSRDLGLPMNGTFVGSIHNMEGALRERFLSELYWYPEYSQQGFKGAYVK
YKGLKWNADILPQQIFEMGKFIDIGMAKLALNMAHLRDVACVHVPEGTTIDKPIQVLVSVTSVGDALPISPRQLVLI
GDKADVTLSQSHVGLARKGFINSVTRVIMGRESRFNHDYIQEAGRDWHEQLSVLNAQAANYTYRSASWGAQSSRINVQ
VVANEKESRTDLRSLLARQKQLLQTYQTVWHNAKDTYSDFSFRNIVADSATAVWKGRGRLEHMCNGADAKQLCRSLLN
ERSKAVSIPVLEMMSPQIKAAHGATVSDLSQDQIIYLRTRGLSDEVARYLMVKGFANEVSNPMVDETAKRRLSRHMSIL
SDPSKKKEPVEEVAAPAA

>jgi|Symmic1|37384|rna26446
MASLTLEDVKRALLNRSNVFVVALWISQACLSVSHKRVSI VMKESPFVLTILQPVWIAVAFGIVWAAMCFRSQEKEESST
DTEAELSES DTSSEEW TAKSIGAVCLLLAQAFPGALSTVFMEYNLQRSHVSFWKLWTLINIWEVVALPPVWVASVPLQGL
TFAQSPKNLLDGFQCLFVGQGSVEGADDEE IWLGTALVVGAMVLFVRVKEAVKFGELTGVMADTKQQLDLHSLIHHSVPS
CRSKQQHKNLVGGS AEICFKGTIRVDAEAGFNMLRFTPRALLPNFARSLLLTKSKVKAMPSLQIQADDVSCSHGAALTQ
LDKDEL FYLASRGLDGRDARRMLLAGFPTDLLDGLKDFAPKVHERVLGKLASMAESDD*

SUFA1 orthologs

>Cre06.g299350 SUFA1
MLVHRQCTPVRVQPPKHLARRAVLLPARRGVAVARAEAAPGLVATEAPPHAITLTPEALVQLRKLRADFKEQSEAMLLRVGVK
QGGCGSGMSYVMDFESQDKVTDHVISYDDGFRLLVDPKSLLYLFGMQLGYSTALIGGGFQFQNPYATDSQQGFKGSFGV*

>jgi|Galsul1|1190|XM_005702494.1
MNTYIINITESALEQLNSISKLNKKDKLYIRIGIKQGGCSGLSYFMNYEKQSNISDKDLVYNYDNFTLVCDNKSILYLYGISL
DYSSSLIDGGFKFLNPNAKQTCGCGKSFS*

>jgi|Cyapar1|19701|g19315.t1
MAFVSVPVPVFPALSVPALSVPRVCMRATTTAAPRRVHAAPGSSSSVFAGVRFEARRSVRARRPNLWLPAAASAEPAAEAAAP
ASDAPEIPEEPMVLAEGETDMAIKISERAMAHLLTLKEKFTEDTCLRFVGRQGGCGSGMSYTM DMVKQGETREDDRVITYPNGF
QVIVDPKSLMYLFGMALDYS DALIGGGDFDKNPATSSCGCGKSFS*

>jgi|Caulen1|96705|g2901.t1
MGHLKKLRAEHQDSTMLFRVGVRS GGCSGLSYVMEFESENKISKDDTVFEYDEGSFRMICDPKSLLYLFGMSLDYSDALIGGG
QFHNPNATETCGCGKSFGV*

>jgi|ChlNC64A_1|20890|e_gw1.5.685.1
PPITLSDAALAHLLKKLREDSGDKQLLLRMGVKSGGCSGMSYVMDFEKAENVKKDDAVMEYDDGAFKLVCDKSKSLLYLFLGLRLD
YSSALVGGGFQFLNPNSTAECGW*

>jgi|Tetstr1|447496|TSEL_034877.t1
MASKFQVLTLTEAAAGRVREIIENSDDKDALGLRVGIKKGGCAGMEYDMSLVEEAKPGDDVIEDKGVRIYVDPSAVLFLLGTEM
DHEVTKFRSGFVFRNPNEVSACGCGESVSLKAADADGFQPANAPG*

>jgi|Chlpri1|5434|rna-gnl|IITBIO|A3770_09p54340_mRNA_A3770_09p54340
MSSRGWTRARSTAGEEQAGHLAPCITLTDDALAHLLKQLKEQKKSLLDQGLQLLLRIGVKQGGCSGMSYFMDFETPDKVVGDDA
VMDLEGDMKLVCDPKSLLYLFGMELDYSHELIGGGFKFTNPNAESTCGCGKSFSV*

>jgi|Pico_ML_1|52567|g3254.t1
MDFAEKESVTDEDSVLEFGDMKIIICDPKSLLYIFGMELDYKSALIGGGFAFRNPNADKTCGCGKSFSV*

>jgi|MicpuC3v2|1188|wlab.210138.1
MGALGRVLRGALASTSRRLVSAPSRRVVGDAFSSSSSSSHRAWSSSIAPFASESAPSPSGATPAATARDDALTEACVRRRLR
ELAASSSGITEGGAESKQPNDDGAPLLRVAVDGGGCSGFQYSFSLDSAPGVKPNDRVFERDGARVVVDVSLTFLKGATVDYV
EEMIKSSFSAISDNPNAESGCGCGSSFVAK*

>jgi|Pracol|19821|PRCOL_00003093-RA
MAASNLEGRGAFLKLGVRTRGCNGMTYTMNYADEKAKFDEFTFDNPNAAGSCGCGESFNPIENPLKIRR*

>jgi|Chlat1|6140|Chrsp41S00446
MSYVMDFEDPANVRDDSLIEHEGGFRMVCDPKSLLYLFGMQLDYS DALIGGGFAFQNP NATSSCGCGKSFS*

>jgi|Klenit1|11643|rna-KFL_004960060
MASTACAMRLGALSAPSINLHPRQYAPHQSI AVLNLNRNTPSGLRSIRS KGRITCSVATPSPPTLPPAISLSEKALAHIKRMR
NESGKDLTLRVGVRQGGCSGMSYVMEFEERANIREEDSLIDQEDFTMVCDPKSLLYLFGMQLDYS DALIGGGFSFQNP NASST
CGCGKSFAA*

>jgi|Mesen1|643|ME000108S10791
MAALTASAGITLSSLSSTALGTSQSVRKCSVDVRRPLALNLTARSQHQSRRRSFVVRSSDTASAISTLTSPAPAVTLSERAL
AHLVKMRSELNKDLLLLRIGVRQGGCSGYSYIMDFEERSNIREGDTIIDHEGFAMVCDPKSLFLFGMQLDYS DALIGGGFNFS
NPNASSTCGCGKSFS*

>jgi|Chabral|353188|rna-CBR_g72684
MGGPRAHSTNIGRSILLRVGEIPTRRDIELAGLEVYMKVATGTAPAGAAAAPAITLTDTALNHLMKLKEGQNNKELCLRVGVR
QGGCSGMSYVMDFEDRAKIRAEDAVIDYEGFTMVCDPKSLLYLFGMQLDYSNALIGGGFSFSPNPASSTCGCGKSFAA*

>jgi|Guith1|90502|estExt_Genewise1Plus.C_1470001
MKSRMLRNPKMTTLLVPLLLLLLAGYCNFLSNPLSLTRISLSTKVC SVRREKIH FAGLRMIEDTSTDECEGPAREGKVGNV
DISNSRWNQDEIKTDGIIRLTQPAISQIQKLRTNRGEQEVVLRVGVVRAGGCSGMSYVMEFEDAKNVDESDELKFEGRVVDV
PKSLMFIFGMELDYS DALIGGGFKFSNPNAASTCGCGQSFGA*

>jgi|Emihu1|61273|e_gw1.3.68.1
MRLILLVLFASLRAASSYLAPPPAALTRTRLAVVSSGGAETALPVRARALLCARPTLLRACPRLDGGAAQDSIISFSERALTQ
IQELREKSGKERACLRMGVRAGGCSGMSYVMDFMEEGEIDEKDTTIDYEQGTIRCVIDPKSLMFLYGLQLDYSDELIGGGFSF
ANPNAEETCGCGKSFGV*

>jgi|Bigna1|50664|estExt_Genewise1.C_880018
MMKEHLLLFTLTITVTFSKSADATRVRLRSGTMNSLRTQRYIKLGESHFAWPQNMIERINVPSRQSRMPLDCKCEQLGVGGV
LSAAEAGIIQISPNAEKRIAELKVSKGLDEAFFRVGVKSGGCSGMSYTM DIVGSDAVEADDTVVEYNGFKCIIIDPKSLLYLFG
MTLDYNDLIGGGFKFFNPNAKDTCGCGSSFGV*

>jgi|Phatr2|14867|e_gw1.16.211.1
LVLRMGVRS GGCSGMSYVMDFAKAADIQEDDQVDEYMSDRIQC VVDSKSM LYLYGLELDYSNELIGGGFQFFNPNAEESCGCG
SSFGV*

>jgi|Ectsil1|29259|rna12779
MRGLTVVAFAMLASMIKGFIVAPTSGSSTGSCRSSRVSGATTIRSHASAALSMSSTGVEDEVMEKIVNIRPKAMDHLQELRRS
QGTEGDI FLRMGVRSGGCSGLSYIMDLVKKEEITEDDMVEEYDGFKCVIDPKSVLYLYGLELDFSDALIGGGFQFQNPNAESS
CGCGKSFGV*

>jgi|Nanocel1779_2|596574|fgenes1_kg.17_#_382_#_TRINITY_DN6595_c0_g1_i1

MHRPFALFSVGHAKAMLMLLLGVLATTQAFLIPSAPFTLSSSTLAPLSSEKNEGTVVAERVITVSDSALEHLKDLRTKQGV
HLYLRMGVRSVSGSGMSYVLDVMKKEDVTEEDHIEAYDEHDFSCVDPKSLLYLFGRLDFDKDLIGGGFAFLNPNQAQESCGC
GSSFGV*

>jgi|Ochro1393_1_4|696524|e_gw1.6.362.1
MTPIQCFATRARNARARPPPIELTERAAGRIKELLSQKPDSVGVRLGVKRRGCNGYSYTMNYADQYESTKDEVVSEHGVQ
VLVDPKAIFFLVGATMDYVEDTLSEFTFSNPNEKGKCGCGESFNV*

>jgi|Auran1|21736
CSGLSYSMDLCKTEDIVPEDHVELWEDDLKVVIDAKSMLYLFGLLELGYSNELIGGGFKFNNPNAEESCGCGTSFGI*

>jgi|Vitbras1|10875|Vbra_585.t1
MWNLFILILGAAAVLPSEAARRFLQSRGSAAFVAPPASVNAAPRRPVLLRRALAVRSSGASTSRRAIDITPKALERIAELRAKKQ
RDGDGDLVLRMGVKSGCGSLSYTMDLIDASDVGPRDTVEAYEDEGFKYVVDPKALLYLFGMQLDYCDDLIGGGFRFDNPNAD
KSCGCGMSFGVPTQLTAMAKEKDGVSQSPGSCCTTKN

>jgi|Symmic1|38402|rna27132 (corrected)
MAAAAAKQIMSVTKSAASRLGTLMARHSDPNAIGVRIGLRQRCNGMSYTMDDYTDKVNKFDEVVEADGGIKVVVDSKAVMFLI
GTEMDFVSNEVGNFVFNNPNKKSECGCGQHDRQSHGLESTTQ

IBA57.2 orthologs

>jgi|Galsul1|6574|XM_005702813.1
MSSNHFLNSQPFESRWTCFLTSTGRILDVALVLVKDSILVVSSIEKKQMLWEHFDKHFPMDNVSVTEENYASFVWIGKQAV
DWIQNWYQQLGMMNGDQEQFQVFDHTNHPIYLLQQSTLEPKWKGYLLLCSAEDIHKVHASMSFQKKGFLLFHMDEKHEWELRI
EMGKGNVLCWSEQYHPLEAGLWHMVSFQKGCYLGQETILRLKTYGGVKRYLVGWLFLEYFVETPSHYVCQKRRVGNITSCKTI
QKTTTRESQPQDMTVVIGLGYLQSEYATLEYEMDKPLERIAIQDLWESTIFFSKVPGQTSTPVKCIVRIRIVFPVWH*

>jgi|Cyapar1|108|g104.t1
MAFSVTPAPTALRPGRALQLSSPAVCSGPTEAREATRLWEQRRRFVLGDQPLRLRATAAAAPHQRSFVPANPVMHVEHPHDVVK
TVEVAGVIDGTPEERLWARQEALGAVAGEPNSDSAWVTAETFGEEKDEDAIRAVGMGVALFDRSSWGRLKLSADCVAFHLNQ
STNDIQKLKSGQGCDCCLVTAQARLIDLALALRTDDNIILLVSPNKQKALAKHFDKYIFFQDKVKVEDLTSKTACFTLIGPSA
RDVLGRMSIDVRDMCGLTIAHKFFSFQRPSSGRMDVMLVRNSGLTAEGFTIVVNADQAADLWSFLEAQGARPLGTRAWNKLVRK
DGRPAAGLELTEEYNPLEAGLWHAVSLDKGCYIGQETLARLNTYKGVKTQIWLGLTVSGAAAGRRPAAGDPIFSEEGEKIGVV
SSVAEVEDKEGLACLAYVRTKVRQEGGRSIDRNI*

>jgi|Cyapar1|24870|g24308.t1
MLGSVVRRCAGAAVPKVATAVRPGIRALASAPAVLKKTALHDLHVKLGGKMVEFAGYAMPVQYPDGI IASHNHCRSAAALFDV
SHMGQLTLNGKDRVAFLESLSVGDFASLPEGSGRLSMFTNEQGGVLDDFIATNAGGYLFVVVNAGCKDADI AHMQAHA AKWRA
QGKDVSLTVIEDRSLVALQGPEAGAVLQRLVKEDLSRMSFMTSREMAVDGIPCRVTRCGYTGEDGFEISVPSAKAVAFERLC
AEKEVKPAGLGPRDSLRLEAGLCCLYGHELSPTISPEAGLTWAIGKRREQGGFLGAAVIQKQIAEGVTRKRVGLLIEGAPAR
EHTPVFDASGKQVGEVTSGGFSPTLKRAIAQAYVATPLSKDGTALQVEVRGKRSPAVVAKMPMVKTTYKAP*

>jgi|Chlre5_6|4731|Cre12.g552850.t1.1
MLSMIQQSGPTVTQARRCGGSKSACSWQRRAGGAAACLSAACGAASSSEASVSGRCHIGVVPLAAPRRRRAGSVVRGPEPPRI
NIDDLMLDVPEIDGDIRSLQVEMGAIFNDAGLATTFGKKKQALQALETGLVLVDQSHWSRLRVSGDDRLLTLLHNQSTQDFKAL
RPGQGADTVFVTATGRCLDLATALVLPSSVMLMVAEGTSDEAARGARPAGAALLERLNKMI FRGDKVAVQDV SERTAQISLMG
PEAEAVLRELAPDALAAVLGAPAGAHVLVGFRGKPVFVAGSGLGPGVPGYTLIADESIGGDVYAAFAAKGAIPMGTDDEWEEA
RVLAGRPLRGAELTEHNPLEAGLYGAVSLNKGCIYQETLAKLHLRDGVNRQLWGLALS GSTAPGTQITSELSKVGVTSTAC
QDAEGEWVGLGYIRSRIEGTQIALEGVRAVGGTPATVTAIPFATRKL SAAAEAPTSSD TVSGRLED AKKKKEEEAAKEAAT
AAKLQAMQERLAAWQAQQKQQT*

>jgi|Caulen1|96351|g2589.t1
MRTTFGRTTIPVSHPTPFHKTNLNAVFRIPRHSPLVSRSQNFDPDLEGDVEYYQRELNGRFNSTGILLDFNDNTRVPNTAPL
LYDRSHFGRLLKIEGKGVFEFLNNRTTNEFRNLHKGSGLDLTLFLQSKGKIIDLATAYVLNGVLLILSPTIQDALMEELD KFI
PGDEVEIRDVRTETRIFSLIGQGSQELLSLLNVQDSVLDGSGYGTNEISGFQNGPVI VAKGNELNRSEGYTLIVDSDVAGDLYK
NLIDAGGVPIGLLDWERLVRVSNRPGLETEL RHASNPFEVGLFHAVSLSKGCF LGQESISRIYTKNTIRKRLWGIQFHSKEAV
VSPEAMIRSEGKDVGRVTSIPVDDVEKDSTQKICLGFRLCKINKTAVDWNNGTEVEIEGFKGTVRHLDTYNNDFETGIGSPIPS
TPEGKTNVESVDKEKKLKEMQERLDAFMKQKQKQ*

>jgi|Tetstr1|440026|TSEL_028387.t1
MAVCTAASIGSGVSSNRTHTHAHTHTHTRPTTSAARRIPLAFRRVASRPSLCSGRRMRGSPRRAAEEDFDLDALLDGLEPPAL
EGEPRIYQEAAGGVFEGGIAVSFGKTDKDVPEAMRESVVLVDRSHWGRLRIGGDDRVSLHNQSTANIAGLAKQGQTDTVLVT
AQARTLDVATALVQESGIMLIVSPGQAAALVERFQKYIFPMDKVEVMDISERCGMLTLVGP GSDELLAELGAGDVVGKPEGSH
TLLNFGGSPVIAAVTTGLRMPGYTLIVGEDKTGELWRQLAMKGAVPMGADAAAAPAGKELTEDYNPLEAGLCHAVSLAKGCYI
GQETIAKVHNLGAGKQQLWGLDAAAACEGEEVLSAGGAKLGAVTSATAKPEGGFALAYLKCKVKGGSVGLAELPVTVGGR
PAKLAPLPYSSRALLPKDLPPSKDAAKEESAAGEDDAAAAAKEAKLKAMQERLAAYQAQMAAGKKEAK*

>jgi|Chlpril1|3704|rna-gnl|IITBIO|A3770_05p37040_mRNA_A3770_05p37040
MRDLVGRSAGRAPTQRCGSPSCSRSSRRGKRRASGRRFGSTRFGRSRVDWSARSLEDIDMLEVDGWEEVTLREIQEEGGAVF
AGGASVPTSFAGDGNESGLTLGDLESAECCVFCDRSDSTVLQVAGGDRLAFLHNQSTASVEGAKPGDVVETAFTVTSKAGLL
DVAKVLVQGGQALVSLASPDGASQIIIGHLDKHIFPRDDVRVSDVSFNLVSVFAVLGAGSSKVMETLGMGHDMGEQEVLVNYEGY
PLCVARTTDLGVPGYTWIVDTSISSDAWSLLTSIENVRPIGSEAYERLRILAGRPAFGKELTLDNPLEAGLYERCVDVEKGC
SIGQEVIKRVHNRNGITRRLGGFSSDKPLRCGDEVLSEGRKVGSLTVCQDGGGSFLGMGYVKVKPGQPFSSEGKSFTVRGEA
VAVRALVYPKYPQESVENKDAREEEESLEEIEEAKKKEEMKARMEEWLKKQSKNK*

>jgi|Pracol1|21076|PRCOL_00004673-RA (no typical signature)
MACASDDGLEALLEELDLDIDLELLEPPAIDGDLGELLSDAGAADVDAATGIALTYHNDVDAFDAARGAEDTSRVAVVNLTH
AARVRVSGADRGAFLRRETHGGGVGEVALVSHGEGAGSVCGAALDAGVVVSQGGQSELLLLPPESRADAAGALGARVRPDEDV
TVADVSGPTAQFALVGEGAADALRSLLGDELVDSPDVAAVLAEGAPAGLSCMAGFQGSPCSLIRAEPLGPSTPQLLVVTDAGV
TAEMFAALVQAGCVPMSNCWEKLRVRCGRPRRGFELGVRGFERTPLEANLWHACSAGGAAEEGEGEPSAGVAEVRQRAAAR
SMEYHLWSLELFPVYVPPGDAVEVDGVEQGMVTSVVQTPGEGELWALAYVKARAGVSDGGEVVRVGTTRRAI AHMPEYGSRRLR*

>jgi|Chlat1|7157|Chrsp57S06830
MASWVGLGPSLPMFLSLPVACASRAAFSSSLATTAVAPSLNGSQRRGLRSSSVCMPPQRRWLPDQASAVRVSSRKGRSGVVANSL
GDALTAPPIDGDIKELQEVAGAVFGDDDIVETYGNDSAALEAMRNLAI SDQSHFGRIRVADADRIRFLHNQSTADFNRLKPG
EGCETVFTPTARTIDLARAFVQANSVLLVSPDMRHQLPLMDKYLFADKVKLQDVTEQCSMLVLVYGPGSHEVITQLQAGD
MVGQPKHSHRLYNVNGSPVVAVGTGLATEGYTLVVDTAVVADLWATLAGFGAVPMGRRAWTHARVLEGRFPVPGAELTDFNP
LEAGLWHTISTDKGCIYQETIARLITYSGTKQYLWGVELESPEAGTLITVDGEKAGKITSCITLDDGVHRGLAYVRTKVG
SGLQVDAAGVKGVTVDPVYVTRTLPA*

>jgi|Mesovir1|195|Mesvi104S01524
MAVFCHATSFSLGGVQGVVAHNSHKVANQNEARSRAYRHCMPARSSGKLCFSFGKASFKL MARKCQSSREI IYRPLVVAHSS
ADDLLAGIEMPPIFGDLRDMQEDAGAVFGDEDAGGPLTFGNDKEALAAARGSGFAVYDKSHYGRIRVVGADRI SFLHNMTTAN
ISRLKPGQGCETLFSVPQARTIDLASLYVQESSVMLSTHTKTAEILKIMDKYIFPMDKVKLMDVSDKCF SFALVGKGAAEAL
TSFGAGDILSKPAGSHALFEFQGNPVM LGVGCGLSVEGFTVVADAPIAGDVWSAFVQAGGVPMGEAA YEALRVAEGRPAPGQE
LTADYNPLEAALYSAVSIDKGCIYQETIARLITYDGVKQQLWGLRLNARVQPGADILDGDQKIGKLTSCCEDIEKPGMFLGL
GYIKRKAEGMGAGVSVMVGTAPATLVDIPYASRSLPTA*

>jgi|Klenit1|9453|rna-KFL_003250090
MNAAGLCCRFESLHFGSLCNSTPSPLSRKICLPFAVGPNSPSADLATRLPQQKGSFSFFTRDRTTFATTTPRKVLQERSGPV
AAAFVDDLPPPIDRDLQELQEELGAEFSPEGVVVTFNNDADALEASKSGFAVVDLSHFGRLRVYDDDRIRFLHNQSTADFNL
LKPGQGGDTVFVTNTARTIDLVTAYIKESSVLLVTSPSQRRELLPYLDKFIFFADKVKLQDVTDQCFMLALVGPQSSQVLEDL
RAGDLVGKPYGTHAMFDFQGSFPMIAVGSGLATDGYTVIADAAAAGDFWKTLMQGGGVPMGARAWERLRILQGRFPVPGAELTP
DYNPLEAGLWNTISLTKGCIYQETIARLITYNVGKQHLWGIQLSEAVEPGTVITLDDAKVGKVTSTVETESGFFALGYIRSK
VCSEPGTLVHVGNVAVGKVEIPFVVRTLDA*

>jgi|Chabra1|335283|rna-CBR_g30837
MDAGVEVVDLHAYIAKIDREFKTQRYDDIDAPLLYVRIETGEATCIA LIDCGASRNYMSQDFMVRAGLGPRVRRKSQPTQVTL
ADGHTHSKIDRCIDNVPMYFAPHASEAVSFDILDTKFDMILGMSWLRSEDHPVNFYRRTVHIRDRNGVLVPCTVAPPHPSVSC
HVVSAA SMRASII RDYIEEMGVCFHLALPPQDASSTDSSSDPRITRLLDAYNYVFEGPHGVVPDRPICHEIILEDGAVPPRG
IYRMSEEEELSVLRAQLDDLLEKGWIRPSSSPYGA PVLFVRKKNKDLRLCIDYCKLNEQTRINAGPLPRIDDLERLGGA KFFS
KLDLKS GYHQL EIRKEDRYKTTRSLEEHVEYLRTVLERLRQAKYKANRDKCE FARQELEYLGHYVTPQ GIRPLADKIEALRVW
PEPTNTTDVRSFMGLAGYYQRFITGYSRIAAPMTRLQSPKVPFVDDDARRSFQTLKTAMLMA PVRSIYDPTLPTRVTTDASD
YGI GAVLEQHDGVDWHPVEYF SHKVPPI NSLDDARKKELLAFDMALKRWRHFLLRRRFTWDIQTKYGAIFDEDSGGNVAVTF
GNDVEAIDA AKTHAVVIDRTHYGRIRVSDERIRFLHNQSTADFIQLKQGQGC DTVFVTPARTIDLATAYIMNSSVI VVTSP
SLRHELVTTLDKYIFFADKVKVEDITDR TAMLSIIGPKSGEILEDMKAEAVVGTPTYGTNMHFELFGE PVTIASGAGLASEGYT
LIMTPDSAGLVNVLGKGA VPMGEAAWEKLRILEGRPSPGKELTGDYNPLEAGLWK TISLTKGCIYQETIARVITYDALKQ
QLWGLELEGLAEPGTIITAANSDEKVGKLT SFAEVDGKFFGLGYIKRTAGGEGMQVSVGGTKGLVVEVPYLTRSLTQ*

>jgi|Guith1|90067|estExt_Genewise1Plus.C_1150008
MLSVRSVAVVLLSTTTTVSFNVRWSAVRPLSRGSLPCPSRRNEPRTRMRAGRQLVACSKDTSEQKDATQSLAAIAEMERR
RQQEEEMRNSGSIGWECLRDTPQSKQELRELQLKFGAEVVPATYPAPEPLMKFMKKGRDRVVDVVS KFSLLPQAEVIAASSSG
VLLADLSHWGALVVTGEDRYKFLSGLCTNRVVGRLPGDVQRACFLSKVGRTVDLCTIAVLSDSLVLVLCSPNRVLQLFQDL DAL
IFPKDKVKCLDVSEGLARFHLVGPKAEFEFLSQMPGITLPEPFCSSPLVLEGKSMNGLVLHGAGCSVPGYTIVAASGDGRVLWE
EFMKVAEPGLRVGQEGWEVLRMVDGVPAAGSELTLEYNPLEAGLWSTVSFDKGCIYQESMARLKTYDGVKQNLWAIQFPAD
KSGAKSSRALVGAKLFAVSACESGQSRRLKLACEGEGRGRKIHRDDHVRPPGALAHPR*

>jgi|Bigna1|83823|fgenes1_pg.116_#_16
MAVRNCPTLLSPSASFSTVSDEEAQETRTCQGGMLEATCHGRQRHLRQLALFKALLLSVVVMVSLGIWSVA AVSVGELRRG
AVRASLLPCR LASYIRGGVRRARST SFAAPNAHGD SKSEPTESKPPPPQITLQLKERIRLQQEAAKRAMMEVKSGATT LA
GKND AELRHVLEGERAAI IDHSNYGIEVTGRDRDLFLHGMSTNDIAALQPGAYVQTVFTTRTGRMIDSTVAIATEESVLLI
VSPGNAATLSEMLDKYILASNVRVRDASYDYAVFRVLGSDVLRRIAPATPMGKAQENG EARVEFWTPNGGGDDSSGIQ NAGGG
KEEEIVAVHGETGGLVLPGMTLLVPTSSVAKVYAEI IKQREEEEEELLQIGLMGWEKLRVHQGV PKLGGEITESYNPLEAQLW
DSLNFIKGCIYQETIARLKYKGVKQFLQRFRLDAMPDGFQTAEDGPENDRDGEI IHSSEGGEKVGVLTS CAAIGGQIWGL
GYVRTKANLEDGSRVVIGETEDAAGISAVMDTPLPRVGDNGFPEWLLQDDEAASPSQ*

>jgi|Ectsil1|18192|rna10292
MAPEFWKSVSSCLLACACIPLAVGLSPWAATTSTRSTSFGSSSALGSSARLWAQRSSSSSSSGSGSSSCGTVSPGPGGRSTP
TCSLDDLRTAQSENGGVIGDVSYYDDGHVLQNVLVSGFEAGTGARQRADAVADAAASGTLVVDKSHWGVIRVEGEDRLRFLHSQ
GTNAFERATVGVVATCFTNNIGRVVDFCEGVVLDLDAVWLISSPHRWQKLLGTMDFIFPMDKTTVSSSLSEELAVFSLAGPKA
AETMAMAKCASCPEPGRSVEWFEFEGEKVLIIGHARPLTRTREEEGVSHDAEGVLSGDGERQGTQQHPYYSVIVPVSVSSKVV
SALSASPSVVAAGEEEWQTLRIKQGFFPGKELTADYNPLEAGLWHAVHFDKGCYIGQESISRNVAYNAVSKALYGVSFEDST
SPEQGTFLFVQETGKSAGVVTSMLEDRDATSHPFGLAYIRTKAGGVGLKLSKEGEPLGTVVQVPYPTRSDAESAMPPAKKEGD
SGDKATAAAVLGLEEEKAKEKARKAAKLEAMQKKIDALKARRKTS*

>jgi|Nanocel1779_2|600710|fgenes1_kg.22_#_427_#_TRINITY_DN6865_c0_g2_i5
MTLLPILGLVLLALCLDSTVGFLQPHQPPHSNNRGAQRKCGSILGMLSVDEAMNMKEACGQAQGVLLDLTAGEWGLLSVTGED
RLRFLHNQLTNSFTDLKPGQVLESSYTSSTGRTIDLVTAYVLQDEVLLSSPSRHEQLLGAFNKYIFPLDRVEVQRRDLSGLF
IWLGPASAAQTVPQLLGNDLTLAPLPGPSQCQYCSSSSFSMVLAEGTGVALSGYITILALRARAASIPAGDVAAHVGRGGMEE
WQALRVLQGRPYDPKELTMDVTPLEAGLWHTVDLNGKCYLQGETLARQANNKGERQKLYAVRGKGLKERKRVIDAEGNRAGVI
TSCMNDMGLAYIRKTVLNPFTDMLFVDGEAQRIQVLDIPYATRAATFNKMTTPKEAAAAAAVNTTAADVAAEAERKAKKLEE
MKRRLEAFQAKSRQT*

GRX3 orthologs

>GRX3 Chlamydomonas reinhardtii (correction in red, GRXS14 ortholog)
MASLQCNMRPATIILTRRPAQVARKSIGVARSQTVRVTASGMADLKKSIDELIASNKVVVFMKGTRQFPMCGFSNTVVQILN
VMDVPYQTVNILEDDAIRSGMKEYSQWPTFPQVYINGDFFGGCDIMMEAYQSGELKEQLEIALNS*

>jgi|Cyapar1|8129|g7870.t1
MAAFACAI PVGVRGPSSVLAQPARSSAAPAPAVRRHAAARAFTASRSQFVGA AVARRSAPAKAFAFRPVEIKAEITPELEK
QIEDLVKNNKVVLFMKGNRLMPQC GFSNTVVQILERCNPYETCDILADPDMRQGMKEYSSWPTFPQVYIDGEFVGGCDI
MLDLFQSGELQSMLEVAMNS*

>jgi|Caulen1|95844|g2143.t1
MKIVCFGRAPHLARFQSRRLKLPIFTSSPSFRQKRGAFTAAAHGPGLPDELKLALDGLISQNKVVVFMKGTKQFPQC GFS
NTVVQILNTVNAQFETVNVLENDVLRAGMKEYSQWPTFPQVYIDGEFFGGCDIMIGNFLLPSIT*

>jgi|ChlNC64A_1|37075|estExt_Genewise1Plus.C_250008
MAAAVA VGHSLAVSPVSQPRGSAQLRAPVRQLAAGWPRSRRACRTQAVGQVPGTPAAPGAGLSPELREAIQQFITDNK
IVAFIKGTKQFPSCGFSNTVVQILNCGAPYVTNVNLEDELLRSGMKEFSQWPTFPQVYIDGEFFGGADILIAAYTSSEL
AEEVERAINN*

>jgi|Tetstr1|427137|TSEL_017326.t1
MALATAPTRATASTARCPVRRPAPATAPRHSSGLAAGLRLRPASGLARRGGQRSQRRADKPLRVSAMGPGVGLSPELRA
ALDEYVNKYKV VFMKGTPQFPQC GFSNTVVQIMNNLVNPFQGVNILEDDSLRSLKEYSAWPTFPQVYIGGEFFGGCDI
MIEEYQNGNLKETVERVLME*

>jgi|Chlpril1|3579|rna-gnl|IITBIO|A3770_05p35790_mRNA_A3770_05p35790
MEARVRARTRAGVGWKRACSCACGGREAGGQGRRREGRGVGRDGI AARNWTGGQGRARTWSRCEAQQGESPSGFTF
ELRDAVDRLISENKIVLFMKGTKMFPQC GFSNTCVQIFNQLGAPFEDINVLEDERIRVGMKEYSQWPTFPQVYIEGEFYG
GCDILLESFQSGELEEYIERVLAS*

>jgi|Pico_ML_1|54442|g4788.t1
MKGTKAFPQC GFSNTCVQILNTFNVPYETVNILEDLSRQGMKEYSNWPTFPQCYIDGEFFGGCDIMIEAYQSGELQETI
EKAMVN*

>jgi|MicpuC3v2|4834|wlab.217373.1 S14
MKGSKDAPRCGFSSTVVQIFTS MNVPFETVDILDDDGLRAGMKIYSNWPTFPQVYIDGNFYGGCDICIGAYKDGLKEAVDVA
LMS*

>jgi|Pracol1|16206|PRCOL_00004792-RA S14
MAPVVLFMKGNREQPRCGFSATSVQVLNACRANFETVDVLD DPSMRDAMKIFSSWPTFPQLYVEGEFYGGCDILMEMYQDGEL
QELIERVNMS*

>jgi|Chlat1|1422|Chrsp12S01982
MALAAAVAVVGTAARPLSVASSGRPQSARRPAPSARLQOSSGSFTPSLRPLQRRRQQQEKRS CRAVAAGEPQPGLTSEL
KEALDDFI AKNKVLFMKGTRFEPQC GFSNTCVQILNIMGAPYETVNILEDLTLRQGMKMYSSWPTFPQVYIDGEFIGGC
DILIESYQSGELKEMLERATVS*

>jgi|Mesovir1|7736|Mesvi7120S00833
MAVGRGGSRKVGASLPTELKQALDDFVN NNKVLFMKGTRQFPQC GFSNTCVQILNTLAVPYETVNILEDLSRSLKVY
SSWPTFPQLYVDGEFVGGCDITI ELYQNGELKEIVEKAMAS*

>jgi|Klenit1|3212|rna-KFL_000610430
MSFMGKTQTVRRTAHSGCRASAGASSGGGSVLTPELRTAVEEFLQEHKVVLFMKGNRQFPQCGFSNTCVQILNMVQVPYE
TVNILENETLRAGMKEYSSWPTFPQYIDGFIGGCDIAIELYQSGELKELVEKALVS*

>jgi|Mesen1|6838|ME000351S05957
MASITGAVAHIAPSATSLCGSKSRSSVGRTHLSSSVSRSCSLTSSSQDLLTVRVGAGRNNQRSISCQASIPDFKAPSAL
TPELKETIDNFLASHKLVLFIKGNRQFPQCGFSDGAVKVANALDKPYETVDILADETLRQAMKEYSQWPTFPQVYLGGEF
VGGFDIFYEALSTKEFQEQVEVVFAS*

>jgi|Guith1|84703|estExt_Genewise1Plus.C_80280
MSKAFLLTILGISITSVSSFSLLPGCTRFVAVRNVANSFVQSPRTAAHTSPLQLRMQLDDSNRKGPMWNAIDKVVKENKI
VLFMKGKNDFPQCGFSNTVVQILNACNARYEAVNVLENELLRQGIKEYSNWPTIPQLYVDGEFLGGCDIAIDMYQSGELA
EMIEKVMAE*

>jgi|Emihu1|69804|e_gw1.107.46.1
METTEDRIKAMVEGNKIFLFMKGSKLFPQCGFSNTAVQILNSYNVEYETFDVLADNAVRDGVKTYSNWPTIPQCYIEGEF
VGGCDLLIEMYQSGELSEMVEKAAAS*

>jgi|Bigna1|63721|fgenes1_kg.58_#_40_#_4739_1_CFAO_EXT
MKGTKDFPQCGFSNTCVGILNRLGVEFETVNVLENEMIRQGMKVYSDWPTFPQVYIGGEFYGGCDIMLQGYESGDLAETI
ENLFS*

>jgi|Phatr2|9035|e_gw1.1.958.1
MADDFLQETPDSTQNRKLNVEEKPVLFLFMKGKNLFPQCGFSNTAVQILESFKIDFETVDVLSDDKIREGVKLYSQWPTI
PQLYVAGEFIGGSDIMIEMYQSGELGEMIEKAKADMV*

>jgi|Ectsil1|32226|rna4822
MMSEDFGEEDSFGARAMQERTHAELQEMVSSHNILAFIKGNRLMPQCGYSGTLVNIQLSLSVPFETVDVLADERIRQGIK
DFSNWPTIPQLYLGGEFIGGADIVIELFQSGELQEMIEVAAAS*

>jgi|Nanoc1779_2|616044|fgenes1_kg.111_#_63_#_TRINITY_DN7382_c1_g10_i1
MKGNRLLFPSCGFSNTAIRILNQLNVEYETVDVLADERIRGEIKRYSAWPTIPQLYLQGEFVGGSDIILEMYQTGELQEMI
EIAAAS*

>jgi|Ochro1393_1_4|696670|e_gw1.6.267.1
MGDVPGLDDKTEKKIQDLITNNKVVLFMKGVNGAPQCGFSDLACRIINTIETPFQSVNVLDDDYIRAGIKVFSSWPTIPQ
LYVDGEFVGGSDIMLEMYESGELAEIEVTNAS*

>jgi|Auran1|28514
MLAGLRVAARAAPRLRAARLLSSGSHSDFETVRAAVPDTPAGVKAMIQEQVDGNTVMLYMGKTPARPQCGFSQQVCRILH
ATGVEFASVNVLEDDALREGIKEFSEWPTIPQLYVGGEFVGGCDIVTESFQSGELDEAFKAAGAK*

>jgi|Auran1|18629
MARTVALIAALAATASAFVAPGAAAPRLSPRRAGDDVEPPPDATTKRIQSMVDDNTVMLFMKGTKIFPQCGFSNMAVQIL
NAIGQDFETYDCLSDENVRSTIKEFSWPTIPQLYVDGEFVGGSDIMLEMYEAGELQEMIEVAAAS*

>jgi|Vitbras1|7343|Vbra_20848.t1
MQPLSACLTAASIFKLAFFVLSCMAAIADAFSLSLPAHSSSLRQRSPSRALRRHTSALYESLLDSAPSPPEADQLDEATR
ERIAALVRENFPVLVFMKGTPQCPQCGFSNTMMQVLQVCGVQRFAVDVLSDERIRQGIKVYSEWPTIPQLYVNGEFLGGS
DTAVEMYSSGELQELLEVAQAS

>CAE7257650.1 purL [Symbiodinium microadriaticum] **bad annotation of the purL protein**
MKGTPVFPQCGFSAAVVQILSHLVKFKGIDVLQDPGIRQGIKDFSNWPTIPQLYVKGEFVGGCDIIREMFQTGELSEFLQTN
GIDIKPAA

GRX6 orthologs

>Cre01.g047800
MQTLSKNAIKASPRPAARMPASVRPFTSSVKPVVSRSVISHATLAETKAVPVMKEGELSQFPGSAGVYAVYDKNGTLQYIGL
SRKVSASVATHMQELPDLTGSIKYNIVDDASREGLTAAKWSVEETVAEAGIIPPGNAPGETKWQSRVARATKPEIRLTAGK
PIQGVTLLELIDRIVKENPIVVFVKGTRQQPQCGFSFRMINMLNTLKADFEVVNVLDEFHNPGLRDAIKNYSQWPTIPQLYIS
GEFVGGADIAEQMVGTGELQTLVRSAAKDPN*

>jgi|Caulen1|93851|g8441.t2
MTFILKSTLQQLQNFKIETQNAFEDFPEAPAVYAVFDANEALQYIGITRKLRLSIMGHSRDLPLDTHSVK
FETKHGATREELTQSWKEMIEDYVKQTGQVPPGNVTGNDQWKTKRVIRQKPEIRLTPGKGIEDLTISLEK
IIDQIVKTNKVAFIKGTRLEPKCGFSYQMLTILNRRNVDEYVVDVLDEVYNPGVREAIKNYSQWPTIPQ

LFVNGEFLGGADILTELQDKGEFEQIVKQL*

>jgi|ChlNC64A_1|31055|estExt_Genewise1.C_90387
MQALAAATACMPMARLAAPRQQQRQRHAPAAASRRQQPCRRSAARRLVMVTAEAGTSGAAAAASQLQAFSQTAPQORVLGA
DGKLAEFPTTAGVYAVYDSSGTLHYLGLSRKIAVSLATHVEALPEGLVHSVVKVLEVPEASKEELQEAWKAWMQEAVAEAG
AIPPGNASGQTLWQQRRPRAAKPEIRLTPGKGVQDLTCSVEDLVDQVVKTCKVAVFVKGTRTQPQCGFSYKVLTLQESG
ADFEVVNVLLDDVYNPGLREAIKNYSAWPTIPQVFGGEFIGGADIVEEMNGSGELKTALQQAGAVRG*

>jgi|MicpuC3v2|4885|wlab.203974.1
MTTISSDFSVALHRCVPVRNTRFCRGCRRVHLPSYSWKVRSATGDQLQYQGSSIIPIVSLSAADFSQVIFPEHPGIYA
VYNTTMEIQYIGISRKISSSVKLHKYEMPRYCGFLRFAAIIPRAKADLQLLWKQWMQEH LAVSGGLLPNGNVRGNSQWSR
RKHLPPKINILLTNGVNSQVNDSTVFQLCREIVESYNIVAFIKGTRKEPECGFSYRVCAILLEELSVDFETVDTLDERHNH
NLRNVLKDFSDWPTIPQVYVRGSLGHDIIEDLYKSGQLKHILVPGSTAEAA*

>jgi|Tetstr1|436275|TSEL_025117.t1
MACALTSPGAATFAAMAVAGVARQTPPPPLAAASQGRRLVAIPRASQGETRRWLATFRSGRPSRRPAALG
AAVCAASTPGMLSINDVEAVPMENGELTAFPPQAGVYAVYSPEGELQYIGLSRKISASIIIGHMQDLDP
LVGAVKVCLLSTDDKQVLMASWKAWEIISGTGKVPNGAPGIKDWQTGGGTRTVKPEVRLMKAGTAVED
LTCPEELIDQVVEKHKVAFIKGTRTEPACGFSFKMLSLNLGAKVEYEVNVFDEVHNPGLREAIKSFS
AWPTIPQLYVDGEFVGGADIIVEMSENGELAEVLGAASS*

>jgi|Chlpril1|532|rna-gnl|IITBIO|A3770_01p05320_mRNA_A3770_01p05320
MRGGVGVSVCRGGGRVFNWVGAWVRGERLVGLRRHGATGGGGGEGGGIPKLEYASSKPLKKAEGLDNWW
PSTPGVYAVFDEREELQYIGLSRRISNSIERHAETELEKMHVSVQVLEMPGASKDSLQQMWKSWIESYAE
TGNIPVGNMKGEGQVWLNRPKVRAKEEIKLTPGKGADDLLVPMDKLIDSLVKTKRVVAFIKGTRTQPECG
FSYQVVTLLNQTHVDYETVNVLDEFYNPGLREAIKVYSDWPTIPQVYVNGEFIGGADIITEMSEKGEVLG
ALQEAE*

>jgi|Pracol1|15470|PRCOL_00005651-RA
AIPFADDGAVNELPDIPGVYAVYNAEDELQYVGLSRKIGTSVSSHAEIDIRLTPGGNKELTMPIDIQDLIG
LLVKDYPILLFMKGTRSAPQCGFSYNTADYDVVNCLDEEYNPGLREAIKEFSTWPTIPQLYIEGEFVGGA
DIVQEMALSGELKTMCDAAAMASKKK*

>jgi|Chlat1|8811|Chrsp90S08147
MVMAAAAAAFAFSVSVCSSTAASNLATRASSKRQGRPGRGAAFPALPAARVGGGAVGGSVSVKGGSS
RGGLTRRAQAVVDQGGVEEPAASVPRLSEAPSVPAKAAEIPAAPGVYAVYDQAGTLQYVGLSRKIQASAH
VEDLPELTDITRFATVGASATRLDQAWQAWVQHLDAVGDVPPGNVSGNKTWTERKPKATKADIRLTPGNK
GSHVSLTVPVETIIDKMVKEVKVLAFAIKGTRTAPQCGFSHRVMQILTESKVDFTINVLDDEHNPGLREA
IKTYSKWPTIPQLYAHGEFIGGADIVEEMYQSGELKQALSKQG*

>jgi|Mesovir1|4225|Mesvi2744S04823
MACLVRGAVNAVVASNSAVTGVVAKQCNVPLVNSPQGQRKLNI PVSQMRSLSLGQCIGRVRASSSRPVNKV
VRATAEPVADAVPALADVPLSVKEQMAGPSSWKIPLPGIYAIYNDKKELQYIGLSRKLSSSIQAHAE
VPDQCIYVKA VPVQVATRTGLQDAWKVWVQQHLDVAGDVPPGNVSGNKTWTERKPKATKADIRLTPGNK
ELTVSMDELDDKLVDIKVLAFAIKGSR TAPQCGFSAKVISILNELKADYETCDVLDEDHNPGLRDAIKLY
SSWPTIPQVYVNGEFLGGADIMVEMYESGELKKLMSKA*

>jgi|Klenit1|10878|rna-KFL_004270110
MAQAAVARISAIPAFHGTNVRRLTERPPTTFVPLQAANTSCMGAAPCKIVTVNQVPKTLRVARCGKAFLGGEQELGRQ
QRQRPRATRPRGAVAAELASLAETEAVPIKTADGNLQKFPAAPGVYAIYDSNQDVQFIGLSRKVSLSLEGHAE
DLPELVS
SAKVAVVQNVDRSALQAAWKAWVEEYIQAGKVPPGNVTGNDTWTKRKARGAKPEIRLTPGAHVKLTPLEEIIDKVVD
VPVVAFAIKGTRTAPQCGFSYRVLTLLNESKVDYEVNVLDEDHNPGLREAIKQYSQWPTIPQVFKGEFIGGADILEEMV
QKNELKPLVSK*

>jgi|Chabra1|333276|rna-CBR_g22937
MASSAGAACAAGGPFATVQTRRTSATTGCSLDGDSRRCLAVTRASCRTLWQSSVSRCECTCPGLATSSGRSGGVTFASA
AAVCRFRSPTAAFSLSFSRRLCSYNGDDYDNEEDDRMHGRRLYKKKTVFLSDGISTTVGRRVSGRKKVLGEGGKLS
DK
IPPAPGVYAIYDKNALQYVGLSRKVSFSLANHAARLPDLVATVKIAIVDGPDRDALQAGKWSWVEEHVSAGGAVPPGNM
QGNNTWKESSRRPPKADRLRTPGAHVPLTVPMETLIGAAIKGTPVIAFAIKGTRSSPQCGFSHRVLMALNKEGV
DYETVNV
LDEEFNPGVREAIKAYSKWPTIPQVFINGELIGGADLLEESVEKGEFKQLLAAIKK*

>jgi|Bigna1|87450|estExt_fgenesh1_pg.C_200153
MKPRSGSRFLILLAFSVSVVYQQSSWREARASDGNSAVLSSPVGTRRYTLPSKTATVTRKLIHQNGI
RARQQKSCNII SPISRQTIVESSGSQMEFPKLPSVPSVSSDLSTVPDQTVYAVLDESQIVRYIGITR
KVASCVELHRRDVSTLTKFIFHITAPDAPKEELQNLWKAWMQDAIELTGGIPDGNMKGNTMWSLKKEAS
KSSKPELRITSGKGLEDLVPLDQILKGVVESNKVVAFAIKGTRMEPACGFSYQVSELTKLDVDFEIVNV
LDEMYNPGVRDAIKEMSDWPTLPQLYVKGEFVGCDIVAEMAENGQLKELVEN*

>jgi|Symmic1|12005|rna10046
MEVPTLSEAPVFTVGDVLSGEPDVSALPEEMGVYAVYDTGDRLQYIGLSRNIQKNIENHAKAIGLPEAT
DLIASVKCIEMPDESKEVLKQTWEFWLKDHLGDGGEIPVGNLPETAPGADPRWRSRGAQAKPSLNLGGVG
GASQAAMEAVKTAVESNPVLLFMKGTAMPQCGFSAARTSGLLREIGVPPFETVNVLDEANNPGVREAVK
DFGQWPTIPQLYVSGQLVGGLGS*

BOL1 orthologs

>Cre03.g180700
MIAFASLRSFASRAASMAEHQLGPIAGAIKSKVEAALSPTHFKLINDSHKHAGHYARDGSTASDAGETHFRLEVTSDAFKGLT
LVKRHQLIYGLLSDEFKAGLHALSMTTKTPAEQ

>jgi|Cyapar1|12126|g11718.t1
MRRKITEGLSPTVLNIKDDSHKHAKHAAMQALGGDPKETHFEVEVVSAAFEKGKPLDEFKAGLHALALKTK
TPAEAAK*

>jgi|Tetstr1|462346|TSEL_007352.t1
MLRGVFSATSAPGKRPSKLAAAIKRKLKKQLDPVVLRVVDNSAKYGAAETYYKVRVVSQVFEGKPIMQ
RHRLVYGILDDELSAGVKRVAVEARSPSEELKAPADEEAQAADAE*

>jgi|ChlNC64A_1|15183|gw1.38.73.1
VQEAFFQASLSVENESHKHAGHSGNPGGGPDAETHFKVEIVSAAFEKGKPVQRHRMVYQLLQAEIDAGVH
ALALKTKTPAE

>jgi|Chlpril1|5968|rna-gnl|IITBIO|A3770_10p59680_mRNA_A3770_10p59680
MRLAVMAAREKGYMGPMARRILEKVSNAFDCEVVSLENQSHHHAGHVGPNPDGADDAETHFKLKVVSDAFE
GKKLLERHRMVNELLKEELRETGGVHALSLDLKPSKAK*

>jgi|Pico_ML_1|54715|g588.t1
MEERPVRQAIIVNKLQEALHPTKLRVEDESHKHAGHAGNPTGRPDAETHFRVEVVSEAFGGRTLVRHRMV
YKLLDREFQAGSTHWRSKLHA*

>jgi|MicpuC3v2|2592|wlab.211568.1
MLRRAMSTAMRPIYGRIQYKLTDALQPATLVIKDESFGHAGHSGNPSGDPDAETHFNVEVVSEAFSGKLP
VARHRMVYALLDEEIKDGVHALSLKTKTPAEAAK*

>jgi|MicpuC3v2|6012|wlab.170979.1
MAATLLAPRALAPAPAPSPRRARAASTKRAASRPRLAPPPLRAVDPETGEWVSVRDTIEEKIVAALAPTY
VLVRDDSEKHALHKGGQHDGHAGVRSKESHFNVTTVVSEQFEGMTPVTRHRTINAILKEEFDAGLHALQLS
TKTPAEYEKSK*

>jgi|Pracol1|14778|PRCOL_00000501-RA
MSSATAAATARLPYRAIVERLGALKPVSLTVTNDSAAHAGHSGNPDGAADAETHFNVEVISEKFDGLRL
IQRHRLIYESLGDLMETIHALSIKAKTPFEESGAQ*

>jgi|Chlat1|2063|Chrsp17S02531
MLTSQVLVLQRAVAAAAGARLTLRPHPSQHQPALRAASSSSSRHRVAMSAAGPMASEMEKKLTEQLRPS
RLEIVNESHMHAGHSGNPGGGPDAETHFRVTVVSDAFEGKSLVARHRLVYGLLDDEMRAGGVHALALT
KTPAEMAKAGGS*

>jgi|Mesovir1|8864|Mesvi907S08677
MQTFCTAASAAGLRVRLPGALGISDPRETCLQTPGSVAVGQRRMMHPVRCLSSTKVPGPYQTSINDKIIK
TFKPSLFFVIENDSHKHAGLQIVSEAFEGLPTRLRRHRMVYEVLAEELDPERVHVSVLKTKTKPKEMEMGPA
PKRKTSPKQ*

>jgi|Mesovir1|4979|Mesvi3629S00501
MGNPEGKADAETHFSLQIVSESFQGVPIVKRHRMVYELLAEEELGPGKVHALALKTKTKPKELEKGNPA*

>jgi|Klenit1|13179|rna-KFL_006780030
MANNAVGRPLFSLDSKLRAAFNPVNLTVNESHMHAGHSGNPSGDPDAETHFRVEMISEAFKKPLVQR
HRQVYDLLKEELEPGKIHALLKTKTPSEAEKEASKRTL*

>jgi|Mesen1|6426|ME000329S05582
MAAASIRRPMFSTIERKLTEALRPTKLVVTDESHLAGHTGNPSGSVDAETHFRVEVVSQEFTEGKPLVAR
HRLVYNLLDHELKNGIHALTLKTKTPVEQP*

>jgi|Chabra1|352485|rna-CBR_g86666
VEAVSPIFQGMPPVARHRLVYSTLTEELQSGVHALSLVLKTPSELSRKA*

>jgi|Guith1|79024|e_gw1.112.19.1
GALSENFLSCRMFIKNIRLLLRSSVKYPQHIIYQSAALAKRSMSTEGKGPIYSSIWNVIDKQLQPTHIELV
DDSAKHAGHEAMKGSTAKETHFRLKVVSSKFEGLSLVKRHQMVYALLNEQFQQGLHALNIVAKTPQEVEK
R*

>jgi|Guith1|97357|estExt_Genewise1.C_990007
MLDDTGFQASDSLQSKMIITDNSMHAGHAGVQGRSGETHFAVEVVSEAFVGLRPVQRHQLVYRLLSKQ
MEGGMENGGIHALQIQARTPEESQS*

>jgi|Emihul|61409|e_gw1.3.105.1
GPVAASITAKLEAAFKPDVLDVVNESASHNVPRGSESHFKVVLVAHAFASVPLLERHRLVNEALADELKG
PVHALSIVAKTPAQWAKASGAVPPSPCLGGSKA*

>jgi|Bigna1|91348|estExt_fgenes1_pg.C_980005
MTRRLLPLVLLALVAAALLATMNAGPTLSFVGGSIGIRNNSVGMVGGPGMAAMKSCRRNTHAGATVEGR
PIYSAMQNKLTCLKVPKKLVITDDSNKHSEHEAMVGKAAESGETHFTVEVVSDAFEGVGLVKRHRMIYEI
LKEELDAGVHALSLKTKTPKEAGL*

>jgi|Phatr2|11230|e_gw1.5.361.1
RPMYDGIINALQKLQPGELRLVDNSQQHAGHAGAKGLDGDETHFDLYITSDAFDGLSLVKRHKLVYMLLG
EIMPKIHALQIKALTLAEADAKK*

>jgi|Pelago2097_1|528065|fgenes1_kg.88_#_10_#_Locus3199v1rpkm57.49
MFVVAARGLRRASRAPVSRLSTQPAEVVMKAKLVAALAPTRVDVDTSGGCGAMYNIEVESAQFKGLST
LKQHKIVYALLHDDIKAMHGLSLKTSASP*

>jgi|Vitbras1|2634|Vbra_11822.t1
MATKGPIYEAIERKLTEKLTPTKLEIVDESHLHAGHAAMKGLDRKQETHFKVMVESAAAFANVPMIKRHRL
IYEILDDEIKAGVHALSISAKPPDPGT

>jgi|Vitbras1|18143|Vbra_17762.t1
MGPVYDAIERKLTEELSPTSLEIVDESAGHADHLGMQEHPKFAGGETHFRIDVTSEMFEGLSRVKRHQL
VYKILDEEIKNAIHAISINAQAPE

BOL4 orthologs

>Cre09.g394701
MASLFAFPALALPRIGHQEACSSSITFARFAARTARRVAPATPVPSAEVSRRSQLLVSRGIPVSGPGATSSSSGGVDGQVTS
VLMESMKAKICEALETDDCSISDVYGDGRHVSIDVSKLFEGKNSMQRQRMVYKAIWLELQETVHAVDAMTTKTPEEAK*

>jgi|Galsul1|3163|XM_005708841.1
MSSSNDTVVERVEQKLREALSPKFLQVTPAFDDPNGSHVSIRVVSEKFENLTSLKRHQLVYKAIWDELQ
PVHAVDSIDAKTPKEVEQP*

>jgi|Cyapar1|13349|g12896.t1
MSAFALNLAAVAFPRGHSVPRIRPAAVAVRSVGINEARSFFGQTIFAAAAPRRTFVAQPPRAQADSAP
AHPTVAQVSKKLTDALAPKKLQVIDNAGDASHIVIDIIVSSQFEGKNVVARHRMIYKAIWDEMQRNQVHAV
DSIIAKTPAEAGL*

>jgi|Caulen1|96281|g2521.t1
MSSELIDSMRSKIQNALNADSVEVVDVEGDGQHVTISVISSTFEGKSSVDRQRMVYKAIWYEMQNVVHAV
DNMITRTPTEASSSQ*

>jgi|Ch1NC64A_1|145612|IGS.gm_9_00419
MAQLFSGRIITAMASASRAAAGSRAWVARPLAAAPLVAARPSQLLIARASSMDNNIGSALMDQMRGKIQA
ALSAELVQVEDMQGDGRHVEIVVVSKEFEGKSAVNRQRMVYKAIWEELQETVHAVDAMVTKTPEEMSG*

>jgi|Tetstr1|458505|TSEL_044911.t1
MARVVLSAAGCRPAVPSRAAQAPRRTAALLPLLSRPGTHSHYAFRQANAVALAIPVSGPGQQPDAGGDS
GGADGQMTAELMATMTSKIGEALETESVNIVDVYGDGRHVSIDVVSELFEGKSSMQRQLVYKAIWFELQ
DAVHAVDSMTTKTPTEVA*

>jgi|Tetstr1|435952|TSEL_024833.t1
MARVVLSAAGCRPAVPSRAAQAPRRTAALLPLLSRPGTHSHYAFRQANAVALAIPVSGPGQQPDAGGDS
GGADGQMTAELMATMTSKIGEALETESVNIVDVYGDGRHVSIDVVSELFEGKSSMQRQLVYKAIWFELQ
DAVHAVDSMTTKTPTEVA*

>jgi|Chlpril1|6774|rna-gnl|IITBIO|A3770_12p67740_mRNA_A3770_12p67740
MLASFATKRGATGGGPGRIRGEHKGVLGDDLGAAGRSRKRSSPRGRSPTTTTGLRQRGWLATSAGDDRRSG
QITEDLMNSMTEKISAGLNPNGDEGERVSVTDASGDGSHVYIEVVS RVFEGKTSVQRQRLVYKAIWEELQ
EAVHAVDGMSCKTPEEV*

>jgi|Pico_ML_1|53206|g400.t1
MRTKIESELEAQVVRVADVSGDGRHVNIDVVSNAFEGKSSVARQRMVYKAIWQELQDVVHAVDAMTCKTP
GEIKN*

>jgi|Pico_ML_1|52689|g3362.t1
MRTKIESELEAQVVRVADVSGDGRHVNIDVVSNAFEGKSSVARQRMVYKAIWQELQDVVHAVDAMTCKTP
GEIKN*

>jgi|MicpuC3v2|9446|wlab.223966.1
MSHHVRRGSGALLAAARQLARARAVRASASQHRHNFVTWQRSETPGGSRAETRGFASSPASRAGEFDRPE
SGLIAGMKSQIQTSLFATFVDVVDMSGDGRHVTIKVVS KAFEGKNPINRQRMVYKAIWEEMQTDKVHAVQ
GIVTQTPEEAQ*

>jgi|Pracol1|21015|PRCOL_00004611-RA
MMWFRLMRARFRGNARLISAAAAAGSGENMSDNNITGAVIEGMKVKITEALEAQSVETQDAYGDNQHV
SIDVVSTMFEGKNSVQRQRLVYKAIWQEMQETVHAVDAMTCKTPPEEAGK*

>jgi|Chlat1|6644|Chrsp49S06122
MSVSAGLRVLGGARRVLSRPVASAVSAATNAGTGPSTSAAGSQLGFVGLRHRSGISAATPSQQQRSRR
SLVTRAEAGGDASAPLDSPTIANMREKISSELEATHSVQDAYGQQRHVSIDVVSPLFEGKTRVARQRMV
YQVIWEELQNTVHAVDSLTTKTPEEAAASSPPKRL*

>jgi|Mesovir1|4462|Mesvi301S05029
MLSTAAMVSPVTVAVLARPAASSASYAPAMANPCRAMPTKALRASGVRGLSLKQSLSAKPLYMHRRNVT
CSIPVSGPIDAPLIQSMRQKIQEALPEKELVVKDVSQDQRHVSIVVADAFEGETSVNRQRKVYKAIWEE
LQQAVHAVDNLVTRTPPEEAASKPAPNFA*

>jgi|Klenit1|3928|rna-KFL_000810225
MAASSSDVGDGGQLTSTMINDMKEKIQEELAVSVVQDISGDGRHVSIDVVS PFEGKSSMNRQRMVYK
AIWQELQGAVHAVDAMTTKTPTPEAGVGQQPRQ*

>jgi|Mesen1|8083|ME000434S07328
MATLCYGAATSTVLALGKLPGLGRDKAQSLSCIKGVANNKSLVSFQPHRFQQRSGQLLQVRAIEVSGP
GAVNSPLMQDMHEKIKQQLAEESVEIKDAYGDGRHVSITVVSPLFEGKRS LDRQRMVYKAIWEELQGT
VHAVDTLVTKAPSEVV*

>jgi|Guith1|100004|au.3_g1106
MSTDASAPEEDTVMSTIEKKITSELPVKLEVIPAYGDPNGSHVTINVVSQAFEGKRAVQRQQM VYKVIW
EEMQGAIHAVDQMKCQTPAEAGM*

>jgi|Emihul1|452727|estExtDG_Genemark1.C_670140
MASALPFTVRAARTSRALPSLSPVAAVGWRS LGVRMQSGINDPTVVDRCTAALEAALEPTSVKVMGAYDD
PNGSHITIECVSPKFEGKRSLARQQMVFKAIWAEMDGGGTVHAVDTMVLKSPSEVKDE*

>jgi|Phatr2|33016|fgenes1_pg.C_chr_2000557
MWTASRAFRPVVRVAWARNTATRQFSTANGPDTTIVQTCQQKIQDALSATDVKVTGAYDDPNGSHISIE
VVAEAFAGKRPVQRQMVYKAIWEELKGPVHAVDAMICKTPDEV*

>jgi|Nanocel1779_2|579356|fgenes1_kg.4_#_1424_#_TRINITY_DN5163_c0_g4_i1
MTMPCRRSPSSVLLFLACFMALTCQGFVSPSSSLISFRRTARTSLTAAADGAVAGEVGAVESCRIKI
QAALSPVELVVKSSNDPNGSHIALRVISAAFEGKNRVQRQVVYKAIWEELSGPIHAVDELVCRTPAEE
GM*

>jgi|Ochro2298_1|397071|gm1.8186_g
MAARSLSLRLLRQAARPHALASASRWATSAPSLSAGLRPTPFLSSTRLFSSEGEDTVVSTCTKKIADLLS
PVHLKVTSSNDPNGSHIQVEVVS AQFEGKNTLARQRLVYKAIWEEMNGPVHAVDSIVAKTPKEVGME*

NFU1 orthologs

>jgi|Chlre5_6|8759|Cre17.g710800.t1.2
MLLRNKTGPAQVGRQAAHATIQRPVLPVPRGLAVRALATAAPPAPIDDKTVPEGHKGHLHNYLYSGDADKEHEQDAYKVRKD
EDEGGVMPVPAYLDRDGEKPLGVYCVYDSAGAAAYVGYSRNMVLAIKGHLARVGPARCAAVRATVYGNKAMASRANLE
RGVASWVEEQLGGVPPGNADPQQRALWEGVAATSAASGDESGPALDVSHMSPAERAAAYEEKRTKMRKAMGEKMEVGDKP

GGQPNEEPPEDLATRRAKLMKAMDRGDWSAVIDQQTQEAVAAQQQDQQQEGSSSSSSSSGGGAAGTDGVVSPFIREATMDE
GDLIKERQQAGSSGRSGRELSVEAVTAALDEVRPYLEADGGDVEVVEVENGVVYLRLQGACSSCPSQSATMKGGIERAIR
NAFGDQVDRDILQLDAKEPAATAEAVDRQLDMLRGAISNLGGSVEVVSVEAGVATLKYKGPVKPGKGIQNMVRDKFKKQLR
EVRLVEA*

>jgi|Galsul1|656|XM_005707587.1
MLLFLHSCGPIQVGCIFYRTRNISHTKKGKLLVVPQCWKASVNKDFKDAFPSSNYIIDDNSIPEGHRSIHETLYSRGEEEH
KTPSKQQVKFRSDGSNWLPVDLFLNGMEDEKVGAVYAVYDANRDIRFIGISRDCVCFALKAHREALSDRDVAWLRLKTWDF
PRRVEMEQQCKDWIAELGFTPPGNITNEESKLWAKSIKEATQLVASDVEPERKSTYEEKKFKLRKAMADPSLIDEESRE
IFERDPTIHAISEGDSKVIEQQSNKVLQDAVTS PFESVEVVANNRVDERSNSNTAESGSFLTVENVNVLVDEVRPYLE
SDGGENVKVLSVDTNRNVVLLLLQGACGTCPSSTTTMKLGIERILRQRFNIGEIVAQSEVASVTTIPLERCESLLEEIRP
AIIIGLGGISVSRVENNQVFLLYQGPDKIKYGIELALKEKLGSSVSIVFE*

>jgi|Caulen1|88997|g4355.t1
MLITNNAVLRPNGVLVPCRIQSCCKLRRSRYRLRTQVSSGGGPPSEALSSQNGVISDSNVPEGHRGLHGALYEGGGAELH
NSSSYEFRDGEDDGEAVMTVSSYLATREGERPLGVYAVYDKSHNIQYIGYSRNIVLALKRHVEKLGEDLCYYLRVMVMFMN
GAEAHDDGGIGYTFREGEDDGCSSLATAGYVAAREGEKPVGVYALYDDKRNQLQYVGYARNVVLA VKAHLARVGEERCWVR
KERRLRLIKAVEGDNWSAVVNEQTNETIEGKPQEIVSPFAQAPIHRSIGQETDHPDEEMTTESVDTALDEVRPYLIADG
GNVEVVNVEDGNVFLRLQGACGTCPSSTATMKMGIERVLKTSFGDQLIEVIQVDSQDTSASIQSVNDQLDVLRLPAIQGYG
GSVEVLKVEDGVCELKYTGPAPIGMGIQAAIRDKFPDIHDVKLIS*

>jgi|ChlNC64A_1|58507|estExt_fgenes3_pg.C_170017
MSARSLGSPPPGASIGSHRRAAALRPGPGRQPSTGASGGRRRAPLAHAGGAPTSPAGPVTDAAVPEGHQGLHGFLYEGEG
GAEAHDDGGIGYTFREGEDDGCSSLATAGYVAAREGEKPVGVYALYDDKRNQLQYVGYARNVVLA VKAHLARVGEERCWVR
VMVFANKAMQSRAALQREADNWLAEAGTLPPNGNAEAEALWAQGVGASGASVDVSLMSADELAQYEERKLKLRKAMGENLH
DDVAGEAEDSRTRRLKLIAAVEGDWSEVIQEQSAEAVVSGEVGSARQQAQQAGGEALAPPAAAPPIVTPFARASVHRAV
GNSSAAAGAGGEAGRPLEMTVEAVDAALDDVRPYLIADGGNVDDVAVEDGRVFLQLQGACGTCPSSTATMKMGIERSLKA
AFGKQLVEVLQVGGQEDNRATAEGVDMHLNMLRGAVGAYGGSVEVVGVEQGVCTLHYKGPEAIGYGLRAAVRDKFPDLVE
VLMMDPDTGEPIKFAQ*

>jgi|Chlpri1|5480|rna-gnl|IITBIO|A3770_09p54800_mRNA_A3770_09p54800
MVVAMTMTMSGPPGRCLRLRAVVSDDNVPEGHANLHNELYGDKSEGKAHAÄVVVNDDEVLEGELEDGTTLFVTSEWLE
KRDGTPKCGLYAIYDEGESLKFGVYSRNIVLAIRGYTEVDAEGKPDFFKHCRVKLISGKRLTSRAYMQQRDSWVEDLTG
AKGGEGEGVALSSGGQGPQPAVKRSEAEAEATHEEQKLMRKAMGDSLAPASGQETEDKKQRRNLNTIKAVEGDWSEVIDR
QTRETRQSPATEEGSAPADRVSPFAKPGSAMPRGGEAGGRSVMADATVNEVLDEVRPYLIADGGNVSVVAVEDGVVKLA
LEGACGSCASATATMSMGIERALRNHFGLSKQVVRVDPGAGGSGDEGVAPSAVDAHLDMRLRSAIENYGGSVRVKSVSDGA
VCVVQYKGPPLSKGIEAAIKDSFPALRQINFVSFD*

>jgi|MicpuC3v2|101|wlab.202714.1
MPSHAVAVAAARALHRAAPPSASTARRVAAPSGGATPRRSRTLPPVAPPRASSSSSSSSSTAAAVDAETVHAIIRSGPPDDG
TAIVPLRAWLEARDGDPAVASVYAIYGADDVMRYVGYAKDASRAVRSHLERQGDDAASKVRVAAFANKATATRANLRAEQ
DRWIGEWVVSRRGAETIEIPVGNQAAGAAARWTLTPAESWENEEKPAPAEAKANEAGTPSVGADGEVISPYAAEAAGADAE
EEALDPNRELLPLTVENVDKALDEVRPYLIADGGNVAVVGIEDGVVAVRMSGACGSCSSSTATLKGGIEKTLRRVFGGEN
VKEVVNLDSDEPGSALTLSKEAVEAHLEKLAGAIHNYGGSVKLLEVI ESERALVLEFSGPVALAQSIASSIKGKFPLVAE
CKIKQV*

>jgi|Praco1|15791|PRCOL_00005187-RA
MPCARAEGAARAARAAAAPRRPRLGRRRPSSSVAAASGAGGVPTPAPTQGVATDDNVPEAHRGLHDFLYGEAGGDVHGG
EARAAA AVAGEGCGPADPALSGAAPEDDGRSCVRADAWLAARDGAKLAGVFAVSDARGDTQFVSYARDVVASVRALRGR
LPRHVA AVRARIFAASMARREELNALADAWLREELPHVPPGNAQGEDASRWRGHDSATALSGAALASYEDNRLRLRRAMG
ENLADAEAGESASLREERRLAML SAVHGDDWSEVIGEQTA EAVGLEDPGGQREQAGGSDTGGAPIAS PFSAGGAGEATAAA
VGTGDGDDGHGAATGGAPDPQSQPQQPALSMSTADECLDAVRPYLVADGGNVEVVG VAGGVLLRLTGACGTCPSAATMS
MGLEKALRATFGEQVQEVVQVDGGGAPVDTSPGALSGVTIASVEAHLEVL RPTVEKYGGSCRVSVSAGGDSVAIEFEGP
APMAMGIEAALKDKFPTLKHVDVDIGG*

>jgi|Chlat1|7988|Chrsp7S07745
MAAATVYLGVGVGVGSETVVSHPRQQQAACSSAAATGASSPQLGLVLRGRRRRRHGSHAGERLSTRAGAAAPAGSTTTE
RSKDG DASSNGLSRRRPTQE QD TVHAPRAAVDFADETYSQQVPEGHKELHGFLYGDGGA EVHGDASSTSVYVPRDGEDDG
AVVVELEQWASVRELAKPAGVYAIYDEADTLQYVGYSRSM AVAVRAHRARAGPHKAAKVRARVVRDPAMITRTRLEAVRD
EWMTDEF GHVPLNGSGEKELWEGVGLAAVMTDAERAAYEEKKLMRKAMGENLYDDVEGEAEDSKQRRNLNLLRAVEGDD
WSGVIDGQTKETVAIPSRTPIPSPAPVEQIVSPFARANAPSVGSTALPMLELT PANVDKVLDEVRPYLIADGGNVEVAGV
HDGVVALRLQGACGTCPSSTATMKMGLEKSLRAHFGDILKEVVQVDKTDITATVQKVD AHLDMRLPAITNYGGSVEVVS
DAAAGTCEVRYSGPPP IGMGVQAAIKDKFPDIRQVTLQ*

>jgi|Mesovir1|4293|Mesvi284S04891
MAAISATAVNAAVCARIDSISAPARSNRNVHGSWRASSIRRLGESARLGSRN TGSMP LVVAASTAGEAMGITENEKARLN
SLGGVMTDET VPEGHQGLHGFLYGEGAEDA HNGKTYAVRPGEDDGAAILPLEAWANARTGERPAGVYAVYATDGT LQY
VGYSRSVVASLRQHLDALGGTMCASVRVKVMTDASLITRARLEAEVEQWVQRLNGAVPPNGAQ AHLWGRKQATPSAMS

PEELREYEEKKLKM RKAMGENLYDDVQGETPDAKTRRMRLQAVEGDDWSSVIDEQTQATIAGGAAAPAAAGPASEQIVS
PFARPGAPDMNGPAGAPSAAPRPMNKD TVDLALEEVRPFLIADGGNVEVVGVEDGIIALRLNGACGTCPSSTSTMKMGIE
RSLRAAFGDQLKEVVQVDKAETGATVESVNAHL DILRPAIVAYGGKVEVVSVD AIRGICEIRFEGPPPIGMGIRAAVKDK
FPELRKIELL*

>jgi|Klenit1|8719|rna-KFL_002800020
MKAVQQQLSTQAFQACFSSSVSGHAAQSPANSRNAPRGVPRRQPRWQRA P VCRAAGEGVTTLPKGPSAGGSPSSAVRPDLT
ATLGAVVSDAAVPEGHKGLHGFLYGDAGADVHDDGA AFQVRGEEDDGSALVGLQQWLAGREAERPPGVYAVYDQGGDLQY
IGYSRNLVLT LKQHRARVGPERSAGVRVKVYANAALVSRARLEEERQAWLAQQPVPPGNGPERDLWEGTGPATAVMTEAE
RGAYEEKKLKM RKAMGENLYDAVEGEDADARTRLRLLA AVEGDDWSGVIDGQTQETLAPPQAQRAAAAAAIEAAKAAV
AAKQIIVSPFARPNAPALGAPAAEK PATVMSVESVDRVLDGVRPYLIADGGNVEVVGVEDGIVALRLQGACGTCPSSTAT
MQMGIEAALKAHFGAALQQVVQVDKQDIGATLEAVNSHLDILRPAITNYGGSVEVLAVDTLKGSADVRFKGPAPIGMGIQ
AAIKDKFPDIKKVNLVE*

>jgi|Mesen1|3635|ME000020S03162
MRAEVAQQSTVTPSTSD EAAQRQRLTNLGGVVTD ETVPEGH KGLHGFLYGDGGADVHDS PNGGYTLQEGEDDGTQVVPFE
QYVGARDAQRFAGVYAIYDQQQSVQYIGFSRSVLSLKGPNAGTAAMSEEEERTQYEEKKLKM RKAMGENLYDDVDGETL
DSKERRMRL LQATEGDDWSSVIDGQTKD TLDPRLATPGTDAAA V GLEAATAEQEEAAPPEAQIVSPFMRPGALPLGE
AAAAAAPVEMS VSSVDKALDQVRPYLIADGGNVEVVG VENG VVFLRLQGACGTCPSSTATMKMGIERSLQATFGEQLKEV
MQVDKIDISASITAVNAHL DMLRPAIQNFGGSVEVISVDSSRGTC EVKYKGPPPIGMGIQAAIKDKFPDIRKVVLLDA*

>jgi|Chabra1|324028|rna-CBR_g4584
MEACMGSAFLQAQSQAGVMVASSSTLVGVAGLMATAQERLQCQTLPGRTVLSSMARSPTLAPSRRVQVGLFGSPQKGS
LSAVTVIRCSKPEAQPSVQT AISPDMAERLSSLGGIVTDAAVPEGHVGLHGFLYGSGGA EVHDGGAKT VVGRPGEDDGS
MLAFEDWADPREGQRPAGVF AIYDGD ESLQYVGYSRNIVVSLRQFRMRMGVEKTMVVRVKVLT DKAMLTRAKLEEERDRW
MGELTGGNLLPPGNAEEKG VWE GTD GSGNGSALIAMTEEERAA YEETKLKLRKAMGDNLYDDVEGEDEDSRQRRNL LRA
VEGDDWSGVIDRQTR ETQAGPAPSAPTSTAASEKPAEAGQIVSPFARPGAPNIGSAGSAVAPQVPLELTAENVDKVLDEV
RPYLIADGGNVEVVT VENG VVALRLQGACGTCPSSTATMKMGIERSLSKFGERLQV IQVDKQDIGATVESVNSHLDIL
RPAITNYGGEVEVM AVNTQKRSCDIRFKGPAPIGMGIQAAIKDKFPDILVNNLLEG*

>jgi|Ectsil1|31965|rna4552
MGRYSASLLVLAVLSLEASSFVFPQLASTAAAVSRSRAPHASRSSSIPHRTAAGRGVVLQSTTDETTETAEEVRLED RDV
PQAHRLGHDFLYSNEEDHGATASDGSVELFDGTQLYEVSKWLKD NAESKHA AVYAVVGKDEKVN FVGVRNVALSLAHL
AAEGESMVHMLKVRSFKF PKRDKMEALKAQWIEECGSTPVGNIKGSEWATGSKDAQNAMSPAQKEAYEENKLKM RKAIAD
PGLVEETE QMSSEDIERRKRLESAVEDDDWSSVIDGQTKLTEE PKVEVDAPTATIISPFS SDGMPVGVGPDGMAAPPLPP
GIAAPVPDAPPTPPASECNDLEFTLENVDKVLDEV RPYLIADGGNVRVMGV D IRRVVKLALQGACGSCPSSTTTMKMGI
ERVLNENFLNMGGVEQVDEASGNAMAEATTAVVEAILEPLRPAMVAMRAKVEVLSVLDGHVKLTYSGHRKVAYGIQMALL
DNPLIQSIDFEYVD*

>jgi|Nanocel1779_2|604355|fgenes1_kg.29_#_125_#_TRINITY_DN4477_c0_g2_i1 (bad
annotation)
MMRRCPTTTAAAMVLL LGGISFFQQQQLVQAFVLVTPKTAPPFLLYHQHGNPPCRLESSASGVS AVTAVAADDNVPEGH
KGLHEFLYGDNDHGATSA AAAAEAEAFTRTQTPLPVA AFLDRFKDVR LA AVYGVYDAKQTLRYVG VTRNVVSLTGHLQV
HGA EVVSSVFA LQSF KFKREEMAAQETWLAELSAAHIVPDGNREGGKGMVVGAKSIRKRLEQSGSGSTAEERAA YEDTKT
KLRTAMADPTLAD ELDLDEEEYGEAEEDERMFRRAQLK KAVEQSDWSDPSNDQEASTTGGVAARAAPAAPAAVIVSPF
NGGDSSSPSVSDGLLALT PENVDIVLEEVRLPHFRWDAADRSSSTEEIGGKLEEILPAIRGLGASATIEDVSEEGVVS
LCYKGPEKVKLG VVYALKGQPGVTDVK

>jgi|Vitbras1|7107|Vbra_20791.t1
MECPNWIFALVLVFLSLWRIRVDARRISAHTRGAPAAFLHLMPQLPTFGTDRRTGRQRGRHAEAAAAAASSQSPVDELIA
EQPSLSPIQTDANVPDGHRLHSFLYDDGKEDDAHAAKPA GSRGRVRFDGVNMYGVDRWLGMHREERVCGVSVLDEDRS
VQYVGVS RNVAVALESHRRQTTPPGTVAFVRIKSWDYPRREEMDALRDEWIRESGRVPPGNEDASLGFAETIKDASTDGMK
DDE RQQYHDNKLKLRAMADNTLIDETPNII EEEIMNREALLHALKDDDSAVIDGQTNATLNPQLINAQ LQEALKIGRD
DTSTDTQGEANVVS PFAPGRQGDDAVAARTAEPAESLPLTVDNVDRVLD D V RPYLQADGGNVMVAVDPTHKTI SLALQG
ACGNCPSSQMTLKMGIERLLKETWPD LNDVVPVSDPSSPHAAPAPLPFGVELIERHLEEV RWGLNGMGASCDVIEAKDGR
AVLRYAGPD PGRITFGIELMLKDAVPNNQLQVVEFVTDEPQTD DRQQATAA

NFU2 orthologs

>jgi|Galsull1|6376|XM_005702978.1
MLGFLFGFC SYKPHHVAFTCFKNRQRF SYQPTVHKIRRLQISADSALPLTKENVEQVLEELRPYLIADGGNVSLTGIDGA
TVRLTLEGACGSCPSSTVTLRMGIETRLKEK IPEIEAVVQEETMGPELNEQNIDSVLDEV RPFLKIAGGKIDL VGIYGTD
SPSPSVSLKMSGGGA AVDSVRLEIIHRLKRNFPKLVNVHYVKA*

>jgi|Cyapar1|6920|g6698.t1
MEASFVVHAPAVRLRSQAASSSRNSAACCRHAPANDISARRAVGPRRLAVEESRSFFAGRQFLRAAVAVRPRALNTTVRM
AAETLPLTPENVEIVLDEV RPYLIADGGNVQLVEIDGRVVRLQLQGACGSCPSSTMTMKMGIERKMKERIPEVAEVNVQ
EAGMELSTENVDKVLEGIRPFLSSEGSTIEVDKIEGSAVTLRMTGSGSLVMSLRMEIMRRVQEKIPSVTRVAFSDPDAK*

>jgi|Chlre5_6|19236|Cre18.g748447.t1.1
MASANILRAQAGASVSPRCATLAASLLPARATSSISRQQLASVSSTPARVWGAAPSTRRGARVVVRVAVEQDEMMELTPE
NVEKVLDEVIRPYLMADGGNVEFMEIDGLVVKLKLQGACGSCPSSTTTMTMGIKRRLMERIPEILDVEQVTEESLGLELNS
DNVETVLNEIRPYLVGTGGGGLLELVAIDGVIVKVKISGPAANVMTVRVAVTQKLREKIPGIAAVQLV*

>jgi|Caulen1|93689|g8283.t1
MLAVCGSVYCLQIIREPLRNSTPKTNAFCLSYGPSSAVLRAPGPVTPLRTHRYKHSALASYRSPGVFASSVGETLELTA
ENVESVLDEIRPYLMADGGNVELVEIDGLTVLLRLQGACGSCPSSTTTMTMGIKRRLMEKIPEILDVEQIIDEQEGLELN
PYNVDAVLDEIRPYLVGTGGGELMFHRLLEGPIVTIRITGPAANVMTVRVAVTQKLREIRIPAIAAVKLV*

>jgi|ChlNC64A_1|144603|IGS.gm_7_00545
MTVGSSGAVCQAAATDTLELTEENVETVLDEVIRPYLMADGGNVEFVEIDGPVVYLRLAGACGSCPSSTTTMTMGIKRRLM
ERIPLGGSRAWPYLVGTGGGGLLELVELDGPIAKVRITGPAANVMTVRVAVTQKLREIRIPAIAAVQLVN*

>jgi|Tetstr1|440027|TSEL_028388.t1
MAAMAATRTVLCGSDGRPTASRTAQRVGCAPPPRSAGLLARAPAPTALRSHRSSRGGVAVAAARSFVLAADTDTLEL
TEENVEIVLDEVIRPYLMADGGDVELVEIDGLVVKLRLKGACGSCPSSTTTMKGIIQRRMQERIPEVMDVEQVMDDEDGGGL
ELTEENVDAVLEEIRPYLVGTGGGELMCKGVDGVIVKVAITGPAAKVMTVRVAVTQKLREKIPGIAAVQLV*

>jgi|Chlpri1|6335|rna-gnl|IITBIO|A3770_11p63350_mRNA_A3770_11p63350
MVATMTATTVARLPKAAARRALREESGVARGEHSSARSCLGRSLSPWPEATPSVSSGGKDRGQSQVLAALGGGRWVCGM
ASAEESYTLELTEENVETVLDEVIRPYLMADGGNVELLEIDGLTVRVKLQGACGSCPSSTTTMKGIIQRKLMEKIPEILEV
EQVMDEIEGLDLTEENVESVLDEIRPYLVGAGGGAIELEEIDGPIVKVHLSGEAASVMTVRVSVTQKLREKIPLIAAVQL
T*

>jgi|Pico_ML_1|50958|g2072.t1
MAEPQVVVDGARTFAAFYDVFAPLAARIAAEDEELLHAFLSSKRVTAPGARQERCSRARTDLELTEENVELVLDEL
PYLMSDGGNVELVEVDGLTVKLRQACGSCPSSTTTMTMGIKRRLMEKIPEILEVEQVEDANAGLELTPENVDSVLGEI
RPYLSGTGGGELELEEIDGPLVKVRLSGPAADVLTVRVAVTQKLREIRIPMIAAVQLV*

>jgi|MicpuC3v2|4144|wlab.213321.1
MATAFARVATATAAHRGSTSAALAASAAPGPRFAPARVPASGPRPRFAPRHATDSPSSTSSTTEKLELTAENVEKVLDEV
RPYLIADGGDVELVEIDGLVVRKLNGACGSCPSSTVTMRMGIERRLLEVIPEIMEVEQVTEEEVGLLENEANVEATLDE
IRPYLKGTGGGELELVDIEEPIVKIKLSGPAASVMTVRVAVTQKLREKMPISIAAVQLL*

>jgi|Pracol|19513|PRCOL_00002783-RA
MADGGNVELVEVDGPVVRVRLNGACGSCPSSTVTMQQGIQRRLMKEIPEISEVEQLIDEQEAGLPLTEENIEATLDEIRP
YLSGTGGGELALDSIDGPIVKIELSGPAADVMTVRVAVTQKLREKIPISIAAVQLV*

>jgi|Chlat1|7859|Chrsp66S00579
MGFAMAVSVGTVVLGSGCGVSSLRGRDGAGAAASSASTSRQSAEATPRQAWLRGGLSGLRHGSSTSQRFRFPTSAGG
RRASTSMTAASFDTLALTEENIESVLDEIRPYLMADGGNVELFEIDGLVVRKLQACGSCPSSTMTMKGIERRLMEKI
PEITEVEQVLDEETGLELNAENVESVLSEIRPYLTGTGGGELQLVKIDGPVVKIRLSGPAAGVMTVRVAVTQKLREKIP
IAAVQLL*

>jgi|Mesovir1|8422|Mesvi827S08310
MLGIGFLCTFTARVRRVAVASSAGLEGLALTAENVESTLDEVIRPYLMADGGNVELVEIDGPIVRLRLQACGSCPSST
MKMGIERRLIEKIPEILEVEQVMEVVKGMELNEANIDIVLDEIRPYLVGTGGGDLQLVKIDGPVVKIRIEGPAADVMTVR
VAVTQKLREKIPMIAAVQLL*

>jgi|Klenit1|2738|rna-KFL_000500450
MDKAQGFKISRHPTPSIGARMSSFAGTRMRSPGVASVRKGIRGAAQQGRAQVMDVEGVPLTEENVEKVMDEIRPYLMAD
GGNVELVEIDGLTVRLRLQACGSCPSSTMTMKGIERRLMEKIPEIQAVEQVLDEETGLPLTDENVENVLSEIRPYLVG
TGGGELTLVKIDGPVVKVQIDGPAANVMTVRVAVTQKLREKIPGIAAVQLL*

>jgi|Mesen1|7241|ME000373S06309
MAGVAMIGSLAGLTSSQVAHCAGTYGSGATCSSRDATAASTALPSHRQLRVLRADDALSGLRVARGWIPLNPLRNMLRQK
GSPEQHRVVAVATQQGEVYTEGLELTEDNVEAVLDEVIRPYLMADGGNVELYEIDGLVVKLKMQGACGSCPSSMVTMKG
ERRLMKEIPEIVGVEQIMDQETGLPLTEENVEKVLSEIRPYLSGTGGGNLTLVKIDGPVVKIAIDGAAAGVMTVRVAVTQ
KLREKIPLIAAVQLL*

>jgi|Chabra1|322077|rna-CBR_g3282
MGISSAAWAAGTLPNVGASQAAGCNSPVGGTTASSSSASVETRCATCHALTARRGGTLQVVAARSNGASACRRLGAGYGE
NDLPLGCSGRHLFPSSSSSSSSFSFSPFRRTTNPSLDGLDKIERDNASLSPLSSAARRRPRTSPVLAMVSATAGTEGLELT
EQNVEKVLDEIRPYLMADGGNVELYEIDGLSVKLLKKGACGSCPSSMTMKGIERRLIERIPEITDVEQVMDDEETGLPL
TNENVEKVLSEIRPYLVGTGGGELTLEKIDGPVVKVRIDGPAAGVMTVRVAVTQKLREKIPMIAAVQLL*

>jgi|Guith1|106027|au.22_g7129
MEWVWRPLVLRSAKRSLLALKDSTDIVSPFDPSVDPLLPKETEESEPTDLDGGLPLTWENVEKVLDTMRPYLMSDGGNVKI
ADIDGGIVRLKLEGACGTCPSSTMTMKMGLERGLREKIPEIVDVVQDLGDGGPELSPDSVEKVLDTVRPFLKVAGGSIEL
FDLRGVGGMQPVIIKMTGTSAALRSVKNEIVQRLQRNFMLAGLRIEWTD*

>jgi|Emihul|452262|estExtDG_Genemark1.C_590098
MASCGRARCLLLSSWLHSAGSLRVPÄVSAVPRVQHRAGAPRCSDPVTSPFEASASNGGETLTGPLPLTAENVDAVLDE
MRPYLMADGGNVAVAEIDGGVVRLELQGACGSCASSAMTMKMLERGLREKIPEIIAVEQVAPDGPQLTEDGIEGVLEEI
RPFLKMAGGDVELVELIETGVAPSCTLRISGSGSTINSVRAEIAQRLKRNFPSLANVMWDAEGP*

>jgi|Bigna1|48123|estExt_Genewise1.C_220195
MAAGLGKLRVSGFLLLSASVVFVLLSARSRNYDEKLQIGTTMKAGAQNRVMQVGRRSTIHQPLKALKMCGDFKRGVVPV
RAIASGYRPSVFTVERLDLTPANVEVVLDEVVPYLQADGGSVDVFEIDGFLVKLRKLGACSSCASSSTTMQLGIKRRLQE
RIEGVGVEVDVGVTKEKMEVSLENIEKILDEIRPYLVGASGSLDYVGGLQNNVKIKITGPAKIMTVRVAIQQKIKEK
FPNMVRVELVA*

>jgi|Phatr2|46593|estExt_fgenesh1_pg.C_chr_100324
MKTTAAFLSLYGLSMASFFVPTAYRHPKTVGTTLLRDQIVSPFDSSERDEGVSATTELDAPPTKLVGPLDLTWDNVEAVL
DEMRYLIQDGGNVIISDIDGPVVKLELQGACGTCPSSTQTMKMLERGLREKIPEIQEVIQAMPEGPELTSEQVDVVL
GVRPFLQVAGGSINMDRIDGVDGLQPTIWLKMEGSSASLSVKLEIAQRLLRHFMVAGLQVQWIV*

>jgi|Ectsil1|20192|rna11414
MPLTMRGSSSFVRGIAVAAVAATSLQNSRAGAFAPPAFVPAGPAYCRSCSSSSSTSALARTAAASSSCSSAQQQARGWGAA
PRRTGAAEPRRMSSTDIESPFATPGMADMGDEEEDPEALLTLTENVETVLDEMRYLMSDGGNVRVVEIDGPVVRLELE
GACGSCPSSTMTMKMGLERRLVQRIPEISEVVQSIPNGPELTVENVEKVLGDGVRPFLSVAGGSINIQLTGVSSIQPVIT
LKMTGSSASLSKIRMEIMQRIQREFMSSSLRVEFAP*

>jgi|Nanocel779_2|560105|estExt_Genewise1Plus.C_300063
MARDLSCGAASTFVLVCLVLVLSLAEAFIPAIPKTPATTRRFSSPETITSPFESGVSSGPNDDFLPLTFENVELVLDEM
PYLMSDGGNVVISEIDGPVVKLELQGACGSCPSSTMTMKMGLERRLREKIPEISEVIQAMPEGPALTTENIDGVLEGVRP
FLMVAGGSISVKEISGLGLQPRVLLHMEGNNQSLFSVRMEITQRI LRNFMVAGLRVDWTDQEGVL*

>jgi|Ochro1393_1_4|919954|fgenesh1_kg.49_#_775_#_TRINITY_DN17531_c0_g1_i2
MTMLFSCIALILFIVVREASSYSHGGLIIRSSRQYRHTYGSWSRRLSSLCMSDSETIISPFDDASRSGDNMPMPSSGDDDD
DDEYDLPLTRDNVEKVLDEMRYLQSDGGDVKLAIEDGPVVKLELVGACGTCPSSSMTMKMGLERKLKEKIPEIAEVVQS
LPDAPELTEEAJETVLGDGVRPFLAVAGGKISVDSIKVGGLQPVIGLMEGSAAALQSVKVEIMQRIQRHFMLSLRIEWV
D*

>jgi|Auran1|6483
EQFPLTWENVELILDELRPYLMSDGGNVRIAGIEGPVVKLELEGACGTCPSSTMTMKMGLERRLKEAIPEISDVVQYLPD
TPDMDEASVDEVLEGVVRPFLSVAGGTVDLVLSLG

>jgi|Vitbras1|20756|Vbra_18723.t1
MISLLAPICIFVCLSSTRAFFQPEPPAFVLPAGITSASRPPASGRQRASHAEIRLHSASPATEAKESPFENLREVEY
RRGAETLPLTPENVETVLDEVVPYLIADGGNVSLKSIEGPDVYLQLEGACGSCPSVTVRMGVERRLRERIPEINSVEA
VSEGAENGMAFTKENVEKILDGVRPFIKMTGGDITVVEANKVGSVAATVALKMTGRSAENRSIRLEIQQLRRNFPLNS
VTFV

>jgi|Vitbras1|20756|Vbra_18723.t1
MISLLAPICIFVCLSSTRAFFQPEPPAFVLPAGITSASRPPASGRQRASHAEIRLHSASPATEAKESPFENLREVEY
RRGAETLPLTPENVETVLDEVVPYLIADGGNVSLKSIEGPDVYLQLEGACGSCPSVTVRMGVERRLRERIPEINSVEA
VSEGAENGMAFTKENVEKILDGVRPFIKMTGGDITVVEANKVGSVAATVALKMTGRSAENRSIRLEIQQLRRNFPLNS
VTFV

>jgi|Symmic1|41120|rna29028
MLAPAARTGQRPARVLSAAPTALAAIEPRDRDFGLALPAIAVAVNCRGGHLRRIVRATVVSFNEAEAKEKEAEK
RESEKKEPKLSLTWDNVQEVLDDELRPYLKSDGGDCKISDIQDSVVKLELIGACSSCSASSVTMKMGIEKTLKERIPEIS
EVVAIDQEQEPLTESGIEEVLNGIRPFLSVSGGSIEIFELVDGEDAKVVVKMGVPLKSMVRVEVQNRIRKRYPAVQNV
GIVGEDGQPPQSA*

HCF101 orthologs

>Cre01.g045902 HCF101
MMKTRAGVQRSGCPAVARTFAATQPSGAAAAAARGVAAWVSGVAPATAAARPAASRRGAHVVTAAAAAPPVATAAAPRT
REEEVLAQLRNVIDPDFGEDIVACGFIKDLAVDEAAGAVAFTLELTTPACPVKEMFQRQSTEFVKALPWVRDVAIKMTSQPAK
PLLPESGRPGGLAKVRHIIAVSSCKGGVGKSTVSVNLAYTLAQMGAKVGIFDADVGPSPPLMVDPEVKVLEMDPATKAIYPT
QYEGVKVVSFGYAGQGSAIMRGPVSGLIQQMLTTADWGLDYLVVDFFPPGTGDIQLTLCQTVAFSAIVTTPQKMAFIDVA
KGIRMFARLVVPCVAVVENMSYFDGEDGKRYFFPGQSGSERIQRDFGLPNLIRFPIVPDLSAAGDGGRLPVSSPASATASTF

MDLGAAVVREVAKMGAAGRAPRQTAHFDQEQDAIVVALPGEPEFLLPPGVVRANDTSATSIDEWGTGQRKREEAPADARPVAVN
PLGNAYAVQITWSDGFNQVASYELLDELKSYAVPRRPVSGGSASNGSVSTSLPGGMQAQAQLA*

>jgi|Cyapar1|4286|g4162.t1
MFASPAVLPAASAASSLNAACTASCSTFAGARSRPTRSVQARRFDLATDSGFRPASFEGSTHYIASSGAQLKRAARRSLG
CGAPACPPAVSPVLASATAAKEEILAQLKNIIDPDLRTDIVSAGFVKELKIDEAAGRVTFTLELTTPACPIKDLFKADA
EKFKELPWVKECEATLSASSNPAFSTEAKGLKQVANI IAVSSCKGGVGKSTVAVNLAYTLAKMGARVGIMDADIYGPSL
PVMVSPASTVVSYPENLIVPLEYEGVKLMSFGFVQKGGQAAMVRGPVSNVLNQLLTTTDWGELDYLVDMPPGTGDIQ
LTLCQLVNITAAVIVTTPQKLSFVDVVKGIQMFDKVSVPSAVVENMAYFLCESCSTPHFPFGKGSKEKLVQQFGFENSE
DLPIQTSVSDASDKGIPYTIMEGADPATLDAKSLASAVVQEISKLFKGGDLDRPAVEFDKAANLIELKYPTRGENLGRV
RRLNPATVRRKCRCAACVDEFTGAQILQPESVPETIKPVSISPVGNYAVSIAWSDGHPSLFPFRALGTIPEAAPAEAAA*

>jgi|Tetstr1|444905|TSEL_003356.t1
MCRLSASSVAQGDMSKCGHSDSDPNGLSVRGGDAARVAQRVAVRFEFRSRRSRACVARRGRAIPALASAAAVSYEGQVLK
QLERIIDPDFGMDVACGFIKDLVDSAAKVSFRMELTTPACPVKHVFESQAEYVVGELDWVKEVALTMDAQTQTHMME
GDGRPGGLKDVAHVIAVSSCKGGVGKSTTTVNLAYTLAQMGAKVGIFDADVYGPSLPTMVSPELRVLQMNQETKAITPVD
YEGVKVVSFGFAGQGAAMRGPMVSGVIQQLLQTTDWGKLDYLLIDFPFGTGDIQLTLCQSVAITASVIVTTPQKLAFID
VAKGIRMFAKLMVPCVAVAENMSYFDGDDGKRYYPFGSGSGERIQRDYGMENTIKFPIVAELSAAGDGGKPLVSDPTGE
VSKVYVYELGATVREVAKLKLMMRRNSVRFDKEIGALVVKLPGEAEFFLLSPALVRRNDTSAASINWTGEKMLDDAAVP
EGVRPETISPLGNAYAVQITWEDGFNQVASFDLLAGLERMALPEGYTGPSGGGSAAQEILSAATASAGSE*

>jgi|Pracol|16810|PRCOL_00006186-RA
MRRSVGADAPVAESGRTAEQEEVLDALRSIIDPDLGDDIVSCGFVKDLDCDRATGAVSFRHLHTTPACPVKDRFQEDAEN
VTLALPWVESVDVALSAAPPKPLIPQMPPEGLQKVSSI IAVYSCKGGVGKSTVSVNLAYSIAQMGARVGIMDADIYGPSLP
TMVSPVTVLEMDATTRSIPKVDYKGVKCVSFGYAGQGQAAMRGPMASGVVAQLLTTTDWGDLDYLLIDMPPGTGDIQL
TMCQTVPLTAAVVVTTPQKLAFVDVAKGVRMFARLRVPCAAVVENMAHFDAGGERHYFPFGKSGERICAEYGVPHLLEVP
IEPALSAAGDCGEPVVSAPTSDTAKVFGELGAAVKEVAKLRSGRTEVLVGFAGAGEGEECGSGGSTPIISVDAPGFDR
ALVLRADVRRSDRGAASVDEWTGEQLTDPASVPEDIEPMEMAALGNAYVQILWPDGLNQVAPLDVLQGLPEVDIPQL*

>jgi|Chlat1|652|Chrsp103S01053
MASVVAGLVQVVPSTLLSSPSCSTTTAAFSFKRAPAAASAAPAAAAASFSGRPLAPPTVGFSSRRRRHRRQHADVVASA
SANPSAAQTDSAASTSDSGLERDVLGALRRVIDPDFGMDIVACGFVKDLICDASTGQVSFRLELTTPACPVKDMFESQAR
EYVGALPWVKDVSVKMSAQPPKPLLPNDVPPGLSRVSSI IAVSSCKGGVGKSTVSVNLAYSILAMMGARVGIFDADVGPS
LPTMVSPVRLRMDPQTKAIQPTIYLVKCVSFGFAGQGSAMRGPMVSGVNVQLLTTTDWGELDYLIDFPFGTGDIQ
LTICQVVPISAAVIVTTPQKLAFIDVAKGVRMFASKLVPCVAVVENIGREEILSVREGVVSDFGIPHLEFQMPIAPTLSAS
GDSGQPAVVEDPTGQVAREFGELSACVVQQVAKLRQAANAVVYDADRKAIVVRLPGSQDTFALHPATVRRNDRSAQSVD
EYTDVAEDIEPASIAPMGNAYVSIITWPDGFVQVAPFEQLAQIERLILDSATVVNTPPSNTNVSSSEADDVSPASAAQAIMK
NAERRREQQTAAA*

>jgi|Mesovir1|2059|Mesvi1495S01046
MPEAKESVLGALRKIIDPDFGMDIVSCGFVKELEVDPATGKVSFVLELTTPACPIKDQFEQQALENVRALPWVTDVKLKM
TAQPPKPLVPTDVPPGLAKVSNII IAVSSCKGGVGKSTVAVNLAYSILAMMGARVGIFDADIYGPSLPTMVSPELRVLQMDA
ETKAIMPTEYLVKLVVSFGFAGQGSAMRGPMVSGVNVQLLTAEWGELDYLVLDFPFGTGDIQLTLCQVSGRQAGPSKK
TTPKKATPKKAARKRPAPSSSSSEEDKGLQEDGSPSTSEYVPPSRKERPPPVVIPPSSDDEDEDVKKGKAALLPK
AHVFCQATGVARPGRPEDLVWTKPAKPEASAPSTKPAKPEASAPPTTKPSKKPAERAEGQLALPALVRLDQHVQHD
AKRVRGADTMGDGGGRVHQIKNAVAYDEYRNVIVLKVPGANEEILLHPATVRRNDQSAASIDEWGTGEQRLRPEMVAEDIK
PQAITPVGNAYVSIQWSDGFNQTTAPVSIQWSIGVNVQDLA*

>jgi|Klenit1|113|rna-KFL_000020280
MAACKALLKPALSCQGFALQQGRRGKLRQPAAPCLRTAYLQTSPPAPLSSQNVAAGWRQILEPPAQGFHTNSRPLGRG
DDRGGARCQAQAQNGAAMGASPEQEVLGKLSQIIDPDFGTDIVSCGFVKEMQIDPNTGVVSFTLQLTTPACPVKDMFES
QARELVSQLPWVKDLSIKMTAQAPRPMTDEDVPPGLRKVANI IAVSSCKGGVGKSTVSVNLAYALHMGARVGIFDADV
GPSLPTMVSPVPLQMRVEGDRIIMPTEYEGVKLVVSFGYAGQGSAMRGPMVSGVNVQLLTTTDWGELDYLVLDFPFG
TGDIQLTLCQAVPITAAVIVTTPQKLSFIDVVKGVRMFAKLKVPICAVVENMCYFEADGKRYFPFGQSGQKVVDQFGIA
SLFQMPIKPELSASGDTGKPQVIVEPAGEVARAFSELGASVVQQTARLRQQVATAVVYDPARKAIVVRPPGSDEEFLLSP
ALVRRNDRSAKSVDWGTGQQLLQYSDVAEDIVPESIRPMGNAYAVAIWPDGLTQVAPYEQLENLERMLEDDDSAFAMPST
TSRVDGSSAADPAAPSSRPSATKAAAGTGSVGAQAPLSDGAARILQEAAARLKEKELAKS*

>jgi|Mesen1|2247|ME000153S01474
MVGTAASMAYASLNVGSYATSAAGARMTAGKALLPARGASLCLAGTLVPLGRHVALRRALSTRRAPGGGASCEAASAAPS
AAPASVEEAKEVLGQLARILDPDFGTDIVSCGFIKNLQVDPQQGSSVFSMELTTPACPVKDDAKEYYVQGLPWVRSLSLT
MTAQPPRPLLPDDIPRGLQKVANI IAVSSCKGGVGKSTVAVNLAYSILGMGARVGIFDADVGEELDYLVDMPPGTGDIQL
TLCQVVPVTAIVTTPQKLAFIDVAKGVRMFASKLVPCVAVVENMCYFDGDDGKRYHPFGTSGSKQVVQFGIPHLEFEL
PIRSELSAAGDCGQPEVVANPAGEVGLFSDLGATVEFWLSPATVRRNDRSAQSIRLKYSDVEESIVPESIRPLGNAYAA
ITWPDGFTQVAPYDQLAELERLVDVRPSGANLEKNKDAALRIMEHARALRDAVRVNP*

>jgi|Chabra1|337696|rna-CBR_g30185
MALRLAQVILWSPSASPGFTRNDGSPPAVASGAASAAAWLGGEHLQGRSSCSGGSGSGSNYGNSSRRSSCSNATCSGWT

GNGGEGSFMFVSRMRAAGSLRNSGGGATASRHLAGSGTAEGAQATAPCTSSWRKGMRAGSCRSQVQRPNASSSQTPSG
RVTKKVAGATAGAVQTSSSIAIVKYSVGGGLGCRWEGESTFHSHLPAVGRTGPTPSSSSTVNLGRKNSRRKNSLCRCEA
PEADGGGGNSAGTATDGSSTTTTTTTDNGSAASTDSSVAPASKVTTSSSAAEKAVLSALAQIIDPDFGMDIVSCGFVKDL
AIDEGTGHVSFRLELTTPACPIKDEFERQARDNVSSLDWKDVNVTMSAQPPRPLIADSM PAGLRKVSNI IAVSSCKGGV
GKSTVAVNLAISLAMMGARVGIFDADVYGPSLPTMVNPEVRLQMDPETKSIFPSEYEGVKIVSMGFAGQGS AIMRGP MVS
GVINQLMTSTEWGELDYLVIDMPPGTGDIHLTLCQLVPLTA AVIVTTPQRLAFIDVVKGRMFAKLK VPCVAVVENMCYFE
AGGKRYFPFGE GSGQQVADMFGIPFLAQLPIRPELSASGDSGRPEVIADPVGEIARAFQEVGVAVVQECAKFRQQVATAV
RYDPQLRAIVVKPPGEEEEFYLHPATVRRNDRSARSVD EWTGEQMLRYSDVMDNIEPEGIRPMGNYAVSISWPDGLTQVA
PFDQLRELERLIDPYAQREALASSVLPNSSERTGKP EEDGEWEKVVEEEEETRAGRAPAMAGEGEKLTGAQTILRSAS
GLQRE RHAAQQD*

>jgi|Emihul1|415743|estExtDG_Genewise1.C_280054
MHMGLHHLHNHATMRIGTSFVRRRCALATARGRARLEEEALAALNGVLDPALARCKRLAVQPDATVAVKLEFRSGAMRRG
AVEAAAAAIEALPWARRSSVESTLSRPRSFMGSNAPPSLRHV GALVGISSCKGGVGKSTVAVNLAYALATQGGRVGLLD
ADVHGPSLPSLVRLPPGSLPLRQDAQTRLLAPAEADGVRLMSYGFVGKGASSGAVPAAMVRGPMVGKTVGQMLSGTDWGE
LDYLLVDLPFGTGDVHLTSLSQTYGLSAAVLVTTTPQRLSTVDVLKGVMDLRELGVPMICLVENTAHFTDDTGRTHLPFGGS
QLEAVRRHAGLRESDFRLPIEPQVGEAVDAGSPLQLLTSQSESAAHALRAAARHLVVAIGERQHGAAGGSGSHLRYPK
RGLVLRVLSGDHEGREFVLSKAALHALRGAGAGDASAPLLDTVEPCETEAGQPGAKLYWCGGGATVVPGLGDFPHLAAAH
G*

>jgi|Auran1|33254
MGVSANRRGESLRVRKTI LAVSSCKGGVGKSTVAVNVALTLAKRGLDVGLADADVHGPSVHVLCGEPKDGAVRLAPEMDA
YGRELLEPF EAHGLKLMSFGYLNDA PAYMRGSRVSGVVQQIVASVAWRDL DVLVDCPPGTGDAQLTLCQVLDMDAAVVV
TTPSRLSFADVVKGV ALFDEVDVPVVAVVENLREFALPR

>jgi|ChlNC64A_1|133105|IGS.gm_1_00241
MSCRASAEAAASNSAPAVVSKEGEVRAALSPIIDPDFGMNIVDGGFFKDLAVDGEAGSVAFRLELTTPACPIKDDFEKAAR
EYVTALPWVKS LDKMDARPPQPLLDDSRPSGLRSVSHVIAVSSCKGGVGKSTTAVNLAYTLAQMGAKVGIFDADVYGP
SLPTMISPEIRVLQMPN PETKAITPVEYEGVKA VSFAGQGS AIMRGP MVSGLIQQLLTTSEWGALDYLVDFPPGTGDI
QLTLCQSVAFSA AVIVTTPQKLAFIDVAKGIRMF AKLMVPCVAVVENMSYFDADGKRYFPFGKSGSERIQHEFGLPNLVR
FP IVPELSAAAGDGGRPVVVQDPAGPTSQAFLELGAAVVREVAKLRRVPQNCVRYDQELGALVVRLPNTEEDFLLDPATVR
RNDQSAASIN EWTGERTLRSDIADDIQPVGVQPVGN YAVQINWQDGFSGQIAAYELLDALPRLSPDDALARQQLRARLAG
IVSGSGSGGEPGEGPSAAQQILQNAQGAPSAQSAPQN*

>jgi|Caulen1|91712|g552.t1
MQQSLSLTKKTQIHHLRNSPATVLKRRTLLCAQSQTSESWSGKNKKSQVLDALSTIIDPDLDENIVEAGFIK DLEISGNE
VSFTVELTTPACPIKA EFQQQCQCEKVGLLDWVQNVETMTANTVRS GAPSSDRPGGLKDVAHILGVYSCKGGVGKSTVAV
NLAISLTQMGAKVGIFDADVYGPSLPTMVT PKHRVLEMDPITKEIQPADYHGVKLV SFGFAGQGKKS PINPHPLLF TCSR
ECYNARGDGKLDYLVIDFPPGTGDIQLTLCQTVALSAAVVVTTTPQKLSFIDVAKGIRMFARLVVPCIAIVENMSFF EADG
KTYYPFGTSGDRIQSEFGLKHLIRFPIVPDL SLSSDDGEPLVVSEPTSQIGRLFTELGVVVVQEI AKLSKKGKSLAIYD
ETRNLCLRVKLPEIEFCLDPKTVRENDTSAKSI SEWTRNEEALSQREGAF EIQSVSGLGN YAVQILWKDGFNQVAADFLLM
SLPHLQPVHEDL*

>jgi|MicpuC3v2|3675|wlab.212701.1
MDLVLSPLTL SRRRTAAQYPRRAAAMAASAAAAA AASTSIPSSSRGRPSLARARLAATRARRAPRRRDVVDVAVRASS
SSSPSITSAPEGSREADVLDALRNVIDPDFGEDVNVCGFVKDLRITDDGDVTF TLELTTPACPVKEEFDRLSKQHVS AVP
WVKSCNVAMTAQEVNDAPDTVEGLRNV RHII IAVSSCKGGVGKSTTVNLA YKLEMGARVGIFDADVYGPSLPTMTSPE
IAVLQMNKETGTITPTEYEGVGVS FGFAGQGS AIMRGP MVSGLINQMLTTTDWGELDYLVIDMPPGTGDVQLTICQVVP
ITAAVVVTTTPQKLAFIDVEKGVRMFSKLRVPCVAVVENMSYFEVDGVRHKPFGE GSGQRICDEYGVPNLFQMPIVP EL
SAGSGKPLVADPAGEVSGAYGAVAAKVVEVAKLSAGPKGLAIDEDDVAGAMETHGVKALRVQLADEGLPFYVRASD
VRKSDKSAAPGEARKAEFLADGVTPIPDDIAPSETNIPVGN YAVQISWPDGFSQVATFAQIAALERLEVKGAGKRAEETA
AAARS*

>jgi|Galsul1|3360|XM_005704422.1
MLFVRS CGISTCSVLSQRNGDALKKCPKCCRKQYYLLDNFRFRNTRQWL GKRLCKLQGMSIQTRLSTRINMCMSDWQKQV
LELLKNIEDPDLKQNI VLG FVQNLERVAKEDGKYDVRFTLQLTTPACPIKEKFQNDAKEWVSLLWVRNVEIDL RANEI
NRAQAGDRPLNKVKHI IAVASCKGGVGKSTVAVNLAFTLT KLGGKVGIMDADIYGPSLPILVQPENKIVQYKDGRIIPLE
YENVKLMSFGYINPESAIMRGP MIANMMNQLLTETDWGSLDYLVIDMPPGTGDIQLTICQTVSLDA AVIVTTPQQLSFQD
VIKGIQMF GKVS VPCVALVENMAYFEPNDIPDKRYLYFGHGKSQKIANDY GIPFVESFPLDPDL CRWSDNGIPAVLALSE
SKISQLYQSLASAVVQQIAKNAFGNGKRIPQVFFDSDKCIIVISCNDQQGI AWNNENKVEWSPWELRNACSCASCVD EFT
GKRHWKSVDRNVKPLIQITAGNYAFSVIWS DGHQSLYPFERVMNPKDWSTTTAVKRASTI*

>jgi|Chlpril1|4217|rna-gnl|IITBIO|A3770_06p42170_mRNA_A3770_06p42170
MSAPLECLRAGGRAGRLAGRTACQRATTSRRCGSQDAWCGAGWGWRRLRKGGRR LCHETRVATACSTGEVEVTPQGGQE
EAEGAELPKGLRGVKHMI AVASCKGGVGKSTTCVNLA YTL SQMGAKVGIFDADVYGPSLPIMVSPDYGKSKLVMDPETKE
IAPVEYENVKLV SFGFTTEGSAIMRGP MAGGLVNQLLTTTNWGDLDYLLIDMPPGTGDIHLTIGQAVPIACAVVVTT PHK
LAYADVAKGIRMFARLDVPCVSVVENMSYFDGDDGKRYFPFGKSGEQIQGEFAI PHLVRMPIKEEVSSGSGDEGVPCVVS
DPLSDVAGSMQELGASVIKEVTMGT KRKHVKYDPMHEMHVSLFEEGSEGGEGRWDSFEIEPLVVRKNDQSAKFIDQL

TGEPINQEFSPHIFPLEIKPLGNYAVQINWSDGFNQIATYDQLRVMERIQ*

>jgi|Nanocel1779_2|579071|fgenesh1_kg.4_#_1139_#_TRINITY_DN2326_c0_g2_i1
MKAALRKKDFVPTIILIVLALLLGAVALQCRAFTPPQQA PMSSSTRRPQVSSFLPPSRRGPSASRSSVLTHTPLPYATLSAE
QSPVDSVSSSAVASSNDENLTQVITQLREVMDPDLGSDVVTLGFIKNLTVAGETGSVRFVLELTTPACPVKDDDLKAQAIR
VVEKLPWVSKAEVILSSRPIDADRPASANIKATGLSLVNNVIAVSSCKGGVGKSTTAVNLAFALDRQGARGVILGADIY
PSLPTMVIPTDEVVEFVESQIRPMVKLTGVDAAGQPKGVKLMSFGYVNPAAAAIMRGPMVIQLINQLLTLSWGELDYL
LDMPPGTGDIPLTLCQSLNITAAVIVTTPQRLSFTDVVKGVDMFDTVNVPCVAVVENMAYADASALDDAASRGASDPQGS
ASTVGSKPWKRLRQEIEEVIKAAEDTKTQAGLVDELLATVLRRLQADRVARPLFGRGHRQRLADMWGLTNTVSLPVL
EVSRSRGDSGVPLVLGQPN SAPAQVFM DLAGR VVREVA KLRHDSRLRPSIT YLPEEHMIAVKMHGYS DTGGAEAVAPPQKMS
PAALRRQRCRCAMC VEEMTGRQILRPEDVPEEIKPVEIGAVGNYAVGITWTDGHKSLFPYRSFVDGYDKMKPAKESR*

>jgi|Ectsil1|21518|rna9702
MHMQLKRPCVVAASVSLQCLCSKAFVQPFSSQHRAKQGIWPHQYSCKSTSLPVQQRARSYHVSMMTETTS DTPPGRKDEV
LAVLSAVMDPDLSDMIVSLGFIKELEISGEDEGRQVTFDVELTTPACPVKAQFQQDCRDLVEALPWVDRAEVTMTAQPV
RDVSDTVPTGLSKVATIIAVSSCKGGVGKSTTAVNLAFALDKQGA KVGI LDADIYGPSLPTMVKPDREEVEFVGNQIRPM
TAHGVKLMSYGFVNQGA AIMRGPMVSQLLSQFVTLTSWGELDYLVIDMPPGTGD IQLTLCQVLNITAAVIVTTPQKLSFT
DVVKGIDLDFDTVNVPSVAVVENMAYDAVDQTVFKTGLESNIEDLLTLDGDELSAAAAREGLASPPQDETALRAAVIAEV
MKKRAS TKQREYIFGKGHQMLADMWGITNTIRMP LADVASSGDSGIPFVVS KPDS DHSESYSQLAEAVVREVA KLKFS
DND RPMLS FQPAEGT VTIETSGGKQVMAAADLRQRCALC VEEFSGKPLLD PASVPENIVPTEFAPIGNYAVSVKWDDG
HSSLYPYKNFAMGYKPRTKRPELSPV*

>jgi|Vitbras1|5621|Vbra_12538.t1
MAHPRQRRQQRQHERPTFLCAVTVACVVSSAAAFVPVQLIGGERGLRPAPAVRQGHHDSSQORRASTLFLTGLSDEAPGV
GGGVDERREVECLKSLKGVDPDLGQDIVTLGFVKGLTIDESTGNVAMTVELTTPACPMKDQFRESITSALEVL PWVNGVD
ITFTSQPIGTRPSPLSDGQRQQHAPGGTSRVRAIIAVSSCKGGVGKSTTAVNLAFALKHMGAKVGI LDADVHGPSLPTM
VRPSEPVVTIEEVP AEDGQTVQILKPIDHDGVKLMSFGFINPMAAVMRGPMVTNLLKQFVG YTDWGELDYL VVDMPPGTG
DIQLTLCQSLDITAAVIVTTPQRLAFTDVVKGIDMFDTVNVPCVAVVENMAHLTDDIDTDR LARDIAAATSMTDQQLRD
QLSIQNE DGESVGREELVRRCVDRALRSPDVRRRKRMLFGKGHRQRLSDMWGIDNAFSLPLPSISRASDSGTFVDT
SDTADSDSETGIGPAEREA KDTYKNIAQSVVQEVAKLQSLSRDKRRPAVLYLSEGNELVVAWQGGQTAI PPAGLRKRCRC
AACVEEGTGRLLDPSSVEDIRPTRLPEVGNALGV DWS DGHKSIYPYRAFVPGYTSRRDTAYEQLES SRDNGDIVLLH
RVPATRRTESEKRETA AAAAAAAAAAAAAAAPVAS

>jgi|Phatr2|1494|gw1.3.92.1
EVLSTLKSVIDPDLGSDIVTLGFVQNLKLDGRDVSFDVELTTPACPVKEQFQLDCQQLVQDL PWTNNIQTMTAQPSVQE
TATLGMSQVGAVIAVSSCKGGVGKSTTAVNLAFSLQRLGATVGIFDADVYGPSLPTMITPQDDTVRFVGRQVAPLQRNGV
RLMSFGYVNDGSAVMRGPMVTQLLDQFLSVTHWGALDYLILDMPPGTGD IQLTTLTQKLNITAAVIVTTPQELS FADVV RG
VEMFDTVNVPCIAVVENMAYYESADPEKIQIFGAGHRDRLSQQWGIEHSFSIPLLNKIAANGDNGTPFVLEFPDSPPAKI
YQELASAVVSEVAKTKFAKSMRPSVQYDAESHLLQVSQNGVGSTDEEHVATLP PAELRRACRCAACVEELTGRQILVPSS
VSDKIAPRNMVPTGNYALSVDWSDGHRSLYPYRQIRAL

>jgi|Symmic1|13436|rna10656
MPAAILLSSKPAAATCSSSSSSSTGPPATSGTSKGKKVLESTQKTDQSESKFSLRMCATGDVRYWITLNEPWCACALG
YGSGEHAPGHREDAGKEPYLA AHMMLAHARAVKCYRHSYQHIHGGSIVSRQKDGTKFAAAARI PRHLPRKPTEPIPSLHG
LAKSLLWLKKAMASGMAFASLGRPWP HQVQPGTAAERLQRSAPSGAPARDGSLRSSVATATFAGLAAGLRNSSAGRRGVS
RAAGTAGPTPIQEVMHAEQVKDPDMGTDIVSSSGFVGQIKAERITGNISLVLEVEMYKAEVEERLSQIPWAKSVEVRVGA
VRQQT SATNPPRPMQMPD SLKNVKTILAVSSCKGGVGKSTVAVNLAFSLYQKG YKVGIFDCDVYGPSLPVMLRFQEDTP
KMEMYQDEQKQKHIRPVIDPNTGIKMVSFGFVGHAAMRGSMVTGVMSQLVTQTDWGELDYLVLDMPPGTGD IHLTLSQV
CQITAAVIVTTPQKLSVIDVERGISMFSQLNVP SVAVVQNMSYMSLPNGERQYIFGQTDAGYAIADTFGIEQVFELPIEA
SVAAAGDSGKPFNGKGDSETAKQMDKLTEAVITEVEKIH DNRQPELSWDGELRILRLKLAGGEELGIDPRELRLRDKGA
GGVTPPPADLSPDEIVDMGNYAVVIKWS DGVVQVAPHKWPGSLSTETREVPCLMSNWR LPEARTLKS DVAGVIPAYLVSE
TCITLNMDWKEPLSDSADQAAQRALDWQLGW FADPIYK GAYPATMRERC GDR LPEFSDEEVAMLGSSDFFGLNHYST
DFVSQGEDGPPAVPNYFADQDVRNVSDPRWQRTDMGWDIVPWGF EKLLSWIQKEYDPTGGILVTENGCAVRENTEAEAVQ
DTARVEYLQGYLAQLHKAMANGAVIKGYLVWSLLDNFEWAFGYAKRFGIVRVDFTTQQRTPKAS AQLISDLCKGGKLVKVP
SRVQASSEFFFPYNGRGKEPEVEPKKVAAPPALSKADAKRMLEEFVMRYQDDH FQSKMVSCFQQYLIHNDEMRL LKARRSL
CMPIQAEIIPKYGF EPTARGVSRVQATLSAPALTEDPD IKQMNELVVYL TGDFPKTKAAEGATV*

>jgi|Ectsil1|22518|rna7047
MLALPYTCRWHPLAGRARARHLVDVSVRRNKPTSCSTSRRGIKVAGAAAQGRRTAAASDTTSTIAALLSPQILQSRRFIS
TAPGPSLATDVSWHVAHYRNGGNARQLPRRRHMSLWGGWGQGTSGGPNS SPEEELARRWRQEKEEEVLAILSGVVEPCTG
KDVELLGFVQDVRIDEVVHGEQPQDNTGEDSAEAPLAISFTLRVPTLALPGRDTLASECEAALLALPWASANALTKVRR
PRWRRTVQRRSTPGSILNRSVGTETTPGGGGGGGGGAPSPGLESVQDVVVCVSSCKGGVGKSTVAVNLAYS LASRGAKVG
LLDADVYGPSLPTLVNPPDDVALRVS PAFDLNLLSPIHRGVACMSFGWVNAKAGVPGAGGHGAAMRGPMVSKVINQLL
LGTDWGELEYLIIDMPPGTGD IQLTLGQALQMSGAVVVTTPQKLSYVDVVKGIDMFAETKVPVLSVENMAYFDCSNGER
HRPFGPGHARELVEECGLASGCVFSLPLSPAVARGSDCGDPVSLSSPDGEEAKVYLSLADGVVRETFRAGKTAADVPEVS
FKSGRGIVLRYISEAEAAEFVIPPFELRTRDPATGEPLASEQAAAVSDDVEPVNISVRGNYAVSISWSDGHRGAIYSYQV
LRKIVEQRNR*

>jgi|Vitbras1|22565|Vbra_10576.t1
MDDALVDGGGEGQPRDRLREEVVHELKKVIDPDLKKDIVTLGFIKLNKIDFTKGAVSFTLELTTPACPVKDQFVQVCTS
LLEALEWVTAGSVQIDITSQKASSSRQQPGSEGIPGLRQVSFIIAVASCKGGVGKSTVAVNLAYTLQHMKAQVGIIDAD
VYGPSLSYLIPLQCDGVRGTPDGLMVPLEHRGVKLMYGFIFKPGDYAALRGPLASGMVRQMITGTAWGELDYLVIDLPPG
TGDHMTVAQSVRVHGAIVVTTTQQQLSLVDVEKGLRVFDQLKIPTLMVVENMSYFTCGTCHTRHEIFQRPPTIARLQDE
YGIDRTHWCRLPIDQRLTVPQEATHTQQRREGDDDAYEFLECPPATFPFVLTVDSDHPVWESFRQIAETVARGLSASRFA
DNRPRVRMTDDMMMEVTIPCASPPPPAPPPAAVSNPAVADNGTSGEGQGATNGHVGALEASELMSGVAGEDVLLVPART
VRLACRCAECIDEFTGKAKLQPWVRPRDVVADRINPTGNYAVTIHWSGDGHESIMAYAVLERVARKHADQQRPKGD

>jgi|Ochro1393_1_4|134025|CE134024_10137
MIESTKLVCLLLCMCIYGVSGFHHRFRRNTGPIICNPVHRTVMPIAPVNKCNRKFLSGVVEFKEGGEQENFRNDVVS
ALQNIIDPNSGTDIVEAGLVGEIFVQNNNVEISLLIDKTSVSIADIEKKLCMLEMSMLEWVSDIDVQLKNREAVETATSD
AVAETNGMAGVKNIIVASSCKGGVGKSTVSVNLAYTSLQGGAKVGILDADIYGPSLPTMTQPLLRDRMYEDNKLIPLY
EGVKLLSLGFINQGAAMRGPMVNQILNQFVLMSSWGDLDYLIIDMPPGTGDIQLTSLQIMNITAAVIVTTPQKLSFVDV
VKGIDMFDTVNVPAAVVENMAEYSTYSFTDDFYSTLAAKIGVVVTSATESTIAAAAAATNTTTTTTNDAAVAIASTIKEA
VEAQRQPRKVFVGHNKLEEMWGINNIVSLPLLEVAASGDSGRPYVLSHPESKLAAQMKALASSLTSELENLAKSTAS
GPTLTFTDASTGMEYRSKKLSSFDLRCDRCACVCEELTGEALLDKSQVSRVTRPLGTAPIGKYAISVDWSDGHKSLYPF
RQIERLAETLASA*

>jgi|Guith1|156868|fgenes2_pm.3_#_26
MSEVLASLAEIDDPVLGESIVSLSAVNLEVEPSGDVSFTVQLSAPDLQGVVKQACQSALSQLPWVKDIEITVTSLPND
SLREASRNQATGLKGKVLHILCASCCKGGVGKSTTAVNLAYSMHKGFKTGILDVDIYGPSLPTMVKPERPFDPRRDI
EIMPVNGFGVKLMMSGFINPVDSFVMRGSRVTPVQQVLVSTTAWGELDYLIIIDMPPGTGDIQLTSLQMETLRIDA
TTPQRLSFVDVVKGIEMFDKVGIPSVAVVENMAFFQNDGMQDNIAQAFANKYSLPQEAVERSELEEILKAKQFLFGQGHKQRL
DMWGIQNSISIPLPDLAKQSDSGMPYVLAFFPDTVSRAYSALAESVISEVSVAKMESGKNVMEPIEYDAPNKQFVIDGN
QRISAKEMRRLCRSPANDPNNVKESVEPVDVFLGRYAIISIQWNDGHQSLMPYRSFVNSYK*

>jgi|Bigna1|87067|estExt_fgenes1_pg.C_160147
MVPPTWSPTACLLQALIIILGGGRES PWDSRKFSLAIAQRLRLPSQRASISLPLSRIRDMSKAGRLVSSNPSELASGPV
RRFHHRHRAHRCRVENTDSENDASPLVDEPHELLESIAIEVPDQAKIGGSKASIDIVKSGMVREALGEKALGIASV
LVELPLPVDQSKTKQALQQCENAVTIEWVNPLIVITAEAGPSAVMSSSPSSASSPVSPFADMGNSEVDTRKGGGLKSV
DKVIIVASCKGGVGKSTTSVNLAFAALRAQGYKGVIVDTDLYGPSLPTMVTPENSEVAFENGGESEIIPVEAHG
VKMMSMGFINPTDSMVVRGARAAPLVEQMLTRTQWGELDFLVIDMPPGTGDIQLTSLQSVSIDAAVVVTTPQRLS
FADVVKGVEMFDKVGIPCVAVVENMAFWQNNELSKRVEALLDEGASSPLPEATRKSLLDLVSAPQYIFGKGHRK
RLAEMWGIENYALPLEPALS EAGDSGCPFYLTGASSNPSIASQYDRLASDVVTEIHNLEKTKAARPEVIFD
ATRNMMVVSITKGGEQIISPVELRRRCRSP TNIKELITDDVKPIDIFPMGNAYQILWNDGHQSLMPYASF
VEGYRTNDNSVNSQLEAA*

PFLA orthologs

>Cre05.g234636.t1.1, Pyruvate formate lyase activase
MLKAALPRIMNAASMSCAATGLQRSGLAMNAATTSTRTGPASGLPKHSWGASARRAFVAPATISERLQPKLSTNYSV
VLPQYEPTDPSGIPFVFGNVHSTESMSAVDGPVRFVLFVQGCAMRCLFCSNPDTWTLLKGGNKTSSKEIAADI
KRVNRYLKPGRGITISGGEAMLQPHFVSTVFQEVHALGLNTTVDTTGQGTKAGNWDVVLPHDTDLVLFCKIHID
PIKYEALTMKQRGALRFADQLAERKIPFYLRYVYIPGYTDGVKDIDKLIWCKQQPTFQGIPELLPYHVLGRNKWEV
MGLPYPLDGTNTPPHEQVRAVIKVFNDNDVPVICA-

>jgi|ChlNC64A_1|57698|estExt_fgenes3_pg.C_90008
MLSLLAQQVTSGLPALLRPAAIPLHLALQRGFACSAAHQASAEPQPAEEVPPNRRFAARGGDFATAAPQTPAEV
PEVVGSHSIESFSAVDGPGVRFVLFVQGCGLRCVFCSNPDTWHMARGKLTSSKDIAKKLERVKPYLSQGDHKG
GITISGGEPLLOPEFTASVLMHAHTRGLTTCIDTTGQGMKHSWDKVLPHLDYALFCIKSPIKEYEWITKRKIG
PALGFVDELEERRQIPYWLRYVMLPGKTDQPEDVAALIQFCRNKRSMQAEVLPYHLLGVEKWASEGKEYPLTGMSS
PSAEVNAFLAPLKEAGIPVLCNKNREEE*

>jgi|Vitbras1|3470|Vbra_11962.t1
MSEIVSYWWQYFVHHLPSHLSVGAVILRPLWRFLTLWPKLLLPATNRRTAALPHPHSHSSHQPPPLSIPPEPEP
SHENTLPKLIVPESVRHASRLSLSSHLDMSRPAPLDIRSVKKTTHSQTSSIGGIASSLTIDHSPHGALRTPQLG
PPGSDKADREMMQLLHDL SIEDVMGSKGELAREAAFMHGDTPRSSEQIPDVTGRVHSVESMTAVDGPGLRYLI
FTQGCPRRCVFCANPDSWDPKGGIEMSSKAVAKQLRKCKPYLKPNGGITCSGGEPLMQPHFIAAVFEEAHLLGL
TTCLDTTGFGNTRSWQITLPHDTDLVLLCLKSFDPHKYKELTKSTHKSAYEFAEELERRGIPWWLRF
LFIPGHTDNDFEYTHCENFVKAHSSCKGVLELLPYHTLGVEKWKNLGLPYVFEDMQPPPKDKVAALKQ
RWEAAGVPVMLAQ

>jgi|Caulen1|90458|g5733.t1
MTEVERMLHTGSEVDPFHSSLVPSVETESPRPAEVFGNVHSLESFSTVDGPGIRYMVFTQGCPSRCLFCCN
PDTWNPDPGQRMSSHDIADKMRRTLPFMKSSGGGITVSGGEALLQPEFVSALFQEAHEMGVDTCLDTTGQGC
LAEKLDVLPHTDLVLLCIKHM RKEVYKKITRRRINPMLKFVTELNRRIPFWVRYVLPGLTMESSDIDL
LCEFCIKQSFIKGIELLPYHTLGVEKWKMGLGYPLADVPPSRNDVEQIIQQVESYGLTVKCDLSTARQTSS*

>jgi|Bigna1|43713|e_gw1.83.28.1

MLGYTSKQQAYSCRPFPPSSVISSVKPPFSFSKTTTRSLSETLAQIGRVHSLYDFTTVDGP GIRYMI FLQGCLRRCAFCFNPD
SWDINGGTRMTADDIMKNLKDHSVGMYPNNGGVTISGGDALLQPDFVAEIFFRAHEELGVTTCCLDTSGLGHESGWNKVLPHTD
LVLLCIKSANPRTFFDLTESTNESMIKFAGHVATAKARDIPLRIRHVLI PGITDTEQETKALVRFARNQPTLEMIEVLPYHTL
GVSKWEEMNLKYRLADARTPTREEKLN FVKSLEDEGFKVLCEE*

>jgi|MicpuC3v2|4310|wlab.223208.1
MDRDDDATPRAAGASAATRRDDAPTRAKKTADEDDERRREKDRKDDPAMKTS SSPRPPAPRDIEDLG DVRGRVHSVDSFTAV
DGHGIRAI VFLQGCEKRCVFCCNPDSWSARS GASMTAKQVFTRIQRNARYYAASGGGITLSGGEPLLQPAFAAELCDLVRARG
LTA AIDTAAGGTSAGWDVVLPRVDLVLLCVKSSIASKHARITGSADDRPFQMLSFLAATRARGIETWIRFVLMSPPERWAPE
DEDLVGGSSGPTPPPTPPPPEDRAAFFRAIATDGDEIAGVAALCDAHANVQGV ELLPYHRFGEYKWKEMGVTPYPMRGGRTPS
RETIASVKKRKFARGVRVIA*

>jgi|Thaps3|263830|thaps1_ua_kg.chr_11a000085
IIGSVHSLETFTSNDGPGIRTLVFLQGCSKRCMYCSNPETQCIVDPFSCPEVAVSDEEVANVLKRYEHFLRPNNGGITFSGGE
PLLQPNFVGSVFKKAKSIGLTTTCVDTSGHGNPAIWNKCLPHTDYVMLCLKGMDMKLASFISGVSESNNQ RAREFAKYVRDNYK
NIKLSIRWVLLKDMTDTDEELDALAAFTKDLSPVFTHV ELLPYHILGKDKEYAMNSLYPMEGMEP

>jgi|Cyapar1|19601|g19218.t1
MVFLQGCRLRCIYQNPDSWNASSCAGKKMTVEEIIAKVKRCLPYLRPNRGGITVSGGDPMLQPEFVRDLFTEIKKLGVTTCL
DTSGIGTSKDAWNHVLVPCDLV MYCPKAFNPETFRKVTRTDARVATTSAEFAEEMMKRNVPMQVRYVILRDHTDTEEEVEH MV
QFCKKHRSHLAAIEILPYHTLGVNKYKVLGIPYSDALEVPPKEKIQAI VQRLKAEGFPVLVP*

>jgi|Chabra1|321999|rna-CBR_g55228
MRFGHSSQSAIGRNCKHLRQHVVWEKAPIWKS LVPKSLVPCERARSRLGEEPVLHRLCVVRKVLTARA AKGARVSVPDGA
ISSTVGESISDSRYVRPFARTATCLLTHSGRCATSVGKEENVEICVGLTRNDTGVRPASKEIGLRQASKDLGGGMAMATSAIS
RGINLLCPSTASASRHVDAVAMRNRPDGAQIQ LHSSPLSFANSSRPLHYRPTSSALSWSLPAATLEREMDSYCLHGLPEGLP
RKRCGGPICGAASASSSSSYTDAAGGAVRELAKRDVVL PKRDSLGPGLDPDMKPEELEDSEGGGIGDIPMVWGRVHSIESFS
AIDGPGIRSVMFLQGCWRRLCFCSNPDTWSPTTKETKRMS SKDIAKLLERNLPYYSGGGGVTLSGGEPLLQPDFCAAIFREAH
ALGISTALDTAGYGRPEEWAKVLPYTNVLLCLKAMD KAVYHKLVLGNSNTVMKFVEA IQKYKVQLVIRFVVLPGYTDTEQE I
EGLIAFAKTQPTLIGVEILPYHTLGVNKWEELGMNYP LKGVSPSPSKEFMMMEVKQRIENEGIKTLM*

>jgi|Tetstr1|421159|TSEL_012202.t1
MANTA AVPRRRWETR SRVARPEAPEQDGARDMTVT FETLQEMHESKEWEHVQAYESHHPPEMVPEIRDASYMYRTYYELDTNQ
DTDWFWLKEEQAI PNIVGTVHSVESFTAVDGP GVRYLIFMQGCAMRCKFCSNPDTWDFKAATGKLIESREMANRIRKVNYL
RPNNGGVTC SGGEAMLQPHFVGALFQEVHEMGLTTCIDTNGQGTKEGNWDIVLPHADYVMFCIKHMDPD KYHALTGLNLSVAL
NFAKELKARKIPFQLRYVLI PGFTDAPEDVDALISWAAEQPTMQFI ELLPYHELGLAKWKELGYKYPMEGVPTPSSEQITEFM
DKCRAQAVPVKCLNFSNEA*

PFO orthologs

>Cre02.g095137.t2.1
MEAHLKAPAVGKICRQGVARALVPAAAGRATNAKRAAVTMATPPEAPGAGTASAPGDYKTAAATTTTTTTTTTGTASNLV
VFSPAVVSSAGKPGKMAPMDGNEATALIAYAVSDVSFIYPITPATNMGEMVDQWASEGRTNVFNVM SVTEMESETGAAGA
LHGALAAGALATTTCSQGLLLMVPNMYK IAGELTPCVLHV TARALAKHALSIFGDHQDVM AVRQTGWAMLC SHSVQEAHDLA
LVSHLATLRASVPFVHFFDGFRTSHEINTIHLIDNEA IKSILAKPIYQDAIKHHRARALNPTH PHQRGT AQGPDVYFQCAEAA
NPFYDRAYDHVAAAMEEVS AVTGRVYKPYEYVGAPDADKVI VLMGSGSAAVEEAVNYLNAQ GKKGVLKVRLYRPWNSKVL LA
QIPATVRKVAVLDRTEKPGSGGEPLYQQIAMTLHEAEMADPALPRRYIVGGRFGLASKEFLPSHAAAVFENLDAKKPLRDFSV
GIMDDVTHRSLPDSRWLPLGTRTL PKGTFECLFWGMGSDGTVGANKEA IKI IASSAGMSAQAYFSYDAHKSGGVTVSHLRF GP
SPIDSPVLVQQADYLA VNHQSYMAKYDTLASLKP GGVLVLTNTVFTSPDSL GKYL PDKVKKQIAALKPQLYVIDAQSVAKASGL
GKHVNMMVMQTVFFNL SGVLPMEKALALLKKSITKAYERKGPEVVAKNHSAVDMAVAALKKLDIPASWSSLPTHV VNPNPAPKG
NTSRWEFIETVAKPMLALEGDKLPVS VFSPEGFVPPGTTVIEKRAIAAQVPIWKAENCTQCNICAFVCPHAAIRPALASPADL
VGAPATFGTIQAKGPGMGDLKYRIQVSPYDCTGCDLCTHACPD DALQSV PINSVLEVE TANWDFAGTLPARTDIMDKATVKGS
QLQPPLMEFSGACEGCGETPYVKLLTQLFGDRLI IANATGCSSIWGGSA PSNPYTTNADGYGP AWANS LFEDNAQFGLGIAMG
TMQRRKTVRKHVQEV LAVEAEKMPLSFGLRNALTRWDEH FDEPEVANV VAKELQPLLEREKDVHPWIRQLYDERDMLHKASI W
IVGGDGWAYDIGFGGLDHVLASGENVNILVLDTEVYSNTGGQ RSKATPKSAVTKFAAGGKERPKKDLGAIAMSYGDVYVASTS
LHANYGQVTKAMAEAEKYGGVSLVLAYAPC IMHGISSGMC SAIDESKQAVETGYWPLYRYNPAVKEDATHH RFQLDAKKLKGD
VEEWLSHENRFQILERKNPEVAHKLHHELD DAVHERFDRMKHMAAGGHMEPGSPPPPA A AAPQAQAPDHGHQHNTKESGCGGH
*

>jgi|Cyapar1|23463|g22964.t1
MSAAAPAAKAPTFFKSLDNEAAASVAYAFSDIAFIYPITPSSPMAEHADEWSAMGKPNLFGQVVQVKEMESEAGAAGALH
GALAAGALGTTFTASQGLLLMIPNMYK IAGELMPCVMHVAARSLAGALC IFGDHSVMAVRQTGWALLSAATVQEAYDL
AIVSHLSTLKS RVPFVHFFDGFRTSHEVQKIDATPLSHFKTFMDMKAI EDHRGRALNPAHPHMRGTAQGS DVFFQNV EAA
NQFYDKVPAIVEQTMNEVASVTGRPFKLVDYYGAPDAELVIVSMGSSCPVIEETIDAL KGQKIGLIKIRLFRPWP HQFL
ETLPKTTKRICVLDRTEAGA QGEPIYLDVCTSLHEGEVTKGLVVGGRYGLGSKEFTPSMVKAIYDNLALPKPKNHFTVG
IVDDVTHTNLPLGPPMRTVPEGT VQCQFWGMGSDGTVGANKDAVKI IGDNSELFTQAYFSYDAHKSGGLTVSHLRF GPKP
ITSTYLVTEADYVAVHLESYVTKNFLATLRPGGAVVLTNVWTA AELEKKLPASVKRQLHELRLARLYI INARAISDSVGL
GKRINMMVMQTVFFYLSGVLPFERAIALLKKSIEKTYGKKGPEIVAMNHRCVDAVAHLVTVP IPDEWKTAQDDTRRRAPG

PVATDFVKNVVKPMLAMEGDSLVPVSAFVPGGVVPPGTTQYEKRAIASKVPTWKADKCSQCNYCAFVCPHAAIRPVLVTEE
ELEGGAPEAFETRKAAGSGEITNFKYRVQVSPYDCTGCELCIHACPDDEALEAAEINDILDTEANWDYFTTVPPRGELFD
RNTVKGSGQFQQLLEFSGACEGCGETPYVKLLTQLFGERMIANATGCCSIWGSAPVNPYTTNKDGRGPAWANS LFEDN
AQFGYGIATGVIQRRAGLRKVIDDVATTPEGQPGFVPMSELSRSAMKAWLDVWQDADKCESASKRLETVLGAECRRPGSS
AMLREIYKQKDLLTKASIWIWGGDWAYDIGYGGDLHVLASGENVNILVMDTEMYSNTEGGQSKSTPMASVAQFAMSGKK
QNKKDLGILAMQYGSVYVASVALGANFSQVTRAMVEAERYPGVSLIIAYAPCMLHGIREGMSYALDESKMAVDGTGYWTLY
RFDPRIKVDAEHTQFQLDSKKIKAELTTFLKHENRFGILARSCPSEAKKLQVNLQEFVIARHEKYKHLAETDAERKANLA
AQFAKLAGGLSSDQWPSSVTVAYGSQTGNAEGVAHVIAQQLRARGVAHVKCEANDLEIGDLP AISHLVVVVSTAGQGE
QPDNIKDFWKALENPALPADYLANTTFAVFGGLDSSYCFCKSAIEIDERLGLQGAQRVLRARGIGDDQDEDKYETGFSNW
VPELYEALKLPEGKKETGVPNKHVKVVLGGKAHLDFDLTKQLAPTPILTEGCARHVRMLSNGLRSLDLDYDRNIRHMVIDL
EGSGLRYGVDALAIYPRNAPDRVADFLGFMGLDPLECIDELQFIEAEGQTATKQKPPVPAHLPLAQLFTDVLDFGRPS
KRIFYERLALYASDPAEKAALERVCSERSAAGGGAFAKLEETPTYADVIRMPSTRAALSIEHLVDILPVMKPRYYSIASS
PHHLPGLKLELCIVVDWKTSGADKFGTCTGYLKDLLDFGADGKASYTIAATMKPSAGMCMFAKPEFPVICAGLTGLAP
MRAMVQERLHQIKAGKKIGETVLF FGCRRFRAKDYL YGEEWEQALKDGNLTHLRVAFSRDQKEKIYIQNKIEQEPALIADL
MLNKNGYFFFCGPAPQAPNDIRAAMEKALQRGAGLTEQQAQKITELIIAGRYVVEAWS*

>jgi|Ch1NC64A_1|134349|IGS.gm_11_00195
MPHPSASPGDRGALPVTCQAAPAGPGRDKEPARGDLLRPHAVAAPPPAPLAQPPAAKPAAARKQAPLDGNEATSRIAYA
TSDVSFIYPITPATPMGEFVDQWSIEGRKNLFGNVMSVTEMESEAGVAGALHGALAAGALATTFTCSQGLLLMIPNMYKI
SGELMPCVLHVSARALAGHALSIFGDHQDVMAVRSTGWTMLSSHSVQEAQDLALVSHLATLAGSVPIVHFFDGFRTSHEI
NKASAAGIDLISDEQMKPLMDELAPATAAHHARAMNPSHPTQRGTAQGPDVYMQAIEAANPYHKAMPGLVQAAMDKVA
TGRQYHLFDYVGHPEAENVVVMAGSGCEALEAAVQNLTRDGHKVGVLKVHLFRPWSPEHLLAALPASAKRVCVLDRTKEH
GSQGEPLLELVASSIQARRQVECVVGGRYGLGSKDFTPMAAIGVFQNLAAPRVEDVRDRFVVGVIVDDVTFSSLPYDAEP
EVLPEITECLFWMGSDGTVGANKEAVKIIANQPGMNAQAYFSYDAHKS GGVTVSHLRF GPKQIRAPWLVDMAHYVG
QQQYWSKYDVLAGLRGGTVLINAPWKSWEVEQFMPVKTKARVAALRPKAIPLLKESIKKAYGKKGDKIVNMNYAAVDK
ALEHLVAIDIPADFGKVERLGAAEHAAASTGSRTAAEFLNNVMPMLAMEGDKLPVSVFTPGGFFPPGTTAVEKRSIAQ
VVPANWANSANCTQCNICSFICPHAAVPALATPEELAGAPAGFATVPIKGGGPALKGFQYRIQVSPTDCTGCELCVHACPD
YALTSTPTAQMLEAETTNWDDFFKQLPVRGELFAKDTVGRGSQFQPLMEFSGACEGCGETPYVKLLTQLFGTRMLIANATG
CSSIWGGSAPANPYTTDFKGRGPAWANS LFEDNAQFGFGIAMGVKQRRAALEAAKVVVGEGAGTPELRAALAEWLPIKD
NGALAGTAAAVAAALPAANGNGNGAVVSEAARGALAYIQNNTDLLDKPSVWIVGGDWAYDIGYAGLDHVMSTGEDV
NVLVDLTAEYSNTGGQSKSSPLGSVLKFAAAGKNRPKKELGLMTQGYPDVYVASISLEANYNQVVKAMSEAEAHKGPS
LVIAYAPCAMHGISSMSQTDAKMAVDSGYWPLYRYKPGATEADGKLTLDSSKKIKGALEDFLKRENRFMVLRTRKDPEAA
GKLHHQMEEHIQLRQERLINMAKGTNGAAPPTPPPAEEEEKK*

>jgi|Tetstr1|465369|TSEL_010055.t1
MPPDGEKAAAGAATGAVVVPRGMYAMDGNEATAVAFAMSEVSFIYPITPSSTMAELADAWAVAGRNVFGQVPEIFEME
SEGGAAGALHGALVSGSLATTFTASQGLLLMVPNMYKIAGELLPCVHVAARAIAQGALSIFGDHSDVMLCRSTGWALLA
SHTVQECVDMALVAHLATLKASVPFVHFFEGFRLSHEVSKIDVIDMLAAKPLVDAAKLARHRSRAMNPNHPNQRGTAQGP
DVFFQAVEAANTYYEKCADVVEEVMMAAVGALTGRHYKLFEEYEGCPDAEHVIVIMGAGAPVVKEAIDVLTARGQKVGLLKV
RLFRPWSPERFVAEVPRTTKRIAVLDRCKESGALGEPLYLDINSTMHAAEQREGQEWLSSIAIIGGRFGLGSKDFTPGMA
LAVFAELDKPKPKNHFTLGIHDDVTFHSLAVLEEPNCVPAGTVECLFWMGADGTVGANKEAIIIGTSTELYAQGYFAY
DAHKS GGVTVSHLRF GSPITSQYLVRADYVAVHHHSYFHKYDWYSAIKPGGKLVNLCEYPADAEVALRFPVDVKRLAE
LGVEVYSINALEISRSVGLGKRINMIMQAVFFYLAVLPFERAVALLKAAINKAYSRKGEVEMNYYAAVDKAVAGLRKI
AYDRAAWAAGTQYVALHARHNDVAPYVREIVMPGDELPLVSAFTDYGGLVGGVPPGTTQFERRQIAASVPKWNVTDK
CTQCNECSFVCPHAAIRPQLTTAEEREHAPEGYHSEQAAGAEASGLNFRVQVSPYDCTGCEMCAHACPQECLEMPADEM
IEAEDEHWEWGKSLPSRGLYLFERGTVKGSQFQPLLEFSGACEGCGETPYVKLLTQLFGERMIVANATGCCSIWGSAPV
NPYTTDAKGRGPAWANS LFEDNAQFGLGIATGIIQRRALAKDNVRELLAAGGGAPEVPEELALALAAWLEDADEAEGSKK
AGEEVEAELEKVRAAGGGELPAPLDEVYAAKDLLTKPSIWVVGDDWAYDIGYGGDLHVLSSGIDLNVLMDTEMYSNTEG
GQRSKSTPMGAVVQFAAGGKRRVKKDLGAMAMAYGDVYTASVCEADFAQVVRAFTEAERYKGTSVIIAFAPCIIQLGLRG
LQSQSLAYAQEAVSTGYWTLRYRHPERTADPNTPLMLDSKKVKGELQSFLDKQNRFFVALKRRRLPQVAAQLSAGLKASLT
AKRCSTESCAQSPREEHKALLESFRKLVEGAEGGAAGGGGGGGAEEVVVLYGSETGNAENLAHTLAANLKARSVRVSC
KECDETAEDLEEIGARGAVVVVISTAGQGEFPKNSATWFQELARARPAGWLSKLRFAVFGLD SAYCHFC LAAAEVER
RLGELGAVPLLP RVGDDQSPDKYETAWMEWQPELFDRLQLAEPERLVGAPPPEYTVQLTKGHHPEWQSAFEVYPFAPIA
TDRRAFPIKLVENKRMAAPDHDRDVRNLVFDISGAKLRYFP GDCLAIYPRNALQPAAALCQWLGLEPDAMTRVKCLDAGR
ASKAKRDLPRGMTVLQLFTDVLDFGRPLRSFYEFYCYVDDPAERAERFPSCKANLDIAHILDVPTIKPRLYSIASSQ
RERVGTCTSFRLSLSCDPEHDDAELNTLMVYVQPTPGGIFHPQDAKVPVMMVGLGTGLAPFRAMVQERKALARAGVELG
ESALYFGCRTRKDDYLFGEWDAAHKEGYLSELRVAFSRDQPKIYVQDKIREDPALVWRLLENEGYFFFCGPARRAPQ
QVREGVEAAFVAAGNLTKQEAEEKVDKLVGLGRYVVEAWA*

>jgi|Vitbras1|12792|Vbra_21448.t1
MKLDIVDGNTAAAHVAYAFSDVAFIYPITPSSTMAELVDLWASQGRKNLFGQKVDCREMQSEAGAAGAVHGS LAAGALT
TFTASQGLLLMIPNMYKIAGELMPCVFHV SARAVAGQALSIFGDHSDVMAIRQCGWMLSSHSVQEAASDMALIAHLATLE
GSVPFVHFFDGFRTSHELQKIEILDYEQMQPLFDDMKSELERHAKALNPTHPHQRGT SQGPDVFFQCVCEASNPFYAKLP
EIVEGVMAKVAPVIGRTYRLFMYYGHEQADRCVVMGAGAPVHLEECVNYLNSKGEKVGLIKVHLFRPWKADALMAAIPDS
VQRIAVLDRTKESGSFGEPLFLDVAATLQDRGDGRVLVGGRYGLGSKDFTPAMAKAVYDNLAAEKPKRRRTVGIKDDVSN
MSLELGPEFRVAPGTVQCMFWMGSDGTVGANKNAIKIIGSNTPLYVQGYFAYDAHKS GGVTVSHLRF GPHPTITSQYLL
QDADYVAVHHQSYVKFKVLQHIRHGNFVLNCPWTRDELETEL PASLKR TMAAKSVQFYTM DAAGVARETGMGRINNI
LMVAF LKLSNVLPFEQAVNLLKSAIQKTYGRKGEAVVHANWRAVDAAALDAIVKVEYDASEWASATDKKVVDVLENAPT'FV

TEVMPVERLEGDTLPVSKFQPGGFMLGTSAFEKRNIALQVPTVDMDKCTQCNYCSFVCPHAAVRPFLITEDEADDAPE
HFVTKKAKGGAEMAGFKYRIQVSPMDCTGCELCVHACPDNALTLYPVDNVIAQENPNWDFALTLP SRGELFDRSTVKGSQ
FQKPLLEFSGACEGCGETPYVKLLTQLFGERMLIANATGCSSIWGGSYPAPYPCVNDKQGQPAWGNLSLFEDNAEYGFGMA
VALNQRRLHELQSVLQGGDLISDALREQLVGWEKEWRHPAKSQAYYETLPPLEKESGKHPLLERIYTLRDLPLPKFS
HWIVGGDGWAYDIGYGGDLHVLASGENVNILVLDTEVYSNTGGQMSKATPLGAIKAFGQSGKKYNKKDLGMMAMAYGNVY
VASCALGANYSQVIRAFVEAESHEGPSLIMCYAPCSEHGVLVSTWAHQADDVKIAVESGYWLLYRYNPELEDKGENPLQL
DAKKLKVDVRQFMERENRFASLMRAEPEVATKLQENMSKFTKERFNRMKSRSQAPGGIDSAQLNATFEALQOGLATGGVQ
AGPGRFDVLFQSETGNAEEVAKRVGGLLKGRGYKVKVRELDEVEVDELAELKFVLVICSTAGQGEFPQNAKDFWTAVSDP
TLDDTVLKDHLHYGVFGLGDSSYCFNQAAKMIDEQFAKLGATRLLPCGLGDDQDEERYDTALGEWLPDFWTEIKAPAPLH
ESDELPPPNFAATPLDAATCTYEQCVPOGAVLLRLTENTRITPDDYDKDVRHLEFDLTGTNFRYGLGDSLAIFPQNDPKE
VEAFCNMYGDADLWVQFSKSDPTQSNARYDALFKKPMTRQLCCECLDLFGRPTRHFYESLQKCTDSEKSRKHLVS
PEGKPEMRKLTDETVHCFDVLQMFPSAKPTLEQLIDMVPLLPKYYSIASSQRYLGGNRLQLCIGIVDWKTPSGKYRVGE
CTGYVTRTPVTETSPPIFYACSLKATTFNLPPNDRVPVIMASMTGLAPFRAFVQHRAFLKRQKQVGPMTMYFGCRYSK
KDYLXGDELQQFAEEGVIDE LRVAFSREQKEKFYVQH HITRDPALLYRRFIEERGYFYLCGQSRQVPIDIRNAVLQALQS
EGGMTDKEATDKLIDMQIRGRYNVEAWS

>jgi|Symmic1|517|rna30934

MRLLEDGPASLFKMEQSASGDMAPSSGGIGYLYFFLRRRMQRALQDETGRTRTLLWDAPGPVLEQRRRLVEEPPGTRVHE
FVVLSSLTSGDGGAECLWLPSETEPNWATILETVVDGNAERDDVFSVLSLEMVQALEVPFKMVATYSHSGRKATEYC
HRDLRARVNAHRKEGAATVEEIEFNKLCDICVAVIPRDGKFATMPFKEARKKRAIVLWCTKHYTEKPPDDATDISDGEVEA
IVAQAAPPPEPGFYRTNTGLLCVCS PANAWWECPWCKPGISKQHTSALRNDSRCFADSSSPTGDTVAKCRHGRLSMGAG
ACRLQALKFGVRRWLLVAQVGTTLQKTSPPFGPYGGLCRQFLMRDQERQEETPQRLEQLGAGSGAADVEYEQFMKAKEVQR
VMFQQENKTIHVQCADTDIEATGTSIACIDGNQAAAHVAYALS DCAFIYPIPTSSPMGEMVDEWAAQGLINCYGGKLSV
TEMQSEAGAALHGAALKAGQGLLLMIPNMYKIAGELLPCVMHVAARALAGQALSIFGDHSDVMACRSTGWAMLASESVE
MAQVNAIVAHVLSMERRVPVQHFFDGFRTSHEVNKIKLIDYNTMKSFINWDAVKEHHD LAMNPRHPHVQGSTSQGPDIFFQ
CVEAGNSFYDGLADLFEEKGKLVQEQTGLHFALYGYEXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXQPCQKRRSVASMLHASAHSEIVVVGGRYGLGSKEFTPNMVLVSFENLAKD
SPKPRFVVGIEDDVTHLSLPVGPWLVNLP EGTTECMFYGLGSDGTGANKSAVKMIALGTELHAQAYFEYDAKKS GGVTI
SHLRFGPKPIHAPYNVRAADYMAIHKSSYVHNYDMTRYLQNAVVCVINCSDVKELEQQLP AKMRRDLAEKKAKLFIIDA
TKIAVRKAGLGRINMIMQTVFFKLF GPKPIHAPYNVRAADYMAIHKSSYVHNYDMTRYLQNAVVCVINCSDVKELEQQLP
PAKMRRDLAEKKAKLFIIDATKIAVKAGLGKRIINMIMQTVFFKLSGRLSAVMPEYEEAVEMLKKS IKKMYGKKGDKVVKMN
IDGVDASIAGIVECEVPAAWASLAVDTEASDAKTSTVAYAKGPRMFPEVQNASQFAAQVQKPCNNLDGNSLPVSAFVPGG
RVPCGTSQYEKRGIAIQVPKVDMDKCTQCNC SLICPHAAVRPFLMTSQELGKAPTSFKEGSRSAIGGGVLDNYQYRIQV
SPWDCTGCELCVRICPADALT LGPAEKAIQEEEEANNWFAVALPDRGDEIDKTTVKGSQFQKPYLEFSGACEGCGETPHVK
LMTQLFQDRLVIANATGCSSIWGGSNPSFPYTTSNKGEGPAWANS LFEDNAEF GFGMRKAFKQRRDY LALQVEDTLADQS
VKMSDELRLQALQFVLV MRKEQMHDLLL PKGRSIIHQIMEKLVPLLEKEKGTHPKIHNLYDLEDMFGRSSFWIVGGDGWAY
DIGYGGDLHDVIAQSEHVNIILVLDTEMYSN TGGQASKATPKGAMAKFAEGGKLTQKKDLGQMAMTYKNVYVASICVHVN PQ
QAVRALIEADAYPGPSLILSYTPCISQGF PMAESI QHCQMAVDSGYWPLYRYNPEAAAHGNNP FQLDSRKIKGDMFKLLA
KENRFAAVMRRDPKHAEELDNKLQENNVQKNHLLQVLNKEDLEGQFSKLVEGLSDAKSSGDSITVLYGSETGNAEEQAKN
LMQDLIARGLKSTVSSLD DFEDEL PNQKIVILVSTCGLGEYPANCKQ TWLKLQSPDL PMSWLAGVKFCV FGLGDSTYS
QFCVAAAGFDTHLGLG GQRMLKRGVGD DRDED RYTGWEAWLP ELWKVLGAPALPLSQDIPAPSYRVDASAGDLAKPPI
SDDDIIPPGANRLTLLTNNVLTGD AKYDRDIRHYEFKIEGTNVAYKTGESLAIWPRNPEDKVMFEFCMMGFDAQAHLRLI
PLEGARNWCPTELSVRQLF SHVLDIFGKPNRKFQTLALFAKDEAEKKQLQSVVENSQEGAALYRDLTHDFAHADV LKK
FASARPP IEHLLQMMFVLKPRSYSIASSPLMHPDKIQLCVVLVDWTVESTKELRLGECTGYMQQKPGAQKIMHADRSSAI
VLPKEPEKPVIMAGMTGLAPWRAVTQERVMQARQGLKVGKCM LFFGARYKQEWLYREEEFESYEKEGVLQMHTAFSREQA
RKIYVQHRIVEAGADVAERMLNQGGHF*

>jgi|Guith1|159505|estExt_fgenes h2_pm.C_220012

MIKNTTWSCIDGNQAAAHYALSEVSFIYPIPTSSPMGEMVDEWAAANGLKNIFGQTL SVTEMQSEAGAAGALHGS LKAG
TFASYTASQGLLLMIPNMYKIAGELLPCVMHVSARTLCAHALNI FGDHSDVMAARTTGWVMLASENPQM VMDQALVSHL
ATMDMRVPVLHFFDGFRTSHEVNKVRVINYDDIKKIFP WDSVKTHHDSALS PMNPHIQGTNQGPDVFFQAAEAANA IHNT
IPAYMQKWKSEKVAEITGRYGFSSYEGHPEAENVIVIMSGAVTVHETVKYLN SHQNQKVGVLKVR LFRPWITERFMAAL
PASVKRIAVLDRVKEVGAIGEPLFCDVCTTLHLSGMDHIK VYGGRYGLGKDFTPGMVLSVFKNLEAAHPKTKFTVGIVD
DVTNLSLNVQEVDMLPSGTIQCL IYGLGSDGTGANKSAIKTIAQNTDNFAQGYFEYDSKKS GGLTVSHLRF GPHPINA
PYLVKNADYIGIHKESYLSRLDMLRNLKENGTVVINCTFGPEKVENMLP PPMKYQLASKKASLYLINAVKVGRETGMGR
INMVMQSVFFKLSRVL PFEKAI ELLKKDVQKMYGKKGEVVKRNWDAIDRSIENLVEVPVPAQWAIIPVTNKEIRKSTL
EFVDEIVEPVNKFEGNDLPVSAFVPGGRVPTGTSAFEKRGIALQVPRVDMDKCTQCNYCSLICPHAAVRPFLFTKAELEG
APEGIKKGSRKAIGGGVLDNYNFRVQVSP L DCTGCELCVRICPADALFFEETDKAVELEAENWN YAISVPNKGEEIDKTT
VKGSQFQQPLLEFSGACEGCGETPYVKLLTQLLGDRLVVANATGCSSIWGASAPSFPYTVNARGE GPWANS LFEDNAEF
GLGMRRAFKQRRNQLMFLVDDALEDKTVPLSDKLRKLLS QYSVMRHEDKLDMLLPKGRSFYNQLREKIIPLLREEAKNHE
KIQELWDQADI FGRHSHWIIIGDGWAYDIGFGLDHLV LASEEHINVLVLDTEMYSN TGGQASKSTPRGAQAKFAETGKLT
AKKDLGQYAMTYKNVYVASICIHVN HQQAIRALLEAEAYPGPSLVICYS PCISQGFMAEAIQQCRDAVESGYWPLYRYN
PLLAEGGNPNPFQLDSKKISGDLFTFLAHENRFASVMRRNDFSQAQHQDRLQRQIADRAKALNVLSDEL PASLKAAGPSE
SVTVLYGSETGNAEEQAKSLFADLKARGTSATLSSLDNDFEELPNQSTVLVVI STCGQGEFPANSHKFWMKLSDP T LPM
SFLEGIKFSV FGLGDSTYSLFCVAAERIDVRLAELGATRILNRGIGDDRDED RYTGWDNWT PQLWNALHVPQKPLERKI
PKPAYKVTRTQGEATPSVSNDKLVPPGSNPLKLMENTLLTPEGYDRDIRHYVFKIKD TNVEYKVGVDLAIYPRNHVDQVE
EFCKMYGLDPNEELNVVSTPEARNQIPEELNVRQLLQCVLDIFGKPNRRFYDTLSL FATDPAEKQKLELIT SEDPDGKAL
YRELSDHMANHADVLKRFPSARPPLEQLMDMIPVIKPRSYSIASAPSMHPDEIELCIVAVDW EVPSTGEKRFQGCTSYLR

TTKPGDTIMCSVKPSSIVLPEDNKAPLLMAGMGTGLAPWRALTQHRIALKQQGIDVGPCTIYYGARKGATEYLYREEFEK
YEKMGVLRMVTAFSRDQFPQKIYVQHRIREDYENVFRQLMKEQGSFYVCGSSRNVPEDIYNAMKEVMMMLGGGMQEADAEAA
LASLKMDGRYTVEAWS*

>jgi|Thaps3|31316|e_gw1.1.16.1
AVANASYRMNDIAYIFFPITPSSVMGELVDEWSATDDGQGKKKFLNLWNQELKVVEMQSEAGAAGALHGALNSGSLATYTT
ASQGLLLYIPNLYKISGELLPTVIHVAARALTGQALSIYGDHSDVMLVRGTGLAMVSSFNVQEAHDMAVITQVATLNSRV
PFLHFMDGFRTSHEISKIQLVSDEQLKELMPWEKVEEHRQRALSPLHPSQRGTAQAPDVFMLVLESSNKHYQAVNSIFGK
AMEDFARVVGRHYKPFYHYFGSTTPTIAVVTMGSSVEVVKETLCKMNMEQACVIGVRLFRPWSPADFCVKVLPKTVKRIA
TLDRTKESGAQGEPLYMDVCASLMSCNRQGI FVAGGRYGLASKDFS PRMALAVVQNLMRKDVGSIQHPFTVGIIDDVTHL
SLNLGKDVNPLPENVTQCVFWGFGSDGTVGGNKETIKIIGDSRTDMSVQAYFEYDAKKSSGWTISHLRFSDSHTLSAPYK
VQDCQANYVACHNESYAAANKFDVVRFSKKGFTFFLNTKIASIDSPDKRIKEFYSSCRALRKVKFYIMDADRLASTYGLK
GKINMICMSAFFRLSGVLPIDEAVALLKKSITKSYSYKGDVVKKNHDILDAVCSDPRYLLEVEVPSPRWRATLTDERRA
YANRHIALIDDEKTRKFMEEIAEPVSHLEGDNIPISKFLSNEMLGGMQPGTTRYEKRRNPNSALIPKWNASACTQCNC
VAVCPHAVIRPFIITRDEAEQAPHPDQFQTVKALGTELAGKRYTLQISPLDCTGCNACVEACPEAPKALEMQDVEDVLK
GGEDNWNFAMTLPERGDLVDKYTVKGSQFQTPLMFESGACSGCGETPYFKMLTQLFGERMIANATGCSTIWGGSFPSPN
YTVSKKTKRGPWANSLFEDNAEYGFGMFSAMKHRRERLIQLVQEYWLECRDEKSDKCTYLFDMQPLFEVLLSNKSNGR
QSKLEMIYSDRDMFPKLSQWIVGGDGWAYDIGYGGDLHVEAFEANDVNVLVVDTEMYSENTGGQASKSTPAGAAVKFAMGG
KQQTKKNMGEIFMTYEHVYVASVCLSNQTQLLQALVEADRHNGPSFIVCYAPCIQQQVRPEGLNDMFDECRFAVDSGYWP
LYRYNPNLINELGKNPFIILDSKKLRREVASFLKREGRFINLKKAHPEIANDLFNKMNYDVHHRMENMLSRAAG

HYDA orthologs

>jgi|Chlre5_6|12867|Cre03.g199800.t1.1
MSALVLKPCAASVIRGSSCRARQVAPRAPLAASTVRVALATLEAPARRLGNVACAAAAPAAEAPLSHVQQALAEALAKPKDDPT
RKHVCVQVAPAVRVAIAETLGLAPGATTPKQLAEGRLRLGFDEVFDTLFGADLTIMEEGSELLHRLTEHLEAHPHSDEPLPMF
TSCCPGWIAMLEKSYPDLPYVSSCKSPQMMLAAMVKSYLEKKGIAPKDMVMVSIMPCTRKQSEADRWFCVDADPTLRQLD
HVITTVELGNIFKERGINLAELPEGEWDNPMGVGSGAGVLFGTGGVMEALRTAYELFTGTPLPRLSLSEVRGMDGKETNI
TMVPAPGSKFEELLKHRAAARAEAAHGTGPGPLAWDGGAGFTSEDGRGGITLRVAVANGLGNAKKLITKMQAGEAKYDFVEIM
ACPAGCVGGGGQPRSTDKAITQKRQAALYNLDEKSTLRRSHENPSIRELYDTYLGEPLGHKAHELLHHTHYVAGGVEEKDEKK*

>jgi|Chlre5_6|18188|Cre09.g396600.t1.1
MALGLLAELRAGQAVACARRTNAPAHPAAVVPCLPSPRAGKFFNLSQKVPSSQSARGSTIRVAATATDAVPHWKLALAEELDKPK
DGGRRVLIAQVAPAVRVAIAESFGLAPGAVSPGKLATGLRALGFDQVFDTLFAADLTIMEEGTELLHRLKEHLEAHPHSDEPL
PMFTSCCPGWVAMMEKSYPELIPFVSSCKSPQMMMAMVKTYLSEKQGIPAKDIVMVSVMPCVRKQGEADREWFVCEPVGVRD
VDHVITTAELGNIFKERGINLPELPDSWDQPLGLGSGAGVLFGTGGVMEALRTAYEIVTKEPLPRLNLSEVRGLDGIKEA
SVTLVPAPGSKFAELVAERLAHKVEEAAAAEAAAVEGAVKPPPIAYDGGQGFSTDDGKGLKLRVAVANGLGNAKKLIGKMVS
GEAKYDFVEIMACPAGCVGGGGQPRSTDKQITQKRQAALYDLDERNTLRRSHENEAVNQLYKEFLGEPLSHRAHELLHHTHYVP
GGAEADA*

>jgi|Cyapar1|17198|g16919.t1
MTPPVLRVCPAPLLTFDQAGRPNSLAALNGIGFASRAVQRQGSQGSMPKPPADPKMQHLIDVVLAAMPTILKDSNDDEEALS
IAKLRMEMKPEDDDHKSDEPSEKVKKDAQVVLRRIEDKRLEKRVHCAQIAPATFVSIABEFDLPGSIKPAQLVQALHKIG
LDHVFDLRFAADVTIMEEASLEKIKSGGKFPMTSCCAGWQRLVEEDYPDLDHVSCKSPMGMNLSALIKSMFAERSGIDV
SQIVVTSIMPCSIKKVEAQRPQLQDPISKTDHVITTREIGQIFRAMNIDWDEL DSGSGPTGQFASWMGAGTGAGAI FGVTTGGVM
EAAVRTAYEFATGKTLEKVEVRDARGLFGVKEGERAPAPRRPPAPDAAPAPRAGALDVPVGVELKFCVVCGLTSIKRIMEMTR
AGECFPHFVEIMMCQNGCISGPGQPKSKDKGIAKKRMQALYTSDERAVVRKSHENPAVKALYQEYLGKPLGEKSHLLHTEFW
PNHVKPENGHGHGHGHENGNNGHGAHDHMEKKALDASFSAITAGMHEEKEEAAHAEAAKPAAVVEYPPPTVEWTLPKAIIRIY
GSETGNAESLARVVAKECAARGIAVKCTEADELEMDLATVSTLIVVVSTAGQGEIPDNAKDFWTGISKKHPPGWLRNLHYS
VMGIGDSSYAHFNRAAVEIEERLAELGAHRFLARGVGDQDEERFETGFEKWAPAYKTLKLPVKDDGSILPSSHVFLTAG
PKEGGARKFDETWKPIVTDPMGQLIRVESNVRIPTDGYDRDVRHIVFEVDQVPEPLKYTVGDALAIYARNDPEEVKAFKLFYA
LEPSDLITFKLNEHFKGDAKAKPPMGQTSLLDYVRPPRPAPLSSPAEACAERGLSGGAQFCDCVDVFGPNLRFFGALAKF
ATDVHQRDELDSIARKTWRGTAAFSVLLKDLGHVTYAGVLQAYPSARPPPLAQLLDMI PAIKPRYYSIASSPHVAVGKLELSIV
IVDWEKPDKSVGFGTCTGYLRQLVDGPHWISVALKPGVGICMPHPSTPIVMAGLGTGLAPMRAMVQERSFQVDHGRVGSV
LFFGCRHSGKGDYLSYDRWDEFVAKGALSHLRVAFSRDGPKKVYIQDKIAAEPELLYEAFVKRGGYFFFCGPAGRVPQDIRAAI
ETAMSEGGKEDGFTREMAKYIDEMRKTSRYIVEAWS*

>jgi|ChlNC64A_1|24040|e_gw1.12.53.1
MATIHINGHTVSVEPETSILTAATQIGIHIPTLCTHPRLPPTPGTCRLCLVEVAGGALKPACATPVCHGLEVTDDSPQVKDSI
RGVLALLKANHPADCMTCDSVGRCEQARPGFWGGAWAGQRMACPRQRMLPMLGCRCLKCGRCVTACGLVQEMDVLGMKGRS
RERHPAVLTEAMDLSKICISCGQCAVMCPVGAITERAEWREVEDQLDAKRKAGRGAGRAGLMVCVTAPAVRVAIGEELGLAPGT
ITTGQMVAARQLGFDYVFDVNFAGADLTIMEEGTELLQRLRHAWGLDLPAEGSGGAGAGPLPMFTSCCPGWVTACEKSFPELL
PHLSTCKSPQMMGAVVKSHFAAKLGKRAQDICLVSMPTAKKYEAERGEMVREGEFPDVDYVITTREFGRLLRERHIPLAS
LPESAFDNPLGEGSGAGVIFGNTGGVMEALRTAYELAAAGQPLPKLEEEAIRGLRGIKQATVTLPTTAPAGMASRQLRVAVAS
GIGQARHLLERMHTGHSPHDFDVEVMACPGGCVGGGQPKSADPLVLLKRMGAVYSIDERSAIRKSHENPSIQKLYKACAEFL
GEPGGSLSHQLLHTTYINRSTASQPTYTAFQRMDEPCNPKLQQAAGAGSSSSALPGTSA*

>jgi|Nanocel1779_2|558506|estExt_Genewise1Plus.C_190151

MLGPFLFAWARIGPTAASAAALRRLPKPRTFAAASFSTTPPAIPTTSPTATFPAEEQLVNVEINGRSIAVPAYATILEAARAA
GAFVPTLCAHPSIPACATCRICLVELETGQGTKKVVGSPASPTDARRVEKLVPM SKLVPACITQVKEGQRIHTHSPDVLENVRF
ILQLMRAKHNPNTCQTCADAGNCTFQDLLYRYQVEDIAPTEIRKNFEHHHEWDGGASIGTPTDSQTPAPSKDTPYGAITFDPDK
CVKCGRCVAVCQKIQQVGGAIGFVGRGADMHLSTMQGLPFQTKCIECQCQVAVCPTGAMRERHDLFHVIDALQNKEVVKVAAV
APSVRVAISEEFDLPPGRISMGIIVVGLKQLGFDYVFDTNFAADLTIMEEGTEFLSRLKKEPGAGPLPMFTSCCPGWINLVEK
NYPYLIPHLSSCKSPQGMMAAVKNVWAPRVGRDPADVNVNSIMPCTAKKGEAHRPADRAATYSAKLDTTVPNNDYVLTREI
GRLFRLHKISLPALESQAPDDPLGESTGA AVLFGATGGVMEAAALRTAYELASGGQPLPKIEMSDIRGLQGVKSARVCIPNPDG
KGP GAVVRVGVVHGTAETRALLERMKSGDESAQFDFVEVMACRGGCIGGGGQSKSDDPFVLQKRIAHVYNLDVESAIRKSHEN
PSIKKLYQEALGEPGGHLSHQLLHTQYTD RSQDTK*

>jgi|Vitbras1|16169|Vbra_650.t1
MLTNRS LQTLASAPLRLPAGSLALALPPLRAISTATVAAKAHDRTHQSSVAASAARRTPLFVSN SRVLGKGLFFHRPPLVTAT
STLPSYAPWTLTWRLPYAAVAAAQTQPKQQEIHVDVGEVARLRDECAVNPECGYKPLAFDPTKCLEWACEKACPVGDKFL
FNQSTCLLCGKCAQVCPTGAISSPPMPIRETLSKLKEEGKVTIAQVAPAIRSVLTEEQICYGLHDLGDFMVFDTNMAADICIM
EEAAELVTRITAPSQESPLPMFTSCCPAWINMV EKHYPHLIPHLSTCKSPQAMLGALVKHHWASQKGV SADDIFMVSVMP CIA
KKDEVQRVQLEHAVD AVITTKELMDQGSFDLEAAKNNTQKLHFDQPFAEHSGGAVIFGASGGVLEAAARTAYKMVTGN EPPHA
SLQPARQQGDVKELVLPVKEGVEIKVAIVNGMDHLSEVLPHMDKYHMI EVMNCEGGCVGGAGNPRALDLKEIQKRQQELYKRD
EKATYFASHETPAVERLYK GELKDEHLRHHLFHTHYNHRFEKDQWHL PATKLIHVKPATTPVSP PAMEHCKTLIPIGFDEIHV
KDRVRVTPEGYKPAQYVFKFTVPSGAAGGPSTYRTGDHFLVPCNPPELVDRLLMWLGKDP MRAHEELVEVTS EDPLRAFP GM
PMTLHRLLSQFVDLNGFPVYYLRRLVSPDTGFVPPEAVAHKNAEA IMSLPKGKYSFSEIVSACDIHPRFYSIASSDTRCPD
EIELLVVDVPGGLCTGHICRQPLEYIQTTP IAGKLRKGVFNYPMDGNTPMVCVTVGSGFAPVRAFLQERKDRGLSGDFVLVHG
CRHASKDFVCQEDIQDGGVSASFYAFSRDDPKKKVYIHDKIAENGAAVWELLQRPDCQFFYCGSSGKI LNQM ETALIKCSEGI
CDGGSII RQMKGEHRYHLEAY

>jgi|Thaps3|264905|thaps1_ua_kg.chr_23000023
IANALVSLRVDGRPVKVAEGATLLDAINTSGSHVPTLCYHPEFQPKAVCRMCLVNVKEANATATAGKLLPACRTKVEEGQEV T
TNSEDIKAFRRRDLQFLLNRHPNDCMRCEAAGNCKLQSLVQEECEDMWPTKTSRGSDEHPHLLLDHTSPSIWRDMSKCI ECG
LCVDACSAQKINAI GFAERGS GMLPITAFDKPLSETGCISCGQCILRCPVGALIERPDWHRVLDVLD DRKRTTIVQTAPATRV
AIGEEFGLEPGSVSTGRMINALRELGF DYVTD TNFSADLTIMEEAHELLQRLQ GKREGALPLFTSCCPGWINYVEINRPDLIP
HLSTTKSPQMQHGA IARNGPMAKQIAAQTSESEEPYIVSIMPCTAKKDESVRPGNRGDI DAVLT TRELAKLIRHRDIPFASLS
NDGEYDSPMGESSGAGAI FGASGGVLEAALRTAADTLGLDGKNVDTLQHEQLRGVDRGIKVASIKGVGSVAAVSSIGSAI ELL
NTDHWKKFLMIEVMACPGGCLGGGGEPKSDDKDILKR RMAGIYSIDKNAPIRKSHENKEVQQLYKDFLSFPLSEISERLLHTS
YAPR

>jgi|Tetstr1|462877|TSEL_007826.t1
MTFSGTVKLSGLDDHITPSQACVVALDGT KIDPTDMSEEVQIHSREDKPAPAMKEGADGKPV TISLQDCLACSGCITTAETVL
LQQQSVDFRSRLAEPGTVTIASVSPQTRASLA EYYGLSAKEAGARLAGYLKSLGVQYVLDLSSARDVSLMEAAAEFCQRFRF
ASSNAAGQMEIDGTEAQSATAGPLPMLASACPGWVCYAEKTHGEYVLPYISTSKSPQAVMGTL LKRQVAACHGMDPAKAFHCS
FQPCFDKKLEASREDFETEAKVKETDCVLSTLEVVEMLNSQGLDLKDCPAGSVDSLLALIGGQPAAGAE DDGALYGWPGALEV
VLTVNDEAVLR FATAYGFRNIQTLMRKIKRKQCEYQVVEIMACPSGCLNGGGQMRPKEGQSVQELIQKLD ETYHHQDV RPRWP
ADNPLVRYVYTSPNLGSCYGPQARNLLHTQYHQREKSVSTAIVDW*

>jgi|Chabra1|335382|rna-CBR_g26339
MLRQLWQRGRGDRTGARCASWAI RRLTTAAEGTMRSVSAHEEAFSSRAGSAGGELRRKMCTAAAIPQ NARQLWTAGCACGGSS
SRAAVSGGSDMGRNRQQQPRALLHSESAAAAAAVEEAKELVGIKVDGVEYRV PKGMTLLQACRDLGIYIPAVCYHPALKP VGT
CRLCLVDTGGWRMEPACTKEVAEGLTVRTH TDEVFDTVGNLALMLRLKHPNACSTCSVNSRCEFQDLLLRYGVEDPPMLEKQK
RVDHHVEEAGIAHEKDESSPALSMDFDKCILCLRCVRACNELQGMN ILGATARGDHETVAPLYGLPLGETSCLSCGQCAAVCP
VGSITEKSHVRQVEELLQMREAESLP PRTEKDWDQEMPKTRLLWGP RQMVSASQEV RASSSSSSSSSSGGGGRGYATTSSSRSS
SSSAGSEAAAAAEATAAAAVAPRTGKRPVMVAQTAPAVRV TIAEAFGMEPGSISTGHMVAA LRKLGF DYVFD TNFSADLTIME
EATELLGRVKS GGPFPMTSCCPGWINMV EKYTPELIPNLSSCKSPQQMLGAVVKSHFAASVGRHPSEVKFVSVMPCVAKKDE
AARKEMHNE DLGGPDV DYVLT TRELASLIERARPRIN FASLKEAEFDSPLGQSSGAALLFGVTGGVMEAAALRTAVAVTAPADA
PPMPKLD FTEVRGLDG VKEANIDVNGTELKVAIAHGGVHLRSLVEDILAGKR TYHFVEMMACRGGCIGGAGNPKDSVDEKLLQ
HRAQAIYKGDEASAIRMSHENKEIAQLYEQFFGEPNSHKAHKL LHTHYTPRKVHKSGEIDRFSEYSEPGPSELCKF*

HYDF orthologs

>jgi|Nanocel1779_2|132847|CE132846_2156
MVLGLIKLPGPYTSKLLRPSRPIYD TSLDRASSLPFFASPCASLCSCTNSTSSSSSSSFSTTSSSYSSGLHRLLLGLVGEMNA
GKSTVLNALTQQATSIVDATPGTTS DIKTAIMEIHGLGPIKLLDTAGVDDTGRLGDKKRNKTWEAMREVD AALVVIEFPNKAK
VATQGMDEEEGRAGRRRVAENGCGGGEARIT AISNSSSRSSSSSRISTITICGRDSRKSSCRENASGAGCLYSL LCLPPSSHFP
CSKHGQATPSLHAHLQY*

>jgi|Cyapar1|10289|g9962.t1
MASPVLARALS RATAPLVFGAASVVRPGMLQVARLVGSALLSRRFLVSSATAFAAHQAPS AHAQAAPKPPGGPLASTQT PRV
NIGIFGSMNVGKSSLMNLITQSNTSIVDSTPGTTADVKAALMEMHALGPAKIFDTAGLDERGLLGEKKREKSVNAAKESDVVV
LVVNPRVGHSEYDMEILDVAKRREKKILVLLNEFGGDPKWSGPGYTERAEYLG LPALRVKLNQPGNMQAVVD FIESEYQRGAV
KSPPLIPPRYMGPDKTVMFVPMDAETPSGRLLRPQARVQEFCLSHYTTTVAFR LDLAAGRSKDAAAVAKERARFEAAVRAAD
PALVITDSQAMDLVHPWLDPAKFPNADLT TFSIVMINDLSGGRIREFADGLQQLARLRQGD KVLIAEACNHNRLSVEVCADIS

TCQLPRKIESMYGKDVVKVDHAFGREFFASELKDYSMVIHCGGCMVDKQKIVARIDDCKELGIPVTNFGVMLSYSMA SPAA AQR
VLRPLAH*

>jgi|Chabral|336334|rna-CBR_g28613, extended N-terminal region
MSDREEQQPDDRILRSARPPIAHLALGPEGDRYLFRQGRERLQQQRRARAAEASRTLVS EAA S SEAMATS AQQTSQAGSSGSV
GSTVAQTQSQSTGGMASQPLVLTDFEIAIQEATLKEAQLQHALGQIKAEKERMIRRRVRMLRREADVSELEEMDLTHATDDVR
IIRSVLLNVVEMQDQQTSLQGIPQSEANINMHNFPSPFSKKALDLEAKIGHGQPPTTDGRRKTLPPNWKAKERIMFVDNDGST
IELDEHFQEGVGSEAGSAEASKGGVVVAVAQKAKLGGGYVPSLSLLQAPVKMSPEIEGVVAKYPDLFEEPIGVVEREVVHAIE
IIPQSSIPKGRIYRMSPGELDELRRQLKELLEKGWIRPSVSPYGPSVLLVPKKEGTLRMCIDYRGLNAITVKNREPLPRIDDL
LDRVQGCYFYSKIDMKSGDQHKTAFTQTRYGLYEFVVMFPGLCNAPGTFQHAMNRI FHDYLDKFVIVYLDNLIFSKTVEEHIA
HLDKVLSSLRQHKFKINGEKCEFGTRVLYLGHEISA E GLKPDNAKAVSIRDWPRPRSVTEMRSFLGMTGYRTFVKNY SIVA
APLTDLTRLDTPWEWTVTTDASQYGIGVVLAKEGPKLRPVEYMSKRMS SQKLAKSTYEKELYAVYKALTHWRHYLLGRFFIL
RTDHQTLKWMRTQSVLSDALKRWIEVIEQYDFDPQYLYKGEYNKVVDALSRPDPFSGALITEFDLTDNVTQSLVEAYREDQFMS
EIIIRRLQAKVKKTSAEFELVNVLLFLEKVG NKRLCVPNSESLSRSLF LGECHDATGNFGYKKTAA NLLQRFWWPTMMRDAQLYV
ETCQVCQRDGPRTQAPLGLLKPLPIPERPGESLSMDFMDTLTITSKSGMRYIYVIVDRFSKFARLVAMSATANTEYVIKMFKEN
WVRDFGLSKSIQHEGPKLLVTD SQAMDIVHPWTLDRSTGKPIIDITTF SIALIQRQTDGRLGLLVKGLEAYKSLRPGDRVLVA
EACNHN RITELCNDIATVQIPQQISKENPKGGVIIDHAFGREFFDLGDVGDLRQYQLAIHCGGCMVDRQKILARITDLEEAGV
PVTNFGVLVLSYAAAPEAFARVLEPYDVEIPKSVFLQSHAQCAA*

HYDEF orthologs

>jgi|Chlre5_6|15537|Cre06.g296750.t1.2
MAHSLSAHSRQAGDRKLGAGAASSRPSRRIVRVAHAHASASKATPDVPVDDLPPAHARA A VAAANRRRAMASAEAAAETL
GDFFLGLGKGGLSPGATANLDREQVLGVLEAVWRRGDNLNLERALYSHANAVTNKYCGGGVYYRGLVEFSNICQNDCSYCGIRNN
QKEVWRYTMPVEEVVEVAKWALENGIRNIMLQG GELKTEQRLAYLEACVRAIREETTQLDLEM RARA A STTTAEAAA SAQADA
EAKRGEPELGVVVSLSVGELPMEQYERLFRAGARRYLR IETSNPDLYAALHPEPMSWHARVECLRN LKKAGYMLGTGVMVGL
PGQTLHDLAGDVMFFRDIKADMIGMGPFITQPGTPATDKWTALYPNANKNSHMKSMFDLTAMNALVRITMGNVNI SATTA LQ
AIIPTGREIALERGANVVMPILTPTQYRESYQLYEGKPCITDTAVQCRRC LDMRLHSVGKTS AAGVWGDPASFLHP I VGPVP
HDLSSPALAAAASADFHEVGAGPWNPIRLERLVEVPDRYPDPDNHGRKKAGAGKGGKAHDSHDDGDHDDHHHHHGAAPAGAAA
GKGTGAAAI GGGAGASRQRVAGAAA SARLCAGARRAGRVVASPLRPAAACRGVAVKAAAAAAGEDAGAGTSGVGSNI VTSFG
IASTTAGHGVPRINIGVGMNAGKSTLVNALAQQEACIVDSTPGTTADVKT V LLELHALGPAKLLDTAGLDEVGG LGDKRRK
ALNTLKECDVAVLVVDTD TAAAAIKSGRLAEALEWESKVMQAHKYNVSPVLLLNVKSRGLPEAQAASMLEAVAGMLDPSKQI
PRMSDLASTPLHERSTITSAFVKEGAVRSSRYGAPLPGCLPRWSLGRNARLLMVI PMAETPGGRLLRPQAQVMEEAIRHWA
TVLSVRDLDAARGKLGPEACEMERQRFDGVIAMMERNDGPTLVVTSQAIDV VHPWTLDRSSGRPLVPIITTF S IAMA YQQNG
GRLDPFVEGLEALETLQDGDRLVISEACNHN RITSACNDIGMVQIPNKLEAALGGKKLQIEHAFGREFFELES GGM DGLK LAI
HCGGCMIDAQKMQQRMKDLHEAGVPVTNYGVFFS WAAWPDALRRALEP WGV EPPVGT PATPAAAPATAASGV*

>jgi|Vitbras1|16622|Vbra_296.t1
MLARPLFWALPSSVEPMRALSILPSRFLSTLASQASSSVPRASTSPFLAIPKAAQINTNNDGAPLYAPMMDAHAAAGVLGDVL
KRGIGLGEVVCLGQNEIHGLMEMCWKEGSRDERVREALFAHANEVTRRFFGNKV FYRGLIEFSNVCQNDCLYCGIRKNQPNVE
RYTISKQQLLSIALWAHDNGYGSVM LQSGELNHPARMKYLT D VITATRR ETVL TNLEKEKRHRHSNQ AEDPAIDELIEEVR
ESFIGSDPRFKDKGIGIALSVGELSEADYRRLHEAGGHRYLLRIESSNPELYSKIHPPTQFFAKRLACLRL LKKVGFQVGSV
MIGIPGQTLLEDLARDIVFRDEDIDMVGMPYITQEDTPVADMWQAQYGHVDKKEHMAEMFDLT TVLISICRITMG NINIAAT
TALQAINANGREVALRRGANVLMPILTPTQFREN YQLYEGKPCITDTAEQCCKELNFR LKSIGKEAAKAVWGDP PHFYQHVR
HHTLARRSFSSKPSPAAPKSPGASELPRINIGFFGAMNAGKSTLMNKLTRSNLSIVDATPGTTADTKITLMELHDIGPTKLFD
TAGIDERGELGEKKRRKTI SVLKESDIALLVDP LQKLSPT EPTLASPDALQWERYLIEMGEKRHKESIVFNIKEQEGATQG
WTRRQIEDGVVRLLEMLDPSSMHRVLC DLGAPEASDRLVKFIQSSPAHGS MYHDDLPCVPESYLTADSVFFLNI PMAETPA
GRLLRPQAMVQESIIRHYGTTVAYRMDLAKARSSDPAEVQEEKSRFLKMLNMV LQQTAGPKLLITDSQAMDIVHPWTLDPVTN
RPLIDITTF SIALHRQSGGQLPTFVEGLERFKHIKQGDRLVLAEGCNHN RITQLCNDIATVQLPGHIRAQGGASIAVDHAFG
REFPDIEDQSDKGLGAYSLVIHCGGCYIDKQKVQARVDDLREMGVPVTNFGLLLSYVAAPEAVDRALQPWGLTSV

>jgi|ChlNC64A_1|142788|IGS.gm_5_00117
MLGSSRALQAAAAA VSMQKALPWENAVTAEEAVKVLGSGFWGTTFAPGDVLDLSVDQMEGLLSATSTGAGSKVLDNALFAHAE
SVTGRFFGRDVYYRGIVEFSNVCENDCGYCGIRKHQRRARRYTMPREVEVEADWAFRHR LGLTMLQSGELNTPQRMKYLT D V
VAAVKERTRALDAEQRLVDPEALPKDAAGVGLCVALSVGELPREQYQQLKDAGADRYLLRIESSNPQLYASIHPPA QARASGP
ARFHKWENRTRCLRD LKDIGFMIGTGVMVGLPGQTLRDLAGDIMFFRDLGADMIGMPYITEAGTPVADMWEQQFGHVDKKK
MQDMLTLTTRMNALARITLGNVNIAATTALQAIDAVGREVALRRGANVLMPILTPTKYREHYTLYEGKPCITDTAEECQKCLN
ARLAMVDKKLKPVGWGDPPSFRDSVHPVPMQAQAAAAA SATGSGRGLRTWAPVSAAAARDGGLHTCSATGQAAAAAPAPGRGA
VLGMGGGPAKGS DVPRTNIGIFGCMNAGKSSLMNRVTRSETSI V DSTPGTTADTKVLMELHDVGP AKLFDTAGIDE EGALGE
KKRRKVL SVIKETDVAVVVVDVGRFLDTPADRLPAALAWERLLLDKAAAAGSSPVLVYNTKLGGAGAGAGPEAAAAAVDR LQA
ALNPTGAIMSRQLQLSREEASDALAEFLQEAAAAAKQHDAPRSLPDEFLSEDAMVFLNIPMAETPSMRLLRPQALVQEEAIR
HFASTIAYRMNLAHARSSDPELVQREARFLRALQPVLAHQGPKIIITDSQARRAGSGRPGCTAVDILHPWTLDPATGDELVP
FTTFSIAMIHRQSRGQLPLFVRGLERFKSLRAGDKVLVAEACNHN RITD I CNDIGMVQIPEKIEAQGGRGVVVDHAFGREFFE
LDAEEHC VNAV LQEAGVPVTNYGLLSYAHSPAALERAMK PWGLRM*

>jgi|Tetstr1|460551|TSEL_005809.t1
MSEDEILKVAHSAFAHKMGTIMLQSGELRTPQRLRFVNRVRRIREETIASELSATGRLPGVPPSQMCKVVPVETADVGLCVA
LSIGELSLPRLQELRESGAARYLLRIESSNPDLFTRIHPPAQTFEQRLRCLQDIKAAGLQLGTGTPPPPPPARPRMALFPFGLV

SASVMIGLPGQTMSDLASDLCCFFRDIGADMIGMGPIYIPEPNTWVADDFHRRHPDLPALQPYLEEMMELTTTMNALARITMGNI
NIAATTALQAINPLGRELALARGANVVMPIPTTEERANYQLYPGKPCVSHSTADGCASCLKMRIASVGKNLQTDGAWADPPHY
RNPVAHSFAAAPLAGRTPATGGEQRRSAHRAAASALPEAGPPPAGPSKGSVDVPRNTNIGVFGCMNAGKSTLVNALTQETSIV
DSTPGTTADTKIALMELHEIGPVKIFDTAGIDEAGELGSKKARKTLAALKETDVSVVVVDAAAVAAAGEGSLAWESDLVAMAR
KYGSMPLLVVNARGGRGSEALAAAAKAAVSPTEDI IALQADLATEAGRSAITGFLQEEILKAKRGAAGVPSLPPQFLSDRAMV
FLNIPMDAETPSMRLLRPQARPRPTGAGPPTPYPSMPClGRGSCGAGRVASRPAALVQEEAIRHFATTVAYRMNLMARSADP
AEVAAERRRRFMRALRPVLEHDGPKIIVTDSQAVDILHPWTLSPDGAPLVPFTTFSIAMVNRMSGGQLRTFVDGVALETKEG
DAVLVAEACNHNRI TDQCNDIGMVQIPQALARLTGGRGFPVVEHAFGREYPEVQEGAGAGLGRFKLAIHCGACMIDRQKMRARL
MDMKEAGVPVINYGLFLSYMQSKQALARAIEPWNL*

>jgi|Tetstr1|436239|TSEL_025083.t1
MEALITRVGSSLLTSLLRPALVAASARLAGPALRSAASLRSHSTDGEMPTFDAAHTRRVNQMLAAKDTLIPEEAAAVLTSHWG
APVLPGQVLSDLTASQLAATLETAHPVGTGNINVDDGHLLRSALYMHAEVSVSTRFFGDKVFQRGVMIGLPGQTMSDLASDLCCF
FREIGADMIGMGPIYIPEPNTWVADDFHRRHPDLPALQPYLEEMMELTTTMNALARITMGNI NIAATTALQAINPLGRELALAR
GANVVMPIPTTEERANYQLYPGKPCVSHSTADGCASCLKMRIASVGKTLQTDGAWADPPHYRNPVAHSFAAAPLAGRTPATGG
EQRRSAHRAAASALPEAGPPPAGPSKGSVDVPRNTNIGVFGCMNAGKSTLVNALTQETSIVDSTPGTTADTKIALMELHEIGP
VKIFDTAGINEAGELGSKKARKTLAALKETDVSVVVVDAAAVAAAGEGSLAWESDLVAMARKYGSMPLLVVNARGGRGSEALA
AAAAKAAVSPTEDI IALQADLATEAGCSAITGFLQEEILKAKRGAAGVPSLPPQFLSDRAMVFLNIPMDAETPSMRLLRPQARP
RPTGAGPPTPYPSMPClGRGSCGAGRVASRPAALVQEEAIRHFATTVAYRMNLMARSADPAEVAAERRRRFMRALRPVLEHDG
PKIIVTDSQAVDILHPWTLSPDGAPLVPFTTFSIAMVNRMSGGQLQTFVDGVALETKEGDAVLVAEACNHNRI TDQCNDIG
MVQIPQALARLTGGRGFPVVEHAFGREYPEVQEGAGAGLGRFKLAIHCGACMIDRQKMRARLMDMKEAGVPVINYGLFLSYMQS
KQALARAIEPWNL*

HYDE orthologs

>jgi|Cyapar1|17193|g16914.t1
MSFDPKLVGAKPRNFKPIEERILPFARFLGKPDLPKPGESLDLSRDELTAALLTTTDAALEASLYAHANSVTERVFGNKVYFRGI
VEFSNVCQKNCEYCGIRKSMPAAQLRRYTMTAEIIVDCAEFCYRQGYGTLMQLSGELDTERRLNFDIVIKQIRERCRELERE
HLGLPADTPLDKLGLAVALSIGELSADKYQRLFDAGAVRYLLRIETSNPALYAKLHPADHKWETRHRCLLDLKRIGFQIASG
VMVGIPGQTYQDLANDLLFLKELDMDMIGMGPIYIYQENTPVGEQWKKEFGSRPKDDHNAWLLSTASRMYSLARILIPDCNITA
TTALQAIHPSGRELGNRGCNVLMPIITPLKYRENYQLYQGKPCVDEGATECRKCLVSRVEWSGKDLALGIWGDPPHFFRRAG
LELPSETPILPKPALTDKKAHH*

>jgi|Tetstr1|460551|TSEL_005809.t1
MSEDEILKVAHSAFAHKMGTIMLQSGELRTPQRLRFVNRVVRRIREETIASELSATGRLPGVPPSQMCKVVPVETADVGLCVA
LSIGELSPLRLQELRESGAARYLLRIESSNPDLFTRIHPAQTTFEQRRLCLQDIKAAGLQLGTGTPPPPPARPRMALFPGLV
SASVMIGLPGQTMSDLASDLCCFFRDIGADMIGMGPIYIPEPNTWVADDFHRRHPDLPALQPYLEEMMELTTTMNALARITMGNI
NIAATTALQAINPLGRELALARGANVVMPIPTTEERANYQLYPGKPCVSHSTADGCASCLKMRIASVGKNLQTDGAWADPPHY
RNPVAHSFAAAPLAGRTPATGGEQRRSAHRAAASALPEAGPPPAGPSKGSVDVPRNTNIGVFGCMNAGKSTLVNALTQETSIV
DSTPGTTADTKIALMELHEIGPVKIFDTAGIDEAGELGSKKARKTLAALKETDVSVVVVDAAAVAAAGEGSLAWESDLVAMAR
KYGSMPLLVVNARGGRGSEALAAAAKAAVSPTEDI IALQADLATEAGRSAITGFLQEEILKAKRGAAGVPSLPPQFLSDRAMV
FLNIPMDAETPSMRLLRPQARPRPTGAGPPTPYPSMPClGRGSCGAGRVASRPAALVQEEAIRHFATTVAYRMNLMARSADP
AEVAAERRRRFMRALRPVLEHDGPKIIVTDSQAVDILHPWTLSPDGAPLVPFTTFSIAMVNRMSGGQLRTFVDGVALETKEG
DAVLVAEACNHNRI TDQCNDIGMVQIPQALARLTGGRGFPVVEHAFGREYPEVQEGAGAGLGRFKLAIHCGACMIDRQKMRARL
MDMKEAGVPVINYGLFLSYMQSKQALARAIEPWNL*

>jgi|Tetstr1|436239|TSEL_025083.t1
MEALITRVGSSLLTSLLRPALVAASARLAGPALRSAASLRSHSTDGEMPTFDAAHTRRVNQMLAAKDTLIPEEAAAVLTSHWG
APVLPGQVLSDLTASQLAATLETAHPVGTGNINVDDGHLLRSALYMHAEVSVSTRFFGDKVFQRGVMIGLPGQTMSDLASDLCCF
FREIGADMIGMGPIYIPEPNTWVADDFHRRHPDLPALQPYLEEMMELTTTMNALARITMGNI NIAATTALQAINPLGRELALAR
GANVVMPIPTTEERANYQLYPGKPCVSHSTADGCASCLKMRIASVGKTLQTDGAWADPPHYRNPVAHSFAAAPLAGRTPATGG
EQRRSAHRAAASALPEAGPPPAGPSKGSVDVPRNTNIGVFGCMNAGKSTLVNALTQETSIVDSTPGTTADTKIALMELHEIGP
VKIFDTAGINEAGELGSKKARKTLAALKETDVSVVVVDAAAVAAAGEGSLAWESDLVAMARKYGSMPLLVVNARGGRGSEALA
AAAAKAAVSPTEDI IALQADLATEAGCSAITGFLQEEILKAKRGAAGVPSLPPQFLSDRAMVFLNIPMDAETPSMRLLRPQARP
RPTGAGPPTPYPSMPClGRGSCGAGRVASRPAALVQEEAIRHFATTVAYRMNLMARSADPAEVAAERRRRFMRALRPVLEHDG
PKIIVTDSQAVDILHPWTLSPDGAPLVPFTTFSIAMVNRMSGGQLQTFVDGVALETKEGDAVLVAEACNHNRI TDQCNDIG
MVQIPQALARLTGGRGFPVVEHAFGREYPEVQEGAGAGLGRFKLAIHCGACMIDRQKMRARLMDMKEAGVPVINYGLFLSYMQS
KQALARAIEPWNL*

>jgi|Chabral|336355|rna-CBR_g28634, **bad annotation?**
MCINYRGLNAITVNAETLPRIDNLLDRVQSCRHFSMIDLKSGYHQIEVLEKLKEANFKINAKKCEWAKTEVLYLGHVLDDEG
IKPEDSKIAAIRAWPTPRTLTELRSFLGLANYRKFAFNFSIIVIPLRLLKKEAIWQWDKCTFALKKLKRALIEYLVLVKA
DPSLSFVVTDDASRYGIGAVLQDDDNQYKPVFEMSTRMPSEKDCYCGIRKHKEGVHRYTIPKEDVSVAMAWFDHKYGTLM
LQAGQLNSPQRMERYLVDLIREIRRTIAKDLEQRRARGVNTDRLKATELGLCVALSVGELSEETYKRLYDAGASRYLLRIETS
NPELFSDLHPGDQMFEEKRLKCLRTAKRAGFQIGTGVMIGLPGQTLRLDLANDIVFFRQEEADMVSAGKWVRCKGWLEPRKHNGG
HSVPHFRPLDTGLQTI FWDVLERGARVQHGNRSRRHSFPAMPHVLDNMAETRKGITAPYTAKEQEKMAALVKENREKERE
KQAKLKAI AEEQATMKKEIGEEMEKKKAAEKEVAAEEEEKERMRTESREGSSGTEKDKDAEMEKKISEWVANLSLGEDEEAQM
YVPEEEKEAFARALAVIGDPLERQEAEEEEKRLWKLRMKREKKRREEANWLTAEEVEKVLEKFREANFKINSKKCEWAKTQVLY

LGHVLDGDGIKPEDSKIAAIRDWPTPRTLTELPSFVGLANYRKFVRNLPTIAAPLRRLKKEAIWQWDKDCSTSAVKKLKRAL
IEYPVLKVADPSLPFVVTTDASQYGIGAVLQQDDDDNGYRPFVEFMSARMPCEKTHSGKDTSPYSEEQQEKMVALVRENKERKEL
EKQAKLKAIAEEHAAMKRLEEMKRIQQEEEEEVAEAEFVIVYLLDILILSKTVEEHVHLDKLLSLRQHKCKINGEKREFG
RTRVVYLGYEIFAEGLKPDDAKVANIRDWPRPQSVIEMRSFLGMTSYCRNFMKNYSIVATPLIDLTRLDTPWETETCEAAFR
HLKHALRHYEVLKLPDPDKPFIVTTDASQYGIGAVLAQQEGKKLRPIEYMSKKMPSQKLAKSTYEKELFAVYKALTHWRHYLL
GRFFILRTHQTLRWMRTRAVLSDTLKHWIEVVEQYDFEPEYIKGEYNKVVDAISRPRPDFSGALITEFNLADNVTQPLVEAYR
EDPFMAEIIIRLEAKDKGTSPEFELVCQSDKPRTQAPL
GLLKPLPITERPGESLSMDFMDTLVTNKSGMRYIYVIVDRFSKFARLVAMSATTNTEYVIKMFENVRVDFGLPKFIVSDRDV
QFTSELWKAAAAEQGMQLQMTSGNHPEANGQAEQMNRAVQHLLRHYIKPNQVDWDKKLTLIASLYNNVVHSATGVSPNSLLPT
FTPVLPLDFFLLPDNQPTAAPGT*

>jgi|Nanocel1779_2|632220|estExt_fgenesh1_pg.C_190091
MLLLKKNVAVTGSTVAFFRKRPRLPVAPLERLASSSPSPPTVTFHRTRPPAATAQCQHDAQASFAPLAHFLGVEKLEPGRII
DLSKEQLMELLTTQDGALVQALYSLAEAAATTAYFGSKVYARGIVEFSNVCTKDCLYCGIRKHKMVHRYTMTKQEIVTCAEFAY
TQGMGTLMQSGELPTPERINFMLDVVRAVRLRTVEMDLEVHPSCSSSLPSASLSVTPNNLAASSQRLAVALSLGELPSETY
KAFYEAGASRYLLRIESSNPVLYHKLHPPSHSFESRRQCLNELKRIGFQVGTGVMVGVPYQDVEDMAGDLLFFKEFDPMIGL
GPYIVQAKTPLGQVYLAERQHSAEELEKERIDLLERTTMYALARLILGNVNIAATTALQALDPQGRELALRRGANVVMPII
TPQRLREEYQIYENKPGIDQDALESTACLDLRVQLAGKELLLDGRWHDPPHYLTRKVR*

HYDG orthologs

>jgi|Chlre5_6|15536|Cre06.g296700.t1.2
MSVPLQCNAGRLLAGQRCGVRRARLNRRVCVPVTAHGKASATREYAGDFLPGTTISHAWSVERETHHRYRNPAAEWINEAAIHK
ALETSKADAGVREILAKAKEKAFVTEHAPVNAESKSEFVQGLTLEECATLINVDSNNVELMNEIFDTALAIKERIYGNR
VVLFAPLYIANHMCNTCTYCAFRSANKGMERSILTDDDLREEVAALQRQGHRRILALTGEHPKYTFDNFLHAVNVIASVKTEP
EGSIRRINVEIPPLSVSDMRRLKNTDSVGTFFVLQFETYHRDTFKVMHPSGPKSDFDFRVLTDQDRAMRAGLDDVGIGALFGLYD
YRYEVCAMLHSEHLEREYNAGPHTISVPRMRPADGSELSIAPPYPVNDADFMKLVAVLRIAVPYTGMILSTRESPEMRSALL
KCGMSQMSAGSRTDVGAYHKDHTLSTEANLSKLAGQFTLQDERPTNEIVKWLMEEGYVPSWCTACYRQGRGTGEDFMNICKAGD
IHFDFCHPNSLLTLQEYLMYADPDLRKKGEQVIAREMGPDAEPLSAQSRRKLERKMKQVLEGEHDVYL*

>jgi|ChlNC64A_1|30311|estExt_Genewise1.C_50128
MLSQAASSLGLASGGLGPLLPAVAAGARLLGRSGLLQQQRGLQWSVEKEPQSYKAVDDIIQDHVIVAAELEKTQQAADKPAR
VRDILTAAKERSFLTNHKGPGPSEYVQGLTYEECATLLNVDPDGTDIMEEIIYNTAFDIKNRIYGNRIVLFAPLYIANYCVNNCR
YCAFRQGNKSLERSALTDKQLREEVAALQQQGHRRLLVLTGEHPKYTFDSFLKAIDTISVRTEPCGNIRINVEIPSLSLSD
MRRLLKATDKIGTYTLFQETYHRPTFKMHMHIAGPKSDYDNRMVTHDRAMRAGLDDVGLGTLFGLYDYKYEVLATLMHANHLERE
YGAGPHTISVPRMRPADGSDVSLAPPYAVDDANFKKLVAAILRIAVPYTGMILSTRESPEMRTELLRVGMSQMSAGSKTDVGAY
HRDDSKATEDNLADLNGQFTLADHRPVQDIVDLMKEGYVPSWCTACYRKGRGTGEHFMKIAKAGNIHNFCHPNSLFTLQEYLN
DYGSEEAQIGQDLIERERVVGLSDSAQNLTKRKLAKVNAGEHDVYI*

>jgi|Vitbras1|16623|Vbra_297.t1
MLSVVRHISAPSAAALPALSCSSRLASPCAARTQRFVRHFGFLWSKELEKGVHYKDANEIINHDLIFKHLEATKGAADKNVIK
DILERAKERSFLKDGPGATGGSEYVRGLELDECATLLNVNPDGTDIMEEIIYNTAFDIKNRIYGNRIVLFAPLYLANYCVNGCAY
CAFRSSNKAIQRTKLSDDQVTEQVSILERQGHRRILALTGEHPAYTFEDFLHALKLITEVRTEPCGSIRRINVEIPSLSVSDM
KRLKAADCVGTYTLFQETYHKPTFHTMHPWGPCKSHYENRVLTHDRAMRGGGLDDVGIGVLFGLYDYRFECEMAMLMHNSHLEEEY
RAGPHTISVPRMRPAMGSGVASDPPYPVDDAHFKKLVAAILRIAVPYTGMILSTREAPEMRAHLLRVGMSQMSAGSSTEVEGGYH
RHATQRTDESLGDLTMQALLPEKYKAAAGDTAMDHEHTDPNLGGQFQLEDTRSSEEIVHDLLESGYVPSWCTACYRKGRGTGE
HFMKIAKAGNIHNFCHPNSLLTLQEWLNDYASPEIKELGNKVIAREKEIDLSDSAKRLDRKMEAVNKQGHHDVYI

>jgi|Nanocel1779_2|598502|fgenesh1_kg.19_#_588_#_TRINITY_DN7538_c0_g1_i7
MLRLAQWCCRPRVSSAAPTSTSPARLNLAIWPHQQRGLAQWSVETERHHKYEDPAKIDEVAINRLLLEETKEKAKDPAHVKA
ILVAAQDRARLQSPPPGHPNADPPLNEFVQGLTLEEAAATLLNVDSADKEMMQLLYNAAALNVKQQIYGNRIVLFAPVYLANWCV
NTCQYCAFRGANKQMQRSMLTMDLRLQEVEALEAIGHKRLLLLTGEHPKYSFDRFLEAVQVVGTVKTPNPCGEIRRINVEIPS
LSVSDMKRLKATQYIGTYTLFQETYHRETFKKMHPPGGPKSDYNHRLTMDRAQLGGLDDVGIGALFGLADYRFEVMGLLMHAH
HLDRTYGTGPHTISIPRMQPALNAPAAEHIPMPVSDDDFKKLVAVIRCAVPYTGMILSTRESAATRSELLRLGISQISAGSRT
DVGAYHRDTTCDANFISNTASSSSTAGRPAKTQLPLAAIAAATAVHEVEEEEDTDEYKRRRGQFSLQDHRPLDTVIKDLLHD
GFIPSFCTACYRKGRGTGAAFMKIAKSGRIQDFCQPNALLSLQEYMEDYASPQTAVGDEVIARESQQMMGDNAKRAYERKRQQL
LRGKRDLFY*

>jgi|Cyapar1|17192|g16913.t1
MLSAAATARAAALAGAAAPRIASSIAPALARTLPSLAKRLMSAKTAGLKGRGSLFSPEREPQHVFKPANEIRAAEIDAALERA
KALAKDESVIDILGRARDRALLKDHGMQPLGGGEYVQGLTVEEAAIILLQVDSNNQELMQMLYDTAFAIKNLIYGNRIVLFA
LYIANYCVNSCTYCSFRAPNTSMPRAVLTDEELRQEVEALQRMGHRRLLLLTGESPRYTFDQFLNALKIASEVRTDPCGSIRR
INVEIPSLSVSDFRRLKETNCVGTYTLFQESYHPETYARAHAPGPKSDYEWVRVQTMdraQTAHVDDVGVLGLHDYKWETL
AMLQHAQHLDQTYGAGPHTVSIIPRMQPAEGAPDAMNIPHPVNDEDFKKLVAVIRCAVPYTGMILSTRENPEMRRQLLRGLVGSQ
MSAGSKTDVGSYHQNQVQKPGAKPVAEHEHFTAGPEPSAQPATCTPKEAKDAAGQFTLQDHRSLDDVVGDLLELGFVPSWCTAC
YRKGRGTGEAFMKIAKKGDIQALCHPNALLTLEEYLIDYASERTREIGKKVLEVETGHIIPSERAKHAFERKLLKIDEGQRDLFY
*

>jgi|Tetstr1|436240|TSEL_025084.t1
MALCAGALSTVLRSAAPLAVGALRVLGCGARSMALWSKELEKGQTYKSADDIIQHDVIERELEATKSSAKDPRAVGAIL
ARERSFLTGYNPAGGSEYVQGLTYKECATLLNVDVANSPLMEQIYETAYNIKERIYGNRIVLFAPLYIANHCVNNCAYCAFR
ADNKNIARSILTDQQLREEVVALEKQGHRRLVLVTGEHPKYTFDQFLNALKIIAETRSEPCGQIRRCNVEIPSLSLSDMKRLK
ATNVVGYTYTLFQETIYHRPTFAEMHISGPKSKYNNRLLTHDRAMRAGLDDVIGIGTLFGLYDYKYEVMAMLMHANHLEAEYGAGP
HTISVPRMRAAEGSDISVAPPYEVDNPNFKKLVAILRIAVPYTGMILSTRESPEMRRELLKVGMSQMSAGSRTDVGAYHRQDS
TETRENLGNMAGQFSLDERPVNQVVDLMEQGYVPSWCTACYRKGRGTGEHFMKIAKAGNIHNFCHPNSSLTLQEYLN DYGTP
AEKAAGQKLIKDESERGFMSDQAVKLMNRKMKKVNAGEHDVYI*

>jgi|Chabra1|346080|rna-CBR_g46723
MGWGSALVRTAGRAVGRRLPRGSLSLGLPAKCMSAATMMQTRGLMWSVEKEQGHTYKPVVEEIIDHNRI TLAMESTMEASKDSSR
IKEILEAARERSFLKDHTPTKSEYVQGLTVEEMATLLNVDPANEEIMNEIYDTAFAIKNRIYGNRIVLFAPLYIANYCVNQCT
YCSFRAPNKTARTVLTDEQIRQEVEALEKQGHRRLVLVTGEHPKYTFDSFLHALHVSEVRTEPCGNIRRNINVEIPSLSVSD
MKRLKATDVVGTYTLFQETIYHQPTFKVMHPAGPKSDYPHRLLTMDRAFRGGVDDVGLGVLFGLYDYRYECLAMVMHANHLEAE
YGAGPHTISVPRMRAEGAETS VNPPYPVDDAHFRKLVAILRIAVPYTGMILSTRESPEMRAELLKVGMSQMSAGSRTDVGAY
HRQADDTEQN LGNLAGQFTLS DHRPVAEIVEHLIRNNYVPSWCTACYRKGRGTGEAFMKIAKAGNIHNFCHPNSSLTLQEY LQ
DYASEDLKKIGEELIAKESSTGLSDSAQRVLD RKM SKITDIQTLTIGYKGHPASLGKGARLVWIAQDQRI AAMFKEATEGSAS
GDNIC*

HCP orthologs

>jgi|Chlre5_6|18301|Cre09.g391650.t1.2
MLRAPVMSSAASRKVAAPAVAARAGCRRVGVMRVFQKAASCDNLHDKNALHERIQKSKELLDAESAMMCYQCEQTKSG
TGCTDIGVCGKTPEVSALQDLLIYSVKGLGSLAHVARTSPAKIEDAAVNTFINGAIFSTLTNVNFADDRFLEFVTDCKRL
HAQLAAKMAAAGVAVPAAETAHQPFWFGSMPLAWNSDTHVALAGVGDMLEVASKTG VKERQHV LGETLAGLQELLMYGL
KGLCAYAHHAELGHTDPAVYADVQAYLHFLCSPAAADVGVLDACFRAGATNFRVMEMLSNAHTDTFGHPVPTPVTLNPN
VPGKAILVTGHDMDHLHMLLEQTAGKGINVYTHGEMPLPAHGYPGLKKYPHLVGHFGGAWYRQKIDFAAFP GAVAVTTNCV
LDPLTAYKDNI FTINETGLSGVPHIRPDATGHKDFTP IINRALQLPGFTPE SVAKMEKKRDVTVGFGHKS VLSVAPQVIQ
AIQEKRLEHIFLVGGCDGSE PQRKYYSKLYQFMPTNTMVLTLGCGKFRI FDQDFGTLPGTDLPRLLDMGQCNDAYSALVV
ATELAKVFKTDVNSLPLSLDL SWFEQKAVAVLLTLLHLGVRNIRLGPRLPAFLTPEAVGVLVDRFGLIPANVADPAADMQ
MMMECK*

>jgi|Chlre5_6|18305|Cre09.g391450.t1.2
MLRAPVMSSAASRKAAPAVAARAGCRRVGAMRVFAFQKSTG SVHDKNALHERIQKSKELLDAESAMMCYQCEQTKSGTG
CTDIGVCGKTPEVSALQDLLIYSVKGLGSLAHVARTSPAKIEDAAVNTFINGAIFSTLTNVNFADDRFLEFVTDCKRLHA
QLAAKMAAAGVAVPAAETAHQPFWFGSMPLAWNSDTHVALAGVGDMLEVASKTG VKERQHV LGETLAGLQELLMYGLKG
LCAYAHHAELGHTDPAVYADVQAYLHFLCSPAAADVGVLDACFRAGATNFRVMEMLSNAHTDTFGHPVPTPVTLNPNP
GKAILVTGHDMDHLHMLLEQTAGKGINVYTHGEMPLPAHGYPGLKKYPHLVGHFGGAWYRQKIDFAAFP GAVAVTTNCVLD
PLTAYKDNI FTINETGLSGVPHIRPDANGHKDFTP IINRAMQLPGFTAESVAKMEKKRDVTVGFGHKS VLSVAPQVIQAI
QEKRLEHIFLVGGCDGSE PQRKYYSKLYQFMPTNTMVLTLGCGKFRI FDQDFGTLPGTDLPRLLDMGQCNDAYSALVVAT
ELAKVFKTDVNSLPLSLDL SWFEQKAVAVLLTLLHLGVRNIRLGPRLPAFLTPEAVGVLVDRFNLI PANVADPGADMKMM
MECK*

>jgi|Chlre5_6|18620|Cre09.g393543.t1.1
MLSRCLGMAGTTLGGSLASGAQSAVSGMFRASGRRATSLQVLAWQLPNLFAGDQQARNAASIKAKMAEANKALES DKMLC
YQCEQTKSGTGCTEIGVCGKTPEVAGLQDLLVYSVKGLASLAHIARNSPAKIEDPAVNTFINGAIFSTLTNVNFADDRFL
EFVSEARAHHARLSAKMAAAGVQVPASATEQQVWFGSMPLP LLWNSQAAALGGVGDMLEVA AKTGIAERQKVLGETLAGL
QELLVYGLKGVCAYAHHAELGFTDPTVYAEIQGALHFLNTPGAKDVGVLDACFKCGATNFKVMEMLSNAHTDTFGHPV
PTPVTLNPNPVGKAILVTGHDMDHLHMLLEQTAGKGINVYTHGEMPLPAHGYPGLKKYPHLVGHFGGAWYRQKIDFAEFPGA
VAVTNCVLDPLQVYKQNI FTINETGLSGVPHIRPDANGHKDFTP IINRALQLPGFTPELIEKRPKKKDVTVGFGHKAVL
SVAPQVIQAIQEKRLEHIFLVGGCDGSE PQRKYYSKLYQYMPNTMVLTLGCGKFRI FDQDFGTLPGTDLPRLLDMGQCND
DSYSALVVATELAKVFKTDVNSLPLSLDL SWFEQKAVAVLLTLLHLGVRNIRLGPRLPAFLTPEAVGVLVDRFNLI PANV
ADPGADMKMMMKNK*

>jgi|Chlre5_6|18619|Cre09.g393506.t1.1
MLARCSMASTTVRGHLVRGSSGSLASVARGTGAVRVMAWEWNP FAGDQQARDASIKAKMAEANKALEHDKMLCYQCEQ
TKSGTGCTEIGVCGKTPEVAGLQDLLVYSVKGLASLAHIARNSPAKIEDPAVNTFINGAIFSTLTNVNFADDRFLEFVSE
ARAHHARLSAKMAAAGVQVPASATEQQVWFGSMPLP LLWNSQAAALGGVGDMLEVA AKTGIAERQKVLGETLAGLQELLV
YGLKGVCAYAHHAELGFTDPTVYAEIQGALHFLNTPGAKDVGVLDACFKCGATNFKVMEMLSNAHTDTFGHPVPTPV
LNPNPVGKAILVTGHDMDHLHMLLEQTAGKGINVYTHGEMPLPAHGYPGLKKYPHLVGHFGGAWYRQKIDFAEFPGA
VAVTNCVLDPLQVYKQNI FTINETGLSGVPHIRPDANGHKDFTP IINRALQLPGFTPELIEKRPKKKDVTVGFGHKAVL
SVAPQVIQAIQEKRLEHIFLVGGCDGSE PQRKYYSKLYQYMPNTMVLTLGCGKFRI FDQDFGTLPGTDLPRLLDMGQCND
SYSALVVATELAKVFKTDVNSLPLSLDL SWFEQKAVAVLLTLLHLGVRNIRLGPRLPAFLTPEAVGVLVDRFNLI PANV
ADPGADMKMMMKNK*

>jgi|Cyapar1|5425|g5263.t1
MPMGLDLSKFALWDEVFHPQSGPRRIVILYGTQTGTSEGFAKAVFASASARNYDVRISCM DRYDCSKLAEEDVIVCLTS
TFYNGEFPDNARSFWAFVNSADRPADSLSKTSFAVFG LGNKNVENFNSAGKMLDERLAALGSKRLVP IGLGDEYDQNGH

ESAYRPWVKSLWAAALGSSEAKRHLPITMNCELSSDKSAPAAAAPAGFASLKVVSNDLLSATGYDREERMLVLDLAGSPF
ASYELGSQILVQPRNSAQKVARTAAALKLDLDAVVSVAVAGQHSQSEEMPIAASLTVRDLLSSYLDDVSSAVSRTLLEGL
SILAADKKEAEEMEKIATDMLPGNEYGRMLTTESFSLADVLERWRSVSITLSELVSNVPRVGRHYSVASSPLVKPKSIE
ILYVLDTWATKKGPAPFGLTWRHMASLKAGDTVLAKIARGGIALPSDHGLPLVGVVALGSGIGVFRGILEHRASLQAKGKK
VGRVTLFYGCRHRTKDYLFKEKDFQEWAKKGLADVVPSPFSHDQAEFIHPGMKIAENPHKVADTLDANGVYVYCGIGGSVPG
IVLDSLIAAVRSKKQLSDSQGAELLAKMRKEKRIVEEAYSRSIDAENALGSVQLRKGGHADAADSKPIAESCGDAKMFC
FQCEQTAFGKGCTTVGICGKTPEVAALQDLLVHRVQVMSWYAHQLRLKLAEEHKNLTAEETEIPAGNRYTLIGMFSTLTNV
NFSAQRMLEYIAECRDFTTKYVAMYKAACKKAGVKERQCPAPVQTGAEADGDIEDAMVEKGRSVGVLSRFRATKNDALVG
CQEMLVYGIKGVCAADHAVMNMKHESENKIYEFVHHGLSFLLKEKESKDLNEVLGMLLKCGEANLVTMALLAQANGTYGEQT
PAVVPVKPVPKGKILISGHELTDEAILKQTEGTGINVYTHGEVLPAGGYPKLKAYKHLVGHFGQAWQRQSVVEFGFFPGA
IVMTTNCITAPREEYKDRIFTAGAVGWPGIPHIANQDFSAAIKRALELPGFSEKDTDFKYPSIASMPRVSQFTTGFGSKL
VISLADKVIIGGIKAGAITRFYVIGGCDGFEGERSYFTDFTKQLPETS SVVLTGCGKYRINHLDFTIGDTGIPRLDLGQ
CND SYAAIEIALALAGALNCTVHDLPLSIVLSWFEQKAI AVL LTLHLGLKNIRVGP NLP AFLRPSVGVLS DMFGLKLV
SDPTADLTDILGGAPKAGAPCVVNSPAA*

>jgi|Cyapar1|8063|g7806.t1
KNDALVGLQEMLVYGIKGCAYADHALMNKFESAHVYAFVHRGLAFLEKEARDMGKVLDMLLDCGEANLVTMDILHKAN
TTYGPQTPYVVPVKPVPKGKILVSGHEMHSIEAILKQTEGKGVNVYTHGEMLPAHGYPKLREFKHLVGHFGQAWQRQSV
FGWFGAIVMTTNCITAPREGYLDRIFTSGAVGWPGVPHIDGNDYTAVIKRALELPGFLEADSEFKYPSIAAMPRVPEFT
VGYGSETVIANADKVIDGIKNGAITRFYVIGGCDGFAGERSYFTDFTRLPETS SVVLTAGCGKYRVNHLEYKTIGATGIP
RLLD MGQCND SYSAIQIAVALAGALNCSIHDLPLSIVLSWFEQKAVAVLLTLHLGLKNMRVGPTLP AFLRPSVIKVLNE
KFGLKLVGDPVSDLTDILGFKPDPEAPT VATMRANP*

>jgi|ChlNC64A_1|29581|estExt_Genewise1.C_20445
MASLRCATRLASLTVRSQSVFARLPGFAGGQAHTAVRASPLLVRAAAAAELAKQPVLDNKMFCYQCEQTSKGTGCTTVG
VCGKTPEMAFLQDLLTYSLKGLGCWADFARQQGVEVPRKVYSLNAAATFATLTNVNFHDARFKEYISTCQSLREGLEKSL
RDKGIDASPPAPVNL PWFDDL AHPAAWKLNQSYLAAASMDQLTQLGKMTSLEHRRHVVDPTLLGLHEMITYGLRGLAAYS
HHAELVGHDRDPEVDEFMARCAYAFLCSEADALDLGKTLGMVDELGA VGKMQLLDKGH TTKFGHPEPTQVRITPKKGKAIL
ISGHDLQDTHDLLAQTEGTGINVWTHGELLPAHGYPALKKRFPHLVGNYGGAWYRQQKDFAEFP GAVLMTTNCIVEPRQS
YADRIFTTGEVWGAVGRHIEGEMGRTKDYSALISKAQELPGFEYEPPESEAKFVTTGFARNAVLGVAGEVVKAVQEGHLK
HIFLITGGCDGSEPD RKYFGRVADATPQD TMVLT LGCGKFRFYDHD FGM L PNTPLPRLVDMGQCNDAYS AVVVASKLAEVF
GTDINGLPLSLDLSWLEQKAVIILLSLHLNVKNIRIGPRAPAFLTPEALKVIVDKWNLQIVDTKNPESDVQKMMAGQ*

>jgi|Tetstr1|441518|TSEL_029748.t1
MAMLLRCSALRAVSAALRHHRVSTAAALLPPGRLMRLRAFGAVTPPLDSPNHGRCFSASAAAAEAHPAEEDKMFCYQCE
QTTKGVGCTTVGVCCKTPETAALQDLLLYRLKGLGIATHAKATAGIVDETINEFFNAAIFSTLTNVNFD SARFGEYIRT
ANRHISSLKERVAAAGAPPLQAPAVPWFQGI PHFPAWDMQADSAEDLEVLDLGHQIGIQARRAMINNDTLLGLHELITY
GLKGAAAYHHHA VMGKRQDINHALQQYLVFISSPEAADAGAVLGKALELGATNLAIMGKLEEHTSTFGHPTPTIVSM
TPKPGKAILVSGHDLSDLKALLDQTEGMGIDITYTHGEMLPAHGYPLSKYPHLKGHYGAWYRQKIDFDKWPGSILITTN
CVLDPPPASYAANLFTTGETGVEGVQH VAPKEFGAVIERAKQLPGFGPEADA EERTHLVGF GREAMLGAAPALLDAIKA
GQLEHIFLITGGCDGSEGSRRYYKNVAQKMPETSAILTLGCAKFRLLGHDYGNLPGTELPRLMDVGQCND SYAAIQIAVAL
AEALGT KDGVNGLPLSLDISWFEQKAVAVLLTLHLGVKDIRLGPALPAFITPDALKILVDTFGLKPADLKNPEEEPGAH
AHARAMSIPVRSRGC GAITYLTNTFTAPS NAPS AVLPQG TGS LTMNKA*

>jgi|Mesovir1|5003|Mesvi365S05478
MATRGALRLLRSSARLGAASTPLANPIRASPAVLAILQNGIHKSEVCHTSTNWLRSFRSSSALAAGGVIDPSTVVGTELDP
KMFCYQCEQALGQKGVKIGQCGKTPETAALQDLLIYAVKGLGSWAHYARTKHGIEDDAVNNFVHAAMFATLTNVNFD SQ
RFMEYLTVVESLRNQIIKQVHAKGDNKGPDAKLPWFDRLP PPAEF SIAEAMGPGVTD ETKMVALAKLASVMHRREVINN
DTLVGIEHELLTYGLKGMMAYYHHAIRLGQHDASVVAFMQEALAFLCSEDSKDLKALDFAFRCCGNNVKT LALLD TAHDK
VLGTPYPAEVRTTPVAGKCILVSGHDLWDLKTLLEQTEGSGINVYTHGEMLPAHGYPEL RKKHKL VGHFGAWMRQQTDF
VKFP GAILMTTNCIMEPSSFYRKRIFTTGEVGVNSAAHITDGKAKGRPETDFSELIECAKKEPGFRATEENTKPVTTGFG
HRFVLSIVDKVLGAIQSGDLKHIYLIIGGCDGSEMDSYFTEVAEGLPKDSIILTMCGKFRNLQNLGTLKQSGLPRLLD
MGQCND SYSAVVVAQALAKALNTDVNSLPLTLIVSWFEQKAVAVLLSLHLGKGVRI GPRLPAFLTPEALALLNQKFDL
KPTNSFDAQTDLKAILAGK*

>jgi|Chabra1|328584|rna-CBR_g19808
MAAATAFSAGSRHAVRGLMRRRAISRLSAVSADISSSSSGSSPGMSAGNSLI PPNFKEFLGARTRTPAINVVPMPQRSTY
RTYRSQAAGTTVTKEKVHEMFCFQCEQTKDGVGCTTVGVCCKDPKTAALQDVLLHSLKGLSQYAVRAKAEGISDPEMDSY
VQEAFATLTNVNFD PKAIQEFINTTIAHRENIAKAYEACQRAGRQPDMLDNAFTKFLPSGDVDALAQQGTLLGLDMRR
AGLNEDI FSLQELLTYGIKGMMAA YAAH AHA VGEADAKIGGFIHEAFDYISKPTTEQTLEELLALCLKCEVNMTAMEILN
KGHIHNFGTPEPTKLRTTPVKGKCILVSGHDLKDYELLKATEGMGINVYTHGEMLPAHGYPEDTLVMTLACGKYRFNKQ
FDSFGTLEGS DLPRLMDIGQCNDAYS AIRIAAALQAMNTDINSLPLSMVLSWFEQKAVAVLLTLHLGVQNI RIGPKLP
AFVSPNILNIIVEKYKLTPISGPEFVEDDIRAMLSLSSSSGA AKAASASPV*

>jgi|Guith1|152637|fgenes2_kg.34_#_20_#_133_1_CCFI_CCFN_EXT A_EXTB
MAFAATSKVMHPMHLSTTSS LAPMSKLQPPTMAACSKSSLKGLKMC SGQRSM AIANQQTLQSEHYPENMACYQCSQTAHG
KGCFNSNGVCGKQMGTSALQDLIVEINKEIGQLANAGKLTPEMRAFFLES MFSTLTNVNFD DSRMV DYIRKQVLLAAAK
GESKSVNDFASEPDHWEVAETYSHAARQKNLDPDAFGLIELATYGIKGACAYMFHAEAVKKQAPDAYPDQEANQVIDGL

LKIMAAMLEPTKHDLLALSLEVGALNVSVKMMLSDSHRALLGIPSPHEVSCCLKPGPCILVSGHDLADLKILLEQTEGKG
VNIYTHGEMPLPAHSYPELRKHKHLAGNWGGAWWEQHGDFKAFPGAILMTSNCLTAPTRSYRDRIFTAGPVGWDGIPHVQG
HDFSQLVDCAVNKAAPATGDEVSTHPAEKVTVGFGYETVEGVTDTLLGAIKGGQLKDVFLVGGCDGREAGRSYFTDLVKQ
SPDSSVILTLGCGKFRFNDQELGNLGE SGIPRLLDIGQCNDAYS AVVI ANGLADKLN VKMSD LPLHFAISWLEQKAVAVL
LSMLHLNMKNIIYLGPNLPAFLTPNLVKALVDNYNLKL TGDAEKDLKEMLQ*

>jgi|Phatr2|12416|e_gw1.8.44.1
MFCRQCEQTADHYACTTQGICGKTSETAACQDALIESIKSVSAWCVQARKEGSVSEQDLEAANIWTLQATFSTLTNVNFS
DDRDIADYIHQQQAI RAELQKKVLAAPTEFIGQLNLAGKSTAELEDFGHTVSI PKRAAAMGNEDAFSLNEIATYGLKGLCA
YAAHCQALGKMDPEVNAAVHEIFTKLASAEPDVGILLATVLRVGQVNGTVLAMLDSAHADLLGVPEPTPCRMTAVEGKAI
LVSGHDMVDLYELLKQTEGTGVNVYTHGEMPLPAHGYPKLKAFDHLKGNYGTAWQNQKFEFASFPGPIIVTTNCIVEPRRM
YKDRLYSMNEVGVDGVQHIPNRDFTGVIEQAQQMKGFVRTIDPPTFHTVGFNHRVVLPLAEQIIDAAKSGTSLSRIFLIGG
CDGSQWDRSYFTDLAEETPDDSLILTLGCAKNRIHQEKLMDAKLANGLPRVIDMGQCNDYS AVVVAT ELAKALDCSVN
DLPLSLCLSHLEQKAAVLLTLLSMGVKNIRLGP SLPAYVTPNVLSILQSEYNLMGTGDAKKDLAAMMEGN*

>jgi|Nanocel779_2|599317|fgenesh1_kg.20_#_499_#_TRINITY_DN7124_c0_g1_i6
MMLSSSHRRFLPLGLLSLQFFSSARPNAAVAVGFVRPLHSSPHPLKCAATNIAPRGAATTAVDAAASANKDMMCFQCEQ
TKHGRSCSPAEGSTQGVCGKTPETAALQDLLIH YTFGVSQYAH SARVHGGLSDVAVDRFLLES LFSTLTNVNFD PDRFVA
FMAEAQEMRDRAKALYEQAVRGKPEANTCLEGPAAMDLRALQGHTDKLETIGHMVGVNYRKDIAANIDAWSLGELLIYGL
KGTAAYADHARILKESNHVYSFIHEAMATLADLPNQTTENLFGMALRCGEINLEVMELLDAGSTGRYGHPTPSEVRTTP
VKGKCI AVSGHDLRDL EEVLKRTAGKGINVYTHGELLPAFGY PKLRETYPHLVANYGGAWQAQKVEFSRFPGP IIVVTNC
LIEPQKAYKDRIYTRGVVGWKGVKHIKDWDSQFFAVIEQALALPGFPEDTPKNVTMTGYGRNAVLGMAGDIVNAINTGDI
KQIYLI GGCDGAENR TYFRDIAVGLPRDSIVLTLGCGKYRMNKI PFEP TGGIPRLDMGQCNDAYS AIKVATALADVY
KCDVNDLPLQY AISWFEQKAVAVFLTMLS LGLKNIKLGPQLPAFLTPNVLQILSDSYGVRQVNLADHDADLQEMINRNPR
*

>jgi|Vitbras1|20179|Vbra_18549.t1
MAAEIEEFQPMNMFCYQCEQTSHGMSCTKTGECGKTAEVAMLQDLLIHKLKELSFFVDAFRANKKSIKFADQPPFPSEE
MSDVVEVPFDVHKMNVFLDLSIFATMTNVNFD DDDRFIDYLRTEGYIAQV KAAANTHLGLTDQFFAKACGKYRQLPSRY
ASVLAEGDIFELTDIAGSTGVEQKQYDV DNEVDVGAQELVVYGLKGAAAYAAHAQLIGKEDSSVYSFAYEALAFLTSPDA
LDLSKCLEMALRVGSINYLVMKLLSEANATFGEPEPTEVPVRPKPGKCILVTGHDL SMLHDVLEATKDRGIQVYTHGELM
PAHTY PKLKAYPHLAGHFGGAWQNQRRQFSKFP GAILVTTNCLMPLIHKSPYKDRIFTTGVC GGEDIQHLESTKDLSKLI
QAAEAAPGFTEADTEFNVRDPVQKHMPPSFHVGHSYQWIVKNAQTVLKL IQEGKIKR FYLVGGCDGSEGERSFYTDLVKQ
LPQTSVVM TLGCGKYRINHLEEGMGKIDGTLPRLLDLGQCNDAFGAIQVALTADALNCGVNDLPLSIVLSWFEQKAVAV
LLTLLFLKVKPLYIGPKLPGLTKNVLDVLSRDFGLVPLATEHEVEAKMSPP EWGTFQQVLVSLAA

>jgi|Symmic1|35194|rna24970
MAKNAEQLALPVKVKCRSTGHLILRGGTLMYNEDMQMVEIEFSKTTQEAYSYKGPFKGEPLALWLG PSTTIMKLSSISSV
KAMDYIEGTRVNCVLIIRLRPEPSDDL MQVAPDAGKEEEVIVQFARVEDRDNDWTGLRYLINALEVTVAKDQVEVPTRSF
SRIKKVRLEEP RAGILVRGRFELASGEEPELEIPESMADGKDLNTFVVDWVQRNCVQPSETTSLYRLVKS LVHRATLESK
TADIIQRINDTHFDKLLKEHPGDPQATLEISKAYLREVGHDI PKLIGQQGTASAMVIQILQRNVEKMKVINDMAYRIHIG
ETDQHVTD FLKIWQGCARFGVEIVLRPQFGFKSEQTNAGKGCTTTGVCKKSPTTAGLQDLCLAKAPGKSFM LCASASWRL
LLVLRTHLRLTNVNF DNARFEAYLKQSAALAEKMEKKLADHSGPYRKDIFGVLEMCVYGLKGMAYFYHAEHLQARSRLA
GEELVKVN DQHPAVQEDKAAAYDEVERTEVYQACNP KELLRLGWVEPPQELYRIGAF LCSAGTRTRRIEPQLHENLKV M
KLLDAGHNAVLGTPEPTQVKQEPKGPAILVSGHDSLILGK LLEQCKGRGVNVYTHGEMPLPAHSYPGLKNLGVRL*

>jgi|Symmic1|10117|rna37234
MVVWAVQKPWLKQNL PYRAAHFPAFPGLGGLGGNPWQQRRPPPTVPADFQVDSNARHF GTCVYYHKWRGYGFLEPTHVG
TVPTNKVFVHVKQLTSDDRFPFLVKGMEVEFSLMVSKDFHRGLNTLRANKVTLVGGESIALQDELDATEKQFVGGQHLRY
TGTLKFYSPRHGFGYVMMDEGYDVEASVPLELRVDHEEVNAAGQQPVHMRDIAVEFGIYRTDRGQYKVNMTLQGGHPLT
QDALENRISMGMGMYRGQIAFWNWRQGYGFIRPDPTAILPEKVVARLVEQAEAAARKRGKRIAEDSLLYFRRPDVT PMLIP
QQGMQVGFQVYIDDKGAGACDDDSIILTLGCGKFR LNGRDYG TIGGIPRLLDV GQCNDAYGAVVIATKLAEALGCGVHDL
PLHFAVSWFEQKAPAGWGGAKTNII*