Supplementary Material
Prediction of several novel families of enzymes involved in oxidative and other complex modifications of bases in nucleic acids

Lakshminarayan M Iyer[1], Mamta Tahiliani[2], Anjana Rao [2]and L. Aravind[1*]

* *Address for correspondence: L. Aravind (aravind@mail.nih.gov)*

[1]*National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 20894, USA*
[2]*Department of Pathology, Harvard Medical School and Immune Disease Institute, 200 Longwood Avenue, Boston, Massachusetts 02115, USA.*

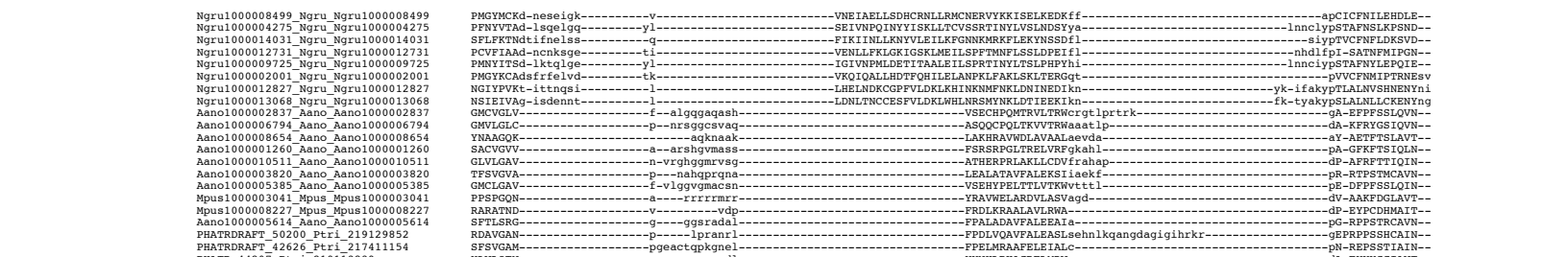
Abstract

Modified bases in nucleic acids present a layer of information that directs biological function over and beyond the coding capacity of the conventional bases. While a large number of modified bases have been identified, many of the enzymes generating them still remain to be discovered. Recently, members of the 2-oxoglutarate- and iron(II)-dependent dioxygenase superfamily, which modify diverse substrates from small molecules to biopolymers, were predicted and subsequently confirmed to catalyze oxidative modification of bases in nucleic acids. Of these, two distinct families, namely the AlkB and the kinetoplastid base J binding proteins (JBP) catalyze in situ hydroxylation of bases in nucleic acids. Using sensitive computational analysis of sequences, structures and contextual information from genomic structure and protein domain architectures, we report five distinct families of 2-oxoglutarate- and iron(II)-dependent dioxygenase that we predict to be involved in nucleic acid modifications. Among the DNA-modifying families, we show that the dioxygenase domains of the kinetoplastid base J-binding proteins belong to a larger family that includes the Tet proteins, prototyped by the human oncogene Tet1, and proteins from basidiomycete fungi, chlorophyte algae, heterolobosean amoeboflagellates and bacteriophages. We present evidence that some of these proteins are likely to be involved in oxidative modification of the 5-methyl group of cytosine leading to the formation of 5-hydroxymethylcytosine. The Tet/JBP homologs from basidiomycete fungi such as Laccaria and Coprinopsis show large lineage-specific expansions and a tight linkage with genes encoding a novel and distinct family of predicted transposases, and a member of the Maelstrom-like HMG family. We propose that these fungal members are part of a mobile transposon. To the best of our knowledge, this is the first report of a eukaryotic transposable element that encodes its own DNA-modification enzyme with a potential regulatory role. Through a wider analysis of other poorly characterized DNA-modifying enzymes we also show that the phage Mu Mom-like proteins, which catalyze the N6-carbamoylmethylation of adenines, are also linked to diverse families of bacterial transposases, suggesting that DNA modification by transposable elements might have a more general presence than previously appreciated. Among the other families of 2-oxoglutarate- and iron(II)-dependent dioxygenases identified in this study, one which is found in algae, is predicted to mainly comprise of RNA-modifying enzymes and shows a striking diversity in protein domain architectures suggesting the presence of RNA modifications with possibly unique adaptive roles. The results presented here are likely to provide the means for future investigation of unexpected epigenetic modifications, such as hydroxymethyl cytosine, that could profoundly impact our understanding of gene regulation and processes such as DNA demethylation.

Contents

1a. Multiple sequence alignment of novel members of the 2OGFeDO superfamily.

```
Secondary Structure                       -EEEE-------------------------------------------------hhhhhhhh--hhhhhhhhH-HHHHHHHHHHHHHh------------------------------------------EEEEEEEE---
TET1_Human                                SWSMYFNgckfgr------------spsprrfridpssplhekn---------LEDNLQSLATRLAPIYKQYAPVAYQNQVEYENVArecrl--------------------------------gs-KEGRPFSGVTAC--
TET1_Mouse                                SWSMYFNgckfgr------------senprkfrlapnyplhekq---------LEKNLQELATVLAPLYKQMAPVAYQNQVEYEEVAgdcrl--------------------------------gn-EEGRPFSGVTCC--
TET3_Human                                SWSMYFNgckyar------------sktprkfrlagdnpkeeev---------LRKSFQDLATEVAPLYKRLAPQAYQNQVTNEEIAidcrl--------------------------------gl-KEGRPFAGVTAC--
TET3_Mouse                                SWSMYFNgckyar------------sktprkfrltgdnpkeeev---------LRNSFQDLATEVAPLYKRLAPQAYQNQVTNEDVAidcrl--------------------------------gl-KEGRPFSGVTAC--
TET2_Human                                SWSMYYNgckfar------------skiprkfkllgddpkeeek---------LESHLQNLSTLMAPTYKKLAPDAYNNQIEYEHRApecrl--------------------------------gl-KEGRPFSGVTAC--
TET2_Mouse                                SWSMYYNgckfar------------skkprkfrlhgaepkeeer---------LGSHLQNLATVIAPIYKKLAPDAYNNQVEFEHQApdccl--------------------------------gl-KEGRPFSGVTAC--
LOC580376_Sea_urchin                      SWSMYYNtckfar------------srtprkfklleanpeqedv---------LSDRFQNMATDLGPLYKRLAPESFNNMVVFEEEGkecrl--------------------------------gk-ETGRPFAGVTAC--
CG2083_Drosophila                         SWSMYYNgckyar------------sktvrkfrlsvk--seeaa---------IEDHMNLIATLLAPVFKQVCPRSYDNQTKYEHEAsdcrl--------------------------------gl-EPGKPFSGVTAC--
v1g22996_Nematostella                     SWNMYFNgckfar------------sksprkykll-dsskeet----------LERILEGIATEIAPVYSKAAPVAFANQTREERNGhecri--------------------------------ghsAVGRPFSGVTCC--
gp2_BPCooper_109392355                    RTFGYAPrrpivw------------reacsitalgrdypq-------------IEQVLESYADQFAAGLGAIDPELVKRGSDQL-ASvlhdw--------------------------------rl-GEAKLWTSGVVN--
FRAAL2749_Faln_111222169                  RNFGFTPrrpvyg------------regcqigrlaaespq-------------EHRVLERWALLLERTLDQIDPSIVARDAQTMTQV-dssw--------------------------------rl-GGGRLWTSGVIN--
GOS_4878734_Mmet_139542046                SIFGSLPriarrn------------dfcrfsahtkkeik--------------NTNIIFSFMNDLINIYKKYLPEQYERDIKVIKESvvedy--------------------------------al-NKKSPFLTCNIN--
GOS_153646_Mmet_134535573                 CIIGSVPrntrmr------------rmhhnrssvhrskaaqt-----------FIKAMVIAGRQSLSVIKELTPELYETHRESVLDRvpeqw-----------------------------------RFCDLFTSSISN--
GOS_9579659_Mmet_135108850                ------------------------------vlratraqpe-------------VFAGLSKVGKYLWGVYQNCFPEVAANFQKFVGGIhdd----------------------------------wk-KTGTPFTTVNVN--
GOS_3272744_Mmet_139110457                ---------------------------------------------------------MLMCCLESENLIKKYMPEQYESQKKLIEETtlp----------------------------------ky-RFGKLFTSSISN--
GOS_4675571_Mmet_139186735                ------------------------------------------------------IKAMLMSCLESEKIIKQYMPEQYASQKKLIEETtlp----------------------------------ey-RFGNLFTSSISN--
GOS_4428895_Mmet_140212139                -----------------------------------------------------------------------------------IEETtlp----------------------------------ky-RFGNLFTSSISN--
JBP1_Tbru_6018041                         GIAGYFDyrgtpv------------elkcrktsftyehtk-------------EWRSVFPMIDYTSAIYKAALPDHWKAQDAAVPDV-------------------------------------vr-IHGSPFSTLTVN--
JBP2_Tbru_72391588                        GILGYYDylnnpt------------krkcrmteftrrnwgk------------IIGPCGELLQLLDQLYKENAPDHYELQRRVIPPE-------------------------------------ym-LFNTVFSTVSVN--
Tc00.1047053506605.229_Tcru_71662347      GIVGYYDylnnpt------------qrkcreteftrknwss------------VVDSCEPFLVALNKLYSECAPTHYKLQRIAIPRH-------------------------------------yq-LFNTVFSTMTVN--
Tc00.1047053510357.10_Tcru_71421637       GIAGYFDyrgspv------------elkcrktsftyenvh-------------SWPNVFPMIDYVSAIYKAVFPEQWAAQDAAVPDI-------------------------------------vr-IHGSPFSTLTVN--
JBP1_Cfas_6018043                         GIAGYFDyrgspv------------elkarktaftyehek-------------KWPAVFPLVDYVSEIYKSVMPEHWAAQDSAIPDI-------------------------------------vr-IHGTPFSTLTIN--
GOS_9294781_Mmet_135380621                GIIGYFDrydrnqlgngktlp----kipcrttkftkefve-------------KWDKCIPFIEEIDKQFSIHIPDRHKVQLERASLT-kd----------------------------------fq-IKNTAFSTITIN--
GOS_3208587_Mmet_139987906                SIIGYADryprip--------------ycrqtaftekhfd-------------MYSQAIPYLQSISKLFEEFLPERWQNQKNEWDKTsed----------------------------------fk-IHGTVFTTITVN--
GOS_7711251_Mmet_136831790                NPIGFYEssnnfs------------klpcrlthftrtnfd-------------KYNYGLPFIQKIDSLFKCLIPDAYKRQLNRANLR-dk----------------------------------fk-IPNTSFSTVTIN--
GOS_9234698_Mmet_135432669                NPIGFYEasknfc------------hlpcrlthftrtnfd-------------DYNKGLPFIQQIDSLFKKLIPDAYKKQLDRANQK-ph----------------------------------lk-IPNTSFSTITIN--
GOS_464085_Mmet_144014002                 NPIGFYEesknfc------------hlpcrlthftrtnfd-------------DYNKGLPFIQQIDSLFKKLIPDAHEKQLDRANKK-ph----------------------------------lk-IPGTSFSTITIN--
GOS_8006756_Mmet_136547457                NIMGYFDrwsislrasfkragmkp-ptrcritsftsrfpe-------------KWENVVPLIQDIDAQYKRLVPKAYANQRKAADSV--k----------------------------------fk-IPNTSFSTVTTN--
GOS_8124242_Mmet_136439712                NIIGYMDtwtiqhkymfsqvgmkeikpavrrsyftqnnyd-------------NWTPMKSLVKHIDAQYKKLAPVQYKKQRAKADET--y----------------------------------fk-IKGTAFTTLTTN--
GOS_398524_Mmet_144068378                 NIIGYFDkrdrnl-------gan--appcrttaftsqqve-------------KWNNVVPLIKNIDLQFKRLIPSNHRIQYDRANKT--d----------------------------------fv-INGTAFSTVTIN--
GOS_137212_Mmet_134552279                 NIFGFFDkwspkqkatfrklgkkp-dvdvrecrfnmdepd-------------KYKKTLPLIKEIDRLYAKLVPVQYKKQKKKARST--h----------------------------------fk-IDNTAFTTITTN--
GOS_1490958_Mmet_143037129                GVVGFMDksamir--------------ycrktaftkkyfd-------------NYQEGLPFVKFVDEQYKKLCPEYYNRQKNIAEGTnqn----------------------------------yv-IPDTSFTTVTVN--
CC1G_03999_Ccin_169848807                 GVACFAPawwavgheksslpg----pstplksksgik----------------FVDEMEESLALVGAILSVVHPELYEIGCHVMNEMrvadesqvkypeqvh--------------------rvlk-TWASPFTALSLI--
LACBIDRAFT_316849_Lbic_170117859          GIVNLSPswfsqghegpedpl----rasadiqepanst---------------YLQNVLLANAIVGAIYAVIQPDLFESGLATFEKLanha---ealddpin--------------------ialr-VWATPFTGLSVI--
CC1G_09735_Ccin_169868221                 GVATFSScwyqqghdgpydsps---psaslkgppgqq----------------FIRDMTESFAVLGGVIAITQPWLFNAGLEVLESLh-rrtfpvdnaetld--------------------ailn-FWSNPFTAFSVI--
CC1G_09735_Ccin_169868221                 GNALRGGtaklelcnlpdgarrpa-lcpiflgpdsamyik-------------FLEEMQESFAVLGGILAVTHPWLFDAGLSILSSLe-hgkidisqpenll--------------------dvls-FWSSPFSKFELV--
CHLREDRAFT_179132_Crei_159485382          GQAVRLQppwqlcprrr--------yvnqiiyeceaehypfg-----------IEEALEEMCTAVWEAAVQIAPYLTKYRDEFLSAWgrq----------------------------------am-YGDTATNLVSMT--
CHLREDRAFT_193135_Crei_159480130          GNALIQNdgkaiskg----------yavahrarvtsnvygka-----------NDVSLQRLAETVWSVVEKRLSFMPAYRDLVITEQgkp----------------------------------fm-LGATATNIISLT--
CHLREDRAFT_180688_Crei_159491750          ARPRVSCkyadqrk-----------yeaaaatedaate---------------PERLIETMCRTVWGAAVEVFPFLEECRDVFLSPVgkg----------------------------------vc-YDGTMTMQVSVT--
```



Species abbreviations: Aano : Aureococcus anophagefferens; Acry : Acidiphilium cryptum; Anid : Aspergillus nidulans; Anig : Aspergillus niger; Ater : Aspergillus terreus; Avin : Az

Back to Contents

2a. Phyletic distribution of the TET/JBP family of proteins

#; Tet/JBP family; bacteriophage gp2 subfamily

gi	gene_name	length	Species	class	GenBank_annotation
111222169	FRAAL2749	277	Frankia alni ACN14a	actinobacteria	hypothetical protein FRAAL2749 [Frankia alni
194100516	Nigel_2	322	Mycobacterium phage Nigel	dsDNA viruses, no RNA stage>caudovirales	gp2 [Mycobacterium phage Nigel].
109392355	Cooperp2	322	Mycobacterium phage Cooper	dsDNA viruses, no RNA stage>caudovirales	gp2 [Mycobacterium phage Cooper].
Sequences from marine metagenomes that are closer to the gp2 subfamily					
139110457	GOS_3272744	133	marine metagenome		hypothetical protein GOS_3272744 [marine meta
139186735	GOS_4675571	137	marine metagenome		hypothetical protein GOS_4675571 [marine meta
139542046	GOS_4878734	201	marine metagenome		hypothetical protein GOS_4878734 [marine meta
140212139	GOS_4428895	107	marine metagenome		hypothetical protein GOS_4428895 [marine meta
142006841	GOS_3046279	235	marine metagenome		hypothetical protein GOS_3046279 [marine meta
134535573	GOS_153646	182	marine metagenome		hypothetical protein GOS_153646 [marine metag
135108850	GOS_9579659	181	marine metagenome		hypothetical protein GOS_9579659 [marine meta

#;Tet/JBP family; JBP subfamily

gi	gene_name	length	Species	class	GenBank_annotation
6018043	JBP1	811	Crithidia fasciculata	euglenozoa>kinetoplastida	J-binding protein [Crithidia fasciculata].
6018045	JBP1	827	Leishmania tarentolae	euglenozoa>kinetoplastida	J-binding protein [Leishmania tarentolae].
146078722	LinJ09.1540	814	Leishmania infantum JPCM5	euglenozoa>kinetoplastida	DNA J-binding protein, putative [Leishmania i
6018041	JBP1	839	Trypanosoma brucei	euglenozoa>kinetoplastida	J-binding protein [Trypanosoma brucei].
157865226	LmjF09.1480	814	Leishmania major strain Friedlin	euglenozoa>kinetoplastida	DNA J-binding protein, putative [Leishmania m
74025708	Tb11.01.5220	839	Trypanosoma brucei TREU927	euglenozoa>kinetoplastida	chaperone protein DnaJ [Trypanosoma brucei TR
72391588	Tb927.7.4650	1077	Trypanosoma brucei TREU927	euglenozoa>kinetoplastida	SNF2 DNA repair protein, putative [Trypanosom
154334093	LbrM14_V2.0040	1098	Leishmania braziliensis MHOM/BR/75/M2904	euglenozoa>kinetoplastida	j-binding protein, putative [Leishmania brazi
146081173	LinJ14.0040	1022	Leishmania infantum JPCM5	euglenozoa>kinetoplastida	j-binding protein, putative [Leishmania infan
157866423	LmjF14.0040	1022	Leishmania major strain Friedlin	euglenozoa>kinetoplastida	j-binding protein, putative [Leishmania major
209407460	jbp2	1098	Leishmania tarentolae	euglenozoa>kinetoplastida	J-binding protein 2 [Leishmania tarentolae].
71421637	Tc00.1047053510357.10	832	Trypanosoma cruzi strain CL Brener	euglenozoa>kinetoplastida	DnaJ chaperone protein, putative [Trypanosoma
71653481	Tc00.1047053506753.120	831	Trypanosoma cruzi strain CL Brener	euglenozoa>kinetoplastida	DnaJ chaperone protein, putative [Trypanosoma
71422266	Tc00.1047053508859.74	1086	Trypanosoma cruzi strain CL Brener	euglenozoa>kinetoplastida	SNF2 DNA repair protein, putative [Trypanosom
71662347	Tc00.1047053506605.229	986	Trypanosoma cruzi strain CL Brener	euglenozoa>kinetoplastida	SNF2 DNA repair protein, putative [Trypanosom
154333008	LbrM09_V2.1540	813	Leishmania braziliensis MHOM/BR/75/M2904	euglenozoa>kinetoplastida	DNA J-binding protein, putative [Leishmania b
Sequences from marine metagenomes that are closer to the JBP subfamily					
139987906	GOS_3208587	293	marine metagenome		hypothetical protein GOS_3208587 [marine meta
136831790	GOS_7711251	346	marine metagenome		hypothetical protein GOS_7711251 [marine meta
135432669	GOS_9234698	420	marine metagenome		hypothetical protein GOS_9234698 [marine meta
144014002	GOS_464085	412	marine metagenome		hypothetical protein GOS_464085 [marine metag
136547457	GOS_8006756	320	marine metagenome		hypothetical protein GOS_8006756 [marine meta
136439712	GOS_8124242	286	marine metagenome		hypothetical protein GOS_8124242 [marine meta
144068378	GOS_398524	329	marine metagenome		hypothetical protein GOS_398524 [marine metag
134552279	GOS_137212	298	marine metagenome		hypothetical protein GOS_137212 [marine metag

143037129	GOS_1490958	255	marine metagenome	hypothetical protein GOS_1490958 [marine meta
135639857	GOS_9010290	279	marine metagenome	hypothetical protein GOS_9010290 [marine meta
137245605	GOS_7342633	288	marine metagenome	hypothetical protein GOS_7342633 [marine meta
140226134	GOS_4085697	262	marine metagenome	hypothetical protein GOS_4085697 [marine meta
142489586	GOS_2357195	297	marine metagenome	hypothetical protein GOS_2357195 [marine meta
140977922	GOS_4016360	255	marine metagenome	hypothetical protein GOS_4016360 [marine meta
141147691	GOS_4034694	235	marine metagenome	hypothetical protein GOS_4034694 [marine meta
137707555	GOS_6884863	252	marine metagenome	hypothetical protein GOS_6884863 [marine meta
137902823	GOS_6711763	252	marine metagenome	hypothetical protein GOS_6711763 [marine meta
134679729	GOS_15194	215	marine metagenome	hypothetical protein GOS_15194 [marine metage
139459875	GOS_3470977	252	marine metagenome	hypothetical protein GOS_3470977 [marine meta
140203033	GOS_4682654	213	marine metagenome	hypothetical protein GOS_4682654 [marine meta
139546231	GOS_4760684	220	marine metagenome	hypothetical protein GOS_4760684 [marine meta
136368507	GOS_8203392	303	marine metagenome	hypothetical protein GOS_8203392 [marine meta
139914055	GOS_5217427	187	marine metagenome	hypothetical protein GOS_5217427 [marine meta
139800537	GOS_4792986	234	marine metagenome	hypothetical protein GOS_4792986 [marine meta
139480713	GOS_3023932	201	marine metagenome	hypothetical protein GOS_3023932 [marine meta
138173944	GOS_3925851	277	marine metagenome	hypothetical protein GOS_3925851 [marine meta
141956139	GOS_4404681	197	marine metagenome	hypothetical protein GOS_4404681 [marine meta
137954862	GOS_6664660	307	marine metagenome	hypothetical protein GOS_6664660 [marine meta
137777329	GOS_6830377	256	marine metagenome	hypothetical protein GOS_6830377 [marine meta
141681140	GOS_3328063	264	marine metagenome	hypothetical protein GOS_3328063 [marine meta
137687150	GOS_6903293	181	marine metagenome	hypothetical protein GOS_6903293 [marine meta
136925938	GOS_7634231	196	marine metagenome	hypothetical protein GOS_7634231 [marine meta
140240419	GOS_3689741	142	marine metagenome	hypothetical protein GOS_3689741 [marine meta
138487664	GOS_3249419	218	marine metagenome	hypothetical protein GOS_3249419 [marine meta
136542980	GOS_8011616	215	marine metagenome	hypothetical protein GOS_8011616 [marine meta
140303641	GOS_5745746	193	marine metagenome	hypothetical protein GOS_5745746 [marine meta
139574841	GOS_3986906	198	marine metagenome	hypothetical protein GOS_3986906 [marine meta
140218070	GOS_4292923	186	marine metagenome	hypothetical protein GOS_4292923 [marine meta
143318242	GOS_1154576	161	marine metagenome	hypothetical protein GOS_1154576 [marine meta
143401206	GOS_1074871	183	marine metagenome	hypothetical protein GOS_1074871 [marine meta
141726655	GOS_8922744	171	marine metagenome	hypothetical protein GOS_8922744 [marine meta
144023211	GOS_454485	156	marine metagenome	hypothetical protein GOS_454485 [marine metag

#;Tet/JBP family; transposon-associated subfamily; note some could be fragmentary

gi	gene_name	length	Species	class	GenBank_annotation
159472010	CHLREDRAFT_190891	494	Chlamydomonas reinhardtii	viridiplantae>chlorophyta	predicted protein [Chlamydomonas reinhardtii]
159491750	CHLREDRAFT_180688	813	Chlamydomonas reinhardtii	viridiplantae>chlorophyta	predicted protein [Chlamydomonas reinhardtii]
159485370	CHLREDRAFT_179120	693	Chlamydomonas reinhardtii	viridiplantae>chlorophyta	predicted protein [Chlamydomonas reinhardtii]
159474434	CHLREDRAFT_191666	1887	Chlamydomonas reinhardtii	viridiplantae>chlorophyta	predicted protein [Chlamydomonas reinhardtii]
159485382	CHLREDRAFT_179132	270	Chlamydomonas reinhardtii	viridiplantae>chlorophyta	predicted protein [Chlamydomonas reinhardtii]
159482446	CHLREDRAFT_181150	345	Chlamydomonas reinhardtii	viridiplantae>chlorophyta	hypothetical protein CHLREDRAFT_181150 [Chlam
159480130	CHLREDRAFT_193135	444	Chlamydomonas reinhardtii	viridiplantae>chlorophyta	hypothetical protein CHLREDRAFT_193135 [Chlam
159470525	CHLREDRAFT_187843	249	Chlamydomonas reinhardtii	viridiplantae>chlorophyta	hypothetical protein CHLREDRAFT_187843 [Chlam
159466571	CHLREDRAFT_180874	604	Chlamydomonas reinhardtii	viridiplantae>chlorophyta	hypothetical protein CHLREDRAFT_180874 [Chlam
Vcar1000013584	Vcar1000013584	842	Volvox carteri	viridiplantae>chlorophyta	estExt_fgenesh4_pg.C_1090014
169845768	CC1G_10133	327	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169855325	CC1G_02065	431	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169861999	CC1G_13531	387	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169868182	CC1G_10451	429	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169860178	CC1G_08110	452	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169857286	CC1G_10221	384	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169866930	CC1G_12863	306	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169857284	CC1G_10220	377	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169848563	CC1G_12955	347	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169869476	CC1G_12709	389	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169867888	CC1G_07251	621	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169862392	CC1G_12947	389	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169869783	CC1G_13149	387	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169842924	CC1G_02774	404	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169851060	CC1G_02483	351	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169862969	CC1G_05589	413	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169860384	CC1G_04140	598	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169850587	CC1G_07039	406	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169869758	CC1G_12677	406	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169850700	CC1G_13537	364	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169864674	CC1G_05497	400	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169864016	CC1G_11683	405	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169852272	CC1G_10670	316	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169857294	CC1G_10225	362	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169842918	CC1G_02771	368	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169862394	CC1G_12948	568	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169862421	CC1G_13133	196	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169850687	CC1G_09395	342	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169863834	CC1G_13154	365	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169865994	CC1G_12619	297	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169848807	CC1G_03999	409	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169869579	CC1G_12549	356	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169866947	CC1G_13115	383	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169856593	CC1G_11467	224	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169866913	CC1G_10425	653	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169862010	CC1G_13092	633	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169868221	CC1G_09735	1008	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
169866816	CC1G_12944	460	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayam
170113031	LACBIDRAFT_310759	403	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170094354	LACBIDRAFT_315476	363	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170095677	LACBIDRAFT_325387	1395	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170107651	LACBIDRAFT_304732	361	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170107366	LACBIDRAFT_330629	740	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170096084	LACBIDRAFT_325753	290	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170118062	LACBIDRAFT_335309	525	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170100483	LACBIDRAFT_298034	291	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170106738	LACBIDRAFT_330369	377	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170096288	LACBIDRAFT_318383	291	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170103707	LACBIDRAFT_329024	359	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170113049	LACBIDRAFT_310603	334	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170110303	LACBIDRAFT_331944	263	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170117440	LACBIDRAFT_316488	184	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170097093	LACBIDRAFT_318984	240	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170114173	LACBIDRAFT_312548	342	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170094478	LACBIDRAFT_324729	393	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170111900	LACBIDRAFT_332549	302	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170108766	LACBIDRAFT_306923	205	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170104721	LACBIDRAFT_302545	218	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170120100	LACBIDRAFT_306943	171	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170105238	LACBIDRAFT_302976	145	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170111465	LACBIDRAFT_309595	219	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170111641	LACBIDRAFT_309600	204	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170105591	LACBIDRAFT_329738	549	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170087510	LACBIDRAFT_321468	292	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170107368	LACBIDRAFT_330630	429	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170105639	LACBIDRAFT_329780	425	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170092087	LACBIDRAFT_313413	435	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170102298	LACBIDRAFT_299476	435	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170117725	LACBIDRAFT_335138	431	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170108014	LACBIDRAFT_306265	363	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170120696	LACBIDRAFT_336317	208	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170119017	LACBIDRAFT_298319	236	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170088597	LACBIDRAFT_308407	112	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170110399	LACBIDRAFT_307783	135	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170111143	LACBIDRAFT_332267	734	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170105509	LACBIDRAFT_329685	384	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170084579	LACBIDRAFT_321006	190	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170106185	LACBIDRAFT_303791	265	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170104027	LACBIDRAFT_329093	236	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170097878	LACBIDRAFT_296330	352	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170117859	LACBIDRAFT_316849	363	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170117727	LACBIDRAFT_316719	363	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170108734	LACBIDRAFT_306859	206	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170100465	LACBIDRAFT_298012	403	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170117012	LACBIDRAFT_334833	2352	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170089437	LACBIDRAFT_292649	1175	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170120448	LACBIDRAFT_336255	329	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170089645	LACBIDRAFT_308573	164	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170108631	LACBIDRAFT_306697	340	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170108732	LACBIDRAFT_306858	176	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
170103709	LACBIDRAFT_329026	359	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82

170114177	LACBIDRAFT_312551	342	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82
215477490	MPER_00118	144	Moniliophthora perniciosa FA553	fungi>basidiomycota	hypothetical protein MPER_00118 [Moniliophtho
215469507	MPER_03117	215	Moniliophthora perniciosa FA553	fungi>basidiomycota	hypothetical protein MPER_03117 [Moniliophtho
215462753	MPER_05994	212	Moniliophthora perniciosa FA553	fungi>basidiomycota	hypothetical protein MPER_05994 [Moniliophtho
215439633	MPER_14848	59	Moniliophthora perniciosa FA553	fungi>basidiomycota	hypothetical protein MPER_14848 [Moniliophtho
215471301	MPER_02368	162	Moniliophthora perniciosa FA553	fungi>basidiomycota	hypothetical protein MPER_02368 [Moniliophtho
215455531	MPER_09421	189	Moniliophthora perniciosa FA553	fungi>basidiomycota	hypothetical protein MPER_09421 [Moniliophtho
215457781	MPER_08327	214	Moniliophthora perniciosa FA553	fungi>basidiomycota	hypothetical protein MPER_08327 [Moniliophtho
215449584	MPER_12620	343	Moniliophthora perniciosa FA553	fungi>basidiomycota	hypothetical protein MPER_12620 [Moniliophtho
215469046	MPER_03319	194	Moniliophthora perniciosa FA553	fungi>basidiomycota	hypothetical protein MPER_03319 [Moniliophtho
215470564	MPER_02669	134	Moniliophthora perniciosa FA553	fungi>basidiomycota	hypothetical protein MPER_02669 [Moniliophtho

#; Metazoan JBPs; note some could be fragmentary

gi	gene_name	length	Species	class	GenBank_annotation
Caps1000002965	Caps1000002965	317	Capitella spI	metazoa>annelida	e_gw1.399.55.1
Caps1000002966	Caps1000002966	96	Capitella spI	metazoa>annelida	gw1.399.56.1
Caps1000002966	Caps1000002966	96	Capitella spI	metazoa>annelida	gw1.399.56.1
Hrob1000003849	Hrob1000003849	324	Helobdella robusta	metazoa>annelida	66924
198433354	LOC100181878	1706	Ciona intestinalis	metazoa	PREDICTED: similar to Protein TET2 [Ciona intestinalis].
156389231	NEMVEDRAFT_v1g22996	256	Nematostella vectensis	metazoa>cnidaria	predicted protein [Nematostella vectensis].
Dpul1000021061	Dpul1000021061	402	Daphnia pulex	metazoa>crustacea	estExt_Genewise1Plus.C_670028
115977068	LOC580376	1938	Strongylocentrotus purpuratus	metazoa>echinodermata	PREDICTED: hypothetical protein [Strongylocentrotus purpur
110761604	LOC412878	2736	Apis mellifera	metazoa>hexapoda	PREDICTED: similar to CG2083-PA [Apis mellifera].
24656270	CG2083	2252	Drosophila melanogaster	metazoa>hexapoda	CG2083 CG2083-PA [Drosophila melanogaster].
91082153	LOC658580	1300	Tribolium castaneum	metazoa>hexapoda	PREDICTED: similar to CG2083-PA [Tribolium castaneum].
Lgig1000002771	Lgig1000002771	416	Lottia gigantea	metazoa>mollusca	e_gw1.5.477.1
195998193	TRIADDRAFT_52488	687	Trichoplax adhaerens	metazoa>placozoa	hypothetical protein TRIADDRAFT_52488 [Trichoplax adhaeren
219503586	BRAFLDRAFT_255656	159	Branchiostoma floridae	metazoa	hypothetical protein BRAFLDRAFT_255656 [Branchiostoma flor
210083868	BRAFLDRAFT_185107	326	Branchiostoma floridae	metazoa	hypothetical protein BRAFLDRAFT_185107 [Branchiostoma flor
219503586	BRAFLDRAFT_255656	159	Branchiostoma floridae	metazoa	hypothetical protein BRAFLDRAFT_255656 [Branchiostoma flor
68439601	LOC570833	769	Danio rerio	metazoa>vertebrata>actinopterygii	PREDICTED: similar to CXXC finger 6 [Danio rerio].
68354704	LOC556972	1246	Danio rerio	metazoa>vertebrata>actinopterygii	PREDICTED: similar to CXXC finger 6, partial [Danio rerio]
68390536	LOC570335	2058	Danio rerio	metazoa>vertebrata>actinopterygii	PREDICTED: similar to CXXC finger 6 [Danio rerio].
113412889	MGC22014	1660	Homo sapiens	metazoa>vertebrata	PREDICTED: similar to CXXC finger 6 [Homo sapiens].
33859755	CXXC6	2136	Homo sapiens	metazoa>vertebrata	CXXC finger 6 [Homo sapiens].
83029316	D230004J03Rik	1713	Mus musculus	metazoa>vertebrata	PREDICTED: hypothetical protein LOC194388 [Mus musculus].
82933099	Cxxc6	2007	Mus musculus	metazoa>vertebrata	PREDICTED: CXXC finger 6 isoform 1 [Mus musculus].
114630791	CXXC6	2136	Pan troglodytes	metazoa>vertebrata	PREDICTED: CXXC finger 6 [Pan troglodytes].
55623030	LOC471263	2142	Pan troglodytes	metazoa>vertebrata	PREDICTED: hypothetical protein [Pan troglodytes].
114578319	LOC459329	1716	Pan troglodytes	metazoa>vertebrata	PREDICTED: similar to MGC22014 protein [Pan troglodytes].
109472258	LOC680576	991	Rattus norvegicus	metazoa>vertebrata	PREDICTED: similar to CXXC finger 6 [Rattus norvegicus].
109467658	RGD1311625_predicted	1947	Rattus norvegicus	metazoa>vertebrata	PREDICTED: similar to CG2083-PA [Rattus norvegicus].
109473779	RGD1561966_predicted	1667	Rattus norvegicus	metazoa>vertebrata	PREDICTED: similar to CXXC finger 6 [Rattus norvegicus].
109510046	Cxxc6_predicted	2006	Rattus norvegicus	metazoa>vertebrata	PREDICTED: similar to CXXC finger 6 [Rattus norvegicus].
47227721	GSTEN:00031114:G:001	2294	Tetraodon nigroviridis	metazoa>vertebrata>actinopterygii	unnamed protein product [Tetraodon nigroviridis].
47223312	GSTEN:00016497:G:001	1615	Tetraodon nigroviridis	metazoa>vertebrata>actinopterygii	unnamed protein product [Tetraodon nigroviridis].
47219959	GSTEN:00033468:G:001	400	Tetraodon nigroviridis	metazoa>vertebrata>actinopterygii	unnamed protein product [Tetraodon nigroviridis].

#; Tet/JBP family; Naegleria-specific subfamily

gi	Domain architecture	gene_name	length	Species	class	GenBank_annotation
Ngru1000008499		Ngru1000008499	584	Naegleria gruberi	heterolobosea	fgeneshNG_pg.scaffold_27000062
Ngru1000004275		Ngru1000004275	322	Naegleria gruberi	heterolobosea	fgeneshHS_pg.scaffold_1463000001
Ngru1000014031		Ngru1000014031	212	Naegleria gruberi	heterolobosea	estExt_fgeneshNG_pg.C_250123
Ngru1000012731		Ngru1000012731	208	Naegleria gruberi	heterolobosea	fgeneshNG_pg.scaffold_168000002
Ngru1000009725		Ngru1000009725	175	Naegleria gruberi	heterolobosea	fgeneshNG_pg.scaffold_42000010
Ngru1000002001	2OGFeDO+chromo	Ngru1000002001	576	Naegleria gruberi	heterolobosea	fgeneshHS_pg.scaffold_4000149
Ngru1000012827		Ngru1000012827	303	Naegleria gruberi	heterolobosea	fgeneshNG_pg.scaffold_267000002
Ngru1000013068		Ngru1000013068	254	Naegleria gruberi	heterolobosea	fgeneshNG_pg.scaffold_121600000

Fasta sequences of predicted JBP/TET family proteins in eukaryotes with incomplete genomes and not in GenBank or reconstructed fasta sequences

```
>gi|Caps1000002966|ref|jgi|Capca1|86523|gw1.399.56.1
EGAKEIRTDNEHNFRENDIGGVAVALTHGSVLFEVAKRELHATTAVKSPNRHEPTRISLVFYQHKNLNLPSHGHDEFLRKTEIWKERREAKQRQE
>gi|Caps1000002965|ref|jgi|Capca1|102094|e_gw1.399.55.1
MNSDSYNEDREGPYYTHLGAGPTVPAIRELMEKRMNITGDALRIEKVIYTGREGKSPQGCPVAKWILRRSSKDEKCMVIVRQRPGHTCPTAIMVIAIVVWDGIPETQATGLYDYLRHTLPDNGHETERRCGTNEKRTCACQGWSDAVGGASFTFGCSWSMYFNGCKYAKSSDSKVHRFRL
>gi|Dpul1000021061|ref|jgi|Dappu1|200491|estExt_Genewise1Plus.C_670028
MEQRTGLAGRAIRFEKIIYTGKEGKTAQGCPIAKWIIRRSSLEEKVLCLIKERRGHRCQTTWLIVISVAWEGLALRDSDYLYGELVYRLNAHGVATNRRCATNEDRTCACQGLDPDTCGASFSFGCSWSMFFNGCKFARSKQQTVRKFRLTDESQEADMGDRLQRFATAIAPLYKRIAPD
>gi|Hrob1000003849|ref|jgi|Helro1|66924
MKFVGAYDEALDGPFYTHLGTAPSIAHIRRIMIDRLMIKDNELRIEKVIYTGKEGKSLQGCPVAKWILRRSGPEEKALVIVRSIKHHYCKNSVIIIGICLWDGLPEEHATQLYDYLRMTLPKCAEATERRCATNEKRTCACQGWNEETGGASFSFGCSWSMYYNGCKFARSASPRKFHLK
>gi|Lgig1000002771|ref|jgi|Lotgi1|106626|e_gw1.5.477.1
MFCRTGAVGKALRIEKVQYSGKEGKSSQGCPIAKWIIRRSGLDEKFLCIVKRRKGHFCDKTYIVINLVAWEGVPADTADELYNHLTSSLTKNGFETERRCGTNEKKTCACQGIDLVKRGASFSFGCSWSMYFNGCKYARSREVRKFKLKDTSKEEELDSRLQSLASIVGPLYETMAPDAH
>gi|Ngru1000008499|ref|jgi|Naegr1|68183|fgeneshNG_pg.scaffold_27000062
MSINTTFNQKTTQSGEPPMMMRMTNSSTPPLTPKNCLPIFVYNDYGKLIREEQQQPTDIITNNNNSMMRSMPTTNRWETNPQTPLSVSPFQPLLPIPNFSHAFIVGNLPPSVSVRRKNRKMSEKPKNNSAPSKIMHQLELSVLNNQRRIAPKGPLADISNIQLPQQESTNKSNNTTPKKP
>gi|Ngru1000004275|ref|jgi|Naegr1|55029|fgeneshHS_pg.scaffold_1463000001
MTTFKQQTIKEKETKRKYCIKGTTANLTQTHPNGPVCVNRGEEVANTTTLLDSGGGINKKSLLQNLLSKCKTTFQQSFTNANITLKDEKWLKNVRTAYFVCDHDGSVELAYLPNVLPKELVEEFTEKFESIQTGRKKDTGYSGILDNSMPFNYVTADLSQELGQYLSEIVNPQINYYISK
>gi|Ngru1000014031|ref|jgi|Naegr1|79850|estExt_fgeneshNG_pg.C_250123
MFLRNILPENTTTEVTNILDKINQRRSKENYYIGSWGKSSSFLFKTNDTIFNELSSQFIKIINLLKNYVLEILKFGNNKMRKFLEKYNSSDFLSIYPTVCFNFLDKSVDENRILHIHPDKEDTGTSLIFYFGKFKGGAISFPELNFKLMVQSADVLLFDGKNNLHAVESLHGKDDVRYSV
>gi|Ngru1000012731|ref|jgi|Naegr1|76485|fgeneshNG_pg.scaffold_168000002
MLEAQHHKLTIYTGMWGHMKPCVFIAADNCNKSGETIVENLLFKLGKIGSKLMEILSPFTMNFLSSLDPEIFLNHDLFPISATNFMIPGNKHRILKPHKDNQDVGLCIIFYFGNYNAPLEFVNKGSVFNTERGDVLLMRGSHFRHVVKPVDNGLLEHVHDPMRISVVLFAHKSLKMNPSY
>gi|Ngru1000009725|ref|jgi|Naegr1|70713|fgeneshNG_pg.scaffold_42000010
MPMNYITSDLKTQLGEYLIGIVNPMLDETITAALEILSPRTINYLTSLPHPYHILNNCIYPSTAFNYLEPQIEKHRIKNAHKDTRDATPSVLFYLGDYDEKEGYLEFPEQNCKVFVKPGDLLLFKGNKYKHQVAPITSGTRLGLVYFAHKACKVMDFYDDYQKESLNKHKQQNQ
>gi|Ngru1000002001|ref|jgi|Naegr1|46005|fgeneshHS_pg.scaffold_4000149
MGPLHVSQHDKKKPKHRRRKKQFLKAQALTRVCWENEKSIDESGKTRVYKMIKEWEFLKGNNIQSNEPILSVYGVNDTIPKEISSNTIIVTKEGMVEMALLKSVLPPSLLEECTQLCREMSEWLATEKDIDKGSFFSGWWTMNMPMGYKCADSFRFELVDTKVKQIQALLHDTFQHILEL
>gi|Ngru1000012827|ref|jgi|Naegr1|76676|fgeneshNG_pg.scaffold_267000002
MDIGIDWRGTHFRHKNHLVKEEVCDRTNWIVLCPNGQVDIAFFPNAIPEELCLEMETVVANSDVDILSCKKAIIDGSWTRYGNGIYPVKTITTNQSILLHELNDKCGPFVLDKLKHINKNMFNKLDNINEDIKNYKIFAKYPTLALNVSHNENYNISKKPYRKHTDGNDIGLGVLTYFGS
>gi|Ngru1000013068|ref|jgi|Naegr1|77085|fgeneshNG_pg.scaffold_1216000001
MTEIVELSNIEPKDQKQAIIGGTWNRYGNSIEIVAGISDENNTLLDNLTNCCESFVLDKLWHLNRSMYNKLDTIEEKIKNFKTYAKYPSLALNLLCKENYNGKVKPYRKHIDPNNNGMDVLMFFGKTFEGGNLIVSYHYTNIDFRMFTLPIQSGDLVFLNSRIYHHKVTKVTSGVRCGLV
>v1g22996_Nematostella
MGAGTEEPPKSKRGEHREKRMSSVPDYDESVPRKKRRRCGVCEPCTRKVNCGDCSNCVNRKTGHQICKYRKCVQLRKTPDSDENEGMSDSNERENKPQPFNTHMDTDEEVESKPASPECKDDKNEQTLPSPVSEMLVATSPVQPCPEVDEPAVIARSCESTLPKPEPESPAKPPEQRVGD
>CG2083_Drosophila
MTLPPPTITITPTTHNHNHQNHNHNHSNDNNSSSCSLFQSKKKRKRCGECVGCQRKDNCGECAPCRNDKSHQICKQRRCEKLTEKKVRKWKMENAGEQARCLDAIIEIAFAVRANGACESENEPGVCSSYLAARLAPGPGDKERNSNLPARIVGFICMALCDTYYAYQINPYGHYTHHWR
>gi|Aano1000001260|ref|jgi|Auran1|61875 ; jgi|Auran1|68521
MSIVEDEEAAADGAEVLSALGDLVEKKGGDREAVVDGWAVRARVRPGDGRVEHVYVDPAGKRHRSRNEVLRCLGLLAPTPAKRARDEAAGAPEAAGDPAPPAAPPPEKRPRAKNPDIDLAAFGWAPPPPPDFAGMLERLRAVDVPRTLARQNIIGEADASAGVRSACVGVVAARSHGVMA
>gi|Aano1000002837|ref|jgi|Auran1|62507
MLRLIGDGNWPTQARPNVTDTGKPVPGMCVGLVFALGQGAQASHVSECHPQMTRVLTRWCRGTLPRTRKGAEFPFSSLQVNYNYAAQKHVDGNNIGPSYIMSIGDHSGGGLWTADQGVIDCKDAWKLFDGNKEHATRPFKGERISFIAFAHGQYNKLEAPVVKMLKGMGFNAARSDGKDL
>gi|Aano1000003820|ref|jgi|Auran1|66742
MGVVAPILLAVLGSVAAAAMLPNDHAFLKKFGGHLRCPYASKWLQHGPERAAELLNVRRLDDALVGSCTYTNPFAGVEACFEMRGSDWTAASAEARCATAMGASMPSVSGELTYGSGCTLPDAMAGWCMVAEGTEASPTALGDDGCDGAATACATWMSGDFVADGLCASEATTDVVGSCT
>gi|Aano1000005385|ref|jgi|Auran1|63084
MSSPSPKKRCPKPWRPLFAAGSRVECRHGGGDVFYPGVVTAVTSDGHYDVDYDDGDEERGVAGSLMRPEAGLKVGDALQKKFPGYGSGVFDGRVAEIFPAEKKVRVAWSDGEETFMTVAAVRKAMDHPPSPAKPVKRAAASAPPAPLPALEAAGGRPKRARKAPPALTYAKMGTNTSSAK
>gi|Aano1000005432|ref|jgi|Auran1|63122
MDGAEALSLAKRACPQFASAGLKPLDPSSITKVETRSVARLWAGMGSVVELRVKAGEDTAATIMVKKVTLPKRCASIGDRRKKASYEVECNFYEKYAAEVAAAGCACPRPLLVERRGTEGVTIVMTKLVGSSAPMDGARSRRVLSGLAKLHALFWGAAADAAVAGGLAAQGEYWYLDTRP
```

Back to Contents

2b. Superalignment of metazoan TET domains with intron-exon boundaries.

The $ refers to the codon that contains the exon-intron boundary and the numbers flanking it refers to the exon number

```
TET1_exon_bounds                                                                              3$4                   4$5                   5$6
TET3_exon_bounds                                                                              2$3                    3$4                  4$5
TET2_exon_bounds                                                                              1$2                   2$3                   3$4
Apis_exons                                                                   1$2                  6$7                                         7$8
Trichoplax_exons                                                            1$2               3$4                   4$5         5$6       6$7
Nematostella_exons                                                          1$2               2$3                   3$4         4$5       5$6
Predicted_sec_str                                                                        ------EE--EE--------EEE-----HHHHHHHHHHHHHH----EEEEEEEEE----------EEEEEEE-----EEEEEEE---------EEEEEE
Notes
Boundaries             <--------------CXXC domain-------------------->     <---C-C---Cys rich domain---------------------------------------C----W-----------?-------H-C--------
v1g22996_Nematostella  RKKRRRCGVCEPCTRKVNCGDCSNCVNRKTGHQICKYRKCVQL-RKT 104 EVAPCNCQ--VFTKEEGPPLYGQMGGAADYDSLRGMLEQRFGIKGKALRIELIIYTNKEGRNAQGCPIARWVIRRSGNDEKVLVLVRKRPGHHCSMALVVTS
TET_Trichoplax         SKKRKACGACGPCLKKENCGSCKNCLHRKTGHQICIYRKCIRL-KKS 123 LDYNCQCQ-----DEDGAPFYTQLGVAGSTEELREMLRDRFRIDESKLRVIEVEYTGVESKTSDGCPRAEWAIRRISKSEKLLALVHRRRGHTCKASVVLMA
CG2083_Drosophila      KKKRKRCGECVGCQRKDNCGECAPCRNDKS-HQICKQRRCEKLTEKK     EIPECECFQSDKNPPEPGTYYTHLGTASSLMDLRREFEERCNLTGRQLRIEKIVYTGKEGKTSQGCPVAKWVIRRADLEEKILVVVKKRPGHRCIAAYIVVC
LOC412878_Apis         KKKRKRCGECIGCQRKDNCGDCAPCRNDKS-HQICKMRRCEKLTEKK     EVPDCNCFPADKCPPEPGSYYTHLGAAASLPDLRNDLERRTGLKGNAIRFEKVIYTGKEGKTTQGCPMAKWILRRSGLEEKILTIVKHRQGHKCPTAWIVVA
Sea_urchin             KKKRKRCGVCKPCLTKQNCGECSSCKNRRTGHQICKMRKCVEL-RKK 500 EAPNCGCLG-KVILRDEAPYYTHLGTGPNLPAIRELVEIRSGFQGSQVRIEKVVYSGKEGKSSTGCPIAKWIIRRSSTDEKILVLVRHRPGHRCDTSYIIIA
TET1_Homo              KKKRKRCGVCEPCQQKTNCGECTYCKNRKNSHQICKKRKCEEL-KKK 787 ELPTCSCLD-RVIQKDKGPYYTHLGAGPSVAAVREIMENRYGQKGNAIRIEIVVYTGKEGKSSHGCPIAKWVLRRSSDEEKVLCLVRQRTGHHCPTAVMVVL
TET1_Pan               KKKRKRCGVCEPCQQKTNCGECTYCKNRKNSHQICKKRKCEEL-KKK 787 ELPTCSCLD-RVIQKDKGPYYTHLGAGPSVAAVREIMENRYGQKGNAIRIEIVVYTGKEGKSSHGCPIAKWVLRRSSDEEKVLCLVRQRTGHHCPTAVMVVL
TET1_Mus               RRKRKACGVCEPCQQKANCGECTYCKNRKNSHQICKKRKCEVL-KKK 753 EAAPCDC-D-GGTQKEKGPYYTHLGAGPSVAAVRELMETRFGQKGKAIRIEKIVFTGKEGKSSQGCPVAKWVIRRSGPEEKLICLVRERVDHHCSTAVIVVL
TET1_Rat               RRKRKACGVCEPCQQKANCGECTYCKNRKNSHQICKKRKCEVL-KKK 753 EAATCQC-D-GGPQKDKGPYYTHLGAGPSVAAVRELMETRYGQKGKAIRIEKIVFTGKEGKSSQGCPVAKWVIRRSGPEEKVICLVRERVDHYCSTAVIVVL
TET1_Canis             KKKRKRCGVCEPCQQKTNCGECTYCKNRKNSHQICKKRKCEEL-KKK 791 DLPTCNCLD-RVIQKDKGPYYTHLGAGPSVAAVREIMENRYGQKGNAIRIEIVVYTGKEGKSSHGCPIAKWVLRRGSDEEKVLCLVRQRTGHHCPTAVMVVL
TET1_Monodelphis       KKKRKRCGVCEPCQKKENCGECVYCKNRKYSHHVCKMRKCEEL-KKK 816 ELPSCSCVE-QIIEKDEGPYYTHLGTGPSVAAVREIMEARYGEKGRAIRIEVVVYTGKEGKSSQGCPIAKWVIRRSSDEEKLLCLVRQRAGHHCQTAVIVIL
TET1_Ornithorhynchus   KKKRKRCGVCEPCQRKANCGECTYCKSKKNSHQICKKRKCEEL-KKK 813 EFPTCNCVE-QIIEKDEGPYYTHLGTGPSVAAVREIMETRYGAKGRAIRIEVVVYTGKEGKSSQGCPIAKWVIRRSSNEEKLLCLVRQRAGHHCQTAVIVIL
TET1_Gallus            KKKRKRCGVCEPCLRKTNCEECSCCRKRKTSHRICKKRKCEEL-KKP 856 ELPTCDCVE-QIIEKDEGPYYTHLGTGPSVAAVREIMENRYGAKGSAVRIEVVVYTGKEGKSSQGCPIAKWVIRRSSDEEKLLCLVRQRAGHHCQTAVIVIL
TET1_Danio             QNKRGHCGVCAPCQCKVNCGECRSCLNQKTDHQICRLRRCIKL-RKK 716 EIPSCDCVE-HIIEKEEGPYYTHLGSGPNVAAVREMMETRYGEKGNAVRVEVVVYTGKEGRSSQGCPIAKWIIRRGSEQEKLLCLVRQRAGHCCQNAVIVIL
TET3_Homo              RKKRKRCGTCEPCRRLENCGACTSCTNRRT-HQICKLRKCEVL-KKK 709 EFPTCDCVE-QIVEKDEGPYYTHLGSGPTVASIRELMEERYGEKGKAIRIEKVIYTGKEGKSSRGCPIAKWVIRRHTLEEKLLCLVRHRAGHHCQNAVIVIL
TET3_Pan               RKKRKRCGTCEPCRRLENCGACTSCTNRRT-HQICKLRKCEVL-KKK 711 EFPTCDCVE-QIVEKDEGPYYTHLGSGPTVASIRELMEERYGEKGKAIRIEKVIYTGKEGKSSRGCPIAKWVIRRHTLEEKLLCLVRHRAGHHCQNAVIVIL
TET3_Macaca            RKKRKRCGTCEPCRRLENCGACTSCTNRRT-HQICKLRKCEVL-KKK 709 EFPTCDCVE-QIVEKDEGPYYTHLGSGPTVASIRELMEERYGEKGKAIRIEKVIYTGKEGKSSRGCPIAKWVIRRHTLEEKLLCLVRHRAGHHCQNAVIVIL
TET3_Canis             RKKRKRCGTCEPCRRLENCGACTSCTNRRT-HQICKLRKCEVL-KKK 711 EFPTCDCVE-QIVEKDEGPYYTHLGSGPTVASIRELMEERYGEKGKAIRIEKVIYTGKEGKSSRGCPIAKWVIRRHTLEEKLLCLVRHRAGHHCQNAVIVIL
TET3_Mus               RKKRKRCGTCDPCRRLENCGSCTSCTNRRT-HQICKLRKCEVL-KKK 717 EFPTCDCVE-QIVEKDEGPYYTHLGSGPTVASIRELMEDRYGEKGKAIRIEKVIYTGKEGKSSRGCPIAKWVIRRHTLEEKLLCLVRHRAGHHCQNAVIVIL
```

```
TET3_Rat                      RKKRKRCGTCDPCRRLENCGSCTSCTNRRT-HQICKLRKCEVL-KKK 716 EFPTCDCVE-QIVEKDEGPYYTHLGSGPTVASIRELMEDRYGEKGKAIRIEKVIYTGKEGKSSRGCPIAKWVIRRHTLEEKLLCLVRHRAGHHCQNAVIVIL
TET3_Bos                      RKKRKRCGTCEPCRRLENCGACTSCTNRRT-HQICKLRKCEVL-KKK 705 EFPTCDCVE-QIVEKDEGPYYTHLGSGPTVASIRELMEERYGEKGKAIRIEKVIYTGKEGKSSRGCPIAKWVIRRHTLEEKLLCLVRHRAGHHCQNAVIVIL
TET3_Xenopus                  NKKRKRCGVCVPCLRKEPCGACYNCVNRSTSHQICKMRKCEQL-KKK 848 EFPTCDCVE-QINEKDEGPYYTHLGSGPTVASIRELMEERFGQKGDAIRIEKVIYTGKEGKSSRGCPIAKWVIRRQSEDEKLMCLVRQRAGHHCENAVIIIL
TET3_Danio                    KKKRKRCGVCGPCMRKENCGTCSNCLNRKIGHQICKLRKCDEL-KRK 963 DFPICDCVD-QVLEKDEGPYYNHLGSGRDIPSVRQLMEDRYGEKGEAVRIEKVVYTGREGKSSQGCPIAKWVLRRSSEKEKVLCVVKQRPGHHCANTVIVVV
TET2_Homo                     -----------------------------------------------     DFPSCRCVE-QIIEKDEGPFYTHLGAGPNVAAIREIMEERFGQKGKAIRIERVIYTGKEGKSSQGCPIAKWVVRRSSSEEKLLCLVRERAGHTCEAAVIVIL
TET2_Homo                     -----------------------------------------------     DFPSCRCVE-QIIEKDEGPFYTHLGAGPNVAAIREIMEERFGQKGKAIRIERVIYTGKEGKSSQGCPIAKWVVRRSSSEEKLLCLVRERAGHTCEAAVIVIL
TET2_Gallus                   -----------------------------------------------     DFPSCSCVE-QIIEKDEGPFYTHLGAGPNVAAIREIMEERFGQKGKAIRIERVVYTGKEGKSSQGCPIAKWVVRRSSQEEKLLCLVRERAGHTCETAVIVIL
TET2_Pan                      -----------------------------------------------     DFPSCRCVE-QIIEKDEGPFYTHLGAGPNVAAIREIMEERFGQKGKAIRIERVIYTGKEGKSSQGCPIAKWVVRRSSSEEKLLCLVRERAGHTCEAAVIVIL
TET2_Canis                    -----------------------------------------------     DFPSCRCVE-QIIEKDEGPFYTHLGAGPNVAAIREIMEERFGQKGKAIRIERVIYTGKEGKSSQGCPIAKWVVRRSCSEEKLLCLVRERAGHTCEAAVIVIL
TET2_Monodelphis              -----------------------------------------------     DFPSCSCVE-QIIEKDEGPFYTHLGAGPNVAAIREIMEERFGQKGKAIRIERVIYTGKEGKSSQGCPIAKWVVRRSCNEEKLLCLVRERAGHTCEAAVIVIL
TET2_Ornithorhynchus          -----------------------------------------------     DFPSCSCVE-HIIEKDEGPFYTHLGAGPNVAAIREIMEERFGQKGKAIRIERVIYTGKEGKSSQGCPIAKWVVRRSSDEEKLLCLVRERAGHTCETAVVVIL
TET2_Danio                    -----------------------------------------------     EIPSCHCVE-QISERDEGPYYTHLGSASNVAGIRKIMEERSGMTGSAIRIEKVLYTGKEGKSGLGCPIAKWVIRRANEDEKILVLVRERAGHSCETSCVVVV
TET2_Rattus                   -----------------------------------------------     DISCEACKN---AEKDEGPYYTHLGAGPNVAAIRTIMEERFGEKGKAIRIEKVIYTGKEGKSSQGCPIAKWVYRRSSTEEKLLCLVRVRAKHTCDTAVIVIV
TET2_Mus                      -----------------------------------------------     QSQNGKCEG-CNPDKDEAPYYTHLGAGPDVAAIRTLMEERYGEKGKAIRIEKVIYTGKEGKSSQGCPIAKWVYRRSSEEEKLLCLVRVRPNHTCETAVMVIA
consensus/95%                 ................................................     p...s.C........-.usaYspLGsussh.slRpbhEpR.s.pGp.lRlE.llaTG+EG+oupGCPhA+WlhRR.s.pEKllslV+.R.sHpC.suhhl..
consensus/90%                 pppp..hs...ss....sC..h..s.p.......C.bb+s.....bp.     -hssCpC......pc-.GPaYTHLGsusslsulRphhEpRaGbcGpAlRIE.llYTGKEGKoSpGCPlAKWVlRR.s.-EKlLslV+pRssHpC.suhhVhh
consensus/100%                ................................................     ......C..........sshYsphGsu.s...lR..hc.R...p.p.lRhb.l.aTs.Eu+ss.GCP.AcWhhRR.s.pEKlhslV+.R..H.C..shhl..
```

Back to Contents

2c. Multiple sequence alignment of the TET1, TET2 and TET3 proteins

TET-1 alignment

```
Intron-exon boundaries_Hs
TET1_Homo                     MSRSRHARPSRLVRKEDVNKKKKNSQLRKTTKGANKNV--ASVKTLSPGKLKQLIQERD-VKKKTEPKPPVPVRSLLTRAGAARM-------NLDRTE-VLFQNPESLTCNGFTMALRSTSLSRRLSQPPLVVAKSKKVPLSKGLE--KQH
TET1_Pan                      MSRSRHARPSRLVRKEDVNKKKKNSQLRKTTKGANKNV--ASVKTLSPGKLKQLIQERD-VKKKTEPKPPMPVRSLLTRAGAARM-------NLDRTE-VLFQNPESLTCNGFTMALRSTSLSRRLSQPPLVVAKSKKVPLSKGSE--KQH
TET1_Mus                      MSRSRPAKPSKSVKTKLQKK-KDIQMKTKTSKQAVRHG--ASAKAVNPGKPKQLIKRRD-GKKETEDKTPTPAPSFLTRAGAARM-------NRDRNQ-VLFQNPDSLTCNGFTMALRRTSLSWRLSQRPVVTPKPKKVPPSK-------K
TET1_Rat                      MSRSRPTKPSKSVKTKLQKK--SSQMKKKISKQAVKHG--ASGKALTPGKLKQLIRRRD-GKKETEVKTPTLPQSLLTRAGAARL-------NQDRNQ-VLFQNPESLTCNGFTMALRRTSLSWRLSQRPVVTPKPKKVPPSK-------K
TET1_Canis                    MSRSRHARPSRLVRREDLNR-KKSTQLKKTCKAANKNV--ASVKTVSPGKLKQLIHERD-VKKKIEPKPPMPVRSLLTRAGAARM-------NLDRTE-VLFQNPESLTCNGFTMALRSTSLSRRLSQPPVIIAKPKKVSPPKNLE--KQH
TET1_Monodelphis              MALTRHGRSSKLVRKEDLSK-KKTSQLNKASKIMTKKV--TLVKALSTEKLQQLMPESNNVKKKSKTKSPAPTRSLLTRAAAARR-------SLDQTD-VLFQNPESLSCNGFTMCLRSTSLIRRLSQPPVVITRPGKIPPSKTLEKENKC
TET1_Ornithorhynchus          MSRSRHTRPSRLVRKEDLRK-KATSQSKKTSKTASKKV--ALVKAVSPGKLKQLMQERDGVKKKPEPKPPVPVKSLLTRAGAARL-------SLDRTD-VLFQNPESLTCNGFTMSLRSTSLTRRLSQSPVAISRPKKAPPPRSSEAEDKC
TET1_Gallus                   --MAHHARPSRLAKKEEIGK-RKTNQGKKKSSQVRKKTTKTSTKAVSSGKGKKPAQEKDVKKKQQEKKPIMSSSGRFLRSSVTRTLSGTPWVSLEKADNILFHSQEAFNINGFTMSLRNRSFSRRFSQAPT-IAKPRKAAAQKKLETKEKH
TET1_Danio                    ----------------------------------------------MRCYILLHNSV--INGSFHRYPYVHMSTLILLQTGSI-------TLKVQ--FLRRPQYC-------------FSCRLSQTPPPSPSLTSAMPRTA----KAP
```

TET-2 alignment

```
Intron-exon boundaries                                                                                 -                                                                         -
TET2_Homo                     MEQDRTNHVEGNRLSPFLIPSP-PICQTEPLATKLQNGSPLPERAHPEVNGDT-KWHSFKSYYGIPCMKGSQNSRVSPDFTQESRGYSKCLQNGGIKRTVSEPSLSGLLQIKKLKQDQKANGERRNF-GVSQERNPGE-SSQPNVSDLSDK
TET2_Mus                      MEQDRTTHAEGTRLSPFLIAPPSPISHTEPLAVKLQNGSPLAERPHPEVNGDT-KWQSSQSCYGISHMKGSQSSHESP---HEDRGYSRCLQNGGIKRTVSEPSLSGLHPNKILKLDQKAKGESNIF-EESQERNHGKSSRQPNVSGLSDN
TET2_Gallus                   MEQDRTIHVDGNRLSPFLISQPSHICQADPSAVKLQNGSPATERPEVEVNGNH-KRLFNKSNYREPHAKGSPNHRISPDLLQEKKACSIYMQNGGIKRTFSEPSLFGLQQSKKVKQDKEVNGEK---------AEPEDNYETPSISNCYSE
TET2_Pan                      MEQDRTNHVEGNRLSPFLIPSP-PICQTEPLATKLQNGSPLPERAHPEVNGDT-KWQSFKSYYGIPCMKGNQNSRVSPDFTQESRGYSKCLQNGGIKRTVSEPSLSGLLQIKKLKQDQKANGERRNF-GVSQERNPGE-SSQPNVSDLSDK
TET2_Rattus                   MEQDRTTHAEGTRQSPFLIASPSPVSHTEPLAVKLQNGSPLAERPHPEVNGDT-KWQSFQSYYGIPHMKGSQSSHESPDFIHEGRGYSGCLQNGGIKRTVSEPSLSGLHLNTKLTLDQKAKGESNNF-GESQEGNPGKSSSHPNVSGLSDD
TET2_Canis                    MEQDRSNHVEGNRLSPFLTPSPSPNCQTEPLALKLQNGSPLTERPYPEVNGDT-KWQSFKSYHGIPHMKRSQNSRVSPDFIQEGRGYSRCLQNGGIKRTVSEPSLSGLHQNKKLKQDQKANGERNNF-GESQESNPGKGSSQLNVSDLSDK
TET2_Monodelphis              MEQDRTDHVEGNRLSPFLIPPPSHICQTESLAVKLQNGSPLTERPHPEVNGEN-KWLSLKSNYGTPPTKGSQNNCISPNFKHEDREYPKCMQNGGIKRTSSEPSLSGFHQHKKVKQDQDVNGQRDDFRGNQEGNSEGE-SNSQPNVSNSWN
TET2_Ornithorhynchus          MEQDRPNHVEGSRLSPFLISASPHVCQTEPLAVKLQNGNPLPEKVHHEVNGDSNKWLPHKSNYGVSHMKGSQYNRTSPDFTQENRELSKCVQNGGIKRTFSEPSLVALHQNKKVKPDQEANGERGVP-VGSQEESQGPGISQPRVSDSCRE
Bos_reconstructed_from_ESTs   MEQDRTNHVEGNRLSPFLIPSPSPICQTEPLARKLQNGSPLTERPYLEVNGDS-KWPSLKSYYGIPHMKGSQKSRVSPDFIQEGRGYSRCLQNGGIKRTVSEPSLSGLHQNKKLKPDQKANGERKSF-GESREGNPDKGSSQPNVSDVSGK
LOC556972_Drer_189514761      ---------------------------------------------------------------------------------METEKTSNETEKSHILAKISATQQTDFLTKHQNGDQSPE-VLSQQKVPANLQGIFDQSMFGMK
LOC693689_Mmul_109075238      MEQDRTNHVEGNRLSPFLIPSP-PICQTEPLATKLQNGSPVPERSHPEVNGDT-KWQSFKSYYGIPHMKGSQNSRVSPDFIQESRGYSKCLQNGGIKRTVSEPSLSGLLQIKKLKQDQKANGERCNF-GVSQERNPGE-SSQPNVSDLSDK
```

TET-3 alignment

```
Hs_intron_exon_boundaries                                                                                                                                                         $
TET3_Homo                     -----------MSQFQVPLAVQPD-----LPGLYDF--------PQRQVMVGSFPGSGLSMAGSESQLRGGGDGRKKRKRCGTCEPCRRLENCGACTSCTNRRT-HQICKLRKCEVLKKKVGLL---KETGSELSPVDGP-VPGQMDSGPV
TET3_Pan                      -----------MSQFQVPLAVQPD-----LPGLYDF--------PQGQVMVGGFPGSGLSMARSESQLRGGGDGRKKRKRCGTCEPCRRLENCGACTSCTNRRT-HQICKLRKCEVLKKKVGLL---KETGSELSPVDGP-VPGQMDSGPV
TET3_Macaca                   -----------MSQFQVPLAVQPD-----LPGLYDF--------PQGQVMVGGFPGPELSMAGSESQLRGGGDGRKKRKRCGTCEPCRRLENCGACTSCTNRRT-HQICKLRKCEVLKKKVGLL---KETGSELSPVDGP-VPGQMDSGPV
TET3_Canis                    -----------MSQFQVPLAVQPD-----LPGLYDF--------PQGQVMVGGFQGPGLPMAGSEPQLRGGGDGRKKRKRCGTCEPCRRLENCGACTSCTNRRT-HQICKLRKCEVLKKKVGLL---KETGSELSPVDGP-VPGQMDSGPV
TET3_Mus                      -----------MSQFQVPLAVQPD-----LSGLYDF--------PQGQVMVGGFQGPGLPMAGSETQLRGGGDGRKKRKRCGTCDPCRRLENCGSCTSCTNRRT-HQICKLRKCEVLKKKAGLL---KETGSELSPVDGP-VPGQMDSGPV
TET3_Rat                      -----------MSQFQVPLAVQPD-----LSGLYDF--------PQGQVMVGGFQGPGLPMAGSETQLRGGGDGRKKRKRCGTCDPCRRLENCGSCTSCTNRRT-HQICKLRKCEVLKKKAGLL---KETGSELSPVDGP-VPGQMDSGPV
TET3_Bos                      -----------MSQFQVPLAVQPD-----LPGLYDF--------PQGQVMVAGYQGPGLPVAGTEPQLRGGGDGRKKRKRCGTCEPCRRLENCGACTSCTNRRT-HQICKLRKCEVLKKKVGLL---KETGSELSPVDGP-VPGQMDSGPV
TET3_Xenopus                  MDTQPAPVPHVLPQDVYEFPDDQES----LGRLRVSEM-PAELNGGGGGGSAAAF--AMELPEQSN---------KKRKRCGVCVPCLRKEPCGACYNCVNRSTSHQICKMRKCEQLKKKRVVPMKGVENCSESILVDGP-KTDQMEAGPV
TET3_Danio                    -MHKLRGGPGPPVQDIYEFSLDEEDPKVFLGARSARVVSPADDYIRSEHEQAGMF--SLSQMEDQEPIEPVVAQKKKRKRCGVCGPCMRKENCGTCSNCLNRKIGHQICKLRKCDELKRKKSP----WETVSVVGSVDGPRGKGQMEIGSH
Drer_intron_boundaries
```

Back to Contents

2d. Multiple sequence alignment of the JBP1-C terminal domain

```
Predicted_secondary_structure          --HHHHHHHHHH------HHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHHH---------EEEE----HHHHHHHHHHHHHHHHHHHHH-------------HHHHHHHHHHHHHHHH---HHHHHHHHHHHEEEEEE-
JBP1_Tbru_6018041                      SEKGDILNEAMNKRHYLDKENLQKMFGEEFVNIWTQSRTHWLQLVKKEWEHQKETNPTR----TRFSWNNTSAMNFAFFDLCDVAKQLMCGAFGDREVNKKEEQSFWGMFAAHLDNACINEIGMLQGSMGMHKLNVKLKDYN
Tb11.01.5220_Tbru_74025708             SEKGDILNEAMNKRHYLDKENLQKMFGEEFVNIWTQSRTHWLQLVKKEWEHQKETNPTR----TRFSWNNTSAMNFAFFDLCDVAKQLMCGAFGDREVNKKEEQSFWGMFAAHLDNACINEIGMLQGSMGMHKLNVKLKDYN
Tb10.70.2400_Tbru_71747344             SARNVVRYVASNYDSFPDGHY-------RVVDYFRVAVEMFLQQVRLEYEKQAVQAVKNGVKKPRYTWGNKSGLAISYATSCDMLKLLY------DVISEE-NRPKWDEVVKQFIYFLISESGIPGGVFSEQTYHTKFLDWR
Tc00.1047053510967.16_Tcru_71406838    GSGSVFRRVVTNYDDFPERRY-------RVIDYFRTAVEMFVTQVRQEYTHQYQQAKKNDRANIRFIWRNKGELAICFAACCDMLKLLY------DRFRPGLERPNWDGIVGQFVLFLMAESRIPAAVLTEQTYKTKFLDWR
Tc00.1047053510289.50_Tcru_71415219    GSGSVFRRVVTNYDDFPERRY-------RVVDYFRTAVEMFVTQVRQEYTHQYLQAKKNDRANIRFIWRNKGELAICFAACCDMLKLLY------DRFRPGLERPNWDGIVGQFVLFLMAESRIPAAVLTEQTYKTKFLDWR
Tc00.1047053510357.10_Tcru_71421637    SELGGALKAAEEKQYFFEEKYLADELGPDLMSMWTQSRAHWLRLVKEDWERLCRRDPER----TKFTWNNSSAMNAAFFDLCEVAKQMMIGLLNKETPSSAENHSFWILFAAHLNYACTTENGMPRDAVGMHKLNVKLKDFH
Tc00.1047053506753.120_Tcru_71653481   SELGGALKAAEEKQYFFEEKYLADELGPDLMSMWTQSRAHWLRLVKEDWERLCRRDPER----TKFTWNNSSAMNAAFFDLCEVAKQMMIGLLNKETPSSAENHSFWILFAAHLNYACTTENGMPRDAVGMHKLNVKLKDFH
LinJ09.1540_Linf_146078722             SETNLMVSTAVEKKKYLNSEFLSHFISAQLLDMWKQARGKWLELVGREWTHMLALNPER----KDFLWRNQSEMNSAFFDLCEVGKQVMLGLLGKEAALPKEEQAFWTMYAVHLNAACAEELNMPHVAMSLRKLNVKLKDFN
JBP1_Ltar_6018045                      SETNLMVSTAVEKKKYLDSEFLLHCISAQLLDMWKQARARWLELVGKEWAHMLALNPER----KDFLWKNQSEMNSAFFDLCEVGKQVMLGLLGKEVALPKEEQAFWIMYAVHLSAACAEELHMPEVAMSLRKLNVKLKDFN
LmjF09.1480_Lmaj_157865226             SETNLMVSTAVEKKKYLNSEFLSHFISAQLLDMWKQARGKWLELVGREWTHMLALNPER----KDFLWKNQSEMNSAFFDLCEVGKQVMLGLLGKEVALPKEEQAFWTMYAVHLNAACAEELHMPHVAMSLRKLNVKLKDFN
LbrM09_V2.1540_Lbra_154333008          SETDLMVSTAAEKRKYLDSEFLSHCISAQLLDMWKQARARWLELVGKEWKHMLTLNPER----KDFLWKNRSEMNSAFFDLCEVGKQVMLGLLDKEAALPKEEQAFWTMYAVHLSAACAEELHMPHDAMSLRKLNVKLKDFN
JBP1_Cfas_6018043                      SESDAVLTTAVEKSKYLERDHLSQCISAELLAMWVEARKHWLRLVATEWARMIATAPER----TDFLWKNKSPMNTAFFDLCEVAKQVMLGLLDKETATPTEERHFWSVYAAHLHRACAERLMMPEEAMSLRKLNVKLKDFS
LmjF36.0390_Lmaj_157876526             DMVKVLFLLSTDKKRYPDGRH-------RTLDYFRVAVEMFVTEVRQEYKRQYQQAQRNGRAMQRFTWKNSGELAICFACCCDNVKLLY------DSLQPGPLKPLWDAFVGQLPPMLIVQSRVPETMLSSQTYHTKYMDWV
LinJ36.2270_Linf_146103463             DMAKVLFLLSTDKKRYPDGRH-------RTLDYFRVAVEMFVTEVRQEYKRQYQQAQRSGRVMQRFTWKNSGELAICFACCCDNVKLLY------DSLQPGPLKPLWDAFVGQLAPMLIVQSRVPETMLSSQTYHTKYMDWV
LbrM35_V2.0480_Lbra_154345472          DMAKVLFFLSTDKKRYPDGRH-------RTLDYFRAAVEMFVTEVRQEYQRQYQQAQRSGRAMQRLTWKNSGELAICFACCCDNVKLLY------DNLQPGPLKPLWDAFVGQLAPMLIVQSRVPEAMLSSQTYHIKYMDWV
consensus/100%                         s....h...s.pbp.a.p.c........p.hsha..u...alpbV..-a.+b...s.pp.....ch.W.Npu.hs.sahs.C-..Kbhh......c.......p..W..hs.ph..hh..b..h....hs.pphp.KhbDa.
```

Back to Contents

2e. Multiple sequence alignment of novel transposase and identical hits

Note: 3 distinct subfamilies are shown of which the first is associated with the TET/JBP-like hydroxylase

```
Secondary Structure labels             Str1    Str2  Str3         Hel-1                                                  Str-4      Hel -4
Secondary_structure_prediction         ------EEEE-E----EEE---eeeee----HHHHHHHHHHHHHH-------------------------------------------EEEEEE-HHHHHHHHH-------------------------------------
CC1G_13152_Ccin_169863830              CDVTGVGAFA-CSRHGAFAPGSVVDFQKGERQMNMDYGLCGAIRSTGASHAR-----------------------------------RLNLLYDIACQYCQHLVKRLKD--------------------------------
CC1G_04139_Ccin_169860382              CDVTGVGAFA-CSRHGCFVPASVVNFHRGERQMYMDYGLSNSIKTTSADKMK-----------------------------------RLILLYDINCQYCRNLKKRFEA--------------------------------
CC1G_07253_Ccin_169867892              CDVTGVGAFA-CMRHGCFAPGSVVDFQKGERQVNMDYGFCEALKTTHTDELS-----------------------------------KVLLVYDINCQYSKHLFERIEE--------------------------------
CC1G_07038_Ccin_169850585              LDATGIGAIA-CMRHGAFVPSSIVDFQKGERHINMDYALSEALKHSNMDKIQ-----------------------------------RVIFVYDINCQYSKKALERLRS--------------------------------
CC1G_12943_Ccin_169866814              LDVTGIGAVA-CMRHGAFAPGSVVDFQKGERQANMDYSLSEALKYTSTDQNK-----------------------------------RVIFAYDINCQFSKNLRKRMSE--------------------------------
CC1G_12134_Ccin_169866995              MDITGVVKID-C-SHCFIKS--VVDLQLGERFPNTDMALVRALLLYKIVSPDGSPHHIIAH--------------------------DVVLSYDIACAYIRKLIDRVCAMFKE----------------------------
CC1G_09579_Ccin_169866269              MRVTGVVKID-C-SHLFVKA--MVDLQLGERFANTDYALLRALQFYKTPHSPSPSPPKALLGPSGALVPSQKEPPQPPPATWKSPYSDIVLSYDIACGYARNLSERATRN-------------------------------
CC1G_02416_Ccin_169850926              IATTGVGIAM-CARHDMRRPESAGDLHKGERFVYMDYFALRTICHNTPD--------------------------------------DVVWSYDIGCQWSQKLEMRTKKYPK-----------------------------
CC1G_08462_Ccin_169856315              YNVTGAAIVI-CSRHGFLRPNGVGDLQKGERYCNIDYIVFSALAAVALL--------------------------------------QVVITYDIACQWVKNFLKRLGKL-------------------------------
CC1G_09553_Ccin_169866217              FDDTGIAALI-C-RHGIPVF-ITNVDTPGEQQ-KFALAMVKRLFSLLPPKA------------------------------------TVVVLYDVGCVVHRSLELYDILP-------------------------------
CC1G_02772_Ccin_169842920              CDVTGVGAIA-CMRHGAFAPGSVVDFQKGERQVNMDFAFSQALKLTHTENIR-----------------------------------RVILAYDINCQYSRRFFERFRA--------------------------------
CC1G_03480_Ccin_169851554              FDAGGVMAAV-C-RHDIPLFAANID-TAGEQQ-KYAVALIEKIASLMPPNA------------------------------------TLVVLYDVGCVLDRSTQLHPIM--------------------------------
CC1G_10453_Ccin_169868186              CDVTGIGAIA-CMRHGCFVPGAMVDFQKGERQINMDYAWLKALKMSGLDGVR-----------------------------------WVLLVYDINCQYCINFMTRIKKYR------------------------------
LACBIDRAFT_298476_Lbic_170101108       MEITGIVNIQ-C-SHVFIKS--SVDLQLGERFANTDFALAHAIRQRRMVMENIDLDFVTSC--------------------------DHLISYDISCHYWRHIEKRFAKN-------------------------------
LACBIDRAFT_397168_Lbic_170117184       FADTGLMAML-CRHDHVLWL--VNMTTAGERQ-YYALVLIKMLFMNLPLSM------------------------------------TVGILYDIGCQLERSCQIWGFL--------------------------------
LACBIDRAFT_299636_Lbic_170102807       YAVSGTVIVI-CSRHALIRRNGAGDLQKGEKYCNVDFIIFAALVGVMLP--------------------------------------WIFITYDIGCQWSTNLQSRMSE--------------------------------
LACBIDRAFT_331097_Lbic_170108453       YAVSGTGIVI-CSRHALVRRNGTGDLQKGEKYCNMDYLIFSALVGICLP--------------------------------------WIFITYDIGCQWSINFRSRMSD--------------------------------
LACBIDRAFT_323718_Lbic_170091880       LRSTGVGMAL-CARHEIVRGGGVADLQKGERYCNMDYILLSALAPLLIA--------------------------------------SVFVSYDIACQFKLKFDERVDK--------------------------------
LACBIDRAFT_304377_Lbic_170107223       FAVTALMALL-C-RHDRVIW-IVNMRSAGEKQ-HYALVLVETLFQHLPQHY------------------------------------TVGLLYDVGCQLHKSCVNWGFL--------------------------------
LACBIDRAFT_325386_Lbic_170095241       CDATGIAAFA-CPRSGAFIPGSVVDFQKGERQLNINYALCEALKNMGLNGLD-----------------------------------DIMIIYDVMCQYHVNLWDRIAN--------------------------------
Ot15g00710_Otau_116055094              MDERSMLLFV-C-HHGFVLRGTALVSTTHERFDMYDVALAGLARWRN----------------------------------------LGDVHIDFACQYGVHLRRRVEALVERDQNSFADLR-------------------
CHLREDRAFT_190650_Crei_159469221       KLVTAAGVVL-C-RHGMLLR--LVNLFHGERHMYATAAAASILAAGTA---------------------------------------VQFWWYDIACRWSKSWQKWHQA--------------------------------
im_7159298_Drer_189534053              LDEEGVEIAV-C-RHGFLLK--GLNMYRGEIFAYPMYLQKEFKD-------------------------------------------AKFLAMDVTCRYVPYLEKVSEA--------------------------------
NEMVEDRAFT_v1g245738_Nvec_156369699    LSETGLEIAA-C-RHGLAQN--AINMMYGEIYGYAHYLQLNHFIPKG----------------------------------------VKFLWYDVVCLYWPWLRRNE----------------------------------
BRAFLDRAFT_68171_Bflo_210127259        LDVTGVFGGV-C-RHDIPLK--FANMFHGERFGYPVYLIKTLLDEASARNN------------------------------------RLRVIYDVACSLTAHLTNSEHDGS------------------------------
BRAFLDRAFT_121844_Bflo_210117047       LDVSGVFGSV-C-RHEIPRC--FLNMIHGERFAYPVCLIEELLADVAGRNI------------------------------------KLKVIYDIACKLHSHLKKTRQ---------------------------------
LACBIDRAFT_297951_Lbic_170100409       RRRTHNEQTLVR-PCGVIYA--RATMYGAEAVSNFLIMVKNAFSVPGAQK-------------------------------------PDHIFYDTNCDAKQQAMNDS----------------------------------
LACBIDRAFT_308276_Lbic_170110907       RRRTHNEQILVA-PCGMILA--RETFYGAEAIATCAEFVKRTFSINHQK--------------------------------------PNHLFFDNNCSLSKHVKGDP----------------------------------
LACBIDRAFT_303619_Lbic_170105878       -------QILVA-PCGMIIA--RETFYYAESITAVVEMIKRTYR-GGFM--------------------------------------PNHIIFDNNCSVAKHVKDDP----------------------------------
CC1G_02394_Ccin_169848471              RAWTHNEQILVC-PCGIILA--RETFLFAESMPACADFIERTFK-QHPF--------------------------------------PDHIIFDNNCNLAKHVKDRD----------------------------------
LACBIDRAFT_308447_Lbic_170089507       RRQTHNEQLIVR-PCGMIVA--RATFYGSETVPQTVAMLKQVFRAPGSM--------------------------------------PDYFIYDNCCGVYNHLQSTNDP--------------------------------
LACBIDRAFT_303252_Lbic_170105635       PATFYCVETICY-PCGVVEA--WAKFAKSESESNILAFMKKVYPTKESH--------------------------------------PDYICIDKACKVLKHMAAQGH--WDE----------------------------
LACBIDRAFT_331992_Lbic_170110479       -----DDELIMR-PCGIILS--RVTFFGSEAISAVNTFVKATFPTPEST--------------------------------------PKYFIFDNNCKLVAHQRVIGD----K----------------------------
CC1G_01061_Ccin_169852634              RRLTHNEELCVA-SCGVILG--RATFYGSEAPNGVRHFLMKLFPTKESL--------------------------------------PRVIWHDNNCKIRSMLENDPDPHLRN----------------------------
LACBIDRAFT_333001_Lbic_170112840       HKHTHNEELCVA-SCGVILG--RATFYGSEAPNGVRTFWMKLFPMKASL--------------------------------------PRVMWHNNNCCIRAMLENDSDPFLKH----------------------------
LACBIDRAFT_299622_Lbic_170102789       PATFYCVETICC-PCGVVEA--WAKFAKSESESNILAFMKKVYPTKESR--------------------------------------PDYICIDKACKVLKHMAAQGH--WDE----------------------------
LOC100151444_Drer_189522959            CGSSGGCSVVMC-PHGVVYS--LKFNLKLESPRDFVDLLLSWQHL------------------------------------------PNVTICDSPGDLATHANTRTPTFHPHEGRLADATLENI--SKANQETLRVSLPW
NEMVEDRAFT_v1g139191_Nvec_156353224    ASSSGGRLTKTC-PHQVVTG--SKFLLRQESARDHVDLLQACRYW------------------------------------------PPVYIADAACSVTQHVSFRYPSL---ANQLWGDSLGCFQVPDHYREPQEVSCPE
LOC479330_Clup_73954202                PHLTGGKIYKVC-PHQVICG--SKYLVRGESARDHVDLLVSSRHW------------------------------------------PPVYVVDMATPVALCADLCYPEL---TNQMWGRNQGCF--SSPAEPPVSVSCPE
smf_Ecab_66710515                      PHLTGGKIYKVC-PHQVVCG--SKYLVRGESALDHVDLLVSSRHW------------------------------------------PPVYVVDMATPVALCADLCYPEL---TNQMWGRNQGCF--SSPTEPPVSVSCPE
A630042L21Rik_Mmus_145279194           AHFTGGKIYKVC-PHQVVCG--SKYLVRGESARDHVDLLASSRHW------------------------------------------PPVYVVDMATPVALCADLCYPEL---TSQMWGKNQGCF--SNPTEPVVSVSCPE
LOC361335_Rnor_109507362               THFTGGKIYKVC-PHQVVCG--SKYLVRGESARDHVDLLASSRHW------------------------------------------PPVYVVDMATPVALCADLCYPEL---TSQMWGKNQGCF--SNPTEPVVSVSCPE
LOC100028908_Mdom_126291205            SHLTGGKIYKVC-PHQVICG--SKYLVRGESARDHVDLLTSSRHW------------------------------------------PPVYVVDLATPVALCADLCYPEL---TSQMWGKNQGCF--SNPTEPPIYMSCPE
```

```
LOC416278_Ggal_118097475          AHHTGGKIYKVC-PHQVVCG--SKYIVRGENARDHVDLLVSSRHW-----------------------------------------PPVYVVDTASSVALCADIYCPDL---TAQLWGKNQGCF--SDPMDPPAYVSCPE
MGC115244_Xlae_148236655          SHLTGGKLYKVC-PHQVVCG--SKYIVRGESPRDHVDLLASSRHW-----------------------------------------PPVYVVDMATQVSLCADLCYPNL---TQQMWGKNQGCF--TDPSQTPKYVSCPE
hbcx_X(Si_187608129               SHLTGGKLYKVC-PHQVACG--SKYIVRGESPRDHVDLLASSRHW-----------------------------------------PPVYVVDMATQVALCADLCYPNL---TQQMWGKNQGCF--SDPLQTPKYVSCPE
LOC568530_Drer_189542986          DLLTAGRITKVC-PHQVVCG--SKYIVRRESARDHLDLLLSSRFW-----------------------------------------PPVYVADCAQKVALCADVSYPQL---AKQMWGENQGCF--SDPTAAPQYVSCSE
SMF_Hsap_160112910                RHFTGGKIYKVC-PHQVVCG--SKYLVRGESARDHVDLLASSRHW-----------------------------------------PPVYVVDMATSVALCADLCYPEL---TNQMWGRNQGCF--SSPTEPPVSVSCPE
LOC509419_Btau_119895718          PHLTGGKVYKVC-PHQVVCG--SKYLVRGESALDHVDLLVSSRHW-----------------------------------------PPVYVVDMATPVALCADLCYPDL---TNQMWGRNQGCF--SSPTEPPVSVSCPE
im:7159298_Drer_189534053         QLSIETLFVSIC--QKKHYL--YRQNDRNKRRQKITQKIAQEKKRLVAEIQRYNQQ------------------------------PNADPVDTNLVVQQLSNKAEE------IMIWPWQEQNT--DGVSIITKKK----
LOC100149527_Drer_189535301       ---RGGWAVIMC-PCGIVYS--IKCNLRAESPRNFADMLLSWKHL-----------------------------------------PNVVIYDFARGLSTHTNLRQPITFTPFEGRLLEPTTENIAKAKSGKLQVTLPWL
CHLREDRAFT_178766_Crei_159484947  KNLSPGLFVVHCLDCGGCLG--FHMLENAESPRALFEVLYTRWPEA----------------------------------------PKVVVYDNNCHAHAYFLNREPEW------------------------------
CHLREDRAFT_194780_Crei_159488509  KNLSPGLFVVHCLECGGCLG--FHMLNNAESPRALFEVLYTRWQDA----------------------------------------PEVVVYDNNCHAHAYFLNREPTW------------------------------
CHLREDRAFT_174339_Crei_159473346  GPRTGGIFTVFC-EHGVCYA--FFVLPRAEGRNEMYSWMVSFLPRA----------------------------------------PEVVVYDFACSLHEYCLNRAPAF------------------------------
LACBIDRAFT_307726_Lbic_170110050  QNLTGGLMCVWC-THSICYG--FHCIPNGEGRNDVFSPIITRWVKA----------------------------------------PKRIIYDFACALGPYCMTREPLF------------------------------
LACBIDRAFT_307932_Lbic_170110668  GSLTGGLMCVWC-PHSVCYG--FHNIPKGEGRNDVFSAIVTRWEKA----------------------------------------PKRIIYDFACALGPYCMTREPQF------------------------------
CC1G_02459_Ccin_169851012         QRLTGGLMVAWC-THGICYG--FHCIAESEGRDDVFSAMVTRWPVA----------------------------------------PKRVVYDFACSLGPYCMLREPEF------------------------------
CNA07690_Cneo_58259409            -PVVTHVIDTFK------QG--PENLALPETFLDLVGVMNARSQKVLR--------------------------------------DLQERADYTCQLPGELLGKARQL------------------------------
BRAFLDRAFT_131287_Bflo_219494211  PSLTPGVFTLFC-EHAICYG--FEVMLQPESPKYPFRILKSRFPTA----------------------------------------PDIIIYDNACQLHSYCLNRDPCF------------------------------
BRAFLDRAFT_95599_Bflo_219471002   PSLTPGIFTLFC-RHGICYG--FDCMTSCESPRHPYQIFRTRFRKA----------------------------------------PNVIVYDNACKLHQYALNREPHF------------------------------
BRAFLDRAFT_91571_Bflo_219462506   WGASGGLLVAAC-PHRVVYA--AKFLLRGESPRDFVDILLSMQWQ-----------------------------------------PTVTIADMPQIIAAHGNGRCPGMFQPHGGRLAEPTPQNIELVTSGNFRKELPCL
consensus/85%                     ...s.......C..p.hh.......h..uE.......hh.p.b...........................................h.hD.ss.h...h..p.................................

Species abbreviations: Bflo : Branchiostoma floridae; Btau : Bos taurus; Ccin : Coprinopsis cinerea; Clup : Canis lupus; Cneo : Cryptococcus neoformans; Crei : Chlamydomonas reinha

Identical hits in Laccaria
170103707  170103709
170094354  170120007
170114173  170114177
```

Back to Contents

2f. Multiple sequence alignment of the HMG domain found in the vicinity of the 2OGFeDO containing transposon

```
Domain boundaries               <-------------------------------?--------------------------------------------------->|-----------------------<HMG domain  begins here--------------
Secondary_structure_prediction  --------------E---E--HHHHHHHHH-HHHHHHHHHHHHHHHHHH-----HHHHHHHHH-----------HHHHHHHH-|H---HHHHHHH-----------------HHHHHHHHH-HHHHHH---------------
CC1G_02072_Ccin_169855339       MARQTSNK----KLKF---KLAANHIEYLSKSQAEWYTAKATRSTAQFV------RKRRRRIVSML-EETSGEAPDPTMKASI-|EEAVAAWFKANCRPRNKAAKI--GTKARSGREVFAHME-WKEVARKKRELMD---------
CC1G_02773_Ccin_169842922       QGSPAD-PAS--SVVF---TFMPSQVAFMRE-ALEEYRELTGAGLKADFS-----ANISATLLQEI-HEVSGVKPSDRRKALL-|EQAVAQWFEENGKVDIRKPKRV-WAFRWTARRVFGSIH-SRIVKFVAKLLLD-------KK
CC1G_03995_Ccin_169848799       MSQPSQ-G----AITFHGWKLTPPQIEILEGLRERYLLALHEPHKVQKVAKAGVQREAVASIRSLH-ETANGKALSTNEKIAL-|TRAVEKWFALYCKKRKAKEKA--GTRDWSVRDVFVRETPSEEIQEKQMELWE---------
CC1G_05495_Ccin_169864670       MARTK-------AEQY---PWTPDDLKLLND-AKAEYKGADAVKRLNVA------DTVARQIQEAKE-AVDGSAMEGKLAQAL-|FTSVRGWLNARCKASRKTPGP--AWKGVKPRLLFIQHN-KLTIKLAHRAKYEKLTGDRYPY
CC1G_05591_Ccin_169862973       MARQPT-R----KLKF---KLTSIHIEYLSRFQSEWYTAKATRSTAQFV------RKRRRRIASMM-EETSGEGADAAMKSSI-|EDAVGAWFKANCRPRAKAAKI--GTKARSGRDVFAHQE-WKEVAKKKRELME---------
CC1G_07037_Ccin_169850583       ------------LRGY---QLTASQVAYLRT-RVDQYRDKKKKAAKEKVA-----ASCARALKEDY-EQMVHQSMSSEMEKSL-|RKAVAHWLDQNALSTKGAPKHI-WGIRWHARLVFQHSN-ARAIMYLARVLAS------PNP
CC1G_07254_Ccin_169867894       VLQGP-------GNSY---QFTASQVAYLKT-RLPQYREGTSKRRDQLV------SQSAKNIISEL-YAITGKKQSREKKKLI-|TRAVKSWFRGRVDNSQVKAKKSVWGGRYYGQLVFKRER-TRIVQYLADIISR-------GD
CC1G_08112_Ccin_169860182       MARQKS-R----KVKF---KLAPNHIDYLSKCQSEWYTAKATRSTAQFI------RKKRRRIAGML-EEVSGETMGSSTKASI-|EDAIASWFKANCRPRGRVAKT--GTKARSGREVFAHRE-WKAIAERKRELME---------
CC1G_09453_Ccin_169867988       MAPSTN-KGS--PVSF---VLHPSQVAWLKT-KLDDYRAVKGKSKRRAFA-----LRIAQLLRNEM-NTISGANMSQAKEEQL-|NLAVQKWFESHKLDKHTSPRK--WTTRWSAHLVLKYRR-TRAITHLATLILN-------DR
CC1G_09747_Ccin_169868245       GVQGP-------GNSY---QFTASQVAYLKT-RLPQYREGTSKRRGQLV------SKSAKNIISEL-YAITGKKQSQEKKKLI-|TRAVKSWFRGRVDNSQAKAKKSVWGGRYYGQLVFKRER-TRIVQYLADIISR-------GD
CC1G_10129_Ccin_169845760       ATLEEH-S----RHGY---SFLPSQLAYMRE-KLGDYDNVQGKARRRFL------RSLTDHLCKEIANQNGGEEVSGDKRAAI-|GKGLRSWFRTNGTRKKREKKT--YGTNWYTRLAYMKVH-SRKVNALAEQIVK-------GG
CC1G_10218_Ccin_169857280       EGETPD-GAAAKANGY---EFTASQIAYLKT-RLPQYREGSVKRRGELV------TMSAKNIVSEL-YAISGKHQSREKKKLV-|VQAVRAWFQKRLESSAVKGKKSVWGGRYYGQLVFKRER-TQIVQYLGNVIVC-------GH
CC1G_10454_Ccin_169868188       MARKKS-R----QTRF---KLSTNHIEYLSKSQSDWYTAKATRSTSQFI------RKKRRRIAEML-EETSGETIDASMKASI-|EEAIASWFRANCRPRGRIAKT--GTKARSGREVFAHKE-WKAIAERKRELME---------
CC1G_10664_Ccin_169852260       MARNSS-A----APIF---TLTRSQMAFLKD-RMQAYRNKKPSARPAYA------RKVAGELKEDL-EEITEGPLSQERQKAL-|RKAVARWFVVNFKYEKGPHKRP-WGIRWHTRLVFQYER-SRAILYVAKMLAE-------NK
CC1G_10773_Ccin_169863827       TASDVG-A----VTGY---TFTKSQLDFMRD-RVDAYVGTASKKRRRFF------KRICDHLGQKI-EKERGKPMLQSEEKKL-|LQGIRTWFKAQTKTTRKSKDD--YSTNWYPRLVYMKSK-RRRTKAMAKRMLK-------GD
CC1G_11470_Ccin_169856599       MAPGK-------QHVR---TWTEDQLKVLND-AKREFHSADAVKRQAVA------DAAAMEIQAALE--GVGERLDSKASQAL-|FTAVRGWLKARCKAKRKADAV--TWKGVKARNVFIRHN-RRSIKETQKALYEQGTGAIFPY
CC1G_11680_Ccin_169864010       ------------ASPY---ALKISQVALLKS-KLDTYRDLKGKARKRFA------HKLAQILRNEI-QDLSGTAMPQAKHEAL-|KRAVTEWFKQNVQRIRGNGKR-MWGIRWHGRLVFQYER-TRAILHIARLLVD-------NA
CC1G_12552_Ccin_169869585       SAEHPT-P----VNGW---TFTPMEQAYMRE-KLEGYNASSGKQRRRFL------RSTTQYICQEN-EKRTGANLTEAEKSAI-|GKALRAWFKAHGNARKTEKNL--YGTKWYGRRVYMRIH-KAKVQALADEIVK------DGD
CC1G_12616_Ccin_169865988       MSSAQF-ATN--PEVV---HLRATEVAYLGS-KLEDYRDAKGKTGKAAIL-----EQCLRTLKGER-EVIKGEPMSRQAASDL-|KAAIEGYIQRNLHTVRGPQRR-PWGIRWTTRLVFQYHH-SQTIMYLAKLIVN-------QK
CC1G_12696_Ccin_169850600       MAPGK-------QPVR---TWTEDQLKVLNN-AKREFHTADNVKRQAVA------DAAATEIQAALE--GAGERFDSKASQAL-|FTAVRGWLKARCKAKRKADAI--TWKGVKARNVFIRHN-RRSIKETQKALYEKATGDIFPY
CC1G_12707_Ccin_169869472       SAVQLT-S----INGW---TFTPMEQAYMRE-KLEGYNTSSGKQRRRFL------RSTTQYICQEN-EKRTGNDLTVAQKSAI-|GKALRAWFKAHGNARRTEKNL--YGTKWYGRRVYMRIH-KAKVQALADEIVK------DGD
CC1G_12859_Ccin_169866922       MAPKK-------QPQH---TWTEEQLKILND-TKPEFHAADAVKRQAVA------DTAASKIQAAM--EEGGQQLDSKASQAL-|FIAVRGWLKARCKAKRKADGI--SWKGIKARNVFIQHN-RRSIKAVQKKLYEKDTGAVFPY
CC1G_12891_Ccin_169869841       SATDFG-V----VAGY---TFTESQLAYMRE-HVEAYIGTAAKKRRRFL------KRIANHLRQAI-EDRNKEPISRAEESKL-|LQGIAAWFKAQTKTSRNAKDQ--YSTLWHPRLVYMRSK-RRKTKALARRMLK-------GD
CC1G_12942_Ccin_169866812       ------------TRGY---TLRDSQVALLQQ-RLEAYRRKRKNKKDKVGRAQVV-RDAVRVLKEDY-ESLARELMSKSHEARL-|KAAVTKWFELNVKYERGPNKK-IWGTRWSARLVFQHER-SRAIVYLAKLLSQ------PQP
CC1G_12950_Ccin_169862398       VARRDD-EDTV-PDGY---QFSASQVAYMQS-RLQQYREGSKGKRSDVV------VAVFKNFVSEI-INESRRPVPKAKRKAI-|KRGVLGWFKQKLAAYKPKQEKAKWTTRWFGLLVFKRER-RDAIKYLAGLLAR-------GE
CC1G_13090_Cicn_169862006       ------------LKAY---SLSVTQTAYLKT-KLEAYRMKRKRKDDRKGRGLVV-KEACRRLKEDW-EAMTELEMSPELEKRM-|KRAVNKWFEENIKYERGPTKK-PWGMRWTARLVFQYER-SRAIGFLAKLMAQ------PAP
CC1G_13128_Ccin_169862411       MAPKK-------PPPM---VWTDDQLKVLND-AKAEFRAADAVKRQIVA------DNIANKLQAEME-KAGGVPLDRTASQSL-|FVAVRGWLNARCKAKRKAEAI--SWKGVKPRILYIRHN-RRTIKEKQKLLFESATGEAFPY
CC1G_13284_Ccin_169869997       MSPPVD---S--TVTF---TLQPSQVAFLTN-KLDEYRAMGAAVKKQGYA-----HRMAQHLKKEM-NAMGATNMSQAREDQL-|KLAVERFFEERKYERSPTKRK--WGIKWHARLVFKYVR-TRAIWHLAKLILA-------NL
LACBIDRAFT_297419_Lbic_170118976 TRQPRP-P----TPGR-K-PFSSEESAILRQ-YLEDFRVKTKEERKKLLM-----YEVYRLMKAKA-----PPLLTEEQWKER-|KTKIKEWFQNHGRCIRMPKFG--YVKLISTRYVVAHLR-QEQIHEVVTRLAK---------
LACBIDRAFT_298016_Lbic_170100467 ETNEYT-A----AAPY---VWETWQTEILNH-YLPFYKANKTVKQRTALL-----NKTFRRMQSK-------RDIPDSDVSGL-|RKAMRQWYGLKAKVRQPKQYK-IRGVNWNFRTVAALLK-KEEIKQAEVELAGGQPV-----
LACBIDRAFT_335151_Lbic_170117741 TRRPRP-P----TPGR-K-PFSSEESAILRQ-YLEDFCVKTKEERKKLLM-----YEVYRLMKAKA-----PPLLTEEQWKER-|KTKIKEWFQNHGRRIRMPKFG--YVKLISTRYVVAHLR-QEQIHEVVTRLAKGVPPGSRV-
LACBIDRAFT_316530_Lbic_170117428 KTVEHV-A----PPAH---IWEDWQTEILEK-YLPEYRSKKSAGARVAIM-----GKVLQKFKAR-------KESAEGDTANL-|RKAVQQWFGVKSRIRKPKEIK-IRSMNWNFRTVASVQK-KMEIKEAQRTLSGNGDI-----
LACBIDRAFT_331947_Lbic_170110305 MGGEHV-T----PTPH---IWEDWQIDILER-YLPDYKSKKTVGAWVAIM-----GKVLTKFKGR-------SESGDGGGAKM-|---RKQWFALKSRVRKPKEIKI-RSMNWNFWTVAAVQK-KLEIKDAQQMLSGDGEI-----
LACBIDRAFT_332251_Lbic_170111125 TGQRRQ-V----ERGR-E-HFSATEEKVLRE-YLDDFTTKSKEERKDLLM-----FKIYRLIKAA-----GPE-LTPDQWRIR-|KKKVKEWYYNHGRRPHVRRYG--YVKPISLKYVIGHLK-QEELYSKVVELAG---------
LACBIDRAFT_329028_Lbic_170103711 TRRPWP-P----TPGR-K-PFSSEESAILRQ-YLEDFRVKTKEERKKLLM-----YEVYRLMKAKA-----PPLLTEEQWKER-|KTKIKEWFQKHGRRIRMPKFG--YVKLISTRYVVAHLR-QEQIHEVVKRLAK---------
LACBIDRAFT_327601_Lbic_170100513 SQTKHH------GSGR---HWTDEESALLEK-HREEYRDANDAARAEIF------SRVLQDMLDIL---PNGKSFKKEERSAV-|KKDIKEWFARYSRKRSQKMLR--MGKSWHARRVFQHEN-KDAIDTEAKRLCAEAIEEGQK-
LACBIDRAFT_330019_Lbic_170106331 SQTKHH------GSGR---HWTDEESALLEK-HREEYRDANDAARAEIF------SRVLQDMLDIL---PNGKSFKKEERSAV-|KKDIKEWFARYSRKRSQKMPR--MGKSWHARRVFQHEN-KDTIDTEAKRLCAEAIEEGQK-
LACBIDRAFT_330403_Lbic_170107125 SQTKHH------GSGR---HWTDEESALLEK-HREEYRDANDAARAEIF------SRVLQDMLDIL---PNGKSFKKEERSAV-|KKDIKEWFARYSRKRSQKMPR--MGKSWHARRVFQHEN-KDTIDTEAKRLCAEAIEEGQK-
LACBIDRAFT_330704_Lbic_170107817 SQTKHH------GSGR---HWSDEESALLEK-HQEEYRDTNDAAHTEIF------SRVLQDMLDIL---PNGKLFKKEERSAV-|KKDIKEWFAQYSRKRSQKMPR--MGKSWHARRVFQHEN-KDAINTEAKRLCAKAIEEGQK-
LACBIDRAFT_316727_Lbic_170117695 SQTKHH------GSGR---HWTDEESALLEK-HRQEYRDANDAARAEIF------SRVLQDMLDIL---PNGKSFKKEERSAV-|KKDIKEWFARYSRKRSQKMPR--MGKSWHARHVFQHEN-KDAIDTEAKRLCAKAIEEGQK-
LACBIDRAFT_330006_Lbic_170105886 SQTKHH------GSGR---HWTDEESALLEK-HQEEYRDANDAARAEIF------SQVLQDMLDIL---PNGKSFKKEERSAV-|KKDIKEWFARYSRKRSQKMPR--MGKSWHTRRVFQHEN-KDAIDTEAKRLCAEAIEEGQK-
LACBIDRAFT_329656_Lbic_170105236 SQTKHH------GSGR---HWTDEESALLEK-HREEYRDANDAACAEIF------SRVLQDMLDIL---PNGKSFKKEERSAV-|KKDIKEWFAWYSRKRSQKMPR--MGKSWHARRVFQHEN-KDAIDTEAKRLCAEAIEEGQK-
LACBIDRAFT_335133_Lbic_170117719 SQTKHH------GSGR---HWTDEESALLEK-HRQEYRDANDAARAEIF------SRVLQDMLDIL---PNGKSFKKEERSAV-|KKDIKEWFARYSRKRSQKMPR--MGKSWHARHVFQHEN-KDAIDTEAKRLCAKAIEEGQK-
LACBIDRAFT_307132_Lbic_170120119 SQTKHH------GSGR---HWTDEESALLEK-HREEYRDANDAARAEIF------SRVLQDMLDIL---PNGKSFKKEERSAV-|KKDIKEWFARYSRKRSQKMPR--MGKSWHARCVFQHEN-KDAIDTEAKRLCAEAIEEGQK-
LACBIDRAFT_335227_Lbic_170117861 SQTKHH------GSGR---HWTDEESALMEK-HREEYRDANDAARAEIF------SRVLQDMLDIL---PNGKSFKKEERSAV-|KKDIKEWFARYSRKRSQKMPR--MGKSWHAQHVFQHEN-KDAIDTEAKRLCAEAIEEGQK-
LACBIDRAFT_335987_Lbic_170119643 SQTKHH------GSGR---HWTDEESALLEK-HREEYRDANDAARAEIF------SRVLQDMLDIL---PNGKSFKKEERSAV-|KKDIKEWFARYSRKRSQKMPR--MGKSWHARCVFQHEN-KDAIDTEAKRLCAEAIEEGQK-
LACBIDRAFT_335999_Lbic_170119682 SQTKHH------GSGR---HWTDEESALLEK-HREEYRDANDAARAEIF------SRVLQDMLDIL---PNGKSFKKEERSAV-|KKDIKEWFARYSRKRSQKMPR--MGKSWHARRVFQHEN-KDTIDTKAKRLCAEAIEEGQK-
LACBIDRAFT_336286_Lbic_170120560 SQTKHH------GSGR---HWTDEESAVLEK-HREEYRDANDAARAEIF------SRVLQDMLDIL---LNGKSFKKEERSAV-|KKDIKEWFARYSRKRSQKMPR--MGKSWHARRVFQHEN-KDAINTEAKRLCAEAIEEGQK-
2ctoA_Maelstrom  \ HMG domains   ------------------------------------------------------------------------------------|----------SSGSSGMPNR--KASRNAYYFFVQEKI-PELRRRGLPVA----------
1K99_Hsap_17942547| in the      -------------------------------------------------------------------------------------|-------------------P--KKPLTPYFRFFMEKR-AKYAKLH-PEM-----------
1i11A    7        |  PDB        -------------------------------------------------------------------------------------|-----------------phI--KRPmNAFMVWAKDER-RKILQAF-PDm----------
2lefA   40        | database    -------------------------------------------------------------------------------------|-----------------MHI--KKPLNAFMLYMKEMR-ANVVAES-TLK----------
1gt0D   59        |             -------------------------------------------------------------------------------------|-----------------DRV--KRPMNAFMVWSRGQR-RKMAQEN-PKM----------
1j46A   54        |             -------------------------------------------------------------------------------------|---------------mqDRV--KRPMNAFIVWSRDQR-RKMALEN-PRM----------
1aab    33        |             -------------------------------------------------------------------------------------|------------gkgdpKKP--RGKMSSYAFFVQTSR-EEHKKKH-PdaSV---------
1e7jA   37        |             -------------------------------------------------------------------------------------|---------------msDKP--KRPLSAYMLWLNSAR-ESIKREN-PGI----------
1j5nA   38       /              -------------------------------------------------------------------------------------|----------trkkkdpNAP--KRALSAYMFFANENR-DIVRSEN-PDI----------
Species Abbreviations: Ccin : Coprinopsis cinerea; Lbic : Laccaria bicolor
```

Back to Contents

2g. Multiple sequence alignment of small alpha helical domain found either in the neighborhood or fused to JBP and/or the transposase

```
Predicted secondary_structure    --HHHHHH---HHHHHHHHH---HHHH----------HHHHHHHH---HHHHHH--------------HHHHH-----------HHHHHHH----H---EEE--HHHHHH-------------HHHHHH---------------
CC1G_09736_Ccin_116498199        -MQCLIDY---ASSRIRAAG---MVWNRHGDSFS---PTERKHFAA---QLYRDL--------------ETGML-----------LYSSGCFRYP----GIIQRPVALGMMFKSTLTG-MLHLGRHASTADYETPNPIPDP--VLP
CC1G_07252_Ccin_169867890        -MQCLIDY---ASSRIRAAG---MVWNRHGDSFS---PTERKHFAA---QLYRDL--------------ETGML-----------LYSSGCFRFP----GIIQRPIALGMMFKSTLTG-MLHLGRHASTADYETPKPIPDP--VLP
CC1G_10219_Ccin_116503602        ---------------LLTSE---MEWVPNFFAFK---EEECKAWA----TLTRDLPGHRAYANR-----QANMWRLMGAHAEKAFGEVKTKGRNP----RIIRRPIALGMLFKSTLSG-TLHLGRHGPMAEYETPEPVPDP--ILP
CC1G_11682_Ccin_169864014        -MQHLLNY---VSSRLRACASALVSLEDRWQYVT--PPGLKPEWWR---QLRVDL--------------NHGLI-----------ILASATLQHP----GKIVTPVFLEHLHRLLIYG-GIDSPQDVALVVTDYPRQAHDY--AVT
CC1G_12550_Ccin_169869581        -WQSLIDY---LSSRHRAAG---VFYVVQYPVIS---KEDRALFLS---ALFHEI--------------SMTAF-----------CLATATIKHP----HSIVAPKCFRDLFQMFYDN-EGQVDPLLIEIKVQDGEGERDR--VLN
CC1G_12678_Ccin_169869760        -AQDLLNY---ASGRLRSLA---IAITNDYGLISR-SPTLKERCWS---QFKQDL--------------NTTLI-----------LYGSMSLLHP----GKIWAPVFLDHVHRLLING-GINDFNAIAEGALNRPMHGFD---VFP
CC1G_13131_Ccin_169862417        -MQQFMYY---LSSRLRGAG---LAYAHQFAKLT---IDQRAEFLR---ILFEEV--------------TWVTT-----------LFGFAAARFP-----TIKPGDTHRELTSLVAYL-RFGDPMPILAACTANNEEVPDR--LLA
CC1G_13153_Ccin_169863832        -----------MSSRLRAAG---LSYLNLRDTIP---EEERMAFRG---SLCGEF--------------IKTMV-----------VLTFATVQGP----DSITPPPVFRKLFHICIAHPEVELDMMWLDVTVKGGEGGPDA--VLS
CC1G_12708_Ccin_169869474        -WQLLIDY---LSSRHRAAG---VFYVVQYAVIS---KEERATFLS---SLFHEI--------------SMTAF-----------CLATATIKYP----HSIVAPKCFRDLFQVFYDN-EGEVDPLLIEIKVQDGEGERDR--VLN
CC1G_12861_Ccin_169866926        -MQQLLYY---LSSRQRAAG---VAYAHQFAELN---MTQKAEFLR---IFYEEL--------------VWHTT-----------IFGFAAARFP-----RVKTGEAHRELVALLTLI-KLGDPMPIYSACKPSHDQVPDR--ILA
CC1G_12949_Ccin_169862396        -TRGLPGH-----RALRACG---VAWNERYPALT---VEKRRAFIN---QFADDL--------------HSAMM-----------LFSSLALRFN----HHVSRPNALTKLTKVALSL-HLTVPS-IVAEAVSVEQEVSDP--ILP
CC1G_13116_Ccin_169866949        LIRCEMGWMLNFFDRKEKIC---LGWADMSGNRP---GHRAYAYRQ--AEMWRLLACEARRTELTAVHQINGMM-----------IYTAAAVRNP----HSITPPPAFRKIFPVCFGDVEGELDMWMLEVAVEGGEGLRDA--KLS
CC1G_05496_Ccin_169864672        -LQQFLYY---LSSRKRSAG---LAYGSHFAQLS---IDQRAEFLK---ILFDDI--------------TWHTM-----------IFGFASLRYP-----EVKPGEAHRNLGRLFMHI-RYGDPTLIVAACSPSYQNVPDR--LLA
CC1G_10425_Ccin_169866913        -FNSLLHY---LGLRV-------LASAQQYLATS---EPEETVFRE---GFEAEL--------------QALVV-----------VIYCAKMRFP-----GITPPEFLQRLLLINLDE-CYGPLCILDGVLASLPDDIDVS--AAD
CC1G_09451_Ccin_169867984        -AQDLLNY---ASARLRHLA---IAITTPHGLIDSDPGLKEKCWE----QFKVDL--------------NSTLI-----------LFGSMSLLHP----GKVWAPVFLDHAHKLLING-GITSPQNVAEVALHRPLHGFD---VFP
CC1G_11469_Ccin_169856597        -MQQFLYY---LSSRQRAAG---TAYAHMFADLD---MNQRAEFLK---IFYEEI--------------VWHTT-----------LFGFAATRFP-----KINPGEAHRELVTLLTVL-KQGDPMQILAACKPSHGRVPDR--ILA
CC1G_11998_Ccin_116510369        -SQTLLDY---ASSRLRHCA---ETLTGRSESIKKTPGLFESCWN----QLKIDIRSA-----------NTTLI-----------LQSIVNITHP----DVITPPSFLVKITRLLVSG-GIDSLQDVASAARDRQPPLRDH--FRP
CC1G_12695_Ccin_116506927        -MQQFLYY---LSSRQRAAG---TAYAHMFADLD---MNQRAEFLK---IFYEEI--------------VWHTT-----------LFGFAATRFP-----KINPGEAHRELVTLLTVL-KQGDPMQILAACKPSHARVPDR--ILA
CC1G_13326_Ccin_169870064        -AQKLVNY---ISARLRHCA---THLTDRAEDIDRVPGLFAACWH----QFKLDF--------------NHGLL-----------VYALVSLQYP----ETISAPAFLKSLHRLLVEG-GIDSCDAVKREANRHDPPYDDF--GVP
LACBIDRAFT_316487_Lbic_170117410 -MEWRVYH---FFSRCKAAW-S-AYSDDKVNWID---DKERQEWWK-LRILHETNRLCMEV-------GKAYCLG----------GVLDLLPWAP----ALLNATNPARWIIKPLSPV--------VLRAQYGPKDQAP----ILA
LACBIDRAFT_333145_Lbic_170113173 -MQMLVYM---MSSRCRAAT---KAWKERPNMSD----NEKEIWLS---LFRGEL--------------DFIWK-----------ALGSHAIEEF-----TYDIPMFLQACVKEYGPP----------DIDWAVVQTAPDC--LVP
LACBIDRAFT_325901_Lbic_170096750 -GQWCIYY---FASRCKAAW---KAFEDHKDDWR--NDPQKHQWWT--IGLRLEG--------------DRLAK-----------AIGYHVAAGF-----IVKVPSWVSFFILRTHPQ-QSDLASLPQQGLLVDPAYMDDPSNIFA
LACBIDRAFT_327556_Lbic_170100793 -GQWCIYY---FASRCKAAW---KAFEDHKDDWR--NDPQKHQWWT--IGLRLEG--------------DRLAK-----------AIGYHVAAGF-----IVKVPSWVSFFILRTHPQ-QSDLASLPQQGLLVDPAYMDDPSNIFA
LACBIDRAFT_330405_Lbic_170106947 -FQEVLCY---WSSRFRA-----ILTSKLENPID---PKFENCWLE---YMSAEL--------------GFIIA-----------TLGMACAYGH-----KFIVPPFLKSCLQGNPLH--------------PQEAFMDDC--IRP
LACBIDRAFT_306696_Lbic_170108481 -FQEVLCY---WSSRFRA-----ILTSKLENPID---PKFKNCWLE---YMSAEQ--------------GFIIA-----------TLGMACAYGH-----KFIVPPFLKSCLQGNPLH--------------PQEAFMDDR--IRP
LACBIDRAFT_304733_Lbic_170107823 -FQEVLCY---WSSRFRA-----ILTLKLENPID---PEFENCWLE---YMSAEQ--------------GFIIA-----------TLGMACAYGH-----KFIVPPFLKSCLQGNPLH--------------PQEAFMDDR--IRP
LACBIDRAFT_335145_Lbic_170117735 -FQEVLCY---WSSRFRA-----ILTSKLENPID---PEFENCWLE---YMSAEQ--------------GFIIA-----------TLSMACAYGH-----KFIVPPFLKSCLQGNPLH--------------PQEAFMDDR--IRP
LACBIDRAFT_316718_Lbic_170117691 -FQEVLCY---WSSRFRA-----ILTSKLENPID---PKFENCWLE---YMSAEQ--------------GFIIA-----------TLSMACAYGH-----KFIVPPFLKSCLQGNPLH--------------PQEAFMDDR--IRP
consensus/100%                   .............................................h......h..................................................h..s.....h...............................s
consensus/95%                    ................b.......................b...a......h...................................b................h..s.....h.........................s.....s
consensus/90%                    ..p........h..R.c..................p.....bp..a......h...b...............................hs.............h..s.h...h..........................D...h.s
consensus/85%                    ..p.hh.a...huuR.+..........p....hs.....bc..a......h..pb..............................hs..shb........h..s.h.p.h..........................D...hbs
Species Abbreviations: Ccin : Coprinopsis cinerea; Lbic : Laccaria bicolor
```

Back to Contents

2h. Multiple sequence alignment of Cys cluster that is often fused to the hydroxylase domain

```
Predicted sec str          --HHHHHHHHHHHHHHH-----------------------------------------HHHHHHHHHH--------HHHHHHH--
CC1G_10424_Ccin_169866911  IQIEARMDFW---------ATWV--GNIEEFRKEVNCGSPLCSV-K--CA-PLETHMSCSACFYDPSRYPSNCGLRDLYLRDLLRS---QAGLSPGQVERFMPAY
CC1G_12620_Ccin_169865996  LQIEVRMDFW---------SSWL--EQNADFGRHVNCGSPHCTQ-G--CT-PLDDFMACEACFSNGQQYPTNCGLRTQFLRDTLRS---RGGLTSRQITDFLPVY
CC1G_13092_Ccin_169862010  FQLEADYSRF---------DRWA--LQFDKGHTFPPCGNPQCSN-G--CSGPVDEGMWCLPCWLDHGAGFEKCKLVDEYLKQTLCE---EGYLTPETAPAFLEDY
CC1G_10224_Ccin_169857292  FQLQSRMEFW---------NTWA--SRVCNGDPTLNCGEEACHA-S--CD-ALLDHPSCVHCFQTPRGITGNCMLQEELIAHFLET---DCETPVEEIQPFLQEY
CC1G_10671_Ccin_169852274  FQLEGFFELW---------EEWAQALRDEDNDQKEGCGNPDCGD-TPPCD-PLLDDMACGLCALQGPEAYDQCGFFKQYMAYHLQH---DAGMPITSIPGFLEQF
CC1G_04135_Ccin_169860374  FETEADLGLW---------REWA---SLDRTGSIRDDCNDCEEGP-R--GN-PLHDYPACNSCF-NSDFGMDPCSFHSHYLEHMLMS---QFYTPVHEIPLVLAEY
CC1G_13157_Ccin_169863840  FQLQGRTDLW---------KAYT----AQLDHQQTQCDKPECQA-AGACE-IMEDYISCRACF----NDNEVCTLEYDYIGHYLNR---DVGIPWDDLVPPLTRY
CC1G_12621_Ccin_169865998  FSVISDNGLW---------CSWT--EALDQRGAIVRCSNTECQL-EG-CT-LLGDVLSCLKCFENDAIGPGRCDLPLLYLRRFMET---KDILPEEKVRPFVTWF
CC1G_10672_Ccin_169852276  LETEGSMQLW---------HLWT-ASRPVLLEEQPTCDDPECVN-GGGCD-PLRDYMSCIRCFDNDDGKHDTCSWRDQYLSVYLALN--NPHLSPEDLKCFLAHY
CC1G_12914_Ccin_169866419  MEAEQDLDAW---------RRYI--PQLEESHPVGVCKSDYCHDNHHECI-PSPHTMACQPCI----STGHQCSFREDFIAHRVAE---AMPWTAKQVAEFVDEY
CC1G_07247_Ccin_169867880  CHDRDRFEDW-FF-------DNIG-----------GPCNHEECIKFPGGCM-LTAEGLGCAPCI----KKGRPCPWKDAYIITVILG---ELEVPYGKAFQFFTKL
CC1G_09678_Ccin_169866372  YMSPHLRHLF-LLLLEK--KNWG--EFLLAHRPLSRCDALACKQHGSDCS-VGDRHLGCRTCI----RNNTLCSHVTDCLLSTLL----QDAIPVAQANELISTF
CC1G_09735_Ccin_169868221  CHDRDRFEGW-FF-------DNIG-----------GPCNQSQCIRFPGGCM-LTKEGLGCAPCI----KKGSPCPWKDAYIIAVILG---ELDVPYGKAFEFFTKL
CC1G_11350_Ccin_169866093  VEALQDQDRW---------NSWT--SQWYTQSNQESCGT--CNN-PDDCV-PIEDAFGCAGCN----VNGYECGFMEEYLKQVREDWKTGSRCCVGPVKEIWNDL
CC1G_04088_Ccin_169860280  FTQNRTLTAA----KAVA-DSTA--VSRSSYSANKVCDNPSCKAKPDSCH-VAPRGRGCLPCN----LALQPCSFVTDYIVQRLGLE--CGLQPDDAIAFLE---
CC1G_01240_Ccin_169852992  SCVRFLTDFL----LNP--KHWK--DIVAANPGLGPCSAPQCASKEVLCL--PSDVLGCTNCV----RLGLTCSHVDTCLIVFISA---KANIDTALARKFLDAL
CC1G_09972_Ccin_169869428  VWIPRIVRFV----LDP--TGWK--AFVASNPYLGQCTALQCSSDKVKCV--LSDEFGCTTCV----RLGLVCSHLETCLTLALAQ---KMNIDAERARQFLRIV
CC1G_10222_Ccin_169857288  LATIQKLVDK-ML-------KEY--------------CTQKECIESPGACM-LTEEGFGCGPCI----ARGATCPWKDAYIIAVIMG---ELDLSYGKAFEFFTKL
CC1G_11349_Ccin_169866091  VRQLFNNKEW---------NDYA----RDNPNPPDRCD-PDCGQ-GL-CR-PEVDLFGCIGCA----TRQTVCSFLRDFIERRLPSLL-HPALPNAVNDGFYQGQ
consensus/100%             ....................................C....p......s.....p...C..C..........C.....hh.............s......h....
consensus/95%              ....................................C....p......s.....p...C..C..........C.....hh.............s......h....
consensus/90%              h........h.........p.h..............Cs...C......C.....c..uC..C..........C.....hl...h.........s...h..hh...
Species Abbreviations: Ccin : Coprinopsis cinerea
```

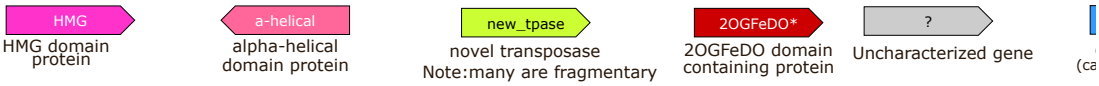
Back to Contents

2i. Gene neighborhoods of the predicted transposase gene associated with JBP

Note the neighborhoods are rendered in the svg format and your browser should support svg to view this. Most current versions of web browsers support svg. However, if that is not so and you see a white screen below without the genes, click here to see the neighborhood

OR if that too doesnt work, click here to access a pdf file

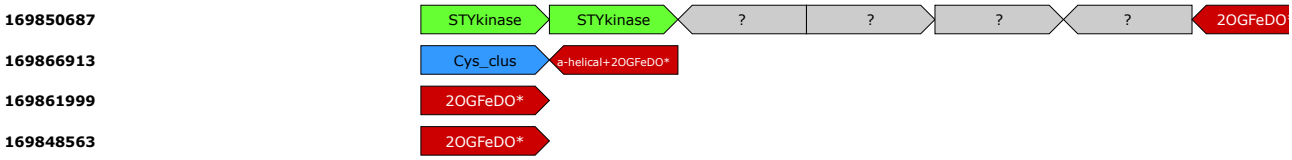
Legend for key genes:



Note:numbers on the left represent Gi numbers of the predicted transposase gene

Coprinopsis gene neighborhoods (click on the boxes to access protein sequences)

Solo 2OGFeDO like genes



2OGFeDO like genes in the vicinity of the predicted transposase gene

Back to Contents

3. Phyletic distribution, domain architectures and alignment of the algal RNA-modification associated family

Gi	Domain architecture	gene name	Len	Organism	class	GenBank Annotat
Aano1000001260	MBD+2OGfeDO	Aano1000001260	484	Aureococcus anophagefferens	stramenopiles	68521
Mpus1000007549	2OGfeDO	Mpus1000007549	1068	Micromonas pusilla	viridiplantae>chlorophyta	56524
Aano1000002837	2OGfeDO+X	Aano1000002837	426	Aureococcus anophagefferens	stramenopiles	62507
Aano1000005385	X+2OGfeDO+X	Aano1000005385	913	Aureococcus anophagefferens	stramenopiles	63084
Aano1000005432	Artificial N+2OGFeDO	Aano1000005432	4491	Aureococcus anophagefferens	stramenopiles	63122
Aano1000008654	LRR+2OGfeDO	Aano1000008654	488	Aureococcus anophagefferens	stramenopiles	65766
Ehux1000028467	LRR+2OGfeDO	Ehux1000028467	422	Emiliania huxleyi	haptophyceae	gm1.2000361
220972248	LRR+2OGfeDO	THAPSDRAFT_7450	548	Thalassiosira pseudonana CCMP1335	stramenopiles	predicted protein [Thalassi
Aano1000006794	dihydro-orotase-likeAmidohydrolase+2OGfeDO	Aano1000006794	1100	Aureococcus anophagefferens	stramenopiles	64773
Aano1000005614	Znf+PseUsynthase+2OGfeDO+Methylase	Aano1000005614	863	Aureococcus anophagefferens	stramenopiles	63253
219129852	ZnF+PseUsynthase+2OGfeDO	PHATRDRAFT_50200	870	Phaeodactylum tricornutum CCAP 1055/1	stramenopiles	predicted prote
217411154	ZnF+PseUsynthase+2OGfeDO	PHATRDRAFT_42626	798	Phaeodactylum tricornutum CCAP 1055/1	stramenopiles	predicted prote
220975450	ZnF+PseUsynthase+2OGfeDO+Methylase	THAPSDRAFT_21769	1310	Thalassiosira pseudonana CCMP1335	stramenopiles	predicted prote
219113839	CysSC+ZnF+RRM+2OGfeDO	PHATR_44207	601	Phaeodactylum tricornutum CCAP 1055/1	stramenopiles	predicted prote
116060216	ZnF+RRM+2OGFeDO	Ot10g03170	412	Ostreococcus tauri	viridiplantae>chlorophyta	unnamed protein
145351687	ZnF+RRM+2OGFeDO	OSTLU_93380	466	Ostreococcus lucimarinus CCE9901	viridiplantae>chlorophyta	predicted prote
Mpus1000003041	RRM+2OGfeDO	Mpus1000003041	462	Micromonas pusilla	viridiplantae>chlorophyta	38430
Mpus1000008227	Methylase+ZnF+RRM+2OGfeDO	Mpus1000008227	1380	Micromonas pusilla	viridiplantae>chlorophyta	58309
Aano1000003820	ZnF+RRM+2OGFeDO+AlkB	Aano1000003820	1794	Aureococcus anophagefferens	stramenopiles	66742
Aano1000010511	RRM+2OGFeDO	Aano1000010511	707	Aureococcus anophagefferens	stramenopiles	72852

```
Multiple sequence alignment
Boundaries                                                                                                         <-core 2OGFeDO begins here-----------------------------------------
Predicted_secondary_structure      -HHHHH-----------------------------------------------------------------------EEEEEE------------HHHHHHHHHHHHHHH-----------------------------EEEH
Aano1000005432_Aano_Aano1000005432   DDAALLQQYASQLGAAARCG---------------LGAARPTPNVVAA--------------------------RACVERMA----------EAGLLADAAAWKRLLRVAAS---------------------TPSLAKPDD
Aano1000002837_Aano_Aano1000002837   LRLIGDGNWPTQARPNV------------------TDTGKPVP--------------------------------GMCVGLVF-ALGQ-GAQASHVSECHPQMTRVLTRWCRGTLPRTRKGA-----------EFPFSSLQVN
Aano1000006794_Aano_Aano1000006794   FRLIKE--WPTQQRPNV------------------APGGYSNPVVE-----------------------------GMVLGLCP-NRSG-GCSVAQASQQCPQLTKVVTRWAAATLPDA---------------KFRYGSIQVN
Aano1000008654_Aano_Aano1000008654   LDALRRTAWTSTRERPRVDADHY------------LTLQRPRTKYYDADGN------------------------YNAAGQKA------QKNAAKLAKHRAVWDLAVAALAEVDAAY----------------AETFTSLAVT
Aano1000001260_Aano_Aano1000001260   LERLRAVDVPRTLARQNIIGEAD------------ASAGVR----------------------------------SACVGVVA-AR-SHGVMASSFSRSRPGLTRELVRFGKAHLPA----------------GFKFTSIQLN
Aano1000010511_Aano_Aano1000010511   LAALEASEIPVQDSRTNV-----------------TRPGAPAPR-------------------------------GLVLGAVN-VRGHGGMRVSGATHERPRLAKLLCDVFRAHAPDP---------------AFRFTTIQIN
Aano1000003820_Aano_Aano1000003820   RDVVKSG-WPASSAARSKLL---------------IEEGRGGG--------------------------------TFSVGVAP---NAHQPRQNALEALATAVFALEKSIIAEKFPR----------------RTPSTMCAVN
Aano1000005385_Aano_Aano1000005385   LTCLGD-KWPTQDRPNV------------------TPDGKPVH--------------------------------GMCLGAVF-VLGGVGMACSNVSEHYPELTTLVTKWVTTTLPE----------------DFPFSSLQIN
Mpus1000003041_Mpus_Mpus1000003041   IATLSTLKWPRASHRVGVASSEYLSLSSLGFE---RETEDRTSTV----------------------------ASPPSPGQNA----RRRRRMRRYRAVWELARDVLASVAGDDV------------------AAKFDGLAVT
Mpus1000008227_Mpus_Mpus1000008227   LRELVRARWPPVSHRPKVNSERYLVVAADVRGE--PRDGEETAE-------------------------------RARATNDV---------VDPFRDLKRAALAVLRWADPE----------------------YPCDHMAIT
Aano1000005614_Aano_Aano1000005614   RRASDSGRWPKTSLARSKL----------------IKPGDDAA--------------------------------SFTLSRGG----GGSRADALFPALADAVFALEEAIAPG--------------------RPPSTRCAVN
PHATRDRAFT_50200_Ptri_219129852      RKIISEGCWPETSIARSNVRRDDLRHDI-------SKDARAGGSFTVVNPTFRDG-----------------LYRDAVGANP-----LPRANRLFPDLVQAVFALEASLSEHNLKQANGDAGIGIHRKRGEPRPPSSHCAIN
PHATRDRAFT_42626_Ptri_217411154      RDADKSGFWPSTTPKRQLVMVSTVVEGKSESCDNPFESGAMPASLSVAHMKARSNKIDKSSAYIYVEGEGGASGSFSVGAMPGEACTQPKGNELFPELMRAAFELEIALCPN-------------------REPSSTIAIN
PHATR_44207_Ptri_219113839           LQILITLRWPAQN----------------------ERQGLSAER-------------------------------YLVLQTNVS--------NDLYYNHLRKACRELMDWADA--------------------EYYYSGIAVT
Ot10g03170_Otau_116060216            RDELFDTKWPAQSQRRSV-----------------KSDNYI----------------------------------VLKRECSP---------SDPYESLKLACEDLMRAVDPS---------------------FPYDSLAIT
OSTLU_93380_Oluc_145351687           KDELLRSKWPARSHRRGV-----------------KSDGYI----------------------------------VLKRELSP---------NDPYAKLKDACERLMRTVDET---------------------FPYDSLAIT
consensus/80%                        bp.l.p..WP.pp......................psu...................................hsh..ss.........sp.h......h..hb..hs.s.........................s.h.ls
Key features                                  * synapomorphic tryptophan
```

Species abbreviations Aano : Aureococcus anophagefferens; Mpus : Micromonas pusilla; Oluc : Ostreococcus lucimarinus; Otau : Ostreococcus tauri; Ptri : Phaeodactylum tricornutum

```
Fasta sequences of proteins not in GenBank and assigned Fake gis
>gi|Aano1000005385|ref|jgi|Auran1|63084
MSSPSPKKRCPKPWRPLFAAGSRVECRHGGGDVFYPGVVTAVTSDGHYDVDYDDGDEERGVAGSLMRPEAGLKVGDALQKKFPGYGSGVFDGRVAEIFPAEKKVRVAWSDGEETFMTVAAVRKAMDHPPSPAKPVKRAAASAPPAPLPALEAAGGRPKRARKAPPALTYAKMGTNTSSAK
>gi|Aano1000005614|ref|jgi|Auran1|63253
MHLQLPLVLLAARVAALHAPPSALTSPRCTARRASAQPVASKPSKDPSTAPPGAHRCRWCGGTFVSRNAVFRHLREAESCRASATAEDARAREAIFARARDFTTTLLVGYVDASGADAERLVAAATGASRTTRASSSKARGRSAALESELRAASRDVISITAPAPVASTADLNARLPPNV
>gi|Aano1000001260|ref|jgi|Auran1|61875 ; jgi|Auran1|68521
MSIVEDEEAAADGAEVLSALGDLVEKKGGDREAVVDGWAVRARVRPGDGRVEHVYVDPAGKRHRSRNEVLRCLGLLAPTPAKRARDEAAGAPEAAGDPAPPAAPPPEKRPRAKNPDIDLAAFGWAPPPPPDFAGMLERLRAVDVPRTLARQNIIGEADASAGVRSACVGVVAARSHGVMA
>gi|Aano1000002837|ref|jgi|Auran1|62507
MLRLIGDGNWPTQARPNVTDTGKPVPGMCVGLVFALGQGAQASHVSECHPQMTRVLTRWCRGTLPRTRKGAEFPFSSLQVNYNYAAQKHVDGNNIGPSYIMSIGDHSGGGLWTADQGVIDCKDAWKLFDGNKEHATRPFKGERISFIAFAHGQYNKLEAPVVKMLKGMGFNAARSDGKDL
>gi|Aano1000005432|ref|jgi|Auran1|63122
MDGAEALSLAKRACPQFASAGLKPLDPSSITKVETRSVARLWAGMGSVVELRVKAGEDTAATIMVKKVTLPKRCASIGDRRKKASYEVECNFYEKYAAEVAAAGCACPRPLLVERRGTEGVTIVMTKLVGSSAPMDGARSRRVLSGLAKLHALFWGAAADAAVAGGLAAQGEYWYLDTRP
```

Back to Contents

4. Phyletic distribution, domain architecture and multiple sequence alignment of the fungal subfamily of AlkB proteins that are fused to SAD and R3H domains

Phyletic distribution:

gi	gene_name	length	Species	class	genbank_annotation
121712916	ACLA_010690	2318	Aspergillus clavatus NRRL 1	fungi>ascomycota	CNH domain protein [Aspergillus clavatus NRRL 1].
67526987	AN3951.2	686	Aspergillus nidulans FGSC A4	fungi>ascomycota	hypothetical protein AN3951.2 [Aspergillus nidulans FGSC A4].
115433188	ATEG_08110	672	Aspergillus terreus NIH2624	fungi>ascomycota	conserved hypothetical protein [Aspergillus terreus NIH2624].
70998340	AFUA_5G07420	465	Aspergillus fumigatus Af293	fungi>ascomycota	hypothetical protein AFUA_5G07420 [Aspergillus fumigatus Af293].
159126373	AFUB_054970	465	Aspergillus fumigatus A1163	fungi>ascomycota	hypothetical protein AFUB_054970 [Aspergillus fumigatus A1163].
119479823	NFIA_079850	508	Neosartorya fischeri NRRL 181	fungi>ascomycota	hypothetical protein NFIA_079850 [Neosartorya fischeri NRRL 181]
211592429	Pc22g14940	696	Penicillium chrysogenum Wisconsin 54-1255	fungi>ascomycota	Pc22g14940 [Penicillium chrysogenum Wisconsin 54-1255].
145242116	An09g03880	699	Aspergillus niger CBS 513.88	fungi>ascomycota	hypothetical protein An09g03880 [Aspergillus niger].
169777417	AO090701000312	370	Aspergillus oryzae RIB40	fungi>ascomycota	hypothetical protein [Aspergillus oryzae RIB40].
212530666	PMAA_037570	477	Penicillium marneffei ATCC 18224	fungi>ascomycota	hypothetical protein PMAA_037570 [Penicillium marneffei ATCC
119182137	CIMG_06114	815	Coccidioides immitis RS	fungi>ascomycota	hypothetical protein CIMG_06114 [Coccidioides immitis RS].
218713509	TSTA_054450	673	Talaromyces stipitatus ATCC 10500	fungi>ascomycota	hypothetical protein TSTA_054450 [Talaromyces stipitatus ATCC
154305295	BC1G_08942	790	Botryotinia fuckeliana B05.10	fungi>ascomycota	hypothetical protein BC1G_08942 [Botryotinia fuckeliana B05.10].
171684467	PODANSg4208	1011	Podospora anserina DSM 980	fungi>ascomycota	unnamed protein product [Podospora anserina].
189192468	PTRG_02240	983	Pyrenophora tritici-repentis Pt-1C-BFP	fungi>ascomycota	conserved hypothetical protein [Pyrenophora tritici-repentis
116193337	CHGG_06386	724	Chaetomium globosum CBS 148.51	fungi>ascomycota	hypothetical protein CHGG_06386 [Chaetomium globosum CBS 148.51]
169600785	SNOG_03244	978	Phaeosphaeria nodorum SN15	fungi>ascomycota	hypothetical protein SNOG_03244 [Phaeosphaeria nodorum SN15].
171683219	PODANSg3583	1068	Podospora anserina DSM 980	fungi>ascomycota	unnamed protein product [Podospora anserina].
156040958	SS1G_11457	615	Sclerotinia sclerotiorum 1980	fungi>ascomycota	hypothetical protein SS1G_11457 [Sclerotinia sclerotiorum 1980].
170093061	LACBIDRAFT_313729	1047	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicolor S238N-H82].
58262770	CNB01220	1146	Cryptococcus neoformans var. neoformans JEC21	fungi>basidiomycota	hypothetical protein CNB01220 [Cryptococcus neoformans var.
134117982	CNBL2390	623	Cryptococcus neoformans var. neoformans B-3501A	fungi>basidiomycota	hypothetical protein CNBL2390 [Cryptococcus neoformans var.
169847183	CC1G_01939	500	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis cinerea okayama7#130].

Common Domain architecture, SAD+ZnR+Cysclus+AlkB

```
Multiple sequence alignment and boundaries of the different domains
Boundaries                        <------------------------------------------------------------------------------------------------------------SAD----------------
CIMG_06114_Cimm_119182137         EEVPNIPPNEARPEPVGTPPAWAV-----DRPELCDALPWFRSTQGGC-----YFKDGIAFGVLIDGDAGD---------------RAHLDH-EVVITRVGGGCI------K-VG-G-KLQLTEDI------------------
ATEG_08110_Ater_115433188         IDTTGADGSSA-SETHGQPPAWAD-----YRPELSDTLPWFRAVQGGA-----YHNGNICWGFLLDADCGI---------------RSYIDD-EIVITRIGGGCT------KDAA-G-NLVPLRDQ------------------
An09g03880_Anig_145242116         AHTPNQPPLPD-QETEGHPPAWAE-----NRPELCDTLPWFRSVQGGV-----YHNGNICWGFLIDADCGI---------------RSYLDD-EVIITRVGGGCT------KDAD-G-NLVLIKDQ------------------
AN3951.2_Anid_67526987            AKVRSIESSPSLAEVDGVPPAWADVREPEDRAALASAIPWFTCYEGGM-----YHSKMVCYGFLLDGDSGN---------------RCHIDD-EVVITRIGGGCS------KDRD-G-NLMLKKDM------------------
BC1G_08942_Bfuc_154305295         LPSLENATLSK-PEAWGEPPVWAE-----KRQQLCSSLPYHKGYESAA-----YRTNGMIAGFLANQGVGP---------------RDVFTE-DIYITSVGGGKS------KGED-G--ETKRVKD------------------
CHGG_06386_Cglo_116193337         VAKVANATADG-PPSWGKPLVWAK-----ARGSLCEALPYFRSFKSSL-----HSSNVVCQGFLIDQEVDQ---------------RDVFGA-QVIISSVGGGRV------KNPG-TGNMIRTKDA-----------------
PODANSg4208_Pans_171684467        VKLAPRVPVAR-PAPRGPPPIHSR-----FRGGLCEALPYYKSYKSSL-----YNIGLLAKGFLIDLEVEK---------------GDVFHQ-SVVISAVGGGRV------REGS-V--MVRGRDA-----------------
LACBIDRAFT_313729_Lbic_170093061  AEPACQIPLYQ-EPLPHHPPIWAE-----TRQEVCESFDWFRSYQGGV-----YHTHGIVKGYLLSAFSSKSGRVHQNSLTLLICSRRDRFEHGGRLIISHGGGKAESIHAHEGRF---APQPAEDQ------------------
CNB01220_Cneo_58262770            LKANELKPAL--SNTHGSCPVWAQ-----TRRGLQAALEYFREPKRTLGASVDISGGGMARGVLFEGAVEA-QGVYWGQ--------GERI---GTIMTSIGHP--------RRQK---PIIPQEDEVAPVLPEPSSSTLSATAGI
SNOG_03244_Pnod_169600785         LFANRKPGLKK-PAPSGNPEVWAA-----GRQELCETLHYYRSYQSGC-----YANGGFARGFMFDKIAHA---------------RDYIDN-DVVISRAGGGLV------KDTDSG-QMKSGKDQ------------------
Species abbreviations: Anid : Aspergillus nidulans; Anig : Aspergillus niger; Ater : Aspergillus terreus; Bfuc : Botryotinia fuckeliana; Cglo : Chaetomium globosum; Cimm : Coccidic
```

Back to Contents

5a. Phyletic distribution and multiple sequence alignment of the R3H domain-associated family of 2OGFeDO proteins

Phyletic distribution of the R3H fused AlkB proteins
1;

Gi	Domain architecture	gene name	Len	Organism	class	GenBank Annotation
67153576	2OGFeDO+X+R3H	AvinDRAFT_5847	698	Azotobacter vinelandii	proteobacteria>gammaproteobacteria	hypothetical protein AvinDRAFT_5
148243704	2OGFeDO+X+R3H	Acry_3179	762	Acidiphilium cryptum JF-5	proteobacteria>alphaproteobacteria	hypothetical protein Acry_3179 [
168699004	2OGFeDO+X+R3H	GobsU_010100005758	779	Gemmata obscuriglobus UQM 2246	planctomycetes	hypothetical protein GobsU_05758
158318848	2OGFeDO+X+R3H	Franean1_7129	777	Frankia sp. EAN1pec	actinobacteria	2OG-Fe(II) oxygenase [Frankia sp
62423906	2OGFeDO+X+R3H	BlinB01002514	785	Brevibacterium linens BL2	actinobacteria	hypothetical protein BlinB010025
Ngru1000002312	2OGFeDO+X+R3H	Ngru1000002312	969	Naegleria gruberi	heterolobosea	fgeneshHS_pg.scaffold_12000094
Ngru1000008142	2OGFeDO+X+R3H	Ngru1000008142	1109	Naegleria gruberi	heterolobosea	fgeneshNG_pg.scaffold_23000035
Ngru1000014932	2OGFeDO+X+R3H	Ngru1000014932	1114	Naegleria gruberi	heterolobosea	estExt_fgeneshNG_pg.C_600108
Ngru1000002448	2OGFeDO+X+R3H	Ngru1000002448	1139	Naegleria gruberi	heterolobosea	fgeneshHS_pg.scaffold_16000074
Ngru1000014072	2OGFeDO+X+R3H	Ngru1000014072	1140	Naegleria gruberi	heterolobosea	estExt_fgeneshNG_pg.C_270061
Ngru1000010562	2OGFeDO+X+R3H	Ngru1000010562	944	Naegleria gruberi	heterolobosea	fgeneshNG_pg.scaffold_52000076
Ngru1000011382	2OGFeDO+X+R3H	Ngru1000011382	972	Naegleria gruberi	heterolobosea	fgeneshNG_pg.scaffold_64000090
Ngru1000005317	2OGFeDO+X+R3H	Ngru1000005317	2591	Naegleria gruberi	heterolobosea	estExt_fgeneshHS_pg.C_760047
Ngru1000002982	2OGFeDO+X+R3H	Ngru1000002982	1252	Naegleria gruberi	heterolobosea	fgeneshHS_pg.scaffold_34000025
170102781	2OGFeDO+X+R3H	LACBIDRAFT_328485	1203	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bicc

164645299	2OGFeDO+X+R3H	LACBIDRAFT_325982	991	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bico
116499177	2OGFeDO+X+R3H	CC1G_09674	946	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	hypothetical protein CC1G_09674
Ppla1000002992	2OGFeDO+X+R3H+TBC	Ppla1000002992	2148	Postia placenta	fungi>basidiomycota	estExt_fgenesh3_pg.C_1210047
46135867	2OGFeDO+X+R3H	FG09449.1	884	Gibberella zeae PH-1	fungi>ascomycota	hypothetical protein FG09449.1 [
85091137	2OGFeDO+X+R3H	NCU09412	934	Neurospora crassa OR74A	fungi>ascomycota	hypothetical protein NCU09412 [N
145254402	2OGFeDO+X+R3H	An18g01550	919	Aspergillus niger CBS 513.88	fungi>ascomycota	hypothetical protein An18g01550
211583340	2OGFeDO+X+R3H	Pc13g02810	770	Penicillium chrysogenum Wisconsin 54-1255	fungi>ascomycota	Pc13g02810 [Penicillium chrysoge
218717189	2OGFeDO+X+R3H	TSTA_016850	914	Talaromyces stipitatus ATCC 10500	fungi>ascomycota	conserved hypothetical protein [
Dpur1000007623	2OGFeDO+X+R3H	Dpur1000007623	952	Dictyostelium purpureum	amoebozoa>mycetozoa>dictyosteliida	estExt_fgeneshDP_pg.C_2240006
Chlo1000004688	2OGFeDO+X	Chlo1000004688	719	Chlorella sp.	viridiplantae>chlorophyta IGS.gm_11_00141	
159485224	2OGFeDO+X	CHLREDRAFT_153458	458	Chlamydomonas reinhardtii	viridiplantae>chlorophyta	predicted protein [Chlamydomonas
159478114	2OGFeDO+X	CHLREDRAFT_175993	953	Chlamydomonas reinhardtii	viridiplantae>chlorophyta	hypothetical protein CHLREDRAFT_
159490748	2OGFeDO+X	CHLREDRAFT_188236	897	Chlamydomonas reinhardtii	viridiplantae>chlorophyta	predicted protein [Chlamydomonas
159475104	2OGFeDO+X	CHLREDRAFT_191918	418	Chlamydomonas reinhardtii	viridiplantae>chlorophyta	predicted protein [Chlamydomonas
158278571	2OGFeDO+X	CHLREDRAFT_189197	1221	Chlamydomonas reinhardtii	viridiplantae>chlorophyta	predicted protein [Chlamydomonas
Chlo1000007748	2OGFeDO+X	Chlo1000007748	536	Chlorella sp.	viridiplantae>chlorophyta fgenesh3_pg.C_scaffold_13000046	
Smoe1000005861	2OGFeDO+X	Smoe1000005861	853	Selaginella moellendorffii	viridiplantae	fgenesh2_pg.C_scaffold_1000000026
Smoe1000018266	2OGFeDO+X	Smoe1000018266	427	Selaginella moellendorffii	viridiplantae	fgenesh2_pg.C_scaffold_0000806
Smoe1000014016	2OGFeDO+X	Smoe1000014016	198	Selaginella moellendorffii	viridiplantae	fgenesh2_pg.C_scaffold_35000267
Smoe1000014051	2OGFeDO+X	Smoe1000014051	304	Selaginella moellendorffii	viridiplantae	fgenesh2_pg.C_scaffold_35000237
Smoe1000017088	2OGFeDO+X	Smoe1000017088	1434	Selaginella moellendorffii	viridiplantae	estExt_fgenesh2_pg.C_240225
Smoe1000010456	2OGFeDO+X	Smoe1000010456	756	Selaginella moellendorffii	viridiplantae	fgenesh2_pg.C_scaffold_11000105
Smoe1000002216	2OGFeDO+X	Smoe1000002216	900	Selaginella moellendorffii	viridiplantae	fgenesh2_pg.C_scaffold_96000020
Smoe1000021895	2OGFeDO+X	Smoe1000021895	1030	Selaginella moellendorffii	viridiplantae	fgenesh2_pg.C_scaffold_1000197
Smoe1000020846	2OGFeDO+X	Smoe1000020846	934	Selaginella moellendorffii	viridiplantae	fgenesh2_pg.C_scaffold_66000023
Smoe1000018380	2OGFeDO+X	Smoe1000018380	1343	Selaginella moellendorffii	viridiplantae	estExt_fgenesh2_pg.C_00688
Smoe1000003520	2OGFeDO+X	Smoe1000003520	904	Selaginella moellendorffii	viridiplantae	fgenesh2_pg.C_scaffold_20000055
Smoe1000003056	2OGFeDO+X	Smoe1000003056	936	Selaginella moellendorffii	viridiplantae	fgenesh2_pg.C_scaffold_46000146
Smoe1000003542	2OGFeDO+X	Smoe1000003542	912	Selaginella moellendorffii	viridiplantae	fgenesh2_pg.C_scaffold_20000036
Smoe1000015653	2OGFeDO+X	Smoe1000015653	870	Selaginella moellendorffii	viridiplantae	fgenesh2_pg.C_scaffold_3000154
Smoe1000021894	2OGFeDO+X	Smoe1000021894	839	Selaginella moellendorffii	viridiplantae	fgenesh2_pg.C_scaffold_1000198
162672410	2OGFeDO+X	PHYPADRAFT_169264	1168	Physcomitrella patens subsp. patens	viridiplantae	predicted protein [Physcomitrell
Psoj1000015557	2OGFeDO+X	Psoj1000015557	614	Phytophthora sojae	stramenopiles	143565
Psoj1000008776	2OGFeDO+X	Psoj1000008776	552	Phytophthora sojae	stramenopiles	135757
Psoj1000015702	2OGFeDO+X	Psoj1000015702	726	Phytophthora sojae	stramenopiles	143738
Psoj1000005084	2OGFeDO+X	Psoj1000005084	851	Phytophthora sojae	stramenopiles	131556
Psoj1000001369	2OGFeDO+X	Psoj1000001369	460	Phytophthora sojae	stramenopiles	127244
Psoj1000002636	2OGFeDO+X	Psoj1000002636	634	Phytophthora sojae	stramenopiles	128746
Psoj1000009248	2OGFeDO+X	Psoj1000009248	437	Phytophthora sojae	stramenopiles	136290
Psoj1000015551	2OGFeDO+X	Psoj1000015551	544	Phytophthora sojae	stramenopiles	143559
Psoj1000012378	2OGFeDO+X	Psoj1000012378	372	Phytophthora sojae	stramenopiles	139870
Psoj1000008778	2OGFeDO+X	Psoj1000008778	458	Phytophthora sojae	stramenopiles	135759
Psoj1000012376	2OGFeDO+X	Psoj1000012376	367	Phytophthora sojae	stramenopiles	139867
Pram1000011751	2OGFeDO+X	Pram1000011751	321	Phytophthora ramorum	stramenopiles	86281
Pram1000006627	2OGFeDO+X	Pram1000006627	367	Phytophthora ramorum	stramenopiles	77035
Pram1000006625	2OGFeDO+X	Pram1000006625	405	Phytophthora ramorum	stramenopiles	77037
Pram1000006622	2OGFeDO+X+2OGFeDO	Pram1000006622	1006	Phytophthora ramorum	stramenopiles	77040
66823093	2OGFeDO+X	DDBDRAFT_0217068	471	Dictyostelium discoideum AX4	mycetozoa>dictyosteliida	hypothetical protein [Dictyostel
Dpul1000008743	2OGFeDO+X	Dpul1000008743	1430	Daphnia pulex	metazoa>crustacea	PASA_GEN_6000037
Dpul1000028680	2OGFeDO+X	Dpul1000028680	857	Daphnia pulex	metazoa>crustacea	SNAP_00037728
Dpul1000019118	2OGFeDO+X	Dpul1000019118	1279	Daphnia pulex	metazoa>crustacea	NCBI_GNO_21300029
Dpul1000004449	2OGFeDO+X	Dpul1000004449	1366	Daphnia pulex	metazoa>crustacea	NCBI_GNO_2900191
Dpul1000023668	2OGFeDO+X	Dpul1000023668	1224	Daphnia pulex	metazoa>crustacea	PASA_GEN_25200007
Dpul1000012014	2OGFeDO+X	Dpul1000012014	917	Daphnia pulex	metazoa>crustacea	PASA_GEN_1000230
Dpul1000004459	2OGFeDO+X	Dpul1000004459	298	Daphnia pulex	metazoa>crustacea	SNAP_00012885
Dpul1000019109	2OGFeDO+X	Dpul1000019109	567	Daphnia pulex	metazoa>crustacea	SNAP_00032397
Dpul1000021114	2OGFeDO+X	Dpul1000021114	944	Daphnia pulex	metazoa>crustacea	SNAP_00020982
Dpul1000004279	2OGFeDO+X	Dpul1000004279	1173	Daphnia pulex	metazoa>crustacea	NCBI_GNO_2900033
Dpul1000004224	2OGFeDO+X	Dpul1000004224	901	Daphnia pulex	metazoa>crustacea	PASA_GEN_5800050
Dpul1000008742	2OGFeDO+X	Dpul1000008742	563	Daphnia pulex	metazoa>crustacea	fgenesh1_pg.C_scaffold_60000011
Dpul1000022212	2OGFeDO+X	Dpul1000022212	764	Daphnia pulex	metazoa>crustacea	SNAP_00025011
Dpul1000012716	2OGFeDO+X	Dpul1000012716	351	Daphnia pulex	metazoa>crustacea	fgenesh1_pg.C_scaffold_49000039
Dpul1000023658	2OGFeDO+X	Dpul1000023658	1338	Daphnia pulex	metazoa>crustacea	NCBI_GNO_25200003
Dpul1000016316	2OGFeDO+X	Dpul1000016316	1226	Daphnia pulex	metazoa>crustacea	NCBI_GNO_3600088
Dpul1000024185	2OGFeDO+X	Dpul1000024185	405	Daphnia pulex	metazoa>crustacea	fgenesh1_pg.C_scaffold_114000034
Dpul1000005753	2OGFeDO+X	Dpul1000005753	1338	Daphnia pulex	metazoa>crustacea	fgenesh1_pg.C_scaffold_20000061
Dpul1000012176	2OGFeDO+X	Dpul1000012176	906	Daphnia pulex	metazoa>crustacea	PASA_GEN_1000276
Ngru1000003599	2OGFeDO+X	Ngru1000003599	845	Naegleria gruberi	heterolobosea	fgeneshHS_pg.scaffold_58000065
Ngru1000013028	2OGFeDO+X	Ngru1000013028	490	Naegleria gruberi	heterolobosea	fgeneshNG_pg.scaffold_833000001
Ehux1000027221	2OGFeDO+X	Ehux1000027221	716	Emiliania huxleyi	haptophyceae	gm1.2500281
169866398	2OGFeDO+X	CC1G_12589	937	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis c
Ppla1000002318	2OGFeDO+X	Ppla1000002318	301	Postia placenta	fungi>basidiomycota	gw1.145.11.1
116499178	2OGFeDO+X	CC1G_09675	528	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis c
164646975	2OGFeDO+X	LACBIDRAFT_315778	950	Laccaria bicolor S238N-H82	fungi>basidiomycota	predicted protein [Laccaria bico
169867452	2OGFeDO+X	CC1G_10968	653	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis c
116499174	2OGFeDO+X	CC1G_09671	543	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	hypothetical protein CC1G_09671
169866338	2OGFeDO+X	CC1G_09661	530	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	hypothetical protein CC1G_09661
212540430	2OGFeDO+X	PMAA_055570	459	Penicillium marneffei ATCC 18224	fungi>ascomycota	conserved hypothetical protein [
169769949	2OGFeDO+X	AO090003000232	427	Aspergillus oryzae RIB40	fungi>ascomycota	hypothetical protein [Aspergillu
66821331	2OGFeDO+2OGFeDO+X	DDBDRAFT_0203201	1654	Dictyostelium discoideum AX4	amoebozoa>mycetozoa>dictyosteliida	hypothetical protein [Dictyostel
169866328	2OGFeDO+2OGFeDO	CC1G_09656	1733	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	hypothetical protein CC1G_09656
Chlo1000002921	2OGFeDO	Chlo1000002921	566	Chlorella sp.	viridiplantae>chlorophyta	IGS.gm_2_00021
Psoj1000015548	2OGFeDO	Psoj1000015548	1469	Phytophthora sojae	stramenopiles	143556
Psoj1000015558	2OGFeDO	Psoj1000015558	1423	Phytophthora sojae	stramenopiles	143566
Ehux1000028524	2OGFeDO	Ehux1000028524	390	Emiliania huxleyi	haptophyceae	gm1.300307Ehux1000028524
66821557	2OGFeDO	DDBDRAFT_0167593	551	Dictyostelium discoideum AX4	mycetozoa>dictyosteliida	hypothetical protein [Dictyostel
66820376	2OGFeDO	DDBDRAFT_0167421	876	Dictyostelium discoideum AX4	mycetozoa>dictyosteliida	hypothetical protein [Dictyostel
66808223	2OGFeDO	DDBDRAFT_0186906	430	Dictyostelium discoideum AX4	mycetozoa>dictyosteliida	hypothetical protein [Dictyostel
Dpur1000005509	2OGFeDO	Dpur1000005509	623	Dictyostelium purpureum	amoebozoa>mycetozoa>dictyosteliida	GID1.0043264
Dpur1000009437	2OGFeDO	Dpur1000009437	1087	Dictyostelium purpureum	amoebozoa>mycetozoa>dictyosteliida	GID1.0047637
Chlo1000007123	2OGFeDO	Chlo1000007123	423	Chlorella sp. viridiplantae>chlorophyta IGS.gm_8_00313		
158280561	2OGFeDO+2OGFeDO+X	CHLREDRAFT_140971	1175	Chlamydomonas reinhardtii	viridiplantae>chlorophyta	predicted protein [Chlamydomonas

Proteins that lack the 2OGFeDO domain

88182770	X+R3H	CHGG_02173	621	Chaetomium globosum CBS 148.51	fungi>ascomycota	hypothetical protein CHGG_02173
85112602	X+R3H+HMG	NCU00833	2202	Neurospora crassa OR74A	fungi>ascomycota	hypothetical protein NCU00833 [N
115385036	X+R3H	ATEG_01700	771	Aspergillus terreus NIH2624	fungi>ascomycota	predicted protein [Aspergillus t
212540412	X+R3H	PMAA_055480	1040	Penicillium marneffei ATCC 18224	fungi>ascomycota	hypothetical protein PMAA_055480
Smoe1000000683	X	Smoe1000000683	185	Selaginella moellendorffii	viridiplantae	fgenesh2_pg.C_scaffold_20000002
Smoe1000018265	X	Smoe1000018265	674	Selaginella moellendorffii	viridiplantae	fgenesh2_pg.C_scaffold_0000807
Smoe1000007796	X	Smoe1000007796	574	Selaginella moellendorffii	viridiplantae	fgenesh2_pg.C_scaffold_36000271
Psoj1000015556	X	Psoj1000015556	482	Phytophthora sojae	stramenopiles	143564
Psoj1000015549	X	Psoj1000015549	368	Phytophthora sojae	stramenopiles	143557
Dpul1000004448	X	Dpul1000004448	2275	Daphnia pulex	metazoa>crustacea	SNAP_00012874
Dpul1000023063	X	Dpul1000023063	636	Daphnia pulex	metazoa>crustacea	SNAP_00024797
Dpul1000004452	X	Dpul1000004452	1491	Daphnia pulex	metazoa>crustacea	SNAP_00012878
Dpul1000026020	X	Dpul1000026020	575	Daphnia pulex	metazoa>crustacea	fgenesh1_pg.C_scaffold_323000004
215449484	X	MPER_12681	438	Moniliophthora perniciosa FA553	fungi>basidiomycota	hypothetical protein MPER_12681
169867436	X	CC1G_10960	502	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis c
169866400	X	CC1G_12590	437	Coprinopsis cinerea okayama7#130	fungi>basidiomycota	predicted protein [Coprinopsis c
215454380	X	MPER_10003	263	Moniliophthora perniciosa FA553	fungi>basidiomycota	hypothetical protein MPER_10003
215449275	X	MPER_12817	645	Moniliophthora perniciosa FA553	fungi>basidiomycota	hypothetical protein MPER_12817
215463051	X	MPER_05861	221	Moniliophthora perniciosa FA553	fungi>basidiomycota	hypothetical protein MPER_05861
171676563	X	PODANSg246	212	Podospora anserina DSM 980	fungi>ascomycota	unnamed protein product [Podospo
154319694	X	BC1G_02328	535	Botryotinia fuckeliana B05.10	fungi>ascomycota	predicted protein [Botryotinia f
171691366	X	PODANSg7647	386	Podospora anserina DSM 980	fungi>ascomycota	unnamed protein product [Podospo
189197891	X	PTRG_04950	505	Pyrenophora tritici-repentis Pt-1C-BFP	fungi>ascomycota	predicted protein [Pyrenophora t
Dpur1000010089	X	Dpur1000010089	304	Dictyostelium purpureum	amoebozoa>mycetozoa>dictyosteliida	fgeneshDP_pg.C_scaffold_38200000

Multiple sequence alignment of the 2OGFeDO domain of the R3H domain-associated family

```
Secondary_structure              HH--H-H------HH--------------EE-EE-----------------------EEEEE-------------EE------HHHHHHHHH-----HHH---------HHH-------HHHHHH--------------
Ngru1000002982_Ngru_Ngru1000002982  FI--L-D------AQ-NK-------GGSKFT-QCGALDS-D---F------G---LPGLQLKLDNEQ-------FKTIGFPLS-NGSASELIQF-----CSKAPYGRGEDTIYD-------ENVRKTWQLDPSRF-------EI
LACBIDRAFT_315778_LBic_164646975  LESVL-Q------QDFAF-------PG-TYH-QVMEHLS-C---P----------HLGLRVDG----------LEDVSLPLN-DQSARAIHAL-----ADSSKNGSG----------------RSFLELGTG-F-------SF
Psoj1000001369_Psoj_Psoj1000001369  ISDVL--------GGASE------HSG-DFS-FGGQADL---LPA----------VPDLFVDG----------VGPVPVPLW-EERAQKLIDK-----CEKSPFGHNMDTKMD-------ENVRKSWQLAPGQV-------QF
Ngru1000002312_Ngru_Ngru1000002312  ILDLL-E------QSCVL-------NG-DYI-TGGKLDS-N---Y-L--------MPAIEILLEDNGK------YKPLPLPIV---YAEQVTEL-----IRNCSFGRTE----N-------VNITKSWHLDPHKF-------RI
TSTA_016850_Tsti_218717189  LLDTL-S-------EI-QS-------NG-TFI-TAKELST-A---I----------IPGLHVPG----------IGAIGLPVT-PDQVKAMIQS-----SRMSPYGKGSETLVN-------ESVRKSWQLDANQF-------SL
Pc13g02810_Pchr_211583340  LREAL--------DAIVS-------SG-TFA-SQGRLPH-----V----------DPQIKVQD----------VGAISLPLQ-ESQLRQMMKQ-----AHQAPYGKGSETLVD-------TAVRNTLELNPGQF-------EL
NCU09412_Ncra_85091137  LRKAL--------DNVKV-------AG-SFA-SFRKLSR-----TP----------PDGLVVDG----------VGDVAMPLS-ETQARQLIAK-----ARQAPYGRGSETIVD-------TSVRNTWEIDAAQF-------SF
NCU09412_Ncra_85091137  LRKAL--------DNVKV-------AG-SFA-SFRKLSR-----TP----------PDGLVVDG----------VGDVAMPLS-ETQARQLIAK-----ARQAPYGRGSETIVD-------TSVRNTWEIDAAQF-------SF
Acry_3179_Acry_148243704  LAALL--------GTVRR-------PG-DFY-VAGEIAF-----S----------PPSLDVAS----------VGPVALPLL-PIQARQLIDA-----AEPAPFGRGEQTIID-------PTVRRCGQIGPDRV-------RL
Franean1_7129_Fsp._158318848  RLARM--------LRGDT-------ET-AFS-VELTAST-----D----------DLSLEVEG----------IGPVTFPVP-YAQARKLLGL-----GQPARFGRGEETLID-------PGVRDTWEIPKNLV-------HA
Ngru1000002448_Ngru_Ngru1000002448  LLTSS-E------TD-VC-------VG-EFS-CGGEVEN-I---P------G---LPGLQICVPSNESGATVE-LKHCSLPITNLNQAQDIISV-----ASKAPYGKGELTLID-------ETVRKSWQLEPSQF-------II
CHLREDRAFT_189197_Crei_159467329  MRASLMGTCYHKYGGRGR-------SC-RFSQVTGAVGG-P---L----------PEVAVRG----------VGKLALPID-AGQAKDLIYA-----GQPAPYGKGTNTVLD-------PNVRKATQIEADRV-------CF
GobsU_010100005758_Gobs_168699004  LDAAL--------KKINR-------PG-SFC-VTGSVRA-----A----------NPGLEVAG----------IGPVGLPLT-AHQAKELAAV-----CKQAPYGKGEETLVD-------TSVRRVWQLDPDHF-------SL
```

```
Ngru1000010562_Ngru_Ngru1000010562  LIQVM-N------ST-DR-------SG-DFS-TGGKLSA-E---Y------S---IPAIEILQESGV-------YKPLPLPIIYEEQAKDLIKN-----CSMSPFGRLDKTIYD-------ESVRKSWQLDPQRF-------KI
CHLREDRAFT_153458_Crei_159485224    LKIAL--------DGVTS-------GG-KFA-VTGTVDS-P---L-----------PEIAVRG-----------VGKLALPID-AGQAEDLVFA-----GMPAPFGKGSKTVLD-------PAVRKAIQIEAARV-------RF
CHLREDRAFT_191918_Crei_159475104    LKVAL--------GGVTS-------GG-KFA-VTGTVDG-P---L-----------PEIAIRG-----------VGKLALPID-AGQAEDLVFA-----GMPAPFGKGSKTVLN-------PAVRKAIQIEAARV-------RF
Pram1000006622_Pram_Pram1000006622  ISEML--------ASADQ------KAG-EYS-FGGVVDT---LPV----------APGLFVEE-----------VGAISVPLT-SEQAEKVIVK-----CEKSPFGHNLDTKMD-------ENVRKSWQLAPDQV-------KI
Ngru1000008142_Ngru_Ngru1000008142  LIEEL-I------PT-VT-------AGGSFS-VGGEFDT-Y---P----------AVGLQLLVPNSENN-SVE-RKYVSLPLVTKQQADDIIQL-----AAKAPYGRGEETIVD-------EKVRKTWQLEPNQF-------AI
Psoj1000015702_Psoj_Psoj1000015702  ISSLL--------AKADD-------NAG-GYS-FGGVADT---LPA----------VPGLYVEG-----------LGDVSVPLC-PEQADKLIAI-----CDKSPFGRKLDTMMD-------ESVRKSWQLAPDQV-------DV
Ngru1000014072_Ngru_Ngru1000014072  LQNTF-T------NITSV-------GG-NYY-YYDRLPD-V---M----K-----VPGIELIIT--EEN--VEKYKPIALPIIYKEQADEIIRI-----GKKSPFGKGEETIHD-------Q-VRNSFELEPHQF-------RI
Ngru1000014932_Ngru_Ngru1000014932  LKKSF-S------QIVSC-------GG-SFV-CSDKF-D-G---V----T-----LPGLELLVSAAQDS--DKKYKAMPFPIVYKEQADEIIRI-----AKKSPYGKGEETIYD-------DNVRNSYELEPNQF-------RI
CC1G_09661_Ccin_169866338           LSDAL-E------SDFDF-------MG-KHY-FARAYPD-YEMFEYKYGEWNYGWEVTIDFGPG--TDVDGSTKSYVLDFPLY-EYKVLCALER-----AFRSHFGRGDKKVVD-------TKVRDTWEIDGSKV-------KL
CC1G_09671_Ccin_116499174           LSDAL-E------SDFDF-------MG-KHY-FARAYPD-YEMAEYNSGEWTRRWKVNIDL-G----TDANKRASCYFDFPLD-DYKVDEVLEH-----ASRSPFGMGDQKVVD-------TKVRDTWEIDGSKV-------KL
Dpul1000012014_Dpul_Dpul1000012014  VEVAI-H------SKLDS--LST--NS-QFA-ASLRLDN-----------------KEFTNRTN-----------CEDMTVPVS-LEKLKKLAHVISKGSTKPDPTDRSTHLL----------KLKKLWQISGQEMLKANGFTAL
Smoe1000003520_Smoe_Smoe1000003520  LKVLW--------STLPR--------Q-KYA-TGGEVDA-SKFAM-----------PASLRVEG-----------AGDLALPLS-GDGFLKLKAV-----SQQAPYGKGEETLLD-------TDVRRAWQVDASKV-------SC
Smoe1000003542_Smoe_Smoe1000003542  LKVLW--------STVPR--------R-KYA-TGGEVDA-SKFAM-----------PASLRVEG-----------AGDLALPLS-NDGFLKLKAV-----SQQAPYGKGEETLLD-------TDVRRAWQVEASKV-------SC
Smoe1000018380_Smoe_Smoe1000018380  LEVLW--------STLPR--------P-KYA-TGGEVDA-TKFAM-----------PTFMQVEG-----------AGDLPLPLS-NDGFLKLKAV-----SQQAPYGKGEETLLD-------TDVRRAWQVEASKV-------SC
Smoe1000015653_Smoe_Smoe1000015653  LKVLW--------SS--R--------P-KFA-SGGEVDG-TKFPM-----------PTSLRVEG-----------AGDLALPLS-SDGFLKLKAV-----SQQAPYGKGDQTLFD-------TTIRRAWQVEASKV-------SC
CC1G_09656_Ccin_116499159           IENAL-G------NDFDF-------EG-SFY-YAKTHPG-A---P-----------NPCLHVQG-----------IGSVGLPLS-EVEAKRLISQ-----AAQAPFGKGDQTVVD-------TSVRNTWEIEPELV-------SF
CC1G_09656_Ccin_169866328           IENAL-G------NDFDF-------EG-SFY-YAKTHPG-A---P-----------NPCLHVQG-----------IGSVGLPLS-EVEAKRLISQ-----AAQAPFGKGDQTVVD-------TSVRNTWEIEPELV-------SF
CC1G_09674_Ccin_116499177           LDKAL-G------GDFSF-------KG-KFY-YASQQPT-A---P-----------NPCLAIEG-----------LGIIGLPLN-EREAKAIIAS-----SKQAPFGKGKKTLVD-------KTIRNTWEIESDKV-------TF
CC1G_09674_Ccin_169866364           LDKAL-G------GDFSF-------KG-KFY-YASQQPT-A---P-----------NPCLAIEG-----------LGIIGLPLN-EREAKAIIAS-----SKQAPFGKGKKTLVD-------KTIRNTWEIESDKV-------TF
Dpul1000012176_Dpul_Dpul1000012176  LRKVE-D----S-EDWNF----------QFA-VGQQISE-ENI-------------LPMFSVGS-----------LGTIPVPVN-IKDMNRLKKM-----SETGYCCNHNMGVVGVKPTGKMKKPKHSWQLDATHI-----I-RH
BlinB01002514_Blin_62423906         LKTHL--------SRVGA-------DI-VHA-ASTRVPP-----D-----------GLNILIDR------------APLSFPIG-EDDIAALISI-----AEPAHFGSGEDTVFD-------PTVRDTWVIDPAHV-------HL
LACBIDRAFT_325982_Lbic_164645299    LEGVL-D------GDFEF-------AG-SYA-CGQTLPQ-A---P-----------NPCLTIAG-----------MGLVGLPLK-ERDARNIIAC-----SSRAPYGHGERTVVN-------ADVRDTWEIEPAKI-------KF
FG09449.1_Gzea_46135867             LLKAL-N------SIEPT-------GD-DVA-ACSHYNI-F---V-----------NPGLTIDG-----------GNLIRLPLE-EDDARTIKGA-----CRQAPFGHG-----D-------NSIRNTWELDASKF-------DL
LACBIDRAFT_325982_Lbic_170097353    LEGVL-D------GDFEF-------AG-SYA-CGQTLPQ-A---P-----------NPCLTIAG-----------MGLVGLPLK-ERDARNIIAC-----SSRAPYGHGERTVVN-------ADVRDTWEIEPAKI-------KF
FG09449.1_Gzea_42553402             LLKAL-N------SIEPT-------GD-DVA-ACSHYNI-F---V-----------NPGLTIDG-----------GNLIRLPLE-EDDARTIKGA-----CRQAPFGHG-----D-------NSIRNTWELDASKF-------DL
CC1G_12589_Ccin_169866398           FEEAM-S------SGMTW-------FG-QYQ-FSTSYSN-A---P-----------DPLLSIQRFSD--------NESLSLPLT-DHDAQTLSTF-----LKSQDSSIRQYSPFE-------GRCGQVFKAASV---------GI
LACBIDRAFT_328485_Lbic_170102781    LDDAL-S------GDFNF-------KG-NAY-YSAKLPH-A---P-----------NPCLSIEG-----------LGMVGLPLS-ERDAKSIISC-----SAQAPFGHGERTVVD-------REVRDTWEIEPSNL-------KF
LACBIDRAFT_328485_Lbic_170102781    LDDAL-S------GDFNF-------KG-NAY-YSAKLPH-A---P-----------NPCLSIEG-----------LGMVGLPLS-ERDAKSIISC-----SAQAPFGHGERTVVD-------REVRDTWEIEPSNL-------KF
An18g01550_Anig_145254402           LCAVL-E------TI-ES-------SG-SFL-T-QDLGL-G---LA----------IPGLNIAG-----------LGNIRLPIS-ADDAKAITQC-----CDRSPYGKGSETLVD-------ESVRKTWQLDPGQF-------SL
Dpul1000004224_Dpul_Dpul1000004224  LLNAI-G-YLTQ-DSMAS-NSSL-----QFT-FSQMLDS-TDSRL-----------PNIINVTE-----------VGDVHVPVS-LLDAERLKFS-----ASPTSFGYGLHNFVQ-------PLLRTSWKIDAEKI-RIL---SL
Ppla1000002318_Ppla_Ppla1000002318  -------------------------G-SFA-FHHSYPK-A---P-----------NPTLYLDG-----------LGHIGLPLS-IRDAEAIKSK-----ANQAPFGMGERTVVD-------KSVRDTWELDASQV-------RF
PHYPADRAFT_169264_Ppat_162672410    ILAAT--------GAVGRERLFERPNA-TFV-ISGTIDHYL---L-----------GS-LGVDG-----------AGVLGLPLS-AADAAKLRMV-----SKQAPHGRGLNTVLD-------TTVRSAWQVDAFKV-------TF
Pram1000011751_Pram_Pram1000011751  -------------------------------------------------------------------------------------------------------------------MD-------EKVRKSWELAPDQV-------QI
Psoj1000008776_Psoj_Psoj1000008776  -----------------------------------------------------------------------------------------M-----CKKSPFGHNLETKMD-------EAVRQSWQLVPSQV-------RF
AvinDRAFT_5847_Avin_67153576        ----------------------------------------------------------------------------------------------------------------------------MRDTWEIPASHV-------RF
Psoj1000008778_Psoj_Psoj1000008778  -----------------------------------------------------------------------------------------M-----CKKSPFGHNLETKMD-------EAVRQSWQLVPSQV-------RF
consensus/80%                       l...h....................s.pa......hs...................l.l...............lshPl....ph..l........s..usaG+s.ps.hs.........lRpsWpl.ssph.......ph
Key features                                                                                                                                                 RxxW
Species abbreviations: Acry : Acidiphilium cryptum; Anig : Aspergillus niger; Avin : Azotobacter vinelandii; Blin : Brevibacterium linens; Ccin : Coprinopsis cinerea; Crei : Chlamy
```

Fasta sequences of proteins from incomplete/unpublished genomes not in GenBank and assigned Fake gis

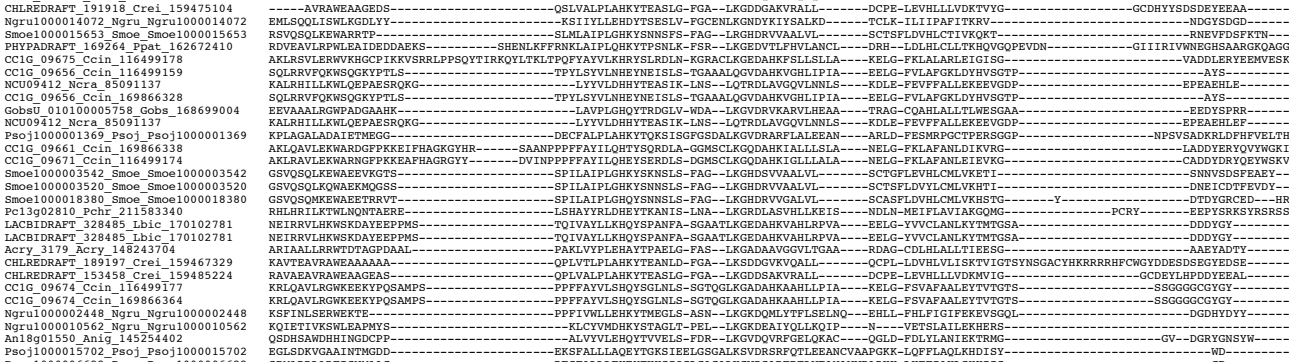
```
>gi|Smoe1000021894|ref|jgi|Selmo1|403361|fgenesh2_pg.C_scaffold_1000198
MAQLNKALWCDREIFATGGEIDVAKFRMPASLHVEGAERISLPLSPLGLAKIKSVSQQAPYGLGERTLVDTDVRRAWQINPDRVSCPEVPNFFSTVPRDLAAHGLAAMGLDSTALELDVKLYKLLLYEAGGHFKFHRDTEKEDGMFATMILQLPTSTGHSGGALAVRQGRQGCEFDFSDG
>gi|Smoe1000021895|ref|jgi|Selmo1|403360|fgenesh2_pg.C_scaffold_1000197
MAVEQLNKALWCDREIFATGGEIDVAKFRMPASLHVEGAERISLPLSPLGLAKIKSVSQQAPYGLGERTLVDTDVRRAWQINPDKVSCPDVPNFFSSVARDLAAHGLAAMGLDSTALELDVKLYKLLLYEAGGHFKFHRDTEKEDGMFATMILQLPTSTGHSGAALAVRQGRQCCEFDFS
>gi|Smoe1000020846|ref|jgi|Selmo1|424217|fgenesh2_pg.C_scaffold_66000023
MSASLLVEGAGSIHLPLSQEGLARLKSVCQQAPYGLGERTLVDTDVRKAWQVNPDKISCPENPSFFSSVVHGLAAAGLAAMGLDPRALELDVKLYKLLLYEAGGHFKFHRDTEKDVCDDDRPAPHSVRAHRRSSRGSSSKYFSTIFFCDCEHQLEEVTGGSRLSLVFSLVRNNVEFIAAE
>gi|Smoe1000018380|ref|jgi|Selmo1|437525|estExt_fgenesh2_pg.C_00688
MSGALAVEMSDFSPAGALPVRRPLGKRRNGDGDGEVDPLLPVPELAPAPKPWKGAPIVAALICVAFLMLQMLPATHYRHPADPTRTWIPIRQEQDQGESGKEVEEVDAREFVTGIKSTSIDLFVCTDEKDLRPLVVVINSTLANTRHTQRIRFHIITTESQREAWLSKLKALFPLAAIDM
>gi|Smoe1000018266|ref|jgi|Selmo1|402801|fgenesh2_pg.C_scaffold_0000806
MLRTLAVEGMAGGAIRLPLSLEDLDKLKAVSQPAPFGIGKRTLGDTSVRQPWQLDPSKISFPDNPEFLAKTAYDIASAALSSMGVHAGSLELGVKLYKMLLYEPGGHFKLHSDKEKEAGMFATMVIQLPTQTGHGGGGLTVRHLSCSKAFEFQSTSGYFYTTFFCHCERKLDEVTRGSRL
>gi|Smoe1000017088|ref|jgi|Selmo1|442467|estExt_fgenesh2_pg.C_240225
MVDARFFPSRKQAWYAGFHGVKPELWWDGFSSRVELNATISDRLGGSVGSWTSGSLPWGYSKQMLLIVGAIDREGHESLTESQCRAAIAEYERYNDCNLEAMQRLIVHFRLPLDKVSRISREEQSWIVRDLYCAGGFPEKVVVMERLMSVKSWACYRNVEVGELLPDQRTKGEEIRLGCV
>gi|Smoe1000015653|ref|jgi|Selmo1|404931|fgenesh2_pg.C_scaffold_3000154
MDYEALLKVLWSSRPKFASGGEVDGTKFPMPTSLRVEGAGDLALPLSSDGFLKLKAVSQQAPYGKGDQTLFDTTIRRAWQVEASKVSCPQMPDFFSAVVLNLAVPAMEELGIDAAALGLDVRLYKLLLYERGGHFKFHKDTEKEDGMFGTMILQLPTQSGHEGGALEIRHGKKGVAFDFS
>gi|Smoe1000014051|ref|jgi|Selmo1|417764|fgenesh2_pg.C_scaffold_35000237
MVDRKALCEALTQEAAGIFACGGSIDDHLPSSISVQGIGEVALPISLDDAFRRQEAAEQAPHGLGNRTLVDTAVRRAWQIDGSKVLSYFSTATLSKPSLLLLFVAWGWIQVEAHLYKLLVYDSDGHFKLHRDTEKEPGMFATLILQLPTRHQHGNLTKTFDFDVPGFNYTAFFADCEHEV
>gi|Smoe1000014016|ref|jgi|Selmo1|417794|fgenesh2_pg.C_scaffold_35000267
MGLGIAKSLADRWWQSRIFPTTLSEPSLLLLFVAWGWIRLCKLLVYESDGHFKLYRDIKKEPGMFATLILQLPAWLPGRCSHGSAWQRHFDAGFKDTAFFADCEHEIQPDANSSPEKLVFCLEHKYTETNLCFSGLKGQDRLVVELLGNVPVLQCCSGVLAETQERRARLPGQDLKELWT
>gi|Smoe1000010456|ref|jgi|Selmo1|409722|fgenesh2_pg.C_scaffold_11000105
MACWSPSGSCRMENEFDVSWGKGWLVRSRLLLLCGPGIRCGALSDASNVMRWVTSWAWIWALDAPQSQRIGLMTQELLKVLWHGVSANTFAVGASASGNSSKISFPDNPEFKTVHDVASAALSNMGVDARAVELGVKLYKIYFYTTFFCDCKHKLDEVTRGPRLCLVFILSLMQEWASCS
>gi|Smoe1000005861|ref|jgi|Selmo1|428254|fgenesh2_pg.C_scaffold_100000026
MPTSLRVEGAGDLALPLSGDGFLKLKAVAQQASYGKGEETLLDTDVRGAWQVEASKVVLAMASPVIEELGIDAAALGLDVRLYTLLLYEQGGHVKFHRDTEKEDGMFATMILQLPTQSGHEGGKLEIRHGKKKSVTFDFSGADSRTGYFQTAFFCDCEHKLAEVTSGTRLCLVFSLVRKF
>gi|Smoe1000003542|ref|jgi|Selmo1|413175|fgenesh2_pg.C_scaffold_20000036
MDHQALLKVLWSTVPRRKYATGGEVDASKFAMPASLRVEGAGDLALPLSNDGFLKLKAVSQQAPYGKGEETLLDTDVRRAWQVEASKVSCPQTPDFFSKVVLDMAVPAMEVLGIDAAALGLDVRLYKLLLYEQGGHFKSHRDTEKEDGMFATMILQLPTRSGHEGGNLTIRHGRKKSVTF
>gi|Smoe1000003520|ref|jgi|Selmo1|413194|fgenesh2_pg.C_scaffold_20000055
MKERKATKRFVDYQALLKVLWSTLPRQKYATGGEVDASKFAMPASLRVEGAGDLALPLSGDGFLKLKAVSQQAPYGKGEETLLDTDVRRAWQVDASKVSCPQIPDFFSKVVLDMAVPAMEEMGIDAAALGLDVRLYKLLLYEQRGHFKSHRDTEKEVGMFATMILQLPTESGHEGGKLEI
>gi|Smoe1000003056|ref|jgi|Selmo1|420503|fgenesh2_pg.C_scaffold_46000146
MAQLNKALWCDREVFATGGAEKIDLPIKSVCQQAPYGLGERTLVDTDVRRAWQINPERVSCPEVPNFFSSVAHDLAAHGLSVMGLDSTALGLDGQDGMFATMILQLPTSTGHSGGALAVRQGRQCCEFDFSDGSSSHYFSTIFFCDCEHQLKEVTSGSRLCLVFSLVRSNVEFIPAENVP
>gi|Smoe1000002216|ref|jgi|Selmo1|427913|fgenesh2_pg.C_scaffold_96000020
MAQLNRGEIDVAKFRMPVSLLINPERVSCPEVPNFLSSVAHDLAAHGLSVMGLDSTALELDVKLYKLLLYEAGGHFKFHRDMEKEDGMFATMILQLPTTTGHSGGAQGRQCCEAQPRPARHLLELVLAVAKEDRREVVAGAPVREIKLPSFVSALPQVRALFQQWAATGDGQADRDAVGK
>gi|Psoj1000015702|ref|143738
MRWRKEHVALSRSAPPHARISPLGDGCKKISSLLAKADDNAGGYSFGGVADTLPAVPGLYVEGLGDVSVPLCPEQADKLIAICDKSPFGRKLDTMMDESVRKSWQLAPDQVDVKNPLWHCGMDKLSETVASRLGYKDVPLQCKLYKLLAYGEGGHFVKHQDTEKEDGMIATMVVQLPSLH
>gi|Psoj1000015557|ref|143565
MPSSDDIEMLSPIEENEEKELGPDEQQQEHSSDEQEDEEIDSDEEEELGPLKRKWPFDGEGRSRDVPLPVGATSKKINTILAQAEANAGEFTFGGVADTLPAVPGLVRKSWQLQPDRVQFKNPVWQTGMDKLAHTVAERLGYKSVSMECKLYKMLVYGEGGHFLKHQDTEKEDGMVATLV
>gi|Psoj1000015551|ref|143559
MTLCEIVASADEYAGEYSFGGRADTLPTAPGLRVDEVSVLLPLDAAQAQKLIAKCDKSLFGHGMETKTDESVRKSWQLEADQVEIKNPLWQTGVDKLTETIASRLGYKGVPLRCVLYKVLVYGEGGHFLNHQDTEKEDDAEHALEKVTKGYRLALVYSICLPASMRHLKRESGRMLSDEL
>gi|Psoj1000015548|ref|143556
MNVRLAAFVAAIATLASGASAASCPFGYDKIASVKQQDRQISTELYDAKSCDVIDYELVKKDLEVLMTNSQDFWPADFGHYGGLFIRLAWHCNGSYRRADGRGGCDGGRIRFNPEHSWADNTNLDKALKLLDPIKKKYGDALSWGDLIVLSGNVAIKSMGGPVLGFCGGRRDDVDGTSSL
>gi|Psoj1000015558|ref|143566
MNVRLAAFVAAIATLASGASAASCPFGYDKIASVKQQDRQISTELYDAKSCDVIDYELVKKDLEVLMTNSQDFWPADFGHYGGLFIRLAWHCNGSYRRADGRGGCDGGRIRFNPEHSWADNTNLDKALKLLDPIKKKYGDALSWGDLIVLSGNVAIKSMGGPVLGFCGGRRDDVDGTSSL
>gi|Psoj1000012378|ref|139870
MSDSDEREVLYDQDPDLEEDFGEKDWPFAGKGDLDDVPAPTGAACTKISNVLGGASEHLGDFSFGGQADLLPPVPSLFVNGVGPVPVPLWEERALKLIDKCAKSPFGHNMDTKLDESVRKSWQLAPDQMAVAIADRLGYKGIPLQCVLFKLLVYGEGGHFVKHQDTEKEDGMIATLVVQP
>gi|Psoj1000012376|ref|139867
MGKELHNQDPDLEEDFGEKDWPFAGKGDLDDVPAPTGAACTKISNVLGGASEHLGDFSFGGQADLLPPVPSLFVNGVGPVPVPLWEERALKLIDKCAKSPFGHNMDTKLDESVRKSWQLAPDQMAVAIADRLGYKGIPLQCVLFKLLVYGEGGHFVKHQDTEKEDGMIATLVVQPPSLHE
>gi|Psoj1000009248|ref|136290
MDLRMPLGLDLLTVCMAAVPEERELEKDFENGKWPFGGTGGPEGVPTPAGAKCVEISDLLARTDGQAGEYSFGGPADSLPAAPGLHVDSVGAVPVPLSDDVAARLIAKCEKSPFGHNMDTKMDDEVRNSWQIGPDKVQLKNPKWSVGMSDMMLEISERLGYKDIKLGCSLYKLLVYGEVR
>gi|Psoj1000008778|ref|135759
MCKKSPFGHNLETKMDEAVRQSWQLVPSQVRFENMTWQGGMEKLCQLIADRLGYRDIPLQCVLYKLLVYGEGGHFVKHQDTEKEDGMIVTLVVQLPSSHEGGDLVVYRGGEVRHRHDFGKADGTAAFLPHYAVHYADAEHALEKVTKGFRLALVYSICLPPTMRHLEKAHDAPLSEDLAG
>gi|Psoj1000008776|ref|135757
MCKKSPFGHNLETKMDEAVRQSWQLVPSQVRFENMTWQGGMEKLCQLIADRLGYRDIPLQCVLYKLLVYGEGGHFVKHQDTEKEDGMIVTLVVQLPSSHEGGDLVVYRGGEVRHRHDFGKADGTAAFLPHYAVHYADAEHALEKVTKGFRLALVYSICLPPTMRHLEKAHDAPLSEDLAG
>gi|Psoj1000005084|ref|131556
METESLEEDFDDGKWPFTGEGEFADVLKPKGVGCAQVSKVLARADEHAGEYSFGGLADTLPAVSGLFVGTVGTIHTPLVADQAEKLIALCKKSPFGHNLDTKMDENVRKSWQLDANLVRFKNPLWQVGMNKLQKVIADRLGYTDVPLECVLYKLLVYGEGGHFVKHQDKEKEDGMIATLV
>gi|Psoj1000002636|ref|128746
MPERFATHSDAENERNSSDEEDELGPVNGKWPFGGVGRSFDVPVPSGATCMKINKVLAQAEANAGEYTFGGVAEALPALPGLVVEDVGQISVPLTESCAKKLIDKCEKSCYGHNFDTKMDENVRKSWQVEPERVQCSLYKMLVYGEGGHFVKHQDTEKEDGMVATLVIQLPSTHEGGDLV
>gi|Psoj1000001369|ref|127244
MSDDERKELFDQDPDLVEDFGYNEWPFAAKGDLNDIPISDVLGGASEHSGDFSFGGQADLLPAVPDLFVDGVGPVPVPLWEERAQKLIDKCEKSPFGHNMDTKMDENVRKSWQLAPGQVQFKNPMWQAGIDKMAVVIADRLGYKGIPLQCVLYKLLVYEEGGHFLKHQDTEKEDGMIATL
>gi|Pram1000011751|ref|86281
MDEKVRKSWELAPDQVQIKNPLWKAGMDTLSEIIAGRLGYKGVSMQCKLYKLLLYGEGGHFVKHQDTEKEDGMVATLVVQPPSEHKGGNLVVYRGGKAAQRHDFGKKDGTAAYLSHYAVHYADAEHALEKVTKGYRLALVFSICLPPNMHHLIRNHDIPLSEELAAAMGRLNSDTDSFAL
>gi|Pram1000006627|ref|77035
MSVRADSPDQVDIKNPLWHSGMDTLSEVIAGRLGYKGVPLQCKLYKLLVYGEGGQFVKHQDTEKEDGMVATLVVQLPSLHEGGDLVVYRGGKEVQRHDFGKKDGTAAYLPQYAVHYADAEHGLEKVTKGYRLVLVYSICLPPAMRHLMKSSDDSLGEELGSFISGMEADENSFALLLAHE
>gi|Pram1000006622|ref|77040
MASVEEDFEEGIWPFGGEGDFDEVLLPKGAACLKISEMLASADQKAGEYSFGGVVDTLPVAPGLFVEEVGAISVPLTSEQAEKVIVKCEKSPFGHNLDTKMDENVRKSWQLAPDQVKINNTQWESGMDKLVTTIAQRLGYENVPLQCSLYKLLVYGEGGHFLKHQDTEKEDGMIATLVIQ
>gi|Pram1000006625|ref|77037
MAWGNRWGGGSWGDRYYEEEQQYDTTQPTEEEVFKKGRWPFGVKGEAEDIPLPKDACKKSPFGQKLSTKMDEKVRKSWELAPDQVQIKNPLWKAGMDTLSEIIAGRLGYKGVSMQCKLYKLLLYGEGGHFVKHQDTEKENGMVATLVVQPPSEHKGGNLVVYRGGKAVQRHDFGKKDGTA
>gi|Ppla1000002992|ref|jgi|Pospl1|128047|estExt_fgenesh3_pg.C_1210047
MTDRPQRDSEPMALAAPAGTYGHVGGKTIIMYPLASSQPSDSEDEEDVDAQLDECQDIESTSITNACVREVEFSGPNFLNAQYRYDNIGMGLKSDGNKESMAQQGYDDDGREEGAVDEGGEEDDKEKEVGDEGMTDIKGEDMEETEEDEEKQDEPLENPSTDIKGDLAQILSEDSSVKGS
>gi|Ppla1000002318|ref|jgi|Pospl1|23394|gw1.145.11.1
GSFAFHHSYPKAPNPTLYLDGLGHIGLPLSIRDAEAIKSKANQAPFGMGERTVVDKSVRDTWELDASQVRFDNPAWDPFLKRVVQEICTTLGVNIAASEPRCELYKLLLYETSSHFLPHVDTEKADGMFATIVVVLPSKFKGGSLHVSHGGLNTFYDCSPKSLTETNVLAWYTDVMHEVK
>gi|Ngru1000014932|ref|jgi|Naegr1|81486|estExt_fgeneshNG_pg.C_600108
MQDDKSSIPEGVVADEETNPTIPGHESTIEQNPSSTEESKMTYPTLHKSLKKSFSQIVSCGGSFVCSDKFDGVTLPGLELLVSAAQDSDKKYKAMPFPIVYKEQADEIIRIAKKSPYGKGEETIYDDNVRNSYELEPNQFRITNTMWQKELNQLLETKIKSGLGIDKYKNVECKLYKLLL
>gi|Ngru1000014072|ref|jgi|Naegr1|79941|estExt_fgeneshNG_pg.C_270061
MEQNETLASSHAETNEHHATNEIKTEEVNDTDSSDQANESDSDHDNSFNMEDLFGRLQNTFTNITSVGGNYYYYDRLPDVMKVPGIELIITEENVEKYKPIALPIIYKEQADEIIRIGKKSPFGKGEETIHDQVRNSFELEPHQFRITNPVWQKELHLLLNSKIKSGLGIEKHKKVVQSP
>gi|Ngru1000013028|ref|jgi|Naegr1|77008|fgeneshNG_pg.scaffold_833000001
VNTKIYKPIRLPIITKNQAFEIIQYFEKAPFGQGDKTIYDETVRKTWQLSPSQFRITNQHWKQYIDNLVEKQVKPGLGIHSSVVVRNSLYKLLLYEEGGHFDFHRDTEKEDKMFATLVVQLPSLYSGGEIIVQHADSEETYEFAKEGSSKPFFFSFYCDCNHKVAKLTSGYRLCLVYNVC
>gi|Ngru1000011382|ref|jgi|Naegr1|73904|fgeneshNG_pg.scaffold_64000090
MTKRKAKSTRSPSPIFDINVLQNMLKARNDIFTLDDYGDDFSNSDDYNSDSDDWMDEGLSYDGPTDPVLRPCDLVAEKLLPAEKNIKKRMSDVLELSGFDGEFWSGGKLHSDYMMPSIEMLDEESGTYRLLPLPIVYEEQVNDLIRSSMNAGEKNLWKLIEPKKFRITNPKWDVLMQELL
>gi|Ngru1000010562|ref|jgi|Naegr1|72292|fgeneshNG_pg.scaffold_52000076
MGNRQSAQRTFFDSDDSDDYACSYDSEEEYTRQGPVASLPIIIRETIAESLPEVNIKKQLIQVMNSTDRSGDFSTGGKLSAEYSIPAIEILQESGVYKPLPLPIIYEEQAKDLIKNCSMSPFGRLDKTIYDESVRKSWQLDPQRFKITNPKWNSLIEDLVNNNVKKDLGIAQEKKIGFSL
>gi|Ngru1000008142|ref|jgi|Naegr1|67363|fgeneshNG_pg.scaffold_23000035
MQNETQQPQEIINEESIDFESNEQEDEDCDVDLNSSPPQIIQDDRTSSPTFEDVISESGDSIALDDTEEENRDDKENKNAKTTSEQTTETREGNNVIDITEESEKCVSIDSSDDEEEEEEINQNDKTTENEEDEENEEESDEDFILESDEEEDYHPSYTTRRMHKLIEELIPTVTAGGSF
>gi|Ngru1000005317|ref|jgi|Naegr1|59810|estExt_fgeneshHS_pg.C_760047
MNTSSLNTNNAEEQFELGIAILNKAWKTKELQMLDEIEGLKKIGQKKQEEIHQINQQHSKLQFEYQSLNTAMKTIQEMNQDLQSERDRLSAQVKALNGELSKLKAFKKQIIQSINLEELEDPSSNGNNTSQFIPPTQTNNSMMSENRSFNSSNINNQSIMQQQTQSQQAQALEGKEFFKQ
>gi|Ngru1000003599|ref|jgi|Naegr1|52657|fgeneshHS_pg.scaffold_58000065
MQAPNGPIIIIDDADEEMMINGDTLPHTMIISSEEEEANSNQIVDSEEEEIEQEKILYKTKVATRIDSQIQELIGGGKVHEALERFCGLVSEGSLQFANFACGGKLDDSMSFLPGIEVRNASGDYSDKQNSMMLSLPIISNEYMNDLKRFSYTCTKTNVNNPYNFMCCVKNNNWSRAWEQ
>gi|Ngru1000002982|ref|jgi|Naegr1|50278|fgeneshHS_pg.scaffold_34000025
METIQVRERIEEKVLAFVRELVSSPSSDIPNEKGFLLKDLTKYPSTFISKFQVMNLVYQLCTLNKFSTKRDLYYCNKPILKDQTQLDSILKDLSSSWRVNRHDLHIVPQSKGEVCGPIAWKEISLIDIENRGQQPTTTIVSASKGCNLPSIPHLQIVEFSIPVEVDKVIIVEKESVFQVL
>gi|Ngru1000002448|ref|jgi|Naegr1|48033|fgeneshHS_pg.scaffold_16000074
MQFNNLLGNTTTPNITTINSGDGISSTIGSSTNDILSTSQLLNNLISHNSQGSSSTNTFNLITLKPMEFPLPNIMSESFESSFVFDNSHRPKPTSPIMFTIPYQRHSLILPDELVIGYSETTKEMADLLTSSETDVCVGEFSCGGEVENIPGLPGLQICVPSNESGATVELKHCSLPITN
>gi|Ngru1000002312|ref|jgi|Naegr1|47425|fgeneshHS_pg.scaffold_12000094
MVKKAKKPSHHDEEEITDEDFMDIMDCLNPRQFEQMSRSLGYTVRDYDHDSDDSYFSDECIVAPTPELETLLIKRDCLAEKLLPPDKNIKKRILDLLEQSCVLNGDYITGGKLDSNYLMPAIEILLEDNGKYKPLPLPIVYAEQVTELIRNCSFGRTENVNITKSWHLDPHKFRITNPKW
>gi|Ehux1000028524|ref|jgi|Emihu1|195256|gm1.300307
MNDDDPTLAALSSAVDSIESRYLTCGQFSMPAPLQITASGGQTATFPIKRGEQLTEIKKLTDVAAPSPFGLGGETVVNTAVRQASQKIRTELVPDAGAVTAKLHKVNIYEKGGHFAEHRDTPRSETHFGSLVLLLPAYHAGGALCVDHAGQRKVVDWGKTGDWAPRYHYGFGADAAKRLA
>gi|Ehux1000027221|ref|jgi|Emihu1|206011|gm1.2500281
MMADPMAESSSNPLMAESSSNPSHAFAVGERVHARVGAEWKMGTVVDLQSSEHGSAVTYEVKLDGTARVADGATLYVSDDANEIRVPCEVGSPANPKNARDSLTDLLQECTGTGSNYCVGGEWTAGVLPEITHQETGEFRYPLSAEQATAFITASQRSGVGTMTDTVINLEARKSWELLP
>gi|Dpur1000009437|ref|jgi|Dicpu1|156983|GID1.0047637
MVFLNNNEYYDYFNLTKLLQKDNNTIGSSISIVKPKHSYYSYGSIDLKNLPPISININNNDNLEFQPPYDNSEIKELVLKLENDKIEQDSQDKKEIYEIPSNVIDIEAFENGNKLMEQILERVKNDFGILKSENIKAKLNKLIIYGEGSAYQIPIENNQHHSDGHNVFGSLQILLPAGES
>gi|Dpur1000007623|ref|jgi|Dicpu1|98697|estExt_fgeneshDP_pg.C_2240006
```

Back to Contents

5b. Multiple sequence alignment of domain C-terminal to the 2OGFeDO domain, called X in Figure 2

Back to Contents

5c. Multiple sequence alignment of the R3H domain fused to the 2OGFeDO domain



Back to Contents

6a. Phyletic distribution and multiple sequence alignment of the DNA glycosylase associated family of 2OGFeDO domains

gi	Domain architecture	Gene name	Len	Species	class
167519270	2OGFeDO+X+coiled_coil?	MONBRDRAFT_31538	1071	Monosiga brevicollis MX1	choanoflagellida
145348837	2OGFeDO	OSTLU_32491	334	Ostreococcus lucimarinus CCE9901	viridiplantae>chlorophyta
145351789	2OGFeDO+MDB4	OSTLU_17228	994	Ostreococcus lucimarinus CCE9901	viridiplantae>chlorophyta
145356086	2OGFeDO	OSTLU_28402	391	Ostreococcus lucimarinus CCE9901	viridiplantae>chlorophyta
145341417	2OGFeDO	OSTLU_28816	292	Ostreococcus lucimarinus CCE9901	viridiplantae>chlorophyta
116060306	2OGFeDO	Ot11g00370	192	Ostreococcus tauri	viridiplantae>chlorophyta
116058658	TIP41?+2OGFeDO (Artificial?)	Ot07g00920	619	Ostreococcus tauri	viridiplantae>chlorophyta
116056119	2OGFeDO	Ot18g02050	644	Ostreococcus tauri	viridiplantae>chlorophyta
116000439	2OGFeDO	Ot01g02080	286	Ostreococcus tauri	viridiplantae>chlorophyta
219123239	2OGFeDO	PHATRDRAFT_47703	719	Phaeodactylum tricornutum CCAP 1055/1	stramenopiles
219110259	2OGFeDO	PHATRDRAFT_43074	401	Phaeodactylum tricornutum CCAP 1055/1	stramenopiles
219122046	2OGFeDO	PHATRDRAFT_37198	434	Phaeodactylum tricornutum CCAP 1055/1	stramenopiles
219109682	UBRbox+2OGFeDO	PHATRDRAFT_42585	444	Phaeodactylum tricornutum CCAP 1055/1	stramenopiles
219127159	2OGFeDO	PHATRDRAFT_49155	345	Phaeodactylum tricornutum CCAP 1055/1	stramenopiles
Ehux1000018843	2OGFeDO	Ehux1000018843	171	Emiliania huxleyi	haptophyceae
Ehux1000006770	2OGFeDO	Ehux1000006770	238	Emiliania huxleyi	haptophyceae
Ehux1000014103	2OGFeDO	Ehux1000014103	243	Emiliania huxleyi	haptophyceae
Ehux1000026600	2OGFeDO	Ehux1000026600	250	Emiliania huxleyi	haptophyceae
Ehux1000020286	2OGFeDO	Ehux1000020286	256	Emiliania huxleyi	haptophyceae
Ehux1000018838	2OGFeDO	Ehux1000018838	270	Emiliania huxleyi	haptophyceae
Ehux1000005452	2OGFeDO	Ehux1000005452	306	Emiliania huxleyi	haptophyceae
Ehux1000023908	2OGFeDO	Ehux1000023908	317	Emiliania huxleyi	haptophyceae
Ehux1000017516	2OGFeDO	Ehux1000017516	336	Emiliania huxleyi	haptophyceae
Ehux1000016671	2OGFeDO	Ehux1000016671	344	Emiliania huxleyi	haptophyceae
Ehux1000030931	2OGFeDO	Ehux1000030931	363	Emiliania huxleyi	haptophyceae
Ehux1000020393	2OGFeDO	Ehux1000020393	366	Emiliania huxleyi	haptophyceae
Ehux1000023178	2OGFeDO	Ehux1000023178	389	Emiliania huxleyi	haptophyceae
Ehux1000001690	2OGFeDO	Ehux1000001690	398	Emiliania huxleyi	haptophyceae
Ehux1000029539	2OGFeDO	Ehux1000029539	447	Emiliania huxleyi	haptophyceae
Ehux1000030432	2OGFeDO	Ehux1000030432	463	Emiliania huxleyi	haptophyceae
Ehux1000028434	2OGFeDO	Ehux1000028434	483	Emiliania huxleyi	haptophyceae
Ehux1000028445	2OGFeDO+Ankyrin_repeats	Ehux1000028445	751	Emiliania huxleyi	haptophyceae
Aano1000001909	Polygal+ZnR+2OGFeDO	Aano1000001909	1238	Aureococcus anophagefferens	stramenopiles
Aano1000005937	2OGFeDO	Aano1000005937	453	Aureococcus anophagefferens	stramenopiles
Mpus1000005868	2OGFeDO	Mpus1000005868	393	Micromonas pusilla	viridiplantae>chlorophyta
Mpus1000002964	2OGFeDO	Mpus1000002964	681	Micromonas pusilla	viridiplantae>chlorophyta
Mpus1000005493	2OGFeDO	Mpus1000005493	253	Micromonas pusilla	viridiplantae>chlorophyta
Mpus1000006879	2OGFeDO	Mpus1000006879	274	Micromonas pusilla	viridiplantae>chlorophyta
Mpus1000009569	2OGFeDO	Mpus1000009569	473	Micromonas pusilla	viridiplantae>chlorophyta
Mpus1000009687	2OGFeDO	Mpus1000009687	169	Micromonas pusilla	viridiplantae>chlorophyta
Mpus1000009687	2OGFeDO	Mpus1000009687	169	Micromonas pusilla	viridiplantae>chlorophyta
Vcar1000003331	2OGFeDO	Vcar1000003331	461	Volvox carteri	viridiplantae>chlorophyta
Chlo1000005889	2OGFeDO	Chlo1000005889	413	Chlorella sp.	viridiplantae>chlorophyta
115375244	2OGFeDO	STIAU_4281	212	Stigmatella aurantiaca DW4/3-1	proteobacteria>deltaproteobacteria
108763376	2OGFeDO	MXAN_2402	212	Myxococcus xanthus DK 1622	proteobacteria>deltaproteobacteria
78184862	2OGFeDO	Syncc9902_1289	199	Synechococcus sp. CC9902	cyanobacteria
116070736	2OGFeDO	BL107_13860	199	Synechococcus sp. BL107	cyanobacteria
78212712	2OGFeDO	Syncc9605_1181	199	Synechococcus sp. CC9605	cyanobacteria

```
Predicted secondary structure       -HHHHH---------HHHHH---------------------HHHHHH---HHHHH---------------HH---------------------HHHHHHHHHH----HHHH-------------EEEEEEEEE-----------
MONBRDRAFT_31538_Mbre_167519270     TSVYPDDLLEALKRDGLLLD---------------------QMPSDNFKFGKR----------GTWWYPLTSKS---------------------------------THAPRFAAE------AVHILKWYQVREKTD-------
OSTLU_32491_Oluc_145348837          ---VHGAAIDGDEGAYFATRARDAFARTAAHRARLADS---CRKIDAL---------------GTLWLGAR------------DAPR--C---LERFVKSVYD----AHARGVTRD------ARSGVEFWARTR----------
OSTLU_17228_Oluc_145351789          VLRYAEEGARERAPEDVARVYDDALD---------------ARALRFLGDAFARRG-------ASTFWIEHDYYAPETGFFSYVHVPSRRENKMDAVVDMVKD----VASRFPGVK------RARYAEWWAHSRPHTD------
Ot07g00920_Otau_116058658           ---VHGATTRGADEETLFRAEALAKFLDVDDALAEK-----CRAIERE---------------GSRWLSSV------------DAPR--C---VERAARAVFE----RYAKGWEYD------ETSGVEFWTQTE----------
OSTLU_28402_Oluc_145356086          --------RRTSGTSSGRTFWV-------------------NAESEPRCA-------------LEYLARETF------------RRA------REAAARAERDAGARDEDERARYD------SKSGAEWWTQVIDEGD------
OSTLU_28816_Oluc_145341417          VGAFVDAYVALADGGYDPATSREGWLDEDAVGG--------DDDATTTTTTTTTTTTLT----SAFWTPRRA----------LAAPR--C---IEALAGGC------AAALGLADD------ALAGAEWWIQDVAWNE------
Ot01g02080_Otau_116000439           LVAFVDAYVKIVDAGEEPTSELWSPS---------------DETFEDGTGDRATRASLT----SAFWCPVDA----------FDAPR--C---VEAAIGALA-----RSGRATAST------RIVGAEWWIQDVAHDE------
PHATRDRAFT_47703_Ptri_219123239     NEMATKLPKETIKDPAMYRQQAKFCGCAQGALLPQTLYQQLCSVFAPEAAYWQESGYAH----RGYYSYFMD------RPSSQQTPR-NV---EDVVMNYVLP----AVETHLGGPTA----GICGFEWWIHTRPIQANL----
PHATRDRAFT_43074_Ptri_219110259     -TCMVNFVAPNTNDCHFLDELL-------------------LDCEVADNGLMP----------RTFWVPSAT---------EDFSPR--C---LESMARDVFT----HHTSGCAYD------GTSGAEWWVQIRPSPEAGRYSM
PHATRDRAFT_37198_Ptri_219122046     --CWAPAHLESFRDDCRTVFS--------------------ARTRDDDQAYSAG---------VTYFCPSQM------------PPR--C---LEALALDIFR----KHTEHLENGVMV---EQSGAEWWTLVLDDDDS----A
PHATRDRAFT_42585_Ptri_219109682     ---------RPTGEDGKYVQQTFSIPALDDDDL--------CSALCLQSEELVKHSK------ETFWLDAST----------DLEHC--C---LEKLAWHIYR----NQINEYKLQGD----LAGGAEWWVQVKSVSLP-----
PHATRDRAFT_49155_Ptri_219127159     SIAFVQQPRSSLSNTAPCRIWDQFLSKDELNCLL-------TAFYDPNAHYWTSHQYAVEPP-SPYFSYLIP-------LDPETGRNEYG---IGSIIRRIQR-----QIAEWKPLV-----QCTYCEMWAHNRPHAT------
Aano1000005937_Aano_Aano1000005937  DEVLRYGAAPAAAPPALAPRQCAAFDDALPGWLFDAL----ARCFAADGGFWPAHGYDVDGGGAGFFSYVHSIADDDSDSDGDDEPA------TRVAARALRG----LFARAMPQL------AARAVEWWAHCRPHGC------
STIAU_4281_Saur_115375244           MNSLVQLTDGALPEPLFRRL---------------------VRRVGALGSERLRTTYQ-----TTFWYGFGD---------------------TNVVEEVILA----LHPRVVGR-------RIAGVEWWLSRMFTTDV-----
MXAN_2402_Mxan_108763376            METLVRLTENAVPPLLFRRL---------------------LRRLGAVGTERLRQTYQ-----TTFWYDFGP---------------------ANVVEDIILT----LRPHIAGK-------RVAGVEWWLSRMYPTDV-----
Syncc9902_1289_Syn_78184862         MLQVVDNLLPPIALQELRDL---------------------CNIHERLKQQHPS---------DALFSWRPDSGRARSSHAPE----------QQAAMDTYLH----QHLRPLVSPWC----QATGAEWWCNTNN---------
BL107_13860_Syn_116070736           MLQVIDNWLPTTLVQSLQQL---------------------CDAHGALRQNQLE---------NQLFSWRPDHGTPRSAHAPE----------QQAVMEHYLS----EFLRPLLAPWS----TATGVEWWCNTNN---------
Syncc9605_1181_Syn_78212712         MLHVIDNLLPASALQDLRDL---------------------CDIYGRLKEQHDG---------DAQFSWRPETGSPRSIHAAA----------QQAVVDRYLD----EALLPLATPFA----QRAGVEWWCNTNN---------
consensus/80%                       ...h.....s...............................sp.....................sha...........................p.hh..hb...........h..........suhEWW.p............
```

Species abbreviations: Aano : Aureococcus anophagefferens; Mbre : Monosiga brevicollis; Mxan : Myxococcus xanthus; Oluc : Ostreococcus lucimarinus; Otau : Ostreococcus tauri; Ptri

Fasta sequences of proteins from incomplete/unpublished genomes not in GenBank and assigned Fake gis

```
>gi|Ehux1000018843|ref|jgi|Emihu1|461717|estExtDG_fgeneshEH_pg.C_70042
MPEVAEAGHAEWWAHCRPHATGHQMHFDSADEGRGARGPRHPLVSCALFLTSSAGVVPGRGARGGAARVSLMVAFWPSLEAREAAEPSAARPFPYGGAPEWAPLFDWPEEEDGDATGGGGGESEGREGDSTRPAPLRLVSPIWQSVQGGGAPAAGLRRMPPYEACFQGF
>gi|Ehux1000006770|ref|jgi|Emihu1|243341|gm1.38500015
MAEAEEARLRRVIAAGARAAARLDTSSTSAKLKQAAGRVAAGATAQHDLALLCSRQGRGSEADRLLRALGMRHRLSVSIFAPPSTATPSPAGLCGAPIVLDGVLPPAVIAGLQRALRPGAAFWEEHGYPTPCFFSYRHDLTASKPPLLLVEQAARLLQPLMERRRALHQLHFDLDEATLL
>gi|Ehux1000014103|ref|jgi|Emihu1|229148|gm1.15300056
MEDDEVDIPVLDGRGRVHGTAGLCDTSLGVWDDAVRLLNPAALRAEAQSAAWPLAASSFWLAADAAPACALEALAAAIFKLHAGSAAGRFDPACSGAEWWANVSDSASVCRERDGYGKVALHFDKDEAVYAAAGVFVHPLLATITHLGEEGAPTLIVPGGQVSAEGEYAASRLPEAVLVP
>gi|Ehux1000026600|ref|jgi|Emihu1|213481|gm1.5700233
MKLVLSLVVFASGHAIQVFDGVFAPPTCEALNVAANARGLGHALYTRSVGPRTALEHAFESFLSEVDDDSPHVEYWSRQEWKHIEAHADVDEALAAGGGPLRYPRHAHVLYLDVGSRVRGPTCVWEAGRGEDGRPPRFAGAMTVVPAVAGRVLRFDGSLQHAVPRPADVWLAPFVISQSG
>gi|Ehux1000020286|ref|jgi|Emihu1|210068|gm1.4100118
MLLHHASGGDLEVSRRVVGAEWWVQHRSLSAPKGFHFDVDQEQQRRGGALRSPALSSILYLTDAGGPTLILDQVATICSWSGHVLLSPEPAEAELIHPLANRFVLFRAACLHGCLPPLGAGCGVEGAEAAEQPKRTTLLVNWWIGERPAPPSCAEAPPSLLDEMRRASAPHYAEALAAGG
>gi|Ehux1000018838|ref|jgi|Emihu1|229943|gm1.15900107
MCLSSGSSEGSGSGSDEQGGGSDSDSEDINVDEQGRPAIGGSIDLSTRVFDGALPSGADPGPLIADCEAVFTARATGAGEGLSSGATFWVAANAKPKTALERLALQIFAHHAAGAEYDPSRSGAEWWTVVIDDTDDGLLLHPHLATVTYLACPAAAAPTLVLERPSPLLATETAPPPCGA
>gi|Ehux1000005452|ref|jgi|Emihu1|246921|gm1.53700004
MCLSSGSSEGSGSGSDEQGGGSDSDSEDINVDEQGRPAIGGSIDLSTRVFDGALPSGADPGPLIADCEAVFTARATGAGEGLSSGATFWVAANAKPKTALERLALQIFAHHAAGAEYDPSRSGAEWWTVVIDDTDDADQGLLLHPHLATVTYLACPAAAAPTLVLERPSPLLATETAPPP
>gi|Ehux1000023908|ref|jgi|Emihu1|204717|gm1.2200174
MQREVAVYRDVLHPSLLAAVVQAVNVYLDSEQPPLPDSMHRTLCATHWAPLTDEEPAAHPIEQLVAALYRLIRRGASPLHADEAAGLAGGEWWVQEQALDDCPKEFHTDQDLCLLPGRDGVQSSVAHPALSSVFYLDGIGGATAVFAQRRCVVARGVSTLEPPLPPHAALAFPTPNQLLL
>gi|Ehux1000017516|ref|jgi|Emihu1|202565|gm1.1700088
MISMIRGFEDGSTTGGASPRDVAPDYSTRVWRSVLPAAAIPRLAEDCASAYLKVKTTTRNRGIEYETNATYWIGASCAPRTALEAVALDIFRHHSRGVRFDASSSGAEWWTQVVDEELDIPFHWDRDYELQKEEGVCLHPHVASVTYLTGLGGAPTLVLPVASPVLAADLPAACAGRVRS
>gi|Ehux1000016671|ref|jgi|Emihu1|206935|gm1.2800087
MLAVGGQAPSIDDILAHRVAGVSGVDYRPRSLWLKLGAPQARALATDCAHVFGAQEGSNWLGATATPRCALECLARAVFERHTAGATFDASRSGAEWWAQVRGGDGARAEGIEFHWDVDEHQCDRARGVHVHPHLSTVTYLSAHGAPTVVLDARSRASSPRATARLHGPLSAGGQLSLRS
>gi|Ehux1000030931|ref|jgi|Emihu1|96145|fgeneshEH_pg.4__375
MARRMKRPVSRREYWRVRAKARRTRRDAQEVSFDSSVLCQSCGFNVYVHERAVDLSAAQMRTLRGDACALRARHGAASFFLPLSAAVDLASEPRHPASPGCSPALLRYVAQLLLSELQDQQDRARLHIAGLLDGGTCGAEFWVQRRAAGEHAINWHWDKDESLLDSSGLTAHPLVSTVTY
>gi|Ehux1000020393|ref|jgi|Emihu1|219081|gm1.8300129
MNWSLGTLPLFGMLHRALALTVLLFVCESCGWLHSTAPSIRVWDGCFAARELDLVRAACAQRGHSFTSVYDRRPEDGPCRGGRTIIEAALNSLLDELDDDSRFVEFWYRDQWKPIDPHRDVDEALCRRIRLGPDGYGHQRCPHWGHVLYLDVASCVRAPTCIFDEDVGDALPRNDGKAPR
>gi|Ehux1000023178|ref|jgi|Emihu1|103706|fgeneshEH_pg.52__85
MVTDASGQPLAAVASTALPCQLLRAWLTELQGSDSPDARAHWWRESHWLPRGEVPGALLRATVAHVDAALPAAVAAAAGRCSGLEFWTQRRPTGSSMQLHWDCDEEHAQRHGELRSPLLSVVLYLDDHGGPTLVLDQRPPEALGRMVRASLVWPHAGAMAVFEGSLLHGVCGLGGAPTAA
>gi|Ehux1000001690|ref|jgi|Emihu1|249508|gm1.78900002
MVTDASGQPLAAVASTALPCQLLRAWLTELQGSDSPDARAHWWRESHWLPRGEVPGALLRATVAHVDAALPAAVAAAAGRCSGLEFWTQRRPTGSSMQLHWDCDEEHAQRHGELRSPLLSVVLYLDDHGGPTLVLDQRPPEALGRMVRASLVWPHAGAMAVFEGSLLHGVCGLGGAPTAA
>gi|Ehux1000029539|ref|jgi|Emihu1|198669|gm1.800424
MYGLQFAEVDHAASWQAAGLIDHFAAVHDDALMPAVFDAVESVAERLLRPDHPGGSKKIETESSFWMPLYEADGSRRAPLNALEAAAHQLHYLAFGDAPTPVIGGEWWLRGEDGDEADRGFRFHFDKDESHLKLRDEIRNPEVSSVTYLGMSGAPTLVLNQTIGHGANEMEPRLAPHGLL
>gi|Ehux1000030432|ref|jgi|Emihu1|96884|fgeneshEH_pg.7__257
MDDDARKLALLQQLAQPDSESEDEGESDDEDEGEGESDDEDEGEGESDASTAQDELSGIVPPLDAHGRVAGCPPVDRSTRTYRDALVLDDATLESLRRDCSVVYSLDGSNWQGADQPPRCLMESLALAVFERHTRGLEVDRTRSGAEWWAQVRTPGDQAEPIQCHWDVDEFCQDALGVAI
>gi|Ehux1000028434|ref|jgi|Emihu1|97804|fgeneshEH_pg.11__166
MAEAEEARLRRVIAAGARAAARLDTSSTSAKLKQAAGRVAAGATAQHDLALLCSRQGRGSEADRLLRALGMRHRLSVSIFAPPSTATPSPAGLCGAPIVLDGVLPPAVIAGLQRALRPGAAFWEEHGYPTPCFFSYRHDLTASKPPLLLVEQAARLLQPLMERHAADLYPGGGGGGGSAA
>gi|Ehux1000028445|ref|jgi|Emihu1|202849|gm1.1700372
MRLTTAAAVALTLQTHGCCSTQDSTTDDRCSATSDAWAAAGLADRFAVVHHHILTPDMVQQLLAAMPSLLSDSAWWMPLVDDAGKRRTPRSAVEEVLYHIYEADFGGQSTAVIGAEWWFRNTTASAGQDAHWDKDQALLEKESRIRFPEVGTVTYLSSGGTPTLVLDQRFGDDACLKVPG
>gi|Aano1000001909|ref|jgi|Auran1|62391
MFSVVAAAVCVAFGAAAPAPRVAFGAASPAPHVYPATGIVVDETASLTVGGVAVPLAFEAGSPARGVGPPPLPSTRARWASDAALPCELRVEGTADRWADAALRAAGRDLDFVLLEHGGAVVVRFSLGPGHAYLVDGRGGRRYYFWVDATAAAARPPGAVQVAAGAAGAGWLEAALRTHT
>gi|Aano1000005937|ref|jgi|Auran1|60487
MAGAAAALRLRARRRALRRGKAPKPSKQSQAPKAPPPKRVKKSRAPPQQSSQAAPQQSQAAPPATQKQKPLQLINEQLKLAKQAGGTPAVAERLRKTLLLAERARARRDSPPLAAAAREAARDYAVVLFQLGRSADAEAVLRAKFPEARFRLSDEVLRYGAAPAAAPPALAPRQCAAFDD
>gi|Mpus1000005868|ref|jgi|MicpuC2|51954
MASGSESESDDLCVVLDADGRVFGAGEGLDYAFRAFDAFAPLVPAGGRSPTSEDENDLTSSSPSTPPSSIDDLVADALLVDSPLAGDAATHWLSARAARAPRCALEELAAAIFAFHTRDLRGPDAFEESSSGVEWWVQARPSLKLHWDKDEELRVATGLWVHPQVSTVTYLTGGGSGGTR
>gi|Mpus1000002964|ref|jgi|MicpuC2|37850
MTAPAAEKRVVRVVEIWDPTVNAWTTASSTSGGRTSKLERALAMAEARGDDDRLEDADAMLTQLAAATRAAMDRAAGRRAREEEEEEEAMPTVPYVHDRHCQEETDDDDDDDDDDDDDDDDERSTRRRGSRTTTNELALTFARTMSARGWIKERMAGHAADGGGGDEPRRSRANAESFYVA
>gi|Mpus1000005493|ref|jgi|MicpuC2|50966
MDNYSQMHFDSDDEGSRASRGGDDKIRHPICSVVIFVNGCSGGPTLVTDQTPSSRRLARRGWLARPKARPPYTTGRVVVFDGTRLHGVVPGRGDVPGVPGDVPGPEGSHRATLMIAFWDKVKIRRAKGPGSSSFWPKGSARPFPPAEALTAALGGETIDGGIPVPAWPSLFMLPSGPEFA
>gi|Mpus1000006879|ref|jgi|MicpuC2|54518
MLTADAPVFYDASARRDTLRGSLVPGAPSPDGEEGGAAGGGPFGSSRCWHRRLGCLPPPGLGLRTALDDDDDDDDAAAPKPPPRATAFVSYPRTGKHVAFDGRWLHGAPEALSAKNMSDGKVGGVGKWRTVSDKRRRTTFLVNVWLDHVPSTAAPLPDDALREVSKEPYDDDDDDDDAAR
>gi|Mpus1000009569|ref|jgi|MicpuC2|63647
MRATRATPVALVLHLDDDDDDDDDDDDDDAARRRASASASVAIDANGRVAGADALDFSFRSFPRFLLLRDDDDAGDPRDGFRAKALLRDATRAFVGESFWLAADAAPRCALESLARAVFDRHVPADVLRAFQQTNADEARFGDGRFGGGCGAEWWTQVRRVDDECREAVAFHWDKDEDLV
>gi|Mpus1000009687|ref|jgi|MicpuC2|64346
MPVERILNPLNGVERAISVLSTSSALRSTGMVAQEFAGAEWWVQDVGPEENPKGFHTDCNFQTKTGNARNENVYQGNACRCAVFHPTVASVLYLVNNGGATAIFGQAKQFFSDYSILQPRLPTEVCIVNPKRNRVLVFEGDRYHAVLHPASSYEDSSCQRVRSSLALP
>gi|Mpus1000009687|ref|jgi|MicpuC2|64346
MPVERILNPLNGVERAISVLSTSSALRSTGMVAQEFAGAEWWVQDVGPEENPKGFHTDCNFQTKTGNARNENVYQGNACRCAVFHPTVASVLYLVNNGGATAIFGQAKQFFSDYSILQPRLPTEVCIVNPKRNRVLVFEGDRYHAVLHPASSYEDSSCQRVRSSLALP
>gi|Vcar1000003331|ref|jgi|Volca1|118201|fgenesh5_synt.33__11
MAPKGSVSKRKGAAAPSAPTPLSLYKRSLLAIRNWTPPTSARADAEALTGLAACCRLLQAQMGLTQPSPAAAGNSPLLAMITTTPITTMNPEAWAALEALEPGMASYSGVVRHSLDLDLGDWQDEPGFLLASAHLEIGRLLGRMFPRKELLEQLHLEAALKLFPDFVAAGAALGQANAAC
>gi|Chlo1000005889|ref|jgi|ChlNC64A_1|144156|IGS.gm_7_00098
MQEATVVAGILPVAHLEALEADADALADCPNYWVPREVAEGAAPALCVAEQAVQQLYRRLVQAQLAPHDGAWQGAEYWCQVSVCNQQEKVQAATLGWVYEGGRGLGFHFDKDEHVMAAEGRMINPILSSVLYLTGSSGDAVRQAPTVLIDQRYSQEQRCALPDNPAWSSLVFPVRNQYCL
```

Back to Contents

6b. Gene neighborhoods of the bacterial versions of the DNA glycosylase associated family

115375244	DNA_glycosylase-><-?||?->DAM_methylase-><-?<-?<-2OGFeDO*||transporter->FKBP->OPT_transporter->	STIAU_4281	212	Stigma
108763376	DAM_methylase-><-2OGFeDO*||transporter->FKBP->OPT_transporter->	MXAN_2402	212	Myxoco
78184862	<-4-diphosphocytidyl-2C-methyl-D-erythritol_kinase<-16S_rRNA_dimethylase<-?||?->?->?->?->2OGFeDO*-><-?||UmuD->UmuC-><-?<-?<-DNAG	Syncc9902_1289	199	Synech
116070736	<-4-diphosphocytidyl-2C-methyl-D-erythritol_kinase<-16S_rRNA_dimethylase<-?||?->?->?->?->2OGFeDO*-><-?||UmuD->UmuC->	BL107_13860	199	Synech
78212712	DNAG->?->?->?-><-UmuC<-UmuD||?-><-2OGFeDO*<-?<-?<-?<-?||?->16S_rRNA_dimethylase->4-diphosphocytidyl-2C-methyl-D-erythritol_kinase->	Syncc9605_1181	199	Synech
33865581	<-UmuC<-UmuD||?-><-2OGFeDO<-2OGFeDO*<-?<-?<-?||?->16S_rRNA_dimethylase->4-diphosphocytidyl-2C-methyl-D-erythritol_kinase->	SYNW1047	116	Synech

Back to Contents

7. Phyletic distribution, domain architectures and multiple sequence alignment of the MOM-family of acetyltransferases

# 1;						
Gi	Domain architecture	Gene name	Len	Species	class	GenBank
94311255		Rmet_2319	94	Ralstonia metallidurans CH34	proteobacteria>betaproteobacteria	hypothe
88811597		NB231_04310	99	Nitrococcus mobilis Nb-231	proteobacteria>gammaproteobacteria	hypothe
167841763		Bpse38_010100034127	121	Burkholderia thailandensis MSMB43	proteobacteria>betaproteobacteria	hypothe
91200323		kuste2620	128	Candidatus Kuenenia stuttgartiensis	planctomycetes	predict
91201384		kuste3681	128	Candidatus Kuenenia stuttgartiensis	planctomycetes	unknown
88810427		NB231_15143	128	Nitrococcus mobilis Nb-231	proteobacteria>gammaproteobacteria	hypothe
91200321		kuste2618	130	Candidatus Kuenenia stuttgartiensis	planctomycetes	predict
38707956		phi1026bp67	148	Burkholderia phage phi1026b	dsDNA viruses, no RNA stage>caudovirales	gp66 [B
88810452		NB231_15268	149	Nitrococcus mobilis Nb-231	proteobacteria>gammaproteobacteria	hypothe
88812041		NB231_06611	149	Nitrococcus mobilis Nb-231	proteobacteria>gammaproteobacteria	hypothe
21225140		SCO6848	152	Streptomyces coelicolor A3(2)	actinobacteria	hypothe
160932974		CLOLEP_01816	152	Clostridium leptum DSM 753	firmicutes	hypothe
87200732		Saro_2719	154	Novosphingobium aromaticivorans DSM 12444	proteobacteria>alphaproteobacteria	hypothe
167464664		Plarl_010100019262	154	Paenibacillus larvae subsp. larvae BRL-230010	firmicutes	hypothe
167462304		Plarl_010100007065	156	Paenibacillus larvae subsp. larvae BRL-230010	firmicutes	hypothe
167725778		BpseD_010100042573	156	Burkholderia pseudomallei DM98	proteobacteria>betaproteobacteria	hypothe
149882812		MPMinI_gp31	157	Microbacterium phage Min1	dsDNA viruses, no RNA stage>caudovirales	conserv
167747664		ANACAC_02385	158	Anaerostipes caccae DSM 14662	firmicutes	hypothe
160899565		Daci_4131	159	Delftia acidovorans SPH-1	proteobacteria>betaproteobacteria	hypothe
154499046		BACCAP_03038	162	Bacteroides capillosus ATCC 29799	bacteroidetes/chlorobi	hypothe
194563444		BDAG_01686	162	Burkholderia dolosa AUO158	proteobacteria>betaproteobacteria	hypothe
166031457		DORFOR_01154	162	Dorea formicigenerans ATCC 27755	firmicutes	hypothe
91209974		UTI89_C0944	165	Escherichia coli UTI89	proteobacteria>gammaproteobacteria	hypothe
167566477		BoklE_010100028224	167	Burkholderia oklahomensis EO147	proteobacteria>betaproteobacteria	hypothe
77747563		XF1762	169	Xylella fastidiosa 9a5c	proteobacteria>gammaproteobacteria	hypothe
110347378		Meso_4592	171	Mesorhizobium sp. BNC1	proteobacteria>alphaproteobacteria	hypothe
9106834		XF_1762	172	Xylella fastidiosa 9a5c	proteobacteria>gammaproteobacteria	conserv
221064772		CtesDRAFT_PD0109	178	Comamonas testosteroni KF-1	proteobacteria>betaproteobacteria	conserv
124262768		Mpe_B0228	203	Methylibium petroleiphilum PM1	proteobacteria>betaproteobacteria	hypothe
153948162		YpsIP31758_2230	204	Yersinia pseudotuberculosis IP 31758	proteobacteria>gammaproteobacteria	hypothe
77972287		YfreA_01003958	204	Yersinia frederiksenii ATCC 33641	proteobacteria>gammaproteobacteria	hypothe
212703055		DESPIG_01094	216	Desulfovibrio piger ATCC 29098	proteobacteria>deltaproteobacteria	hypothe
157325133		LiPA51l_gp109	217	Listeria phage A511	dsDNA viruses, no RNA stage>caudovirales	gp178 [
206603078		CGL2_11277194	225	Leptospirillum sp. Group II '5-way CG'	nitrospirae	Hypothe
163811796		DeVSDRAFT_1191	230	Dehalococcoides sp. VS	chloroflexi	hypothe
194434336		Sd1012_3704	231	Shigella dysenteriae 1012	proteobacteria>gammaproteobacteria	conserv
194435184		Sd1012_3850	231	Shigella dysenteriae 1012	proteobacteria>gammaproteobacteria	conserv
194430197		EcB171_5686	235	Escherichia coli B171	proteobacteria>gammaproteobacteria	protein
121582984		Pnap_4406	236	Polaromonas naphthalenivorans CJ2	proteobacteria>betaproteobacteria	hypothe
82776100		SDY_0774	238	Shigella dysenteriae Sd197	proteobacteria>gammaproteobacteria	putativ
91202340		kuste4638	238	Candidatus Kuenenia stuttgartiensis	planctomycetes	unknown
56693176		LP65_gp128	238	Lactobacillus phage LP65	dsDNA viruses, no RNA stage>caudovirales	hypothe
167041282		ALOHA_HF4000005H07ctg1g5	239	uncultured marine microorganism HF4000_005H07		hypothe
188494561		Ec53638_4358	241	Escherichia coli 53638	proteobacteria>gammaproteobacteria	conserv
9633544		Mup55	241	Enterobacteria phage Mu	dsDNA viruses, no RNA stage>caudovirales	Mom [En
157158693		EcE24377A_1038	241	Escherichia coli E24377A	proteobacteria>gammaproteobacteria	hypothe
209919755		ECSE_2564	241	Escherichia coli SE11	proteobacteria>gammaproteobacteria	phage D
77972128		YfreA_01004070	242	Yersinia frederiksenii ATCC 33641	proteobacteria>gammaproteobacteria	hypothe
163811804		DeVSDRAFT_1199	243	Dehalococcoides sp. VS	chloroflexi	hypothe
218885858		DvMF_0754	243	Desulfovibrio vulgaris str. 'Miyazaki F'	proteobacteria>deltaproteobacteria	putativ
82777163		SDY_1915	245	Shigella dysenteriae Sd197	proteobacteria>gammaproteobacteria	phage M
15834252		ECs4998	245	Escherichia coli O157:H7 str. Sakai	proteobacteria>gammaproteobacteria	putativ
160944077		FAEPRAM212_01578	248	Faecalibacterium prausnitzii M21/2	firmicutes	hypothe
77408584		SAN_1056	257	Streptococcus agalactiae COH1	firmicutes	conserv
221137281		SauraJK_010100001345	257	Staphylococcus aureus subsp. aureus str. JKD6008	firmicutes	hypothe
15923400		SAV0410	257	Staphylococcus aureus subsp. aureus Mu50	firmicutes	hypothe
148550644		Swit_5206	258	Sphingomonas wittichii RW1	proteobacteria>alphaproteobacteria	hypothe
163743125		RG210_14530	259	Phaeobacter gallaeciensis 2.10	proteobacteria>alphaproteobacteria	hypothe
167041313		ALOHA_HF4000005H07ctg2g12	263	uncultured marine microorganism HF4000_005H07		hypothe
110347123		Meso_4324	270	Mesorhizobium sp. BNC1	proteobacteria>alphaproteobacteria	Rv0623-
154498653		BACCAP_02643	279	Bacteroides capillosus ATCC 29799	bacteroidetes/chlorobi	hypothe
29833717		SAV_7175	68	Streptomyces avermitilis MA-4680	actinobacteria	hypothe
91204625		kusta0108	82	Candidatus Kuenenia stuttgartiensis	planctomycetes	predict
91203148	HTH+MOM	kustd2042	293	Candidatus Kuenenia stuttgartiensis	planctomycetes	unknown
116625178	HTH+MOM	Acid_6122	303	Solibacter usitatus Ellin6076	fibrobacteres/acidobacteria	hypothe
91201013	HTH+MOM	kuste3310	316	Candidatus Kuenenia stuttgartiensis	planctomycetes	hypothe
94270826	HTH+MOM	MldDRAFT_2064	291	delta proteobacterium MLMS-1	proteobacteria>deltaproteobacteria	hypothe
91204228	HTH+MOM	kustc1136	271	Candidatus Kuenenia stuttgartiensis	planctomycetes	unknown
116622055	HTH+MOM	Acid_2941	291	Solibacter usitatus Ellin6076	fibrobacteres/acidobacteria	hypothe
91203380	HTH+MOM+Tn5	kustc0288	743	Candidatus Kuenenia stuttgartiensis	planctomycetes	unknown
91200740	HTH+MOM+Tn5	kuste3037	743	Candidatus Kuenenia stuttgartiensis	planctomycetes	hypothe
91201548	HTH+MOM+Tn5	kuste3845	768	Candidatus Kuenenia stuttgartiensis	planctomycetes	unknown
91201057	HTH+MOM+Tn5	kuste3354	881	Candidatus Kuenenia stuttgartiensis	planctomycetes	conserv
91202068	HTH+MOM+Tn5	kuste4366	768	Candidatus Kuenenia stuttgartiensis	planctomycetes	unknown
91203017	HTH+MOM+Tn5	kustd1911	1491	Candidatus Kuenenia stuttgartiensis	planctomycetes	unknown
91200453	HTH+MOM+Tn5	kuste2750	838	Candidatus Kuenenia stuttgartiensis	planctomycetes	unknown
91201316	MOM+Tn5	kuste3613	185	Candidatus Kuenenia stuttgartiensis	planctomycetes	hypothe
91201320	MOM+Tn5	kuste3617	475	Candidatus Kuenenia stuttgartiensis	planctomycetes	conserv
198282266	MOM+TNPA	Lferr_0119	439	Acidithiobacillus ferrooxidans ATCC 53993	proteobacteria>gammaproteobacteria	hypothe
88812543	MOM+TNPA	NB231_08365	620	Nitrococcus mobilis Nb-231	proteobacteria>gammaproteobacteria	putativ
91201359	MOM+TNPA	tnpA	407	Candidatus Kuenenia stuttgartiensis	planctomycetes	similar
91204611	MOM+TNPA	kusta0094	407	Candidatus Kuenenia stuttgartiensis	planctomycetes	putativ
206603429	MOM+TNPA	CGL2_10706007	454	Leptospirillum sp. Group II '5-way CG'	nitrospirae	Hypothe
91201010	MOM+TNPA	kuste3307	206	Candidatus Kuenenia stuttgartiensis	planctomycetes	unknown
91201147	MOM+TNPA	kuste3444	442	Candidatus Kuenenia stuttgartiensis	planctomycetes	conserv
91201297	MOM+TNPA	kuste3594	442	Candidatus Kuenenia stuttgartiensis	planctomycetes	conserv
91202934	MOM+TNPA	kustd1828	404	Candidatus Kuenenia stuttgartiensis	planctomycetes	conserv
91200450	MOM+TNPA	kuste2747	442	Candidatus Kuenenia stuttgartiensis	planctomycetes	conserv
91203734	MOM+TNPA	tnpA	442	Candidatus Kuenenia stuttgartiensis	planctomycetes	similar
91202346	MOM+TNPA	kuste4644	442	Candidatus Kuenenia stuttgartiensis	planctomycetes	conserv
91200332	MOM+TNPA	kuste2629	463	Candidatus Kuenenia stuttgartiensis	planctomycetes	putativ
91201224	MOM+TNPA	tnpA	442	Candidatus Kuenenia stuttgartiensis	planctomycetes	similar
91201388	MOM+TNPA	tnpA	230	Candidatus Kuenenia stuttgartiensis	planctomycetes	similar
29826702	MOM+E9colicin+E9colicin	SAV_162	335	Streptomyces avermitilis MA-4680	actinobacteria	hypothe
84662609	ZnR+ZNR+ZNR+VSR-NUC+MOM	OP1_ORF16	510	Xanthomonas phage OP1	dsDNA viruses, no RNA stage>caudovirales	hypothe
170676315	ZnR+ZNR+ZNR+VSR-NUC+MOM	VIP0039	487	Salmonella phage E1		hypothe
19552983	ZnR+ZNR+ZNR+VSR-NUC+MOM	NCgl1708	606	Corynebacterium glutamicum ATCC 13032	actinobacteria	hypothe
149276370	ZNF+VSR-NUC+MOM	PBAL39_01582	272	Pedobacter sp. BAL39	bacteroidetes/chlorobi	hypothe
113951996	ZnR+VSR-NUC+MOM	hef	544	Enterobacteria phage U5	dsDNA viruses, no RNA stage>caudovirales	Hef [Ba
118197759	ZNF+VSR-NUC+MOM	YS40_139	534	Thermus phage phiYS40	dsDNA viruses, no RNA stage>caudovirales	hypothe
Eukaryotic MOMs						
219114737	MOM+Spermidine_synthase	PHATRDRAFT_44275	824	Phaeodactylum tricornutum CCAP 1055/1	stramenopiles	predict
219126128		PHATRDRAFT_48772	397	Phaeodactylum tricornutum CCAP 1055/1	stramenopiles	predict
Ehux1000027900		Ehux1000027900	321	Emiliania huxleyi haptophyceae gm1.1000315		

Fasta sequence of the Emiliania protein that contains a MOM domain

>gi|Ehux1000027900|ref|jgi|Emihu1|199615|gm1.1000315
MVRPSDFASTVAAWLQGEGFRTRSCSVCSSRPLLLVRGAHSTVALHLVPTPRSSDAVLPARVPIALSDSFATRRQTLVHLWEDQWDEHPRIVRSRLLAKLGRSERVMARQTVARRIDAATLDAFLVSNHLWGSTKARYRYGLFRKRDERLVAVASFSARWNMRREPGATARASHELIRYC

Multiple sequence alignment of the MOM domain

163811804	Dehalococcoides sp. VS	SLRVKPIPFTAAKR----ILERNHY-LHSMPGGTKLAFGV----FPE--------G-RLEGAITFG--S------GPANAYKMVSEASP---SDC-LTLSRLWLSDSLPSNS--ESRVLGVT-
163811796	Dehalococcoides sp. VS	SIMVKPLPFVAAKA----ILEKYHY-LHSMPGGTKLAFGV----FLG--------D-RLYGAITFG--S------GPANVYRMVEGAKP---SDC-LTLSRLWLSDELPSNS--ESRVLGIA-
167041282	microorganism HF4000_005H07].	SLRVRPISHRAAKG----LLLRNHY-LHTMPGGTRLAFGV----FSG--------D-RLMGAATLG--V------GPFNVHRLVAGATR---EDC-LVLTRLWLADDLPKNS--ESRALAVI-
167041313	microorganism HF4000_005H07].	SIQVRPVSVKIAKE----IIVRNHY-LHTMPGGTKLAFGI----FSG--------K-HLMGTVTLG--V------GPFNAHRLVSGATR---GDC-ICLTRLWLADDLPKNS--ESRVLAVV-
15834252	Escherichia coli O157:H7 str. Sakai	ELRVETISCWLARL----VIVNKHY-SHRFVNNSYLHLGI----FSE--------R-ELVGVMQWGYAL------NPNSGARVVTGTQN---REY-MELNRLWLHDCMPRNS--ESRAISYA-

```
82777163   Shigella dysenteriae Sd197                      ELHVETISCWLART----IIQTKHY-SRRFVNNSYLHLGV----FSG--------R-DLVGVLQWGYAL------NPNSGRRVVLETDN---RGY-MELNRMWLHDNMPRNS--ESRAISYA-
194430197  Escherichia coli B171                           ELRVETISCWLART----IIQTKHY-SRRFVNNSYLHLGV----FSG--------R-DLVGVLQWGYAL------NPNSGRRVVLETDN---RGY-MELNRMWLHDNMPRNS--ESRAISYA-
212703055  Desulfovibrio piger ATCC 29098                  DFYVAQIPCATARA----VIIEHHY-SHRVVNNSYLHLGV----YRE--------G-AFVGVLQWGYAL------NPARAGKVVANTTQ---GQY-LELNRMWLSDDAPRNS--ESRAISYA-
218885858  Desulfovibrio vulgaris str. 'Miyazaki F'        -FYVAAIPSADARA----AIIKHHY-SHRVVNNSYVHVGV----YYQ--------D-NFAGVLSLGYAL------NPARADKVVAGSAQ---GDY-LELNRMWLADVCPRNS--ESQALSYT-
9633544    Enterobacteria phage Mu                         ELRVETISCWLART----IIQTKHY-SRRFVNNSYLHLGV----FSG--------R-DLVGVLQWGYAL------NPNSGRRVVLETDN---RGY-MELNRMWLHDDMPRNS--ESRAISYA-
209919755  Escherichia coli SE11                           ELRVETISCWLART----IIQTKHY-SRRFVNNSYLHLGV----FSG--------R-DLVGVLQWGYAL------NPNSGRRVVLETDN---RGY-MELNRMWLHDDMPRNS--ESRAISYA-
188494561  Escherichia coli 53638                          ELRVETISCWLART----IIQTKHY-SRRFVNNSYLHLGV----FSG--------R-ELVGVLQWGYAL------NPNSGRRVVLETDN---RGY-MELNRMWLHDDMPRNS--ESRAISYA-
157158693  Escherichia coli E24377A                        ELRVETISCWLART----IIQTKHY-SRRFVNNSYLHLGV----FSG--------R-DLVGVLQWGYAL------NPNSGRRVVLETDN---RGY-MELNRMWLHDDMPRNS--ESRAISYA-
194434336  Shigella dysenteriae 1012                       ELRVETISCWLART----IIQTKHY-SRRFVNNSWLHLGV----FSG--------R-DLVGVLQWGYAL------NPNSGRRVVLETDN---RGY-MELNRMWLHDDMPRNS--ESRAISYA-
82776100   Shigella dysenteriae Sd197                      ELRVETISCWLART----IIQTKHY-SRRFVNNSWLHLGV----FSG--------R-DLVGVLQWGYAL------NPNSGRRVVLETDN---RGY-MELNRMWLHDDMPRNS--ESRAISYA-
77972128   Yersinia frederiksenii ATCC 33641               DLHVATIPVWLARL----VIIHFHY-SKRIVNNSYLHLGI----FSG--------R-DLVGVMQWGYAM------NPSSGGRVVEGTGN---RQY-MELNRLWVHDDMLTNT--ESRAISYA-
153948162  Yersinia pseudotuberculosis IP 31758            TLTVDWATHQAASF----ACLNWHY-AKAVPVGKLVKVGA----WED--------G-KFIGVVIFSRGA------NNHIGQPYSLQQDQ-----V-CELTRVALRQHI--SP--VSQILAKA-
194435184  Shigella dysenteriae 1012                       ELRVETISCWLART----IIQTKHY-SRRFVNNSWLHLGV----FSG--------R-DLVGVLQWGYAL------NPNSGRRVVLETDN---RGY-MELNRMWLHDDMPRNS--ESRAISYA-
77972287   Yersinia frederiksenii ATCC 33641               SLTVDWATHQAASF----ACLNWHY-AKAVPVGKLVKVGA----WED--------G-RFIGAVIFSRGA------NNHIGQPYSLQQDQ-----V-CELTRVALRQHI--SP--VSQILAKA-
160944077  Faecalibacterium prausnitzii M21/2             EIQIRVIPAKIANP----FIRAHHY-SGKVVSNSQLHFGA----FLD--------G-RLHGVLSYGPSL------DKKKIIGLVEGTTW---DGF-LELNRMAFDDYLPRNS--ESYCIAKT-
15923400   Staphylococcus aureus subsp. aureus Mu50        PLILKEIPKDEAIS----FIRQYHY-SKILPRLCKYFLGI----FSE--------E-KLLGVVELGWGT------QPLQTIRKLFPDSSLQTTDY-LEIGKMCFLPEMNQTNYFGSQALSAL-
163743125  Phaeobacter gallaeciensis 2.10                  PLALAWIRYHDALW----ACKTWHY-TKTMPTGKTVKIGV----FEA--------G-RFIGAVVFSRGA------TPRIGSPCGLEQVE-----V-CELTRVALDNHK--AP--VTQIISEA-
77408584   Streptococcus agalactiae COH1                   PLILKEIPKDEAIS----FIRQYHY-SKILPRLCKYFLGI----FSE--------E-KLLGVVELGWGT------QPLQTIRKLFPDSSLQTTDY-LEIGKMCFLPEMNQTNYFGSQALSAL-
157325133  Listeria phage A511                             DYEVKRIETKEAKP----FILGLHY-SKRMPSVSY-AFGI----FLN--------D-ELLGICTIGKPA------SHSLCKGVCGEEYA---YKV-FELNRLCMKEKL------EANVLSYF-
198282266  Acidithiobacillus ferrooxidans ATCC 53993       RIEVRPILPTERER-WNTLMQAHHYRGFRTMAGRTLRYVA----TLD--------D-RWVALLGWQAAAYQCQAREQWIGWSWVLRRQR-L-HLI-ANNARFLILPRESFPNL-ASRILSLN-
56693176   Lactobacillus phage LP65                        -LEFKQITYDDAKP----FLLEKHY--AHRMPSISYAYGL----YLD--------D-TLHGVLTFGTPA------SPSLQKGFFGNEYK---STV-IELNRLYVDDMV------ADKVYDAT-
91201147   Candidatus Kuenenia stuttgartiensis             SITVRPVSQNDQAN-WDTLMRQHHYLGFRSLVGESIRYVA----ESQ--------G-QWLALIGWAAAALKCTVRDKWIGWPPFLKSQR-L-KLI-ANNSRFLILPQIHVPNL-ASRILSLN-
9106834    Xylella fastidiosa 9a5c                         SLHLLPVSLRCANA----FVQEHHR-HHRPVQGAKFALAV----ALS------ATD-GICGVAIVGRPV----ARHLDNGWTLEVTRLC-T-DGA-PNACSKLYGAAWK-----AAKALGYT-
91201297   Candidatus Kuenenia stuttgartiensis             SITVRPVSQNDQAN-WDTLMRQHHYLGFRSLVGESIRYVA----ESQ--------G-QWLALIGWAAAALKCTVRDKWIGWPPFLKSQR-L-KLI-ANNSRFLILPQIHVPNL-ASRILSLN-
77747563   Xylella fastidiosa 9a5c                         SLHLLPVSLRCANA----FVQEHHR-HHRPVQGAKFALAV----ALS------ATD-GICGVAIVGRPV----ARHLDNGWTLEVTRLC-T-DGA-PNACSKLYGAAWK-----AAKALGYT-
91201224   Candidatus Kuenenia stuttgartiensis             SITVRPVSQNDQAN-WDTLMRQHHYLGFRSLVGESIRYVA----ESQ--------G-QWLALIGWAAAALKCTVRDKWIGWPPFLKSQR-L-KLI-ANNSRFLILPQIHVPNL-ASRILSLN-
91202934   Candidatus Kuenenia stuttgartiensis             SITVRPVSQNDQAN-WDTLMRQHHYLGFRSLVGESIRYVA----ESQ--------G-QWLALIGWAAAALKCTVRDKWIGWPPFLKSQR-L-KLI-ANNSRFLILPQIHVPNL-ASRILSLN-
206603429  Leptospirillum sp. Group II '5-way CG'          TLVVRPVEPSERSR-WDEEMRTHHYLGFRAMPGESIRYVA----LLE--------G-KWVALLGWGSAAWSNGHRDRMIGWTIPQRARR-L-GYL-ANNLRFLILPGIRLPHL-ASKVLALC-
91202346   Candidatus Kuenenia stuttgartiensis             SITVRPVSQNDQAN-WDTLMRQHHYLGFHSLVGESIRYVA----ESQ--------G-QWLALIGWAAAALKCTVRDKWIGWPPFLKSQR-L-KLI-ANNSRFLILPQIHVPNL-ASRILSLN-
91200332   Candidatus Kuenenia stuttgartiensis             SITVRPVSQNDQAN-WDTLMRQHHYLGFHSLVGESIRYVA----ESQ--------G-QWLALIGWAAAALKCTVRDKWIGWPPFLKSQR-L-KLI-ANNSRFLILPQIHVPNL-ASRILSLN-
91203734   Candidatus Kuenenia stuttgartiensis             SITVRPVSQHNQAN-WDTLMRQHHYLGFRSLVGESIRYVA----ESQ--------G-QWLALIGWAAASLKCTVRDKWIGWPPFLKSQR-L-KLI-ANNSRFLILPQIHVPNL-ASRILSLN-
91200450   Candidatus Kuenenia stuttgartiensis             SITVRPVSQHNQAN-WDTLMRQHHYLGFHSLVGESIRYVA----ESQ--------G-QWLALIGWAAAALKCTVRDKWIGWPPFLKSQR-L-KLI-ANNSRFLILPQIHVPNL-ASRILSLN-
160932974  Clostridium leptum DSM 753                      -LEICPISLKEANA----FVEQHHR-HHKPVTGHKFSIGC-TD----------G-E-KIVGVAIVGRPV----SRYLDDSWTLEVNRLC-T-DGT-RNACSMLYAAAWR-----TARAMGYH-
91201010   Candidatus Kuenenia stuttgartiensis             SITVRPVSQNDQAN-WDILMRQHHYLVFHSLVGESIRYVA----ESQ--------G-QWLALIGWAAAALKCTVRDKWIGWPPFLKSQR-L-KLI-ANNSRFLILPQIHVPNL-ASRILSLN-
38707956   Burkholderia phage phi1026b                     RLIVTPITLEEANA----FVSTHHR-HHRPVVGHKFSVAV----AAG--------D-DIRGVAIVGRPV----ARGNDNGLTLEITRCC-T-DGT-RNACSALYAAAWR-----VARALGYR-
206603078  Leptospirillum sp. Group II '5-way CG'          SLVVRPIYASERER-WDSLMREHHSLGFGWFAGKSIKYVA----LLD--------G-EWVALLGWATSALKVSSRDKWIGWTEELKSER-L-PYV-ANNSRFLVLPGKRVKNL-ASKSLALS-
194563444  Burkholderia dolosa AUO158                      -LVIVPISLEEANA----FVAEHHR-HHAPVVGHKFSIAV----ADDMLMGRDDCM-GVCGVAIVGRPV----ARGNDDGWTLEVTRCC-T-DGT-RNACSALYGAAWR-----AARALGYA-
154499046  Bacteroides capillosus ATCC 29799               -LRLRPISLRDANE----YVRQHHR-HHKPVAGHKFSIGC-EA----------D-G-ELVGVIIAGRPV----SRYLDDGFTLEVTRLC-T-NGA-KNACSFLYGAAAR-----AAAAMGYK-
167725778  Burkholderia pseudomallei DM98                  SLVIVPISLEEANV----FVAEHHR-HHAPVVGHKFSIAV----ADDMLMGRYDQT-GVCGVAIVGRPV----ARGNDDGWTLEVTRCC-T-DGT-RNACSTLYGAAWR-----AARALGYV-
91202340   Candidatus Kuenenia stuttgartiensis             PINLEQVRRTPFEKIFNGLVSEYHYLGYTQPIGEHLKYMA----FSG--------G-RPVACLAWGSAPWYIGARDRFIGWSKKARERN-L-HLI-ANNLRFLILPWVQVSCL-ASYLLASN-
116625178  Solibacter usitatus Ellin6076                   P-ELRQVRRTAEEPLFNSLIEHHHYLGYEQPVGEHLKYVV----WAQ--------G-RPIACLAWSSAPRHLGSRDRYIGWSAEARRRN-I-RGI-AYNTRFLILPWVQIPHL-ASHILGRV-
94270826   delta proteobacterium MLMS-1                    PIVLRQVRRTPQEAIFNGLIEHHHYLGYRQPVGEHLKYIA----VAR--------N-RPVACLAFSSVAYRLAARDSFIGWNDEARKQN-L-HLL-AYNTRFLIPDWVRVPHL-ASHLLGAA-
148550644  Sphingomonas wittichii RW1                      AYSVDIIDHAVART----FIAEHHY--LPSYPAAQVAVGL-FG-LRA----------VLAGVAVFAVPA-TDAVITRHTGFTDPAKG---------CVLARLILLDRVPQNG--ESFFCARA-
116622055  Solibacter usitatus Ellin6076                   PIELRQARRTTEEPLFNSLIEHHHYLGYEQPVGEHLKYVV----WAK--------S-RPIACLAWSSAPRHLGSRDRYIGWSAEARRRN-I-RGI-AYNTRFLILPWVQIPHL-ASHILGRV-
167566477  Burkholderia oklahomensis EO147                 SLIVTPITLEEANA----FVSIHHR-HHRPVIGHKFSVAV----AVN--------E-AIRGVAIVGRPV----ARGNDNGLTLEVTRCC-T-DGT-RNACSALYAAAWR-----AARALGYR-
121582984  Polaromonas naphthalenivorans CJ2               RLTIAPVSFSTAKN----FVAPHHR-HCPAPAGWRFWTGI----FNG--------D-ELIGCVMVGRPV----ARALDPARVVEVNRLC-V---------RTDIAPELPWNA--CSQLYGWA-
91204611   Candidatus Kuenenia stuttgartiensis             ------------------MRQHHYLGFRSLVGESIRYVA----ESQ--------G-QWLALIGWAAASLKCTVRDKWIGWPPFLKSQR-L-KLI-ANNSRFLILPQIHVPNL-ASRILSLN-
91201359   Candidatus Kuenenia stuttgartiensis             ------------------MRQHHYLGFRSLVGESIRYVA----ESQ--------G-QWLALIGWAAASLKCTVRDKWIGWPPFLKSQR-L-KLI-ANNSRFLILPQIHVPNL-ASRILSLN-
91203017   Candidatus Kuenenia stuttgartiensis             DVEIIRVTTRIHSKLWRSMLEAHHYLGSGPLCGAQLRYLV----RSE--------HYGWIGGLSYSACARRVESRDEWIGWTEEARKRN-H-TFV-INNSRYLIAPTVRVKCL-ASHVLAKC-
154498653  Bacteroides capillosus ATCC 29799               PMEVAPVSLKEALD----FVRRNHR-HCAAPQGHKFSVGL----TAD--------G-ALIGVVIASTPK----ARALNDGRTLELNRIC-C-DPVYRNAVSKVSGAAIR-----AARAMGYR-
167464664  Paenibacillus larvae subsp. larvae BRL-230010  QIHVEVSRSTELDR----WIAERHY-LKSTPAGAKLRLWI----LDE----R---R-NRIGAMMWGRPI-----------SRNLDQSH-I-----LELTRLFCIEDTDPFV--ESKALSLA-
91201013   Candidatus Kuenenia stuttgartiensis             DIVKTPVESRI----WNEMIHRWHYLGYKRLVGAQLRYQI----NSR--------Y-GVLGALGFGASAWNVQAREDFIGWSRQIREQN-L-NLI-VNNCRFLILPWIKSRNL-ASRVLGLV-
160899565  Delftia acidovorans SPH-1                       -LQLRPITRAEADA----FIREHHR-HHGCPVGALWRQAV-HD----------DAG-RLVGVAVVGRPV----ARALDDGLTVEVTRLC-T-DGA-TNACSLLYAATRR-----AAEAKGYR-
167462304  Paenibacillus larvae subsp. larvae BRL-230010  RISLEVSHSTKLDK----WIAGRHY-LKSTPAGARLRLWI----LDD----Q---G-NRIGAMMWGRPV-----------ARNLEQTS-L-----LELTRMYLIDDTIPNA--ESKALSLA-
91204228   Candidatus Kuenenia stuttgartiensis             DIVKTPVESRI----WNEMIHRWHYLGYKRLVGAQLRYQI----NSR--------Y-GVLGALGFGASAWNVQAREDFIGWSRQIREQN-L-NLI-VNNCRFLILPWIKSRNL-ASRVLGLV-
167747664  Anaerostipes caccae DSM 14662                   ----VPIELKAANE----YVAQNHR-HHSPVYRDKFRVAC----TEH--------G-RICGVIQCGRPV----SRHLDDGKTLEVTRCC-T-DGT-ANACSFLYGRVAR-----IAKEMGYE-
149276370  Pedobacter sp. BAL39                            QCTIVTISQQEANV----FLDQHHL---QFSVKAKYRFGL----MKS--------G-QLVAVALFSG-----------GRPMPRISAS---YLS-YELVRFASISGYT-----VSGGFTKL-
91203148   Candidatus Kuenenia stuttgartiensis             -----VVKDPQERYLWEYLFQHYHYLGNPRLVGEHLRHIV----RIG--------N-QVVACLGWASAAWKVKDRDRFIEWDETTKRTH-L-HLI-ANNVRFLILPWVKIKHL-ASKVLSLA-
91201057   Candidatus Kuenenia stuttgartiensis             DVEIIRVTTRFHSKLWRSMLEAHHYLGSGPLCGAQLRYLV----SSE--------HYGWIGGLSYSACARRVESRDEWIGWTEEARKRN-H-MFV-INNSRYLIAPTVRVKCL-ASHVLAKC-
118049792  Comamonas testosteroni KF-1                     RLHLLPVSLRTANV----FVLAHHH---RPVQAAKFALAV----TLV------DSD-LIRGVAIVGRPV----ARHLDDGWTREVTRLC-T-DST-PNACSKLYGAAWQ-----AAKSLGYI-
91200453   Candidatus Kuenenia stuttgartiensis             DVEIIRVTTRFHSKLWRSMLEAHHYLGSGPLCGAQLRYLV----RSE--------YYGWIGGLSYSACARRVESRDEWIGWTEEARKRN-H-TFV-INNSRYLIAPTVRVKCL-ASHVLAKC-
91209974   Escherichia coli UTI89                          RVQISPVTLRVAKA----FISRHHR-HNKPPVGHKFSIGL-RN----------DAG-ELIGVATAGRPV----ARHLDDGLTLEVNRTC-T-TGE-RNANSALYGAVWR-----AAKAMGYQ-
91202068   Candidatus Kuenenia stuttgartiensis             DVEIIRVTTRIHSKLWRSMLEAHHYLGSGPLCGAQLRYLV----RSE--------HYGWIGGLSYSACARRVESRDEWIGWTEEARKRN-H-TFV-INNSRYLIAPTVRVKCL-ASHVLAKC-
91201548   Candidatus Kuenenia stuttgartiensis             DVEIIRVTTRFHSKLWRSMLEAHHYLGSGPLCGAQLRYLV----RSE--------YYGWIGGLSYSACARRVESRDEWIGWTEEARKRN-H-TFV-INNSRYLIAPTVRVKCL-ASHVLAKC-
91203380   Candidatus Kuenenia stuttgartiensis             DVEIIRVTTRFHSKLWRSMLEAHHYLGSGPLCGAQLRYLV----SSE--------HYGWIGGLSYSACARRVESRDEWIGWTEEARKRN-H-MFV-INNSRYLIAPTVRVKCL-ASHVLAKC-
91200740   Candidatus Kuenenia stuttgartiensis             DVEIIRVTTRFHSKLWRSMLEAHHYLGSGPLCGAQLRYLV----SSE--------HYGWIGGLSYSACARRVESRDEWIGWTEEARKRN-H-MFV-INNSRYLIAPTVRVKCL-ASHVLAKC-
149882812  Microbacterium phage Min1                       -LRVEPISLRAANA----FVQEHHR-HHKPARGHKFSVSV-VD----------EAG-RIHGVGIAGRPV----SRHLDASGHLEVVRVC-T-DGT-RNVCSMLYGSLRR-----AGVALGYQP
88812543   Nitrococcus mobilis Nb-231                      RIECVPVVDSEPARRWQALMTAERFLGAGPLVGARLCYLV----ESE--------HFGEIGALAFSAAAWRLAARDRLIGWDEPARRTQ-L-GRI-VCNSRFLIRPTLRIPQL-ASHVLGRC-
87200732   Novosphingobium aromaticivorans DSM 12444      -LTFVPISIRSAKA----FVREHHR-HNPSVAGARTAIGL----ELD--------G-RLVGVGLLGNPK----------ARELAADRTC-------AEAVRVCVSHEAPKG---ASSKINSR-
167841763  Burkholderia thailandensis MSMB43               SLIIAPISLDEANA----FVATHHR-HHRPVAGHKFSIAVVDPLSLDSIVA--QDP-QVCGVAIVGRPV----ARGNDDGLTLEVTRCC-T-DGT-RNACSALYGAAWR-----AARALGYL-
124262768  Methylibium petroleiphilum PM1                  ------VEQRDAKA----FVRAHHR-HCPPPAGWKFGGAI----RNG--PGD---D-GVIGVVMVGRPV----AKAINQTKVMEVNRLC-I---------REDIADGLQWNA--CSLAYGWA-
170676315  Salmonella phage E1                             KCAIREVICKDYRA----FMNENHI---QGYAIATHRYGL----YYG--------D-DLVACMGFKQLK----------------------DGV-WDCVRYATSCRVPGG---HSKLFKFM-
29826702   Streptomyces avermitilis MA-4680                RYTVDVIPDEEPAK---GFVLAHHY--SGSYPAVTVQFGL-YD-VVD---GE---R-RLCGVAVFGVPV-----STAVLTKPLPELRPY---TES-LVCSRFVLLDECPANS--ESWFLANC-
110347378  Mesorhizobium sp. BNC1                          DVQFRAIERPDLNL---LLVEWEHRMGPVRRPNRGWSHGL----FHD--------G-QPVAVVATDRLI-RERVAGFDRAQAIELSRLA---AAR-PDLTRVVLR---------LWRAVVFP-
21225140   Streptomyces coelicolor A3(2)                   PLHLVPVRSREAKD----FVRTWHR-HHPPPAGQIFAVG-----AAD------ETG-TLRSVAIVGRPV----TRHLDDGATLEVTRTA-S-DGT-RNANSLLYGAAWR-----AAKALGYR-
91201320   Candidatus Kuenenia stuttgartiensis             ------------------MLEAHHYLGSGPLCGAQLRYLV----RSE--------HYGWIGGLSYSACARRVESRDEWIGWTEEARKRN-H-TFV-INNSRYLIAPTVRVKCL-ASHVLAKC-
84662609   Xanthomonas phage OP1                           DCYVEQISVADANV----FLAQWH--PQGKGGHNAVPFGL-KT-KKD--------N-ELVAVMTFAK--------DAYRRNPATANRDE-Q-WGN-WDLTRYTTSASVRGG---ASKLFKHF-
88812041   Nitrococcus mobilis Nb-231                      DFIVRPVERHEESR-YQAQMAAHHYLGALPKIGETLWYVA----TWR--------G-RWLAQIGLSAAALKCGVRDDWIGWGFRTQLDR-L-KLI-ANNTRCLILPEGHYPNL-GSRVLALV-
88810452   Nitrococcus mobilis Nb-231                      DFIVRPVERHEESR-YQAQMAAHHYLGALPKIGETLRYVA----TWR--------G-RWLAQIGLSAAALKCGVRDDWIGWGFRTQLDR-L-KLI-ANNTRCLILPEGHCPNL-GSRVLALV-
118197759  Thermus phage phiYS40                           KTKIEEVDYTTANE----FFEENHI---SGSASFKYAFGL----YYD--------D-EIVAAASFAKPR-----YNENFSWELVRFASK-K----------------------NTIVIGAL-
166031457  Dorea formicigenerans ATCC 27755                PIEARPMELKEAQD----YINTYHR-HHKAAHRDKFRIAA----TID--------G-KVVGVVQVGRPV----SRVLDDGNTLEVLRLC-T-TGE-KDICSFLYSRAAR-----IAKEMGYS-
110347123  Mesorhizobium sp. BNC1                          ---------------------------HRPTRGWSYGL----RHE--------G-RLLAVVATDTLI-RERVAGLTRHDAVELSRLC---AAK-PDLCRVALR---------LWRSFAFP-
consensus/85%                                              .h.h..l...........hh...Hb......ss..h.h.h..................hhuhh.hu.s...........s......p.p........p.sphhh.s.h.......uphhub..
```

Back to Contents

8. Table of distribution of Methylases and predicted DNA-modifying hydroxylases

	<---methylases								--->	<--------Hydroxylases		
	RIP	SNF+methylase	DNMT1	DNMT2	DNMT3	Ptri1000000241	LmjF25.1200	Aureococcus-specific	PHYPADRAFT_104049	TET/JBP	TET_assoc_tpase	SAD+ALKB
Eurotiomycetes \Asco	+	+	-	-	-	-	-	-	-	-	-	+
Soradariomycetes |mycetes	+	-	+	-	-	-	-	-	-	-	-	-
Leotiomycetes |	+	+	+	-	-	-	-	-	-	-	-	+
Dothideomycetes |	+	+	+	-	-	-	-	-	-	-	-	+
Saccharomycotina |	-	-	-	-	-	-	-	-	-	-	-	-
Taphrinomycotina /	-	-	-	+	-	-	-	-	-	-	-	-
Agaricomycotina \Basidio	-	+	(+)	(+)	-	-	-	-	-	(+)	(+)	+
Ustilagomycotina /mycetes	-	-	-	-	-	-	-	-	-	-	-	-
Batrachochytrium>Chytrid	-	-	-	+	-	-	-	-	-	-	-	-
Encephalitozoon	-	-	-	-	-	-	-	-	-	-	-	-
Chlorophytes	-	+	+	+	-	-	-	-	-	+	+	-
multi plants	-	-	+	+	+	-	-	-	(+)	-	-	-
red alga Cyanidioschyzon	-	-	-	-	+	-	-	-	-	-		-
Animals	-	-	+	+	+	-	-	-	-	+	(+)	-
Monosiga	-	-	-	+	-	-	-	-	-	-	-	-
Amoebozoa	-	-	-	+	-	-	-	-	-	-	-	-
Kinetoplastids	-	-	-	-	-	-	+		-	+	-	-
Naegleria	-	-	+	+	-	-	-	-	-	+	-	-
Thalassiosira	+	(+)frag	+	+	+	+	-	-	-	+	-	-
Phaeodactylum	-	+	-	+	+	+	-	-	-	+	-	-
Aureococcus	-	+	-	-	-	+	-	+	-	+	-	-
Oomycetes	-	-	-	-	-	-	-	-	-	-	-	-
Ciliates	-	-	-	-	-	-	-	-	-	-	-	-
Apicomplexans	-	-	-	(+)	-	-	-	-	-	-	-	-

Back to Contents