Title: Bacterial contribution to genesis of the novel germ line determinant *oskar*

Authors: *Leo Blondel*[1]*, Tamsin E. M. Jones*[2,3] *and Cassandra G. Extavour*[1,2*]

Affiliations:

1. Department of Molecular and Cellular Biology, Harvard University, 16 Divinity Avenue, Cambridge MA, USA
2. Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge MA, USA
3. Current address: European Bioinformatics Institute, EMBL-EBI, Wellcome Genome Campus, Hinxton, Cambridgeshire, UK

* Correspondence to extavour@oeb.harvard.edu

Abstract: New cellular functions and developmental processes can evolve by modifying existing genes or creating novel genes. Novel genes can arise not only via duplication or mutation but also by acquiring foreign DNA, also called horizontal gene transfer (HGT). Here we show that HGT likely contributed to the creation of a novel gene indispensable for reproduction in some insects. Long considered a novel gene with unknown origin, *oskar* has evolved to fulfil a crucial role in insect germ cell formation. Our analysis of over 100 insect Oskar sequences suggests that Oskar arose de novo *via* fusion of eukaryotic and prokaryotic sequences. This work shows that highly unusual gene origin processes can give rise to novel genes that can facilitate evolution of novel developmental mechanisms.

One Sentence Summary: Our research shows that gene origin processes often considered highly unusual, including HGT and de novo coding region evolution, can give rise to novel genes that can both participate in pre-existing gene regulatory networks, and also facilitate the evolution of novel developmental mechanisms.

Main Text: Heritable variation is the raw material of evolutionary change. Genetic variation can arise from mutation and gene duplication of existing genes (*1*), or through *de novo* processes (*2*), but the extent to which such novel, or "orphan" genes participate significantly in the evolutionary process is unclear. Mutation of existing cis-regulatory (*3*) or protein coding regions (*4*) can drive evolutionary change in developmental processes. However, recent studies in animals and fungi suggest that novel genes can also drive phenotypic change (*5*). Although counterintuitive, novel genes may be integrating continuously into otherwise conserved gene networks, with a higher rate of partner acquisition than subtler variations on preexisting genes (*6*). Moreover, in humans and fruit flies, a large proportion of novel genes are expressed in the brain, suggesting their participation in the evolution of major organ systems (*7, 8*). However, while next generation sequencing has improved their discovery, the developmental and evolutionary significance of novel genes remains understudied.

The mechanism of formation of a novel gene may have implications for its function. Novel genes that arise by duplication, thus possessing the same biophysical properties as their parent genes, have innate potential to participate in preexisting cellular and molecular mechanisms (*1*). However, orphan genes lacking sequence similarity to existing genes must form novel functional molecular relationships with extant genes, in order to persist in the genome. When such genes arise by introduction of foreign DNA into a host genome through horizontal gene transfer (HGT), they may introduce novel, already functional sequence information into a genome. Whether genes created by HGT show a greater propensity to contribute to or enable novel processes is unclear. Endosymbionts in the host germ line cytoplasm (germ line symbionts) could increase the occurrence of evolutionarily relevant HGT events, as foreign DNA integrated into

52 the germ line genome is transferred to the next generation. HGT from bacterial endosymbionts
53 into insect genomes appears widespread, involving transfer of metabolic genes or even larger
54 genomic fragments to the host genome (see for example *9, 10-12*).
55
56 Here we examined the evolutionary origins of the *oskar* (*osk*) gene, long considered a novel gene
57 that evolved to be indispensable for insect reproduction (*13*). First discovered in *Drosophila*
58 *melanogaster* (*14*), *osk* is necessary and sufficient for assembly of germ plasm, a cytoplasmic
59 determinant that specifies the germ line in the embryo. Germ plasm-based germ line
60 specification appears derived within insects, confined to insects that undergo metamorphosis
61 (Holometabola) (*15, 16*). Initially thought exclusive to Diptera (flies and mosquitoes), its
62 discovery in a wasp, another holometabolous insect with germ plasm (*17*), led to the hypothesis
63 that *oskar* originated as a novel gene at the base of the Holometabola approximately 300 Mya,
64 facilitating the evolution of insect germ plasm as a novel developmental mechanism (*17*).
65 However, its subsequent discovery in a cricket (*15*), a hemimetabolous insect without germ
66 plasm (*18*), implied that *osk* was instead at least 50 My older, and that its germ plasm role was
67 derived rather than ancestral (*19*). Despite its orphan gene status, *osk* plays major developmental
68 roles, interacting with the products of many genes highly conserved across animals (*13, 20, 21*).
69 *osk* thus represents an example of a novel gene that not only functions within pre-existing gene
70 networks in the nervous system (*15*), but has also evolved into the only animal gene that has
71 been experimentally demonstrated to be both necessary and sufficient to specify functional
72 primordial germ line cells (*22, 23*).
73

74 The evolutionary origins of this remarkable gene are unknown. Osk contains two biophysically
75 conserved domains, an N-terminal LOTUS domain and a C-terminal hydrolase-like domain
76 called OSK (*20, 24*) (Figure 1a). An initial BLASTp search using the full-length *D.*
77 *melanogaster osk* sequence as a query yielded either other holometabolous insect *osk* genes, or
78 partial hits for the LOTUS or OSK domains (E-value < 0.01; Supplementary files: BLAST
79 search results). This suggested that full length *osk* was unlikely to be a duplication of any other
80 known gene. This prompted us to perform two more BLASTp searches, one using each of the
81 two conserved Osk protein domains individually as query sequences. Strikingly, in this BLASTp
82 search, although we recovered several eukaryotic hits for the LOTUS domain, we recovered no
83 eukaryotic sequences that resembled the OSK domain, even with very low E-value stringency
84 (E-value < 10; see Methods section "*BLAST searches of oskar"* for an explanation of E-value
85 threshold choices; Supplementary files: BLAST search results).
86
87 To understand this anomaly, we built an alignment of 95 Oskar sequences (Supplementary files:
88 Alignments>OSKAR_MUSCLE_FINAL.fasta; Supplementary File 1A and B) and used a
89 custom iterative HMMER sliding window search tool to compare each domain with protein
90 sequences from all domains of life. Sequences most similar to the LOTUS domain were almost
91 exclusively eukaryotic sequences (Supplementary File 1C). In contrast, those most similar to the
92 OSK domain were bacterial, specifically sequences similar to SGNH-like hydrolases (*20, 24*)
93 (Pfam Clan: SGNH_hydrolase - CL0264; Supplementary File 1D; Figure 1b). To visualize their
94 relationships, we graphed the sequence similarity network for the sequences of these domains
95 and their closest hits. We observed that the majority of LOTUS domain sequences clustered
96 within eukaryotic sequences (Figure 1c). In contrast, OSK domain sequences formed an isolated

97 cluster, a small subset of which formed a connection to bacterial sequences (Figure 1d). These
98 data are consistent with a previous suggestion, based on BLAST results (*17*), that HGT from a
99 bacterium into an ancestral insect genome may have contributed to the evolution of *osk*.
100 However, this possibility was not formally addressed by previous analyses, which were based on
101 alignments of full length Osk containing only eukaryotic sequences as outgroups (*15*). To
102 rigorously test this hypothesis, we therefore performed phylogenetic analyses of the two domains
103 independently. A finding that LOTUS sequences were nested within eukaryotes, while OSK
104 sequences were nested within bacteria, would provide support for the HGT hypothesis.
105
106 Both Maximum likelihood and Bayesian approaches confirmed this prediction (Figure 2a, Figure
107 2–figure supplements 1 and 2), and these results were robust to changes in the methods of
108 sequence alignment (Figure 2-figure supplements 6-10). As expected, LOTUS sequences from
109 Osk proteins were related to other eukaryotic LOTUS domains, to the exclusion of the only three
110 bacterial sequences that met our E-value cutoff for inclusion in the analyses (Figs. 2a, Figure 2–
111 figure supplements 1 and 2; see Methods and Supplemental Text). LOTUS sequences from non-
112 Oskar proteins were almost exclusively eukaryotic. (Supplementary File 1C); only three bacterial
113 sequences matched the LOTUS domain with an E-value < 0.01. Osk LOTUS domains clustered
114 into two distinct clades, one comprising all Dipteran sequences, and the other comprising all
115 other Osk LOTUS domains examined from both holometabolous and hemimetabolous orders
116 (Figure 2a). Dipteran Osk LOTUS sequences formed a monophyletic group that branched sister
117 to a clade of LOTUS domains from Tud5 family proteins of non-arthropod animals (NAA). NAA
118 LOTUS domains from Tud7 family members were polyphyletic, but most of them formed a
119 clade branching sister to (Osk LOTUS + NAA Tud5 LOTUS). Non-Dipteran Osk LOTUS

domains formed a monophyletic group that was related in a polytomy to the aforementioned (NAA Tud7 LOTUS + (Dipteran Osk LOTUS + NAA Tud5 LOTUS)) clade, and to various arthropod Tud7 family LOTUS domains.

The fact that Tud7 LOTUS domains are polyphyletic suggests that arthropod domains in this family may have evolved differently than their homologues in other animals. The relationships of Dipteran LOTUS sequences were consistent with the current hypothesis for interrelationships between Dipteran species (*25*) Similarly, among the non-Dipteran Osk LOTUS sequences, the hymenopteran sequences form a clade to the exclusion of the single hemimetabolous sequence (from the cricket *Gryllus bimaculatus*), consistent with the monophyly of Hymenoptera (*26*). It is unclear why Dipteran Osk LOTUS domains cluster separately from those of other insect Osk proteins. We speculate that the evolution of the Long Oskar domain (*27, 28*), which appears to be a novelty within Diptera (Supplementary Files: Alignments>OSKAR_MUSCLE_FINAL.fasta), may have influenced the evolution of the Osk LOTUS domain in at least some of these insects. Consistent with this hypothesis, of the 17 Dipteran *oskar* genes we examined, the seven *oskar* genes possessing a Long Osk domain clustered into two clades based on the sequences of their LOTUS domain. One of these clades comprised five Drosophila species (*D. willistoni*, *D. mojavensis*, *D. virilis*, *D. grimshawi* and *D. immigrans*), and the second was composed of two calyptrate flies from different superfamilies, *Musca domestica* (Muscoidea) and *Lucilia cuprina* (Oestroidea).

In summary, the LOTUS domain of Osk proteins is most closely related to a number of other LOTUS domains found in eukaryotic proteins, as would be expected for a gene of animal origin,

and the phylogenetic interrelationships of these sequences are largely consistent with the current species or family level trees for the corresponding insects.

In contrast, OSK domain sequences were nested within bacterial sequences (Figure 2b, Figure 2–figure supplements 3, 4). This bacterial, rather than eukaryotic, affinity of the OSK domain was recovered even when different sequence alignment methods were used (Figure 2- figure supplements 7-11). The only eukaryotic proteins emerging from the iterative HMMER search for OSK domain sequences that had an E-value < 0.01 were all from fungi. All five of these sequences were annotated as Carbohydrate Active Enzyme 3 (CAZ3), and all CAZ3 sequences formed a clade that was sister to a clade of primarily Firmicutes. Most bacterial sequences used in this analysis were annotated as lipases and hydrolases, with a high representation of GDSL-like hydrolases (Supplementary File 1D). OSK sequences formed a monophyletic group but did not branch sister to the other eukaryotic sequences in the analysis. Within this OSK clade, the topology of sequence relationships was largely concordant with the species tree for insects (*29*), as we recovered monophyletic Diptera to the exclusion of other insect species. However, the single orthopteran OSK sequence (from the cricket *Gryllus bimaculatus*) grouped within the Hymenoptera, rather than branching as sister to all other insect sequences in the tree, as would be expected for this hemimetabolous sequence (*29*).

Importantly, OSK sequences did not simply form an outgroup to bacterial sequences. To formally reject the possibility that the eukaryotic OSK clade has a sister group relationship to all bacterial sequences in the analysis, we performed topology constraint analyses using the Swofford–Olsen–Waddell–Hillis (SOWH) test, which assigns statistical support to alternative phylogenetic

topologies (*30*). We used the SOWHAT tool (*31*) to compare the HGT-supporting topology to two alternative topologies with constraints more consistent with vertical inheritance. The first was constrained by domain of life, disallowing paraphyletic relationships between sequences from the same domain of life (Figure 2–figure supplement 5a). The second required monophyly of Eukaryota but allowed paraphyletic relationships between bacterial and archaeal sequences (Figure 2–figure supplement 5b). We found that the topologies of both of these constrained trees were significantly worse than the result we had recovered with our phylogenetic analysis (Figure 2–figure supplement 5), namely that the closest relatives of the OSK domain were bacterial rather than eukaryotic sequences (Figure 2b, Figure 2–figure supplements 3 and 4).

OSK sequences formed a well-supported clade nested within bacterial GDSL-like lipase sequences. The majority of these bacterial sequences were from the Firmicutes, a bacterial phylum known to include insect germline symbionts (*32, 33*). All other sequences from classified bacterial species, including a clade branching as sister to all other sequences, belonged either to the Bacteroidetes or to the Proteobacteria. Members of both of these phyla are also known germline symbionts of insects (*9, 34*) and other arthropods (*35*). In sum, the distinct phylogenetic relationships of the two domains of Oskar are consistent with a bacterial origin for the OSK domain. Further, the specific bacterial clades close to OSK suggest that an ancient arthropod germ line endosymbiont could have been the source of a GDSL-like sequence that was transferred into an ancestral insect genome, and ultimately gave rise to the OSK domain of *oskar* (Figure 3).

While multiple mechanisms can give rise to novel genes, HGT is arguably among the least well understood, as it involves multiple genomes and ancient biotic interactions between donor and host organisms that are often difficult to reconstruct. In the case of *oskar*, however, the fact that both germline symbionts (*36*) and HGT events (*9*) are widespread in insects, provides a plausible biological mechanism consistent with our hypothesis that fusion of eukaryotic and bacterial domain sequences led to the birth of this novel gene. Under this hypothesis, this fusion would have taken place before the major diversification of insects, nearly 500 million years ago (*29*).

Once arisen, novel genes might be expected to disappear rapidly, given that pre-existing gene regulatory networks operated successfully without them (*1*). However, it is clear that novel genes can evolve functional connections with existing networks, become essential (*37*), and in some cases lead to new functions (*38*) and contribute to phenotypic diversity (*5*). Even given the growing number of convincing examples of HGT from both prokaryotic and eukaryotic origins (see for example *39, 40-42*), some authors suspect that the contribution of horizontal gene transfer to the acquisition of novel traits has been underestimated across animals (*43*). Moreover, the functional contribution of genes horizontally transferred specifically from bacteria to insects has been documented for a range of adaptive phenotypes (see for example *44, 45, 46*), including digestive metabolism (*10, 11, 47*), glycolysis (*48*) complex symbiosis (*12*) and endosymbiont cell wall construction (*49*). *oskar* plays multiple critical roles in insect development, from neural patterning (*15, 50*) to oogenesis (*51*). In the Holometabola, a clade of nearly one million extant species (*52*), *oskar*'s co-option to become necessary and sufficient for germ plasm assembly is likely the cell biological mechanism underlying the evolution of this derived mode of insect germ line specification (*15, 17, 19*). Our study thus provides evidence that HGT can not only

211 introduce functional genes into a host genome, but also, by contributing sequences of individual
212 domains, generate genes with entirely novel domain structures that may facilitate the evolution
213 of novel developmental mechanisms.

Methods

BLAST searches of Oskar

All BLAST (*53*) searches were performed using the NCBI BLASTp tool suite on the non-redundant (nr) database. Amino Acid (AA) sequences of *D. melanogaster* full length Oskar (EMBL ID AAF54306.1), as well as the AA sequences for the *D. melanogaster* Oskar LOTUS (AA 139-238) and OSK (AA 414-606) domains were used for the BLAST searches. We used the default NCBI cut-off parameters (E-value cut-off of 10) for searches using OSK and LOTUS as queries, and a more stringent E-value threshhold of 0.01 for the search using full length *D. melanogaster* Oskar as a query. We chose an E-value threshold of 10 for LOTUS and OSK to capture potentially highly divergent homologs of the two domains, especially for the OSK domain, where we were looking for any viable candidate for a homologous eukaryotic domain. All BLAST searches results are included in the Supplementary files: BLAST search results.

Hidden Markov Model (HMM) generation and alignments of the OSK and LOTUS domains

101 1KITE transcriptomes (*54*) (Supplementary File 1A) were downloaded and searched using the local BLAST program (BLAST+) using the tblastn algorithm with default parameters, with Oskar protein sequences of *Drosophila melanogaster, Aedes aegypti, Nasonia vitripennis* and *Gryllus bimaculatus* as queries (EntrezIDs: NP_731295.1, ABC41128.1, NP_001234884.1 and AFV31610.1 respectively). For all of these 1KITE transcriptome searches, predicted protein sequences from transcript data were obtained by in silico translation using the online ExPASy translate tool (https://web.expasy.org/translate/), taking the longest open reading frame. Publicly available sequences in the non-redundant (nr), TSA databases at NCBI, and a then-unpublished

transcriptome (*55*) (kind gift of Matthew Benton and Siegfried Roth, University of Cologne) were subsequently searched using the web-based BLAST tool hosted at NCBI, using the tblastn algorithm with default parameters. Sequences used for queries were the four Oskar proteins described above, and newfound *oskar* sequences from the 1KITE transcriptomes of *Baetis pumilis, Cryptocercus wright,* and *Frankliniella cephalica*. For both searches, *oskar* orthologs were identified by the presence of BLAST hits on the same transcript to both the LOTUS (N-terminal) and OSK (C-terminal) regions of any of the query *oskar* sequences, regardless of E-values. The sequences found were aligned using MUSCLE (8 iterations) (*56*) into a 46-sequence alignment (Supplementary files: Alignments>OSKAR_MUSCLE_INITIAL.fasta). From this alignment, the LOTUS and OSK domains were extracted (Supplementary files: Alignments>LOTUS_MUSCLE_INITIAL.fasta and Alignments>OSK_MUSCLE_INITIAL.fasta) to define the initial Hidden Markov Models (HMM) using the hmmbuild tool from the HMMER tool suite with default parameters (*57*). 126 insect genomes and 128 insect transcriptomes (from the Transcriptome Shotgun Assembly TSA database: https://www.ncbi.nlm.nih.gov/Traces/wgs/?view=TSA) were subsequently downloaded from NCBI (download date September 29, 2015 ; Supplementary File 1A). Genomes were submitted to Augustus v2.5.5 (*58*) (using the *D. melanogaster* exon HMM predictor) and SNAP v2006-07-28 (*59*) (using the default 'fly' HMM) for gene discovery. The resulting nucleotide sequence database comprising all 309 downloaded and annotated genomes and transcriptomes, was then translated in six frames to generate a non-redundant amino acid database (where all sequences with the same amino acid content are merged into one). This process was automated using a series of custom scripts available here: https://github.com/Xqua/Genomes. The non-redundant amino acid database was searched using

the HMMER v3.1 tool suite (*57*) and the HMM for the LOTUS and OSK domains described above. A hit was considered positive if it consisted of a contiguous sequence containing both a LOTUS domain and an OSK domain, with the two domains separated by an inter-domain sequence. We imposed no length, alignment or conservation criteria on the inter-domain sequence, as this is a rapidly-evolving region of Oskar protein with predicted high disorder (*20, 24, 60*). Positive hits were manually curated and added to the main alignment, and the search was performed iteratively until no more new sequences meeting the above criteria were discovered. This resulted in a total of 95 Oskar protein sequences, (see Supplementary File 1B for the complete list). Using the final resulting alignment (Supplementary Files: Alignments>OSKAR_MUSCLE_FINAL.fasta), the LOTUS and OSK domains were extracted from these sequences (Supplementary Files: Alignments>LOTUS_MUSCLE_FINAL.fasta and Alignments>OSK_MUSCLE_FINAL.fasta), and the final three HMM (for full-length Oskar, OSK, and LOTUS domains) used in subsequent analyses were created using hmmbuild with default parameters (Supplementary files: HMM>OSK.hmm, HMM>LOTUS.hmm and HMM>OSKAR.hmm).

Iterative HMMER search of OSK and LOTUS domains

A reduced version of TrEMBL (*61*) (v2016-06) was created by concatenating all hits (regardless of E-value) for sequences of the LOTUS domain, the OSK domain and full-length Oskar, using hmmsearch with default parameters and the HMM models created above from the final alignment. This reduced database was created to reduce potential false positive results that might result from the limited size of the sliding window used in the search approach described here. The full-length Oskar alignment of 1133 amino acids (Supplementary files:

Alignments>OSKAR_MUSCLE_FINAL.fasta) was split into 934 sub-alignments of 60 amino acids each using a sliding window of one amino acid. Each alignment was converted into a HMM using hmmbuild, and searched against the reduced TrEMBL database using hmmsearch using default parameters. Domain of life origin of every hit sequence at each position was recorded. Eukaryotic sequences were further classified as Oskar/Non-Oskar and Arthropod/Non-Arthropod. Finally, for the whole alignment, the counts for each category were saved and plotted in a stack plot representing the proportion of sequences from each category to create Figure 1b. The python code used for this search is available at https://github.com/Xqua/Iterative-HMMER.

Sequence Similarity Networks

LOTUS and OSK domain sequences from the final alignment obtained as described above (see "*Hidden Markov Model (HMM) generation and alignments of the OSK and LOTUS domains*"; Supplementary files: Alignments>LOTUS_MUSCLE_FINAL.fasta and Alignments>OSK_MUSCLE_FINAL.fasta) were searched against TrEMBL (*61*) (v2016-06) using HMMER. All hits with E-value < 0.01 were consolidated into a fasta file that was then entered into the EFI-EST tool (*62*) using default parameters to generate a sequence similarity network. An alignment score corresponding to 30% sequence identity was chosen for the generation of the final sequence similarity network. Finally, the network was graphed using Cytoscape 3 (*63*).

Phylogenetic Analysis Based on MUSCLE Alignment

For both the LOTUS and OSK domains, in cases where more than one sequence from the same organism was retrieved by the search described above in "*Iterative HMMER Search of OSK and*

306 *LOTUS domains*", only the sequence with the lowest E-value was used for phylogenetic analysis.
307 For the LOTUS domain, the first 97 best hits (lowest E-value) were selected, and the only three
308 bacterial sequences that satisfied an E-value < 0.01 were manually added. For *oskar* sequences,
309 if more than one sequence per species was obtained by the search, only the single sequence per
310 species with the lowest E-value was kept for analysis, generating a set of 100 sequences for the
311 LOTUS domain, and 87 sequences for the OSK domain. Unique identifiers for all sequences
312 used to generate alignments for phylogenetic analysis are available in Supplementary Files 1C,
313 1D. For both datasets, the sequences were then aligned using MUSCLE (*56*) (8 iterations) and
314 trimmed using trimAl (*64*) with 70% occupancy. The resulting alignments that were subject to
315 phylogenetic analysis are available in Supplementary Files:
316 Alignments>LOTUS_MUSCLE_TREE.fasta and Alignments>OSK_MUSCLE_TREE.fasta. For
317 the maximum likelihood tree, we used RaxML v8.2.4 (*65*) with 1000 bootstraps, and the models
318 were selected using the automatic RaxML model selection tool. The substitution model chosen
319 for both domains was LGF. For the Bayesian tree inference, we used MrBayes V3.2.6 (*66*) with
320 a Mixed model (prset aamodel=Mixed) and a gamma distribution (lset rates=Gamma). We ran
321 the MonteCarlo for 4 million generations (std < 0.01) for the OSK domain, and for 3 million
322 generations (std < 0.01) for the LOTUS domain. For the tree comparisons (Fig2 supplementary
323 figures 8-9), the RaxML best tree output from the MUSCLE and PRANK alignments were
324 compared using the tool Phylo.io (*67*).
325
326 ***Phylogenetic analysis based on PRANK alignment***
327 For the OSK domain, the raw full length sequences obtained from the HMMER search were
328 aligned to each other using HMMER HMM based alignment tool: hmmalign, with the same

HMM used to do the search, namely OSK.hmm (supplementary data: Data/HMM/OSK.hmm). Starting from this base alignment, we used the default alignment method option offered by PRANK (version: v.170427) ((*68*). We then used PRANK to realign those sequences, which in turn led to a usable alignment for phylogenetic analysis. This alignment was trimmed using the same parameters as described in *Hidden Markov Model (HMM) generation and alignments of the OSK and LOTUS domains* above. The final alignment is available in supplementary data: Alignment/OSK_prank_aligned.fasta. We then performed a phylogenetic analysis of this alignment using RAXML with the same parameters described in *Phylogenetic Analysis Based on MUSCLE Alignment* above. The resulting tree is presented in Figure 2 Supplementary figures 7 and 8.

For the LOTUS domain, the raw full length sequences obtained from the HMMER search were aligned to each other using the HMMER HMM based alignment tool: hmmalign, with the same HMM used to do the search, namely LOTUS.hmm (Supplementary data: Data/HMM/LOTUS.hmm). Starting from this base alignment, we then used PRANK with default options to realign those sequences. This alignment was trimmed using the same parameters as described in the *Hidden Markov Model (HMM) generation and alignments of the OSK and LOTUS domains*. The final alignment is available in supplementary data: Alignments/LOTUS_prank_aligned.fasta. We then performed a phylogenetic analysis using RAXML with the same parameters described above in *Phylogenetic Analysis Based on MUSCLE alignment*. The resulting trees are presented in Figure 2 Supplementary figures 6 and 9.

Phylogenetic Analysis Based on T Coffee alignment

For the LOTUS and OSK domain, the raw full length sequences obtained from the HMMER search were aligned to each other using T-Coffee with its default parameters (*69*). This alignment was trimmed using the same parameters as described in *Hidden Markov Model (HMM) generation and alignments of the OSK and LOTUS domains* above. The final alignment is available in supplementary data: Alignment/LOTUS_tcoffee_aligned.fasta Alignment/OSK_tcoffee_aligned.fasta. We then performed a phylogenetic analysis of this alignment using RAXML with the same parameters described in *Phylogenetic Analysis Based on MUSCLE Alignment* above. The resulting trees are presented in Figure 2 Supplementary figures 10 and 11.

Visual Comparison of Phylogenetic Trees

To compare the trees obtained with different alignment tools, we used Phylo.io (*67*). The trees were imported in Newick format, and the Phylo.io tool generated the mirrored and aligned versions of the trees represented in Figure 2 Supplementary figures 8, 9, 12 and 13. The color of the branches is the tree similarity score, where lighter colors represent a higher number of topological differences. Exactly, it is a custom implementation of the Jacard Index by Phylo.io.

Statistical Analysis of Tree Topology

To statistically evaluate our best-supported topology of the OSK and LOTUS trees, we compared constrained topologies to the highest likelihood trees using the SOWHAT tool (*31*). SOWHAT automates the stringent SOWH phylogenetic topology test (*30*), and compares the log likelihood between generated trees. We defined three constrained trees to test our results, one requiring

monophyly of all domains of life, a second requiring only eukaryotic monophyly, and the last one requiring monophyly of the oskar LOTUS domain (Supplementary Files: Data>Trees>constrained_kingdom_tree.tre, constrained_eukmono_tree.tre & constrained_lotus_mono_tree.tre). We then ran SOWHAT using its default parameters, 1000 bootstraps, and the two constrained trees against the OSK or LOTUS alignment used to generate the phylogenetic trees (Supplementary Files: Alignments>OSK_MUSCLE_TREE.fasta & LOTUS_MUSCLE_TREE.fasta). All best trees generated by SOWHAT are available in (Supplementary Files: Data>Trees>SOWHAT_*_test.tre).

Data Availability

All sequences discovered using the automatic annotation pipeline described in (M&M HMM and oskar search) are annotated as such in Supplementary File 1B.

Code Availability

All custom code generated for this study is available in the GitHub repository https://github.com/extavourlab/Oskar_HGT, commit ID 6f6c4c50dfb9391567d70f9eea922f3876a4e153.

References

1. J. S. Taylor, J. Raes, Duplication and divergence: the evolution of new genes and old ideas. *Annual review of genetics* **38**, 615-643 (2004).
2. D. Tautz, T. Domazet-Loso, The evolutionary origin of orphan genes. *Nat. Rev. Genet.* **12**, 692-702 (2011).
3. P. J. Wittkopp, G. Kalay, Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* **13**, 59-69 (2011).
4. H. E. Hoekstra, J. A. Coyne, The locus of evolution: evo devo and the genetics of adaptation. *Evolution* **61**, 995-1016 (2007).
5. S. Chen, B. H. Krinsky, M. Long, New genes as drivers of phenotypic evolution. *Nat. Rev. Genet.* **14**, 645-660 (2013).
6. W. Zhang, P. Landback, A. R. Gschwend, B. Shen, M. Long, New genes drive the evolution of gene interaction networks in the human and mouse genomes. *Genome Biol.* **16**, 202 (2015).
7. Y. E. Zhang, P. Landback, M. Vibranovski, M. Long, New genes expressed in human brains: implications for annotating evolving genomes. *BioEssays* **34**, 982-991 (2012).
8. S. Chen *et al.*, Frequent recent origination of brain genes shaped the evolution of foraging behavior in Drosophila. *Cell Reports* **1**, 118-132 (2012).
9. J. C. Dunning Hotopp *et al.*, Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science (New York, NY)* **317**, 1753-1756 (2007).
10. R. Acuna *et al.*, Adaptive horizontal transfer of a bacterial gene to an invasive insect pest of coffee. *Proc Natl Acad Sci U S A* **109**, 4197-4202 (2012).
11. D. B. Sloan *et al.*, Parallel histories of horizontal gene transfer facilitated extreme reduction of endosymbiont genomes in sap-feeding insects. *Mol. Biol. Evol.* **31**, 857-871 (2014).
12. F. Husnik *et al.*, Horizontal gene transfer from diverse bacteria to an insect genome enables a tripartite nested mealybug symbiosis. *Cell* **153**, 1567-1578 (2013).
13. R. Lehmann, Germ Plasm Biogenesis--An Oskar-Centric Perspective. *Curr. Top. Dev. Biol.* **116**, 679-707 (2016).
14. R. Lehmann, C. Nüsslein-Volhard, Abdominal Segmentation, Pole Cell Formation, and Embryonic Polarity Require the Localized Activity of *oskar*, a Maternal Gene in *Drosophila*. *Cell* **47**, 144-152 (1986).
15. B. Ewen-Campen, J. R. Srouji, E. E. Schwager, C. G. Extavour, *oskar* Predates the Evolution of Germ Plasm in Insects. *Curr. Biol.* **22**, 2278-2283 (2012).
16. C. G. Extavour, M. E. Akam, Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. *Development* **130**, 5869-5884 (2003).
17. J. A. Lynch *et al.*, The Phylogenetic Origin of *oskar* Coincided with the Origin of Maternally Provisioned Germ Plasm and Pole Cells at the Base of the Holometabola. *PLoS Genetics* **7**, e1002029 (2011).
18. B. Ewen-Campen, S. Donoughe, D. N. Clarke, C. G. Extavour, Germ cell specification requires zygotic mechanisms rather than germ plasm in a basally branching insect. *Curr. Biol.* **23**, 835-842 (2013).
19. E. Abouheif, Evolution: oskar Reveals Missing Link in Co-optive Evolution. *Curr. Biol.* **23**, R24-R25 (2012).

20. M. Jeske *et al.*, The Crystal Structure of the *Drosophila* Germline Inducer Oskar Identifies Two Domains with Distinct Vasa Helicase- and RNA-Binding Activities. *Cell Reports* **12**, 587-598 (2015).
21. M. Jeske, C. W. Muller, A. Ephrussi, The LOTUS domain is a conserved DEAD-box RNA helicase regulator essential for the recruitment of Vasa to the germ plasm and nuage. *Genes Dev.* **31**, 939-952 (2017).
22. A. Ephrussi, R. Lehmann, Induction of germ cell formation by *oskar*. *Nature* **358**, 387-392 (1992).
23. J. Kim-Ha, J. L. Smith, P. M. Macdonald, *oskar* mRNA is localized to the posterior pole of the *Drosophila* oocyte. *Cell* **66**, 23-35 (1991).
24. N. Yang *et al.*, Structure of *Drosophila* Oskar reveals a novel RNA binding protein. *Proc. Natl. Acad. Sci. USA* **112**, 11541-11546 (2015).
25. A. H. Kirk-Spriggs, B. J. Sinclair, Eds., *Manual of Afrotropical Diptera*, (South African National Biodiversity Institute, Pretoria, South Africa, 2017), vol. 1.
26. R. S. Peters *et al.*, Evolutionary History of the Hymenoptera. *Curr. Biol.* **27**, 1013-1018 (2017).
27. N. F. Vanzo, A. Ephrussi, Oskar anchoring restricts pole plasm formation to the posterior of the *Drosophila* oocyte. *Development* **129**, 3705-3714 (2002).
28. T. R. Hurd *et al.*, Long Oskar Controls Mitochondrial Inheritance in *Drosophila melanogaster*. *Dev. Cell* **39**, 560-571 (2016).
29. B. Misof *et al.*, Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**, 763-767 (2014).
30. D. L. Swofford, G. J. Olsen, P. J. Waddell, D. M. Hillis, in *Molecular Systematics 2nd Edition,* D. M. Hillis, C. Moritz, B. K. Mable, Eds. (Sinauer Associates, Inc., Sinauer, MA, 1996), chap. 11, pp. 407-453.
31. S. H. Church, J. F. Ryan, C. W. Dunn, Automation and Evaluation of the SOWH Test with SOWHAT. *Syst. Biol.* **64**, 1048-1058 (2015).
32. D. Wheeler, A. J. Redding, J. H. Werren, Characterization of an ancient lepidopteran lateral gene transfer. *PLoS ONE* **8**, e59262 (2013).
33. S. T. Chepkemoi *et al.*, Identification of Spiroplasmainsolitum symbionts in Anopheles gambiae. *Wellcome Open Research* **2**, 90 (2017).
34. E. Zchori-Fein, S. J. Perlman, S. E. Kelly, N. Katzir, M. S. Hunter, Characterization of a 'Bacteroidetes' symbiont in Encarsia wasps (Hymenoptera: Aphelinidae): proposal of 'Candidatus Cardinium hertigii'. *International Journal of Systematic and Evolutionary Microbiology* **54**, 961-968 (2004).
35. E. Zchori-Fein, S. J. Perlman, Distribution of the bacterial symbiont Cardinium in arthropods. *Mol. Ecol.* **13**, 2009-2016 (2004).
36. K. Bourtzis, T. A. Miller, Eds., *Insect Symbiosis*, (CRC Press, Boca Raton, FL, 2006), vol. 3, pp. 304.
37. S. Chen, Y. E. Zhang, M. Long, New genes in *Drosophila* quickly become essential. *Science* **330**, 1682-1685 (2010).
38. G. Cornelis *et al.*, Ancestral capture of syncytin-Car1, a fusogenic endogenous retroviral envelope gene involved in placentation and conserved in Carnivora. *Proc. Natl. Acad. Sci. USA* **109**, E432-441 (2012).
39. F. Husnik, J. P. McCutcheon, Functional horizontal gene transfer from bacteria to eukaryotes. *Nat Rev Microbiol* **16**, 67-79 (2018).

40. I. Di Lelio *et al.*, Evolution of an insect immune barrier through horizontal gene transfer mediated by a parasitic wasp. *PLoS Genetics* **15**, e1007998 (2019).
41. N. Wybouw, Y. Pauchet, D. G. Heckel, T. Van Leeuwen, Horizontal Gene Transfer Contributes to the Evolution of Arthropod Herbivory. *Genome Biol. Evol.* **8**, 1785-1801 (2016).
42. D. G. Quispe-Huamanquispe, G. Gheysen, J. F. Kreuze, Horizontal Gene Transfer Contributes to Plant Evolution: The Case of Agrobacterium T-DNAs. *Front Plant Sci* **8**, 2015 (2017).
43. L. Boto, Horizontal gene transfer in the acquisition of novel traits by metazoans. *Proc Biol Sci* **281**, 20132450 (2014).
44. A. C. Wilson, R. P. Duncan, Signatures of host/symbiont genome coevolution in insect nutritional endosymbioses. *Proc. Natl. Acad. Sci. USA* **112**, 10255-10261 (2015).
45. S. Lopez-Madrigal, R. Gil, Et tu, Brute? Not Even Intracellular Mutualistic Symbionts Escape Horizontal Gene Transfer. *Genes (Basel)* **8**, (2017).
46. N. A. Provorov, O. P. Onischuk, Microbial Symbionts of Insects: Genetic Organization, Adaptive Role, and Evolution. *Mikrobiologiya* **87**, 151-163 (2018).
47. M. Shelomi *et al.*, Horizontal Gene Transfer of Pectinases from Bacteria Preceded the Diversification of Stick and Leaf Insects. *Sci Rep* **6**, 26388 (2016).
48. Z. Zeng *et al.*, Bacterial endosymbiont Cardinium cSfur genome sequence provides insights for understanding the symbiotic relationship in Sogatella furcifera host. *BMC Genomics* **19**, 688 (2018).
49. D. C. Bublitz *et al.*, Peptidoglycan Production by an Insect-Bacterial Mosaic. *Cell* **179**, 703-712 e707 (2019).
50. X. Xu, J. L. Brechbiel, E. R. Gavis, Dynein-Dependent Transport of *nanos* RNA in *Drosophila* Sensory Neurons Requires Rumpelstiltskin and the Germ Plasm Organizer Oskar. *J. Neurosci.* **33**, 14791-14800 (2013).
51. A. Jenny *et al.*, A translation-independent role of oskar RNA in early Drosophila oogenesis. *Development* **133**, 2827-2833 (2006).
52. J. A. Rees, K. Cranston, Automated assembly of a reference taxonomy for phylogenetic data synthesis. *Biodiversity Data Journal* **5**, e12581 (2017).
53. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410 (1990).
54. H. Aspöck *et al.* (2018).
55. M. A. Benton, N. J. Kenny, K. H. Conrads, S. Roth, J. A. Lynch, Deep, Staged Transcriptomic Resources for the Novel Coleopteran Models Atrachya menetriesi and Callosobruchus maculatus. *PLoS ONE* **11**, e0167431 (2016).
56. R. C. Edgar, MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004).
57. S. R. Eddy. (2007).
58. M. Stanke, R. Steinkamp, S. Waack, B. Morgenstern, AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.* **32**, W309-312 (2004).
59. I. Korf, Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
60. A. Ahuja, C. G. Extavour, Patterns of molecular evolution of the germ line specification gene *oskar* suggest that a novel domain may contribute to functional divergence in *Drosophila*. *Dev. Genes Evol.* **222**, 65-77 (2014).

61. U. Consortium, The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res.* **37**, D169-174 (2009).
62. J. A. Gerlt *et al.*, Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): A web tool for generating protein sequence similarity networks. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1854**, 1019-1037 (2015).
63. P. Shannon *et al.*, Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498-2504 (2003).
64. S. Capella-Gutierrez, J. M. Silla-Martinez, T. Gabaldon, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009).
65. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
66. J. P. Huelsenbeck, F. Ronquist, MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754-755 (2001).
67. O. Robinson, D. Dylus, C. Dessimoz, Phylo.io: Interactive Viewing and Comparison of Large Phylogenetic Trees on the Web. *Mol. Biol. Evol.* **33**, 2163-2166 (2016).
68. A. Loytynoja, Phylogeny-aware alignment with PRANK. *Methods Mol. Biol.* **1079**, 155-170 (2014).
69. C. Notredame, D. G. Higgins, J. Heringa, T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205-217 (2000).

Acknowledgments:

We thank Sean Eddy, Chuck Davis, and Extavour lab members for discussion.

Funding: This work was supported by funds from Harvard University.

Competing interests: The authors declare no competing interests.

Data and materials availability: All data is available in the main text or the supplementary materials.

Figure Legends

Figure 1. **Sequence analysis of the Oskar gene. a**, Schematic representation of the Oskar gene. The LOTUS and OSK hydrolase-like domains are separated by a poorly conserved region of predicted high disorder and variable length between species. In some dipterans, a region 5' to the LOTUS domain is translated to yield a second isoform, called Long Oskar. Residue numbers correspond to the *D. melanogaster* Osk sequence. **b**, Stackplot of domain of life identity of HMMER hits across the protein sequence. For a sliding window of 60 Amino Acids across the protein sequence (X axis), the number of hits in the Trembl (UniProt) database (Y axis) is represented and color coded by domain of life origin (see Methods: Iterative HMMER search of OSK and LOTUS domains), stacked on top of each other. **c, d** EFI-EST-generated graphs of the sequence similarity network of the LOTUS (**c**) and OSK (**d**) domains of Oskar (*62*). Sequences were obtained using HMMER against the UniProtKB database. Most Oskar LOTUS sequences cluster within eukaryotes and arthropods. In contrast, Oskar OSK sequences cluster most strongly with a small subset of bacterial sequences.

Figure 2. **Phylogenetic analysis of the LOTUS and OSK domains. a**, Bayesian consensus tree for the LOTUS domain. Three major LOTUS-containing protein families are represented within the tree: Tudor 5, Tudor 7, and Oskar. Oskar LOTUS domains form two clades, one containing only dipterans and one containing all other represented insects (hymenopterans and orthopterans). The tree was rooted to the three bacterial sequences added in the dataset. **b**, Bayesian consensus tree for the OSK domain. The OSK domain is nested within GDSL-like domains of bacterial species from phyla known to contain germ line symbionts in insects. The ten non-Oskar eukaryotic sequences in the analysis form one clade comprising fungal Carbohydrate Active Enzyme 3 (CAZ3) proteins. For Bayesian and RaxML trees with all accession numbers and node support values see Figure 2–figure supplement 1-4.

Figure 3. Hypothesis for the origin of *oskar*. Integration of the OSK domain close to a LOTUS domain in an ancestral insect genome. **a**, DNA containing a GDSL-like domain from an endosymbiotic germ line bacterium is transferred to the nucleus of a germ cell in an insect common ancestor. **b**, DNA damage or transposable element activity induces an integration event in the host genome, close to a pre-existing LOTUS-like domain. **c**, The region between the two domains undergoes *de novo* coding evolution, creating an open reading frame with a unique, chimeric domain structure. **d**, In some Diptera, including *D. melanogaster*, part of the 5' UTR of *oskar* has undergone *de novo* coding evolution to form the Long Oskar domain.

The Supplementary Information for this paper consists of the following elements:

Supplementary File 1

PDF DOCUMENT: FILE NAME BLONDEL_JONES_EXTAVOUR_HGT_ SUPPLEMENTARY_FILE_1_CLEAN

1. Supplementary Tables
    a. Supplementary File 1A: List of genomes and transcriptomes used for automated *oskar* search.
    b. Supplementary File 1B: List of *oskar* sequences used in the final alignment.
    c. Supplementary File 1C: List of sequences used for phylogenetic analysis of the LOTUS domain.
    d. Supplementary File 1D: List of sequences used for phylogenetic analysis of the OSK domain.

Source Data 1

ZIPPED FOLDER: DOWNLOAD HERE

https://www.dropbox.com/sh/zqf6kpo0kzav7xp/AAC5WPVm9lrDrZHqZg_RlsiTa?dl=0

1. Subfolder **Alignments**: All sequences identified and analyzed in this study, in FASTA format and with corresponding Alignments
2. Subfolder **BLAST search results**: Results of BLASTP searches with full length Oskar, OSK or LOTUS domains as queries
3. Subfolder **Data**: Necessary files for running the different IPython notebooks:
    a. Subfolder ***HMM***: HMM models used for iterative searching for sequences similar to full-length Oskar, LOTUS and OSK domains
    b. Subfolder ***Taxonomy***: Conversion table for UniProt ID to taxon information. (uniprot_ID_taxa.tsv)
    c. Subfolder ***Trees***: Contains the tree files obtained from
        i. RaxML phylogenetic analyses of the OSK and LOTUS domains aligned with MUSCLE, T-Coffee or PRANK
        ii. MrBayes phylogenetic analyses of the OSK and LOTUS domains aligned with MUSCLE
        iii. SOWHAT analyses.

Scripts

All scripts used herein are hosted on GitHub at https://github.com/extavourlab/Oskar_HGT

Supplementary Figure Legends

Figure 2 Supplements

Figure 2–figure supplement 1: LOTUS Domain RaxML MUSCLE Tree. Phylogenetic tree of the HMMER sequences retrieved from the UniProt database using the LOTUS alignment HMM model. The top 97 hits were selected for phylogenetic analysis, and the only three bacterial sequences found to be a match were added to the alignment manually. The resulting 100 sequences were aligned using MUSCLE with default settings. The sequences were filtered to contain only one sequence per species (best E-value kept) yielding 100 sequences for analysis. Finally, the tree was created using RaxML v8.2.4, using 1000 bootstraps and model selection performed by the RaxML automatic model selection tool. See "Phylogenetic Analysis" in Methods for further detail. Sequences are color-coded as follows: Purple = Oskar; Red = Non-Oskar Arthropod; Green = Non-Arthropod Eukaryote; Blue = Bacteria. Names following leaves display the UniProt accession number followed by the species name and the UniProt protein name.

Figure 2–figure supplement 2: LOTUS Domain Bayesian MUSCLE Tree. Phylogenetic tree of the HMMER sequences retrieved from the UniProt database using the LOTUS alignment HMM model. 100 sequences were chosen for analysis as described for Supplementary Figure 1. The tree was created using Mr Bayes V3.2.6 using a Mixed model (prset aamodel=Mixed) and a gamma distribution (lset rates=Gamma). The algorithm was allowed to run for 3 million generations to achieve a std < 0.01. See "Phylogenetic Analysis" in Methods for further detail. Sequences are color-coded as follows: Purple = Oskar; Red = Non-Oskar Arthropod; Green = Non-Arthropod Eukaryote; Blue = Bacteria. Names following leaves display the UniProt accession number followed by the species name and the UniProt protein name.

Figure 2–figure supplement 3: OSK Domain RaxML MUSCLE Tree. Phylogenetic tree of the HMMER sequences retrieved from the UniProt database using the OSK alignment HMM model. The top 95 hits were selected for phylogenetic analysis, and the only five non-Oskar eukaryotic sequences found to be a match were added to the alignment manually. The resulting 100 sequences were aligned using MUSCLE with default settings. The sequences were filtered to contain only one sequence per species (best E-value kept), yielding 87 sequences for analysis. Finally, the tree was created using RaxML v8.2.4, using 1000 bootstraps and model selection performed by the RaxML automatic model selection tool. See "Phylogenetic Analysis" in Methods for further detail. Sequences are color-coded as follows: Purple = Oskar; Red = Non-Oskar Arthropod; Green = Non-Arthropod Eukaryote; Blue = Bacteria. Names following leaves display the UniProt accession number followed by the species name and the UniProt protein name.

Figure 2–figure supplement 4: OSK Domain Bayesian MUSCLE Tree. Phylogenetic tree of the HMMER sequences hit on the UniProt database using the OSK alignment HMM model. 87 sequences were chosen for analysis as described for Figure 2 – figure supplement 3.The tree was created using Mr Bayes V3.2.6 using a Mixed model (prset aamodel=Mixed) and a gamma distribution (lset rates=Gamma). The algorithm was allowed to run for 4 million generations to achieve a std < 0.01. See "Phylogenetic Analysis" in Methods for further detail. Sequences are

color-coded as follows: Purple = Oskar; Red = Non-Oskar Arthropod; Green = Non-Arthropod Eukaryote; Blue = Bacteria. Names following leaves display the UniProt accession number followed by the species name and the UniProt protein name.

Figure 2–figure supplement 5: SOWHAT constrained trees and results. Two trees constrained by alternative relationships that would be expected under vertical transmission of sequences were designed and tested against our result supporting a putative HGT event of the OSK domain. (**a**) The first tree (right) is constrained by domain of life, requiring bacterial and eukaryotic sequences to be monophyletic, and disallowing sister group relationships of subsets of eukaryotic sequences and bacterial sequences. Our unconstrained tree topology (left) outperformed this topology with a p-value of 0.002 (95% confidence interval upper: 0.007 lower: 0.0002). (**b**) The second tree requires monophyly of Eukaryota. Our unconstrained tree topology (left) outperformed this topology with a p-value of 0.009 (95% confidence interval upper: 0.017 lower: 0.004). (**c**) The third tree tested whether the LOTUS domain split observed in the tree generated with the MUSCLE alignment was significantly different from a tree where the LOTUS sequences formed a monophyly. The unconstrained tree (left) outperformed this topology with a p-value of 0.037 (95% confidence interval upper: 0.05 lower: 0.026).

Figure 2–figure supplement 6: LOTUS Domain RaxML PRANK Tree. Phylogenetic tree of the same sequences used for the previous LOTUS trees. The sequences were aligned using PRANK and the tree generated with RaxML as described in ***Phylogenetic Analysis Based on PRANK alignment***. Sequences are color-coded as follows: Purple = Oskar; Red = Non-Oskar Arthropod; Green = Non-Arthropod Eukaryote; Blue = Bacteria. Names following leaves display the UniProt accession number followed by the species name and the UniProt protein name.

Figure 2–figure supplement 7: OSK Domain RaxML PRANK Tree. Phylogenetic tree of the same sequences used for the previous OSK trees. The sequences were aligned using PRANK and the tree generated with RaxML as described in ***Phylogenetic Analysis Based on PRANK alignment***. Sequences are color-coded as follows: Purple = Oskar; Red = Non-Oskar Arthropod; Green = Non-Arthropod Eukaryote; Blue = Bacteria. Names following leaves display the UniProt accession number followed by the species name and the UniProt protein name.

Figure 2–figure supplement 8: OSK Tree PRANK Comparison. Comparison of the tree obtained with RaxML starting from the MUSCLE alignment (left) versus the PRANK alignment (right) for the OSK domain. Similarity scores for the branching events are color coded from yellow to blue (see figure color bar legend). The OSK (purple) clade and CAZ3 (green) clade have been colored and compacted for readability as they do not have any internal branching changes. Node color is blue if the leaf is a sequence, and red if this is a compacted group of sequences.

Figure 2–figure supplement 9: LOTUS Tree PRANK Comparison. Comparison of the tree obtained with RaxML starting from the MUSCLE alignment (left) versus the PRANK alignment (right) for the LOTUS domain. Similarity scores for the branching events are color coded from yellow to blue (see figure color bar legend). The LOTUS (purple) clades have been colored and

compacted for readability as they do not have any internal branching changes. Node color is blue if the leaf is a sequence, and red if this is a compacted group of sequences.

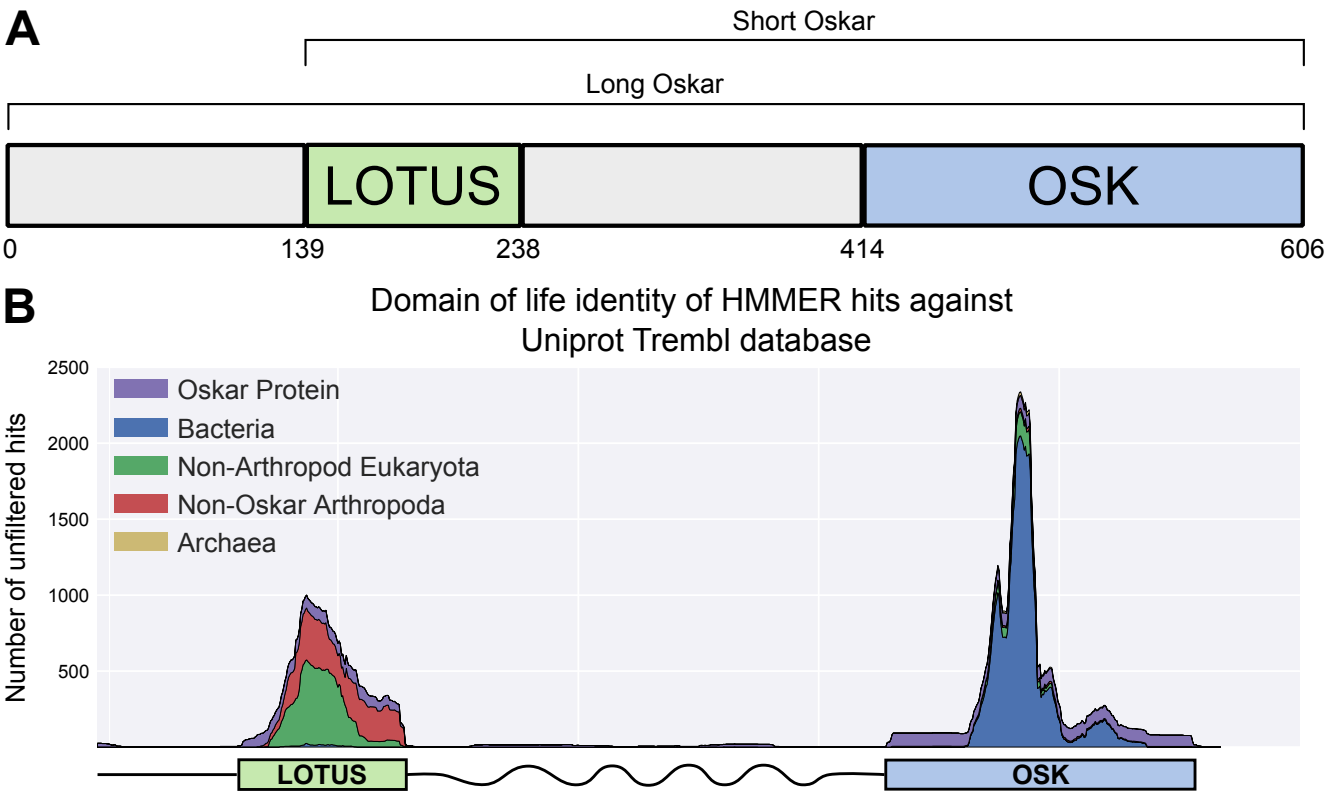
Figure 2–figure supplement 10: LOTUS Domain RaxML T-Coffee Tree. Phylogenetic tree of the same sequences used for the previous LOTUS trees. The sequences were aligned using T-Coffee and the tree generated with RaxML as described in ***Phylogenetic Analysis Based on* T-Coffee *alignment***. Sequences are color-coded as follows: Purple = Oskar; Red = Non-Oskar Arthropod; Green = Non-Arthropod Eukaryote; Blue = Bacteria. Names following leaves display the UniProt accession number followed by the species name and the UniProt protein name.

Figure 2–figure supplement 11: OSK Domain RaxML T-Coffee Tree. Phylogenetic tree of the same sequences used for the previous OSK trees. The sequences were aligned using T-Coffee and the tree generated with RaxML as described in ***Phylogenetic Analysis Based on* T-Coffee *alignment***. Sequences are color-coded as follows: Purple = Oskar; Red = Non-Oskar Arthropod; Green = Non-Arthropod Eukaryote; Blue = Bacteria. Names following leaves display the UniProt accession number followed by the species name and the UniProt protein name.

Figure 2–figure supplement 12: OSK Tree T-Coffee Comparison. Comparison of the tree obtained with RaxML starting from the MUSCLE alignment (left) versus the T-Coffee alignment (right) for the OSK domain. Similarity scores for the branching events are color coded from yellow to blue (see figure color bar legend). The OSK (purple) clade and CAZ3 (green) clade have been colored and compacted for readability as they do not have any internal branching changes. Node color is blue if the leaf is a sequence, and red if this is a compacted group of sequences.

Figure 2–figure supplement 13: LOTUS Tree T-Coffee Comparison. Comparison of the tree obtained with RaxML starting from the MUSCLE alignment (left) versus the T-Coffee alignment (right) for the LOTUS domain. Similarity scores for the branching events are color coded from yellow to blue (see figure color bar legend). The LOTUS (purple) clades have been colored and compacted for readability as they do not have any internal branching changes. Node color is blue if the leaf is a sequence, and red if this is a compacted group of sequences.

Figure 1
Blondel, Jones & Extavour



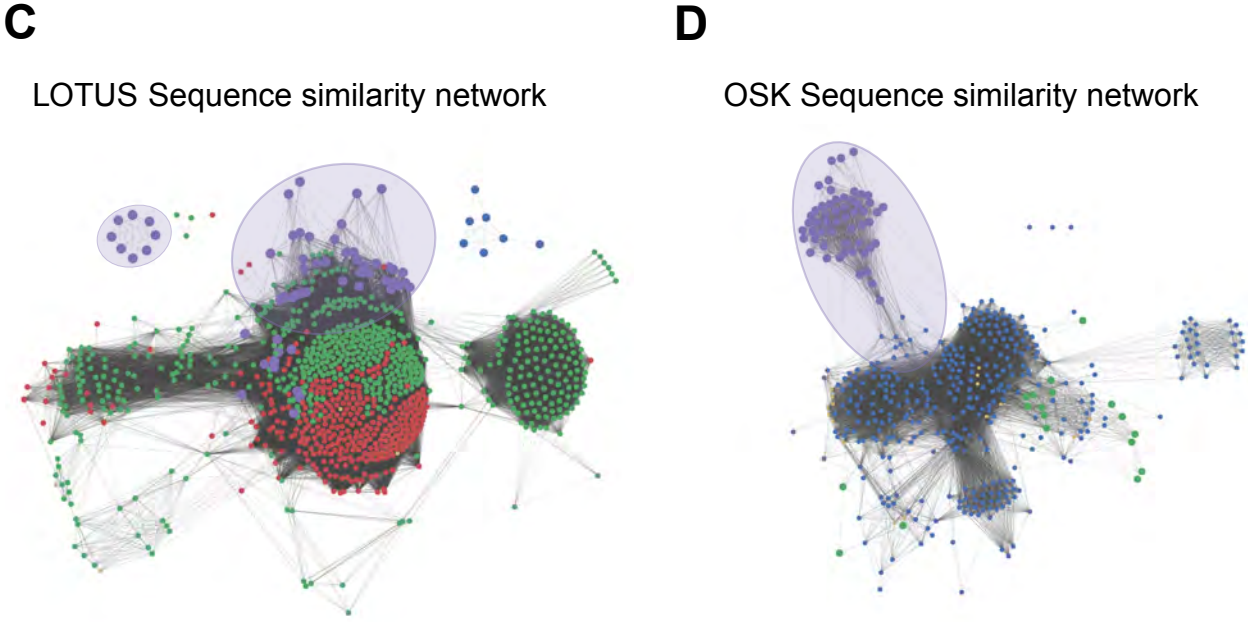
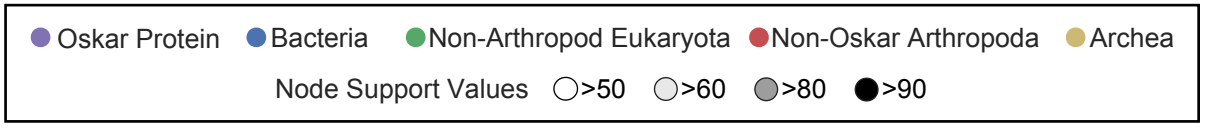
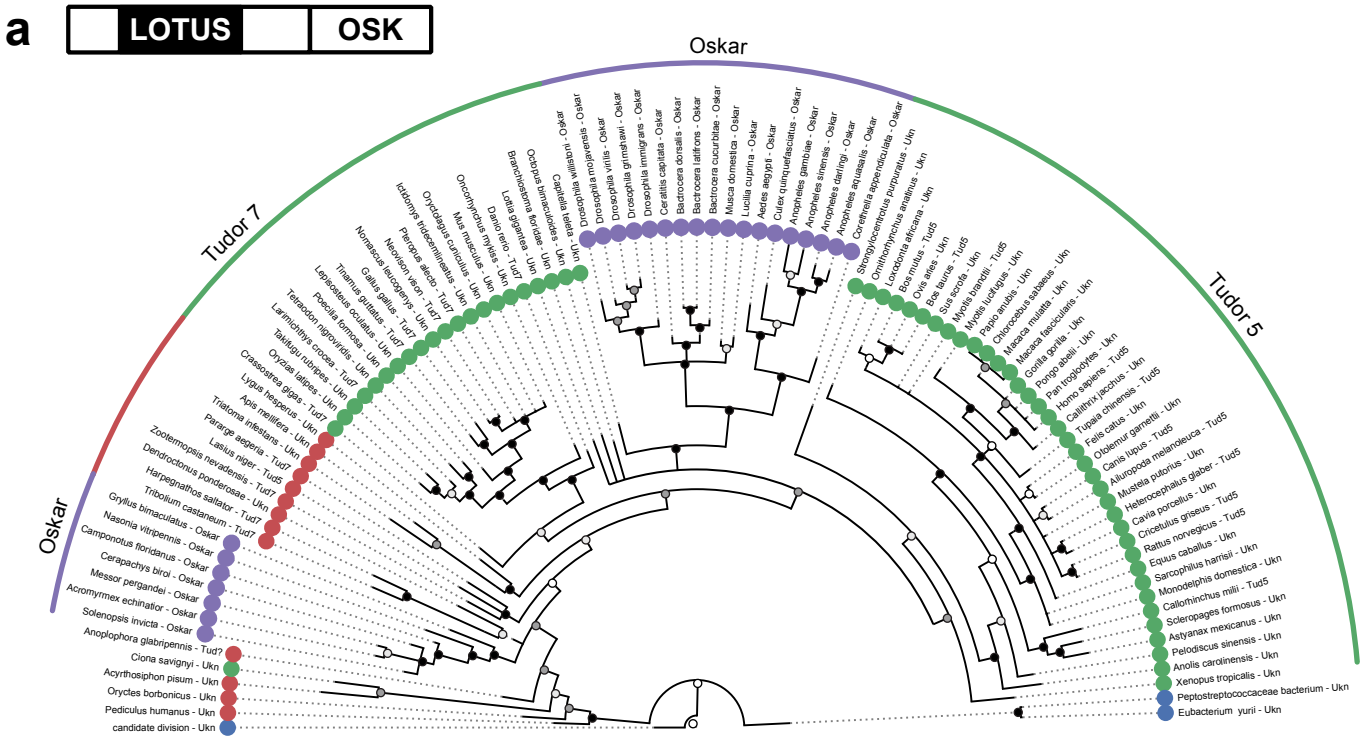
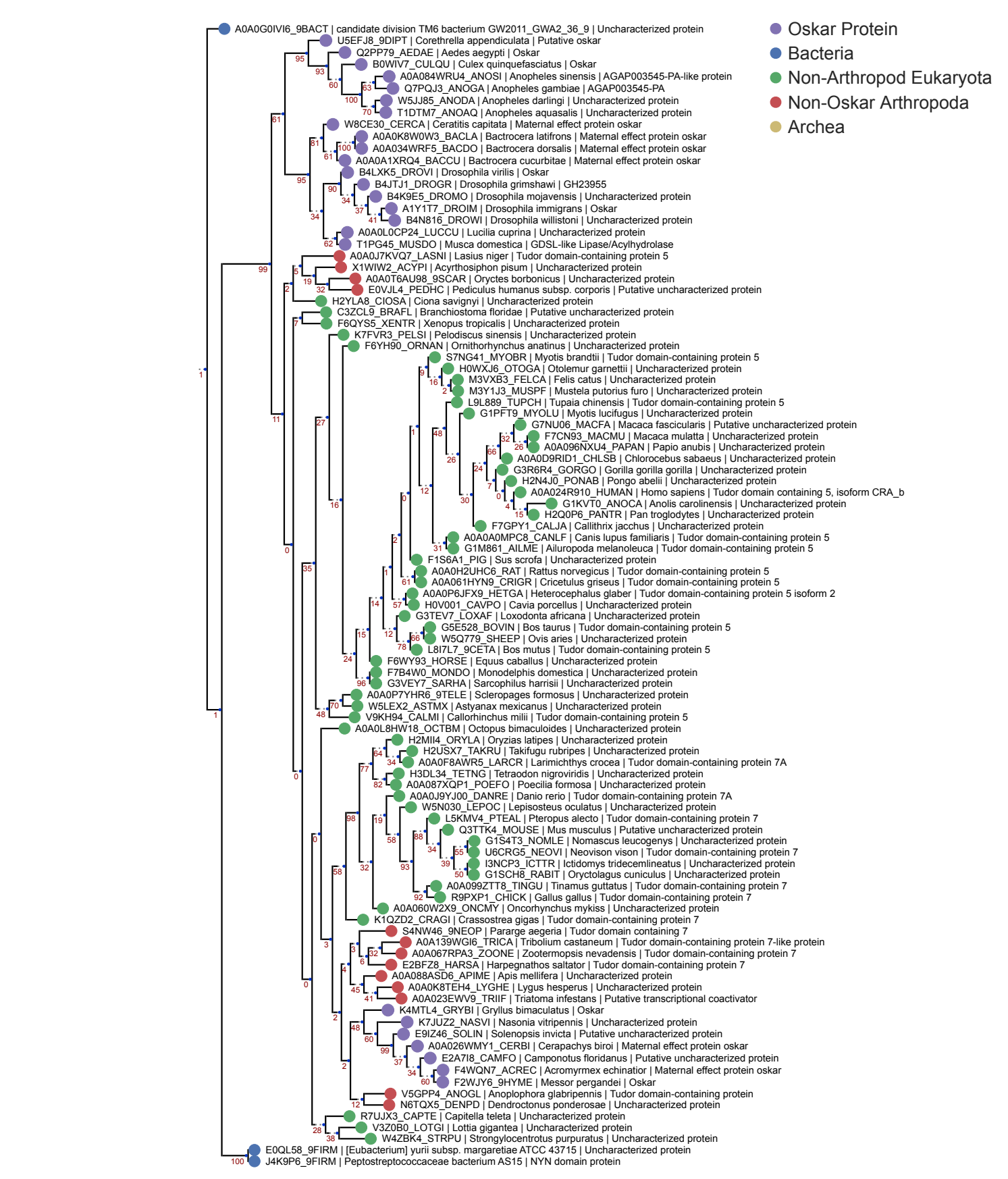
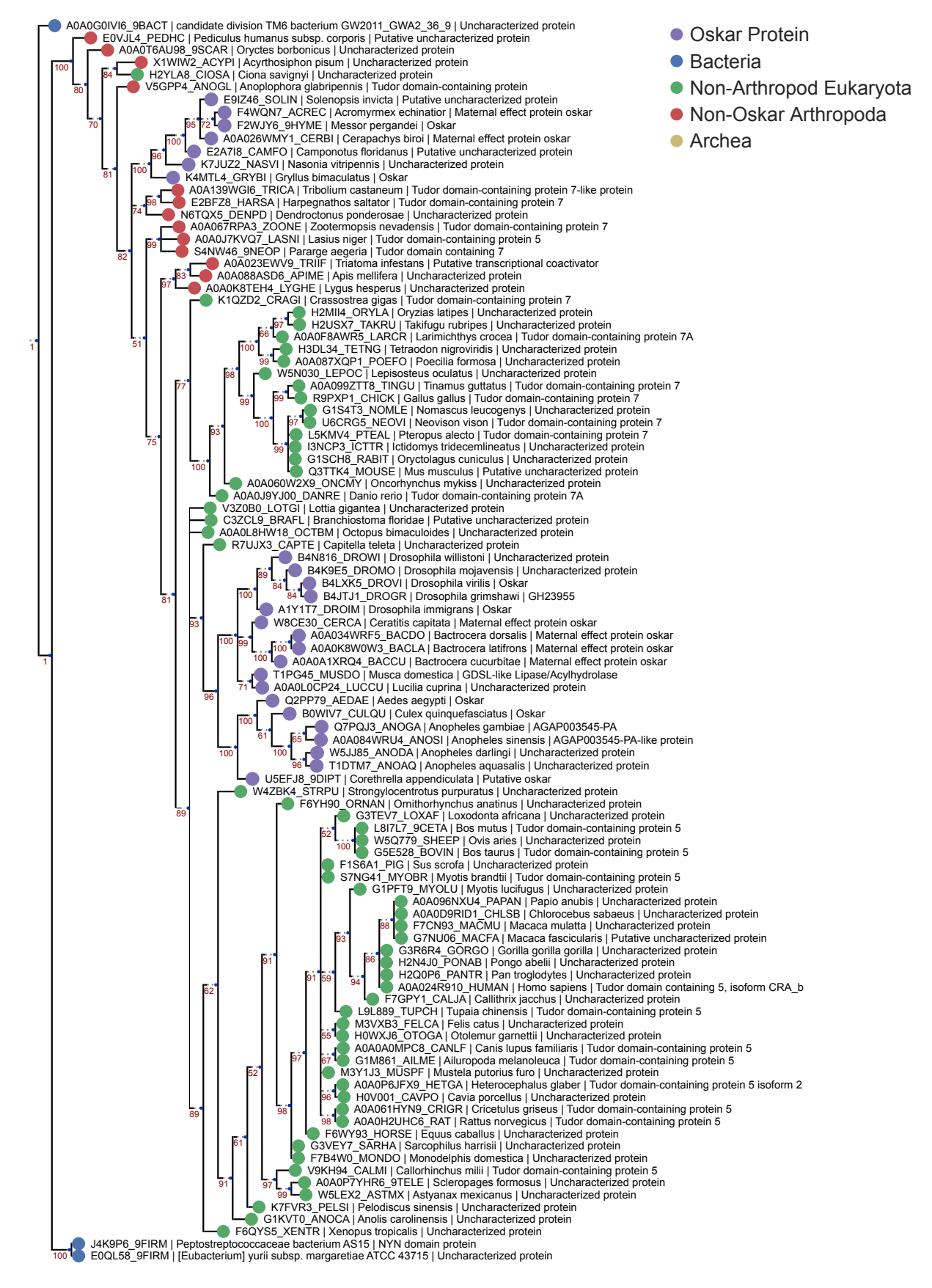


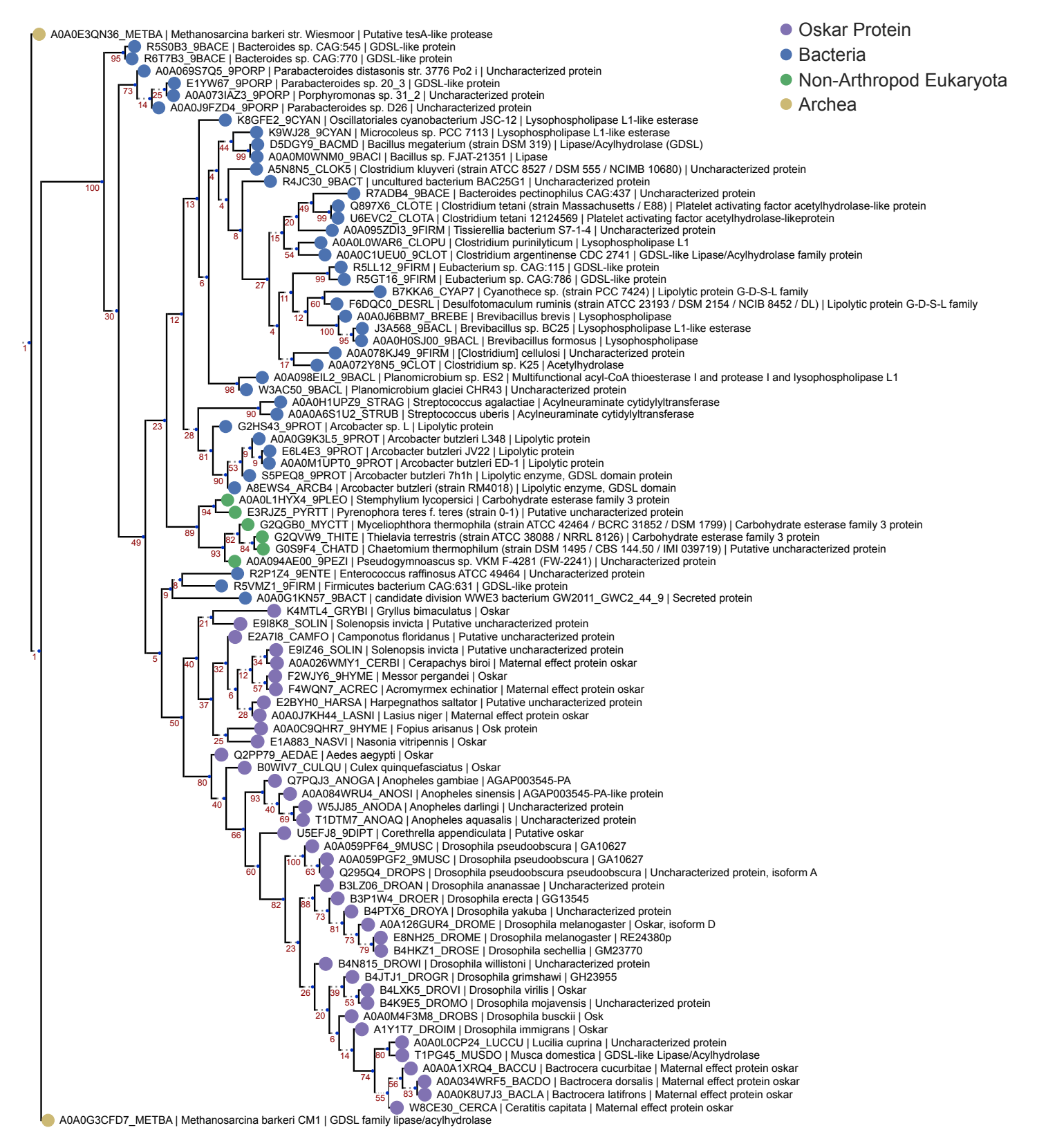
Figure 2
Blondel, Jones & Extavour

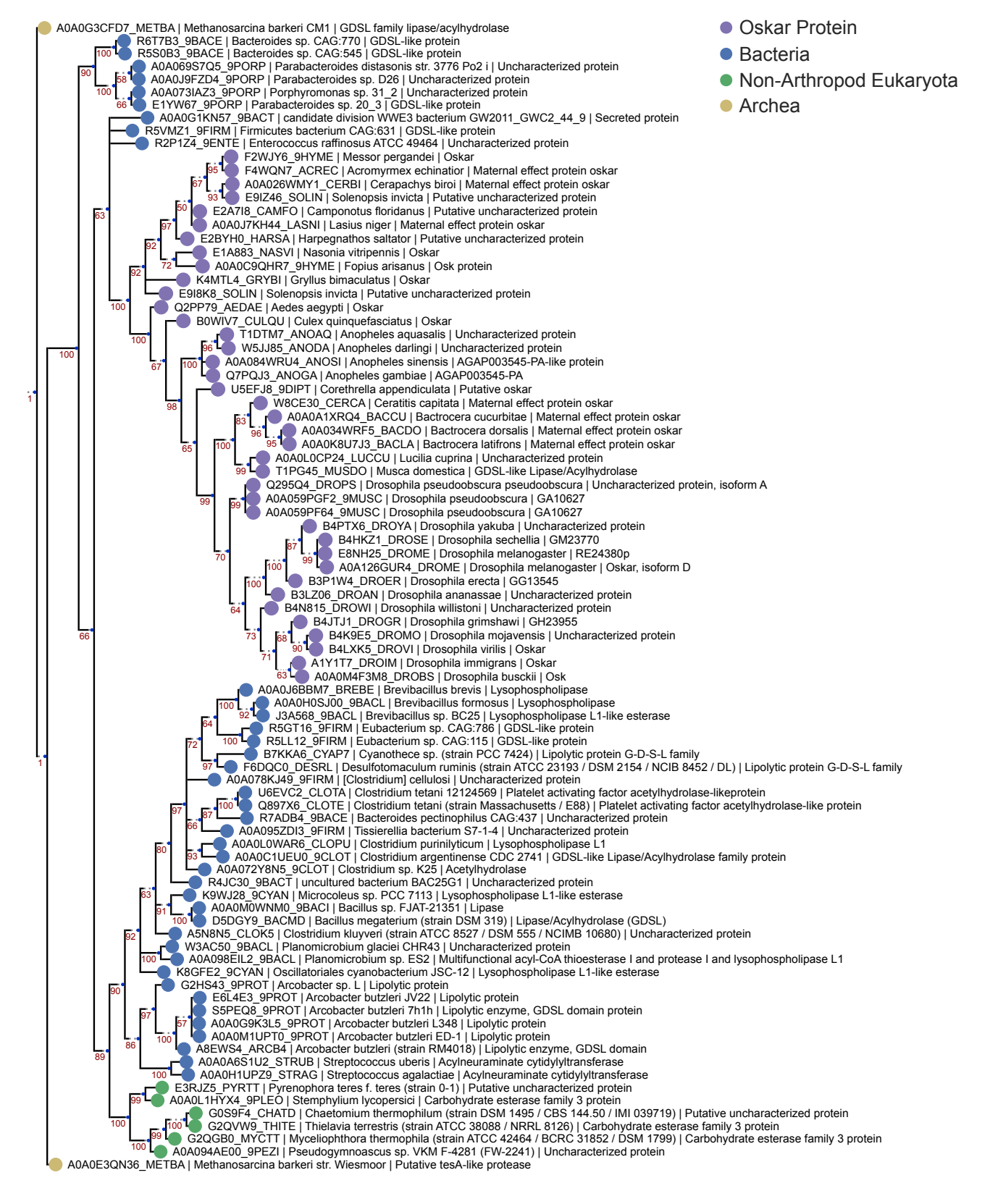


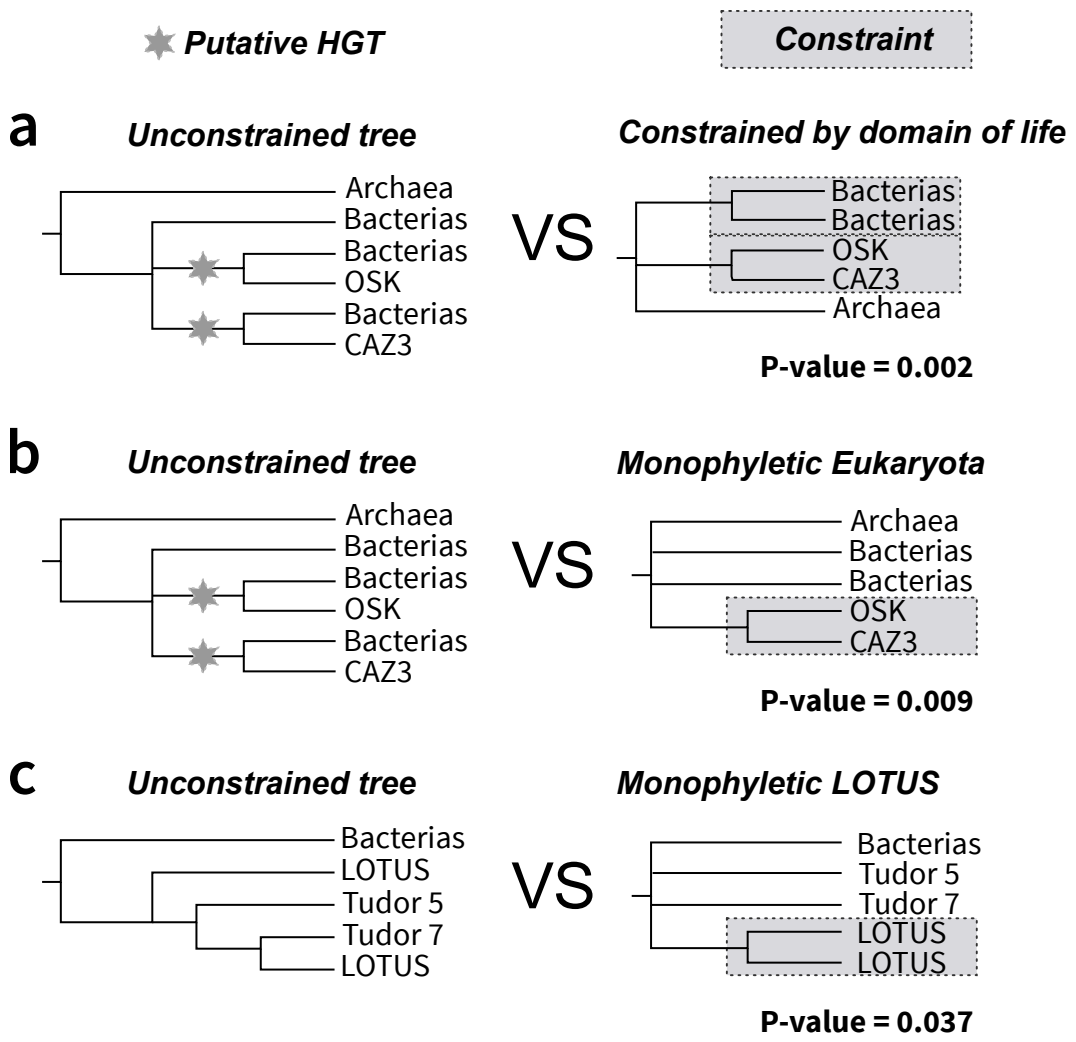


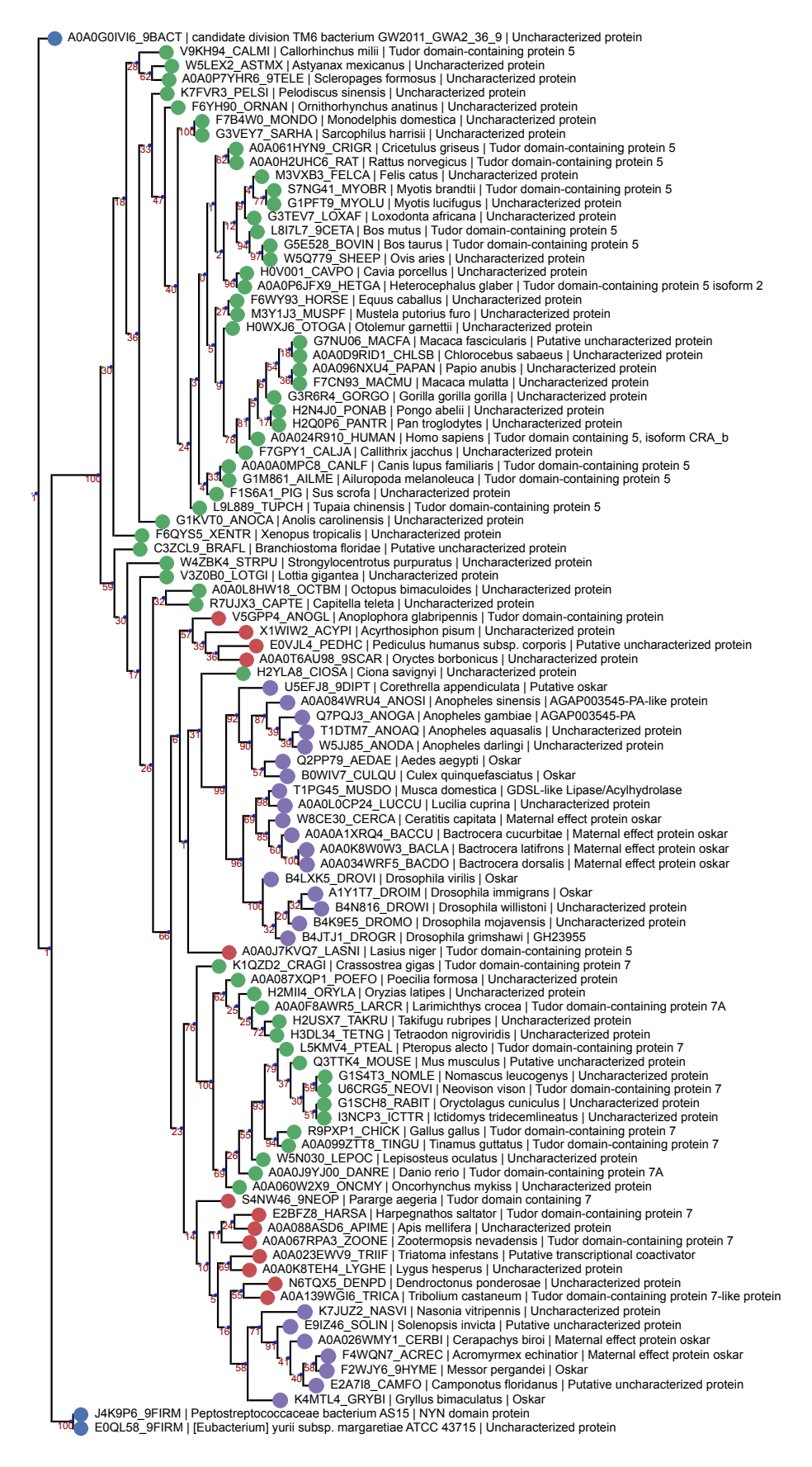

Oskar Protein
Bacteria
Non-Arthropod Eukaryota
Non-Oskar Arthropoda
Archea
A0A0G0IVI6_9BACT | candidate division TM6 bacterium GW2011_GWA2_36_9 | Uncharacterized protein
U5EFJ8_9DIPT | Corethrella appendiculata | Putative oskar
Q2PP79_AEDAE | Aedes aegypti | Oskar
B0WIV7_CULQU | Culex quinquefasciatus | Oskar
A0A084WRU4_ANOSI | Anopheles sinensis | AGAP003545-PA-like protein
Q7PQJ3_ANOGA | Anopheles gambiae | AGAP003545-PA
W5JJ85_ANODA | Anopheles darlingi | Uncharacterized protein
T1DTM7_ANOAQ | Anopheles aquasalis | Uncharacterized protein
W8CE30_CERCA | Ceratitis capitata | Maternal effect protein oskar
A0A0K8W0W3_BACLA | Bactrocera latifrons | Maternal effect protein oskar
A0A034WRF5_BACDO | Bactrocera dorsalis | Maternal effect protein oskar
A0A0A1XRQ4_BACCU | Bactrocera cucurbitae | Maternal effect protein oskar
B4LXK5_DROVI | Drosophila virilis | Oskar
B4JTJ1_DROGR | Drosophila grimshawi | GH23955
B4K9E5_DROMO | Drosophila mojavensis | Uncharacterized protein
A1Y1T7_DROIM | Drosophila immigrans | Oskar
B4N816_DROWI | Drosophila willistoni | Uncharacterized protein
A0A0L0CP24_LUCCU | Lucilia cuprina | Uncharacterized protein
T1PG45_MUSDO | Musca domestica | GDSL-like Lipase/Acylhydrolase
A0A0J7KVQ7_LASNI | Lasius niger | Tudor domain-containing protein 5
X1WIW2_ACYPI | Acyrthosiphon pisum | Uncharacterized protein
A0A0T6AU98_9SCAR | Oryctes borbonicus | Uncharacterized protein
E0VJL4_PEDHC | Pediculus humanus subsp. corporis | Putative uncharacterized protein
H2YLA8_CIOSA | Ciona savignyi | Uncharacterized protein
C3ZCL9_BRAFL | Branchiostoma floridae | Putative uncharacterized protein
F6QYS5_XENTR | Xenopus tropicalis | Uncharacterized protein
K7FVR3_PELSI | Pelodiscus sinensis | Uncharacterized protein
F6YH90_ORNAN | Ornithorhynchus anatinus | Uncharacterized protein
S7NG41_MYOBR | Myotis brandtii | Tudor domain-containing protein 5
H0WXJ6_OTOGA | Otolemur garnettii | Uncharacterized protein
M3VXB3_FELCA | Felis catus | Uncharacterized protein
M3Y1J3_MUSPF | Mustela putorius furo | Uncharacterized protein
L9L889_TUPCH | Tupaia chinensis | Tudor domain-containing protein 5
G1PFT9_MYOLU | Myotis lucifugus | Uncharacterized protein
G7NU06_MACFA | Macaca fascicularis | Putative uncharacterized protein
F7CN93_MACMU | Macaca mulatta | Uncharacterized protein
A0A096NXU4_PAPAN | Papio anubis | Uncharacterized protein
A0A0D9RID1_CHLSB | Chlorocebus sabaeus | Uncharacterized protein
G3R6R4_GORGO | Gorilla gorilla gorilla | Uncharacterized protein
H2N4J0_PONAB | Pongo abelii | Uncharacterized protein
A0A024R910_HUMAN | Homo sapiens | Tudor domain containing 5, isoform CRA_b
G1KVT0_ANOCA | Anolis carolinensis | Uncharacterized protein
H2Q0P6_PANTR | Pan troglodytes | Uncharacterized protein
F7GPY1_CALJA | Callithrix jacchus | Uncharacterized protein
A0A0A0MPC8_CANLF | Canis lupus familiaris | Tudor domain-containing protein 5
G1M861_AILME | Ailuropoda melanoleuca | Tudor domain-containing protein 5
F1S6A1_PIG | Sus scrofa | Uncharacterized protein
A0A0H2UHC6_RAT | Rattus norvegicus | Tudor domain-containing protein 5
A0A061HYN9_CRIGR | Cricetulus griseus | Tudor domain-containing protein 5
A0A0P6JFX9_HETGA | Heterocephalus glaber | Tudor domain-containing protein 5 isoform 2
H0V001_CAVPO | Cavia porcellus | Uncharacterized protein
G3TEV7_LOXAF | Loxodonta africana | Uncharacterized protein
G5E528_BOVIN | Bos taurus | Tudor domain-containing protein 5
W5Q779_SHEEP | Ovis aries | Uncharacterized protein
L8I7L7_9CETA | Bos mutus | Tudor domain-containing protein 5
F6WY93_HORSE | Equus caballus | Uncharacterized protein
F7B4W0_MONDO | Monodelphis domestica | Uncharacterized protein
G3VEY7_SARHA | Sarcophilus harrisii | Uncharacterized protein
A0A0P7YHR6_9TELE | Scleropages formosus | Uncharacterized protein
W5LEX2_ASTMX | Astyanax mexicanus | Uncharacterized protein
V9KH94_CALMI | Callorhinchus milii | Tudor domain-containing protein 5
A0A0L8HW18_OCTBM | Octopus bimaculoides | Uncharacterized protein
H2MII4_ORYLA | Oryzias latipes | Uncharacterized protein
H2USX7_TAKRU | Takifugu rubripes | Uncharacterized protein
A0A0F8AWR5_LARCR | Larimichthys crocea | Tudor domain-containing protein 7A
H3DL34_TETNG | Tetraodon nigroviridis | Uncharacterized protein
A0A087XQP1_POEFO | Poecilia formosa | Uncharacterized protein
A0A0J9YJ00_DANRE | Danio rerio | Tudor domain-containing protein 7A
W5N030_LEPOC | Lepisosteus oculatus | Uncharacterized protein
L5KMV4_PTEAL | Pteropus alecto | Tudor domain-containing protein 7
Q3TTK4_MOUSE | Mus musculus | Putative uncharacterized protein
G1S4T3_NOMLE | Nomascus leucogenys | Uncharacterized protein
U6CRG5_NEOVI | Neovison vison | Tudor domain-containing protein 7
I3NCP3_ICTTR | Ictidomys tridecemlineatus | Uncharacterized protein
G1SCH8_RABIT | Oryctolagus cuniculus | Uncharacterized protein
A0A099ZTT8_TINGU | Tinamus guttatus | Tudor domain-containing protein 7
R9PXP1_CHICK | Gallus gallus | Tudor domain-containing protein 7
A0A060W2X9_ONCMY | Oncorhynchus mykiss | Uncharacterized protein
K1QZD2_CRAGI | Crassostrea gigas | Tudor domain-containing protein 7
S4NW46_9NEOP | Pararge aegeria | Tudor domain containing 7
A0A139WGI6_TRICA | Tribolium castaneum | Tudor domain-containing protein 7-like protein
A0A067RPA3_ZOONE | Zootermopsis nevadensis | Tudor domain-containing protein 7
E2BFZ8_HARSA | Harpegnathos saltator | Tudor domain-containing protein 7
A0A088ASD6_APIME | Apis mellifera | Uncharacterized protein
A0A0K8TEH4_LYGHE | Lygus hesperus | Uncharacterized protein
A0A023EWV9_TRIIF | Triatoma infestans | Putative transcriptional coactivator
K4MTL4_GRYBI | Gryllus bimaculatus | Oskar
K7JUZ2_NASVI | Nasonia vitripennis | Uncharacterized protein
E9IZ46_SOLIN | Solenopsis invicta | Putative uncharacterized protein
A0A026WMY1_CERBI | Cerapachys biroi | Maternal effect protein oskar
E2A7I8_CAMFO | Camponotus floridanus | Putative uncharacterized protein
F4WQN7_ACREC | Acromyrmex echinatior | Maternal effect protein oskar
F2WJY6_9HYME | Messor pergandei | Oskar
V5GPP4_ANOGL | Anoplophora glabripennis | Tudor domain-containing protein
N6TQX5_DENPD | Dendroctonus ponderosae | Uncharacterized protein
R7UJX3_CAPTE | Capitella teleta | Uncharacterized protein
V3Z0B0_LOTGI | Lottia gigantea | Uncharacterized protein
W4ZBK4_STRPU | Strongylocentrotus purpuratus | Uncharacterized protein
E0QL58_9FIRM | [Eubacterium] yurii subsp. margaretiae ATCC 43715 | Uncharacterized protein
J4K9P6_9FIRM | Peptostreptococcaceae bacterium AS15 | NYN domain protein

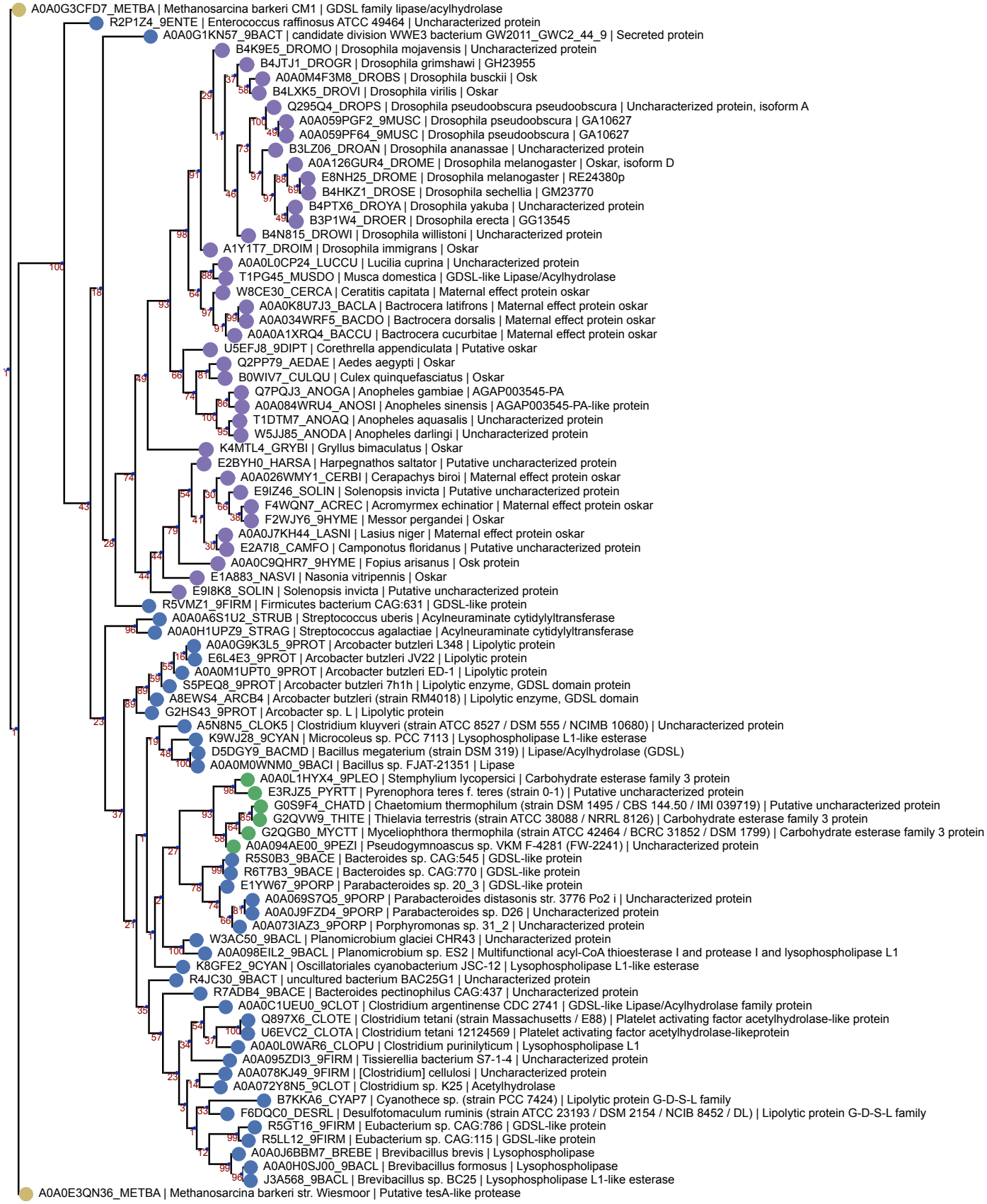

Oskar Protein
Bacteria
Non-Arthropod Eukaryota
Non-Oskar Arthropoda
Archea
A0A0G0IVI6_9BACT | candidate division TM6 bacterium GW2011_GWA2_36_9 | Uncharacterized protein
E0VJL4_PEDHC | Pediculus humanus subsp. corporis | Putative uncharacterized protein
A0A0T6AU98_9SCAR | Oryctes borbonicus | Uncharacterized protein
X1WIW2_ACYPI | Acyrthosiphon pisum | Uncharacterized protein
H2YLA8_CIOSA | Ciona savignyi | Uncharacterized protein
V5GPP4_ANOGL | Anoplophora glabripennis | Tudor domain-containing protein
E9IZ46_SOLIN | Solenopsis invicta | Putative uncharacterized protein
F4WQN7_ACREC | Acromyrmex echinatior | Maternal effect protein oskar
F2WJY6_9HYME | Messor pergandei | Oskar
A0A026WMY1_CERBI | Cerapachys biroi | Maternal effect protein oskar
E2A7I8_CAMFO | Camponotus floridanus | Putative uncharacterized protein
K7JUZ2_NASVI | Nasonia vitripennis | Uncharacterized protein
K4MTL4_GRYBI | Gryllus bimaculatus | Oskar
A0A139WGI6_TRICA | Tribolium castaneum | Tudor domain-containing protein 7-like protein
E2BFZ8_HARSA | Harpegnathos saltator | Tudor domain-containing protein 7
N6TQX5_DENPD | Dendroctonus ponderosae | Uncharacterized protein
A0A067RPA3_ZOONE | Zootermopsis nevadensis | Tudor domain-containing protein 7
A0A0J7KVQ7_LASNI | Lasius niger | Tudor domain-containing protein 5
S4NW46_9NEOP | Pararge aegeria | Tudor domain containing 7
A0A023EWV9_TRIIF | Triatoma infestans | Putative transcriptional coactivator
A0A088ASD6_APIME | Apis mellifera | Uncharacterized protein
A0A0K8TEH4_LYGHE | Lygus hesperus | Uncharacterized protein
K1QZD2_CRAGI | Crassostrea gigas | Tudor domain-containing protein 7
H2MII4_ORYLA | Oryzias latipes | Uncharacterized protein
H2USX7_TAKRU | Takifugu rubripes | Uncharacterized protein
A0A0F8AWR5_LARCR | Larimichthys crocea | Tudor domain-containing protein 7A
H3DL34_TETNG | Tetraodon nigroviridis | Uncharacterized protein
A0A087XQP1_POEFO | Poecilia formosa | Uncharacterized protein
W5N030_LEPOC | Lepisosteus oculatus | Uncharacterized protein
A0A099ZTT8_TINGU | Tinamus guttatus | Tudor domain-containing protein 7
R9PXP1_CHICK | Gallus gallus | Tudor domain-containing protein 7
G1S4T3_NOMLE | Nomascus leucogenys | Uncharacterized protein
U6CRG5_NEOVI | Neovison vison | Tudor domain-containing protein 7
L5KMV4_PTEAL | Pteropus alecto | Tudor domain-containing protein 7
I3NCP3_ICTTR | Ictidomys tridecemlineatus | Uncharacterized protein
G1SCH8_RABIT | Oryctolagus cuniculus | Uncharacterized protein
Q3TTK4_MOUSE | Mus musculus | Putative uncharacterized protein
A0A060W2X9_ONCMY | Oncorhynchus mykiss | Uncharacterized protein
A0A0J9YJ00_DANRE | Danio rerio | Tudor domain-containing protein 7A
V3Z0B0_LOTGI | Lottia gigantea | Uncharacterized protein
C3ZCL9_BRAFL | Branchiostoma floridae | Putative uncharacterized protein
A0A0L8HW18_OCTBM | Octopus bimaculoides | Uncharacterized protein
R7UJX3_CAPTE | Capitella teleta | Uncharacterized protein
B4N816_DROWI | Drosophila willistoni | Uncharacterized protein
B4K9E5_DROMO | Drosophila mojavensis | Uncharacterized protein
B4LXK5_DROVI | Drosophila virilis | Oskar
B4JTJ1_DROGR | Drosophila grimshawi | GH23955
A1Y1T7_DROIM | Drosophila immigrans | Oskar
W8CE30_CERCA | Ceratitis capitata | Maternal effect protein oskar
A0A034WRF5_BACDO | Bactrocera dorsalis | Maternal effect protein oskar
A0A0K8W0W3_BACLA | Bactrocera latifrons | Maternal effect protein oskar
A0A0A1XRQ4_BACCU | Bactrocera cucurbitae | Maternal effect protein oskar
T1PG45_MUSDO | Musca domestica | GDSL-like Lipase/Acylhydrolase
A0A0L0CP24_LUCCU | Lucilia cuprina | Uncharacterized protein
Q2PP79_AEDAE | Aedes aegypti | Oskar
B0WIV7_CULQU | Culex quinquefasciatus | Oskar
Q7PQJ3_ANOGA | Anopheles gambiae | AGAP003545-PA
A0A084WRU4_ANOSI | Anopheles sinensis | AGAP003545-PA-like protein
W5JJ85_ANODA | Anopheles darlingi | Uncharacterized protein
T1DTM7_ANOAQ | Anopheles aquasalis | Uncharacterized protein
U5EFJ8_9DIPT | Corethrella appendiculata | Putative oskar
W4ZBK4_STRPU | Strongylocentrotus purpuratus | Uncharacterized protein
F6YH90_ORNAN | Ornithorhynchus anatinus | Uncharacterized protein
G3TEV7_LOXAF | Loxodonta africana | Uncharacterized protein
L8I7L7_9CETA | Bos mutus | Tudor domain-containing protein 5
W5Q779_SHEEP | Ovis aries | Uncharacterized protein
G5E528_BOVIN | Bos taurus | Tudor domain-containing protein 5
F1S6A1_PIG | Sus scrofa | Uncharacterized protein
S7NG41_MYOBR | Myotis brandtii | Tudor domain-containing protein 5
G1PFT9_MYOLU | Myotis lucifugus | Uncharacterized protein
A0A096NXU4_PAPAN | Papio anubis | Uncharacterized protein
A0A0D9RID1_CHLSB | Chlorocebus sabaeus | Uncharacterized protein
F7CN93_MACMU | Macaca mulatta | Uncharacterized protein
G7NU06_MACFA | Macaca fascicularis | Putative uncharacterized protein
G3R6R4_GORGO | Gorilla gorilla gorilla | Uncharacterized protein
H2N4J0_PONAB | Pongo abelii | Uncharacterized protein
H2Q0P6_PANTR | Pan troglodytes | Uncharacterized protein
A0A024R910_HUMAN | Homo sapiens | Tudor domain containing 5, isoform CRA_b
F7GPY1_CALJA | Callithrix jacchus | Uncharacterized protein
L9L889_TUPCH | Tupaia chinensis | Tudor domain-containing protein 5
M3VXB3_FELCA | Felis catus | Uncharacterized protein
H0WXJ6_OTOGA | Otolemur garnettii | Uncharacterized protein
A0A0A0MPC8_CANLF | Canis lupus familiaris | Tudor domain-containing protein 5
G1M861_AILME | Ailuropoda melanoleuca | Tudor domain-containing protein 5
M3Y1J3_MUSPF | Mustela putorius furo | Uncharacterized protein
A0A0P6JFX9_HETGA | Heterocephalus glaber | Tudor domain-containing protein 5 isoform 2
H0V001_CAVPO | Cavia porcellus | Uncharacterized protein
A0A061HYN9_CRIGR | Cricetulus griseus | Tudor domain-containing protein 5
A0A0H2UHC6_RAT | Rattus norvegicus | Tudor domain-containing protein 5
F6WY93_HORSE | Equus caballus | Uncharacterized protein
G3VEY7_SARHA | Sarcophilus harrisii | Uncharacterized protein
F7B4W0_MONDO | Monodelphis domestica | Uncharacterized protein
V9KH94_CALMI | Callorhinchus milii | Tudor domain-containing protein 5
A0A0P7YHR6_9TELE | Scleropages formosus | Uncharacterized protein
W5LEX2_ASTMX | Astyanax mexicanus | Uncharacterized protein
K7FVR3_PELSI | Pelodiscus sinensis | Uncharacterized protein
G1KVT0_ANOCA | Anolis carolinensis | Uncharacterized protein
F6QYS5_XENTR | Xenopus tropicalis | Uncharacterized protein
J4K9P6_9FIRM | Peptostreptococcaceae bacterium AS15 | NYN domain protein
E0QL58_9FIRM | [Eubacterium] yurii subsp. margaretiae ATCC 43715 | Uncharacterized protein

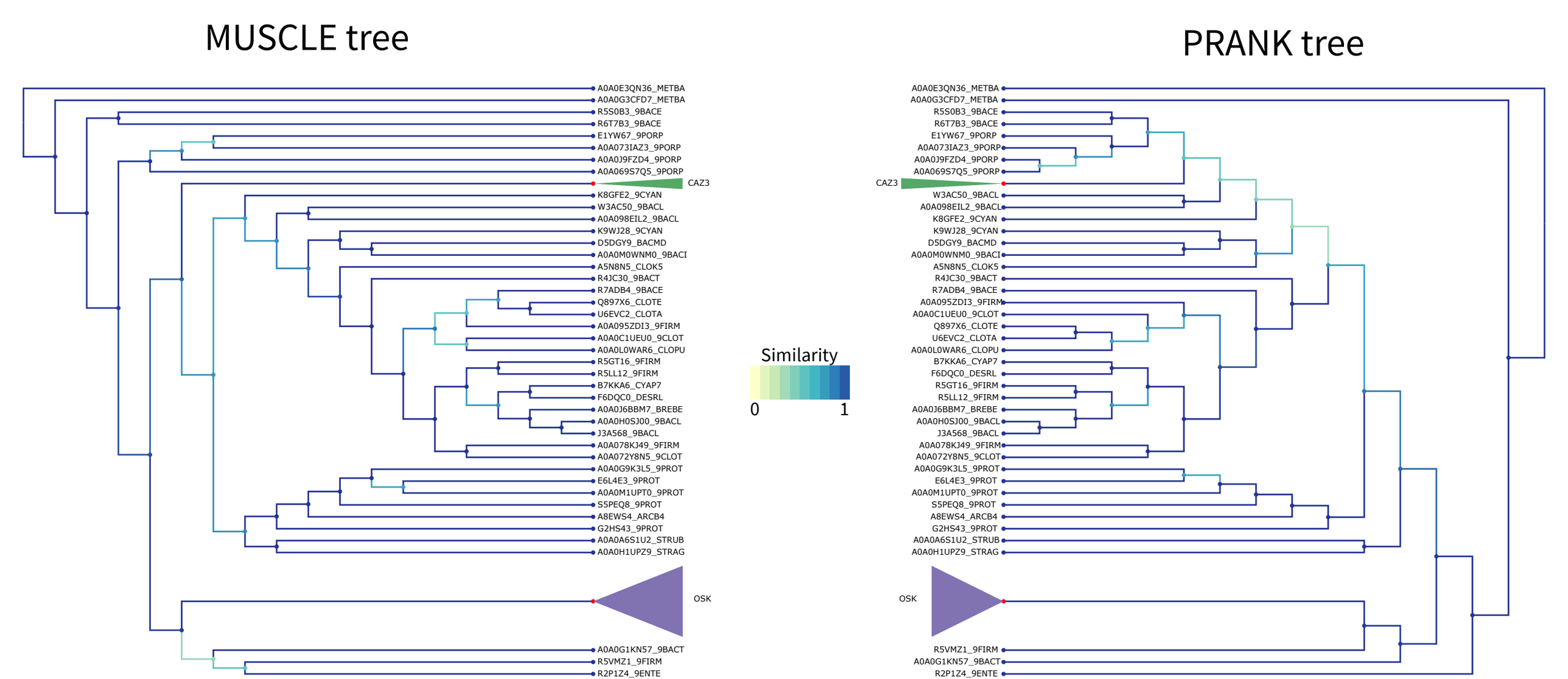

Oskar Protein
Bacteria
Non-Arthropod Eukaryota
Archea
A0A0E3QN36_METBA | Methanosarcina barkeri str. Wiesmoor | Putative tesA-like protease
R5S0B3_9BACE | Bacteroides sp. CAG:545 | GDSL-like protein
R6T7B3_9BACE | Bacteroides sp. CAG:770 | GDSL-like protein
A0A069S7Q5_9PORP | Parabacteroides distasonis str. 3776 Po2 i | Uncharacterized protein
E1YW67_9PORP | Parabacteroides sp. 20_3 | GDSL-like protein
A0A073IAZ3_9PORP | Porphyromonas sp. 31_2 | Uncharacterized protein
A0A0J9FZD4_9PORP | Parabacteroides sp. D26 | Uncharacterized protein
K8GFE2_9CYAN | Oscillatoriales cyanobacterium JSC-12 | Lysophospholipase L1-like esterase
K9WJ28_9CYAN | Microcoleus sp. PCC 7113 | Lysophospholipase L1-like esterase
D5DGY9_BACMD | Bacillus megaterium (strain DSM 319) | Lipase/Acylhydrolase (GDSL)
A0A0M0WNM0_9BACI | Bacillus sp. FJAT-21351 | Lipase
A5N8N5_CLOK5 | Clostridium kluyveri (strain ATCC 8527 / DSM 555 / NCIMB 10680) | Uncharacterized protein
R4JC30_9BACT | uncultured bacterium BAC25G1 | Uncharacterized protein
R7ADB4_9BACE | Bacteroides pectinophilus CAG:437 | Uncharacterized protein
Q897X6_CLOTE | Clostridium tetani (strain Massachusetts / E88) | Platelet activating factor acetylhydrolase-like protein
U6EVC2_CLOTA | Clostridium tetani 12124569 | Platelet activating factor acetylhydrolase-likeprotein
A0A095ZDI3_9FIRM | Tissierellia bacterium S7-1-4 | Uncharacterized protein
A0A0L0WAR6_CLOPU | Clostridium purinilyticum | Lysophospholipase L1
A0A0C1UEU0_9CLOT | Clostridium argentinense CDC 2741 | GDSL-like Lipase/Acylhydrolase family protein
R5LL12_9FIRM | Eubacterium sp. CAG:115 | GDSL-like protein
R5GT16_9FIRM | Eubacterium sp. CAG:786 | GDSL-like protein
B7KKA6_CYAP7 | Cyanothece sp. (strain PCC 7424) | Lipolytic protein G-D-S-L family
F6DQC0_DESRL | Desulfotomaculum ruminis (strain ATCC 23193 / DSM 2154 / NCIB 8452 / DL) | Lipolytic protein G-D-S-L family
A0A0J6BBM7_BREBE | Brevibacillus brevis | Lysophospholipase
J3A568_9BACL | Brevibacillus sp. BC25 | Lysophospholipase L1-like esterase
A0A0H0SJ00_9BACL | Brevibacillus formosus | Lysophospholipase
A0A078KJ49_9FIRM | [Clostridium] cellulosi | Uncharacterized protein
A0A072Y8N5_9CLOT | Clostridium sp. K25 | Acetylhydrolase
A0A098EIL2_9BACL | Planomicrobium sp. ES2 | Multifunctional acyl-CoA thioesterase I and protease I and lysophospholipase L1
W3AC50_9BACL | Planomicrobium glaciei CHR43 | Uncharacterized protein
A0A0H1UPZ9_STRAG | Streptococcus agalactiae | Acylneuraminate cytidylyltransferase
A0A0A6S1U2_STRUB | Streptococcus uberis | Acylneuraminate cytidylyltransferase
G2HS43_9PROT | Arcobacter sp. L | Lipolytic protein
A0A0G9K3L5_9PROT | Arcobacter butzleri L348 | Lipolytic protein
E6L4E3_9PROT | Arcobacter butzleri JV22 | Lipolytic protein
A0A0M1UPT0_9PROT | Arcobacter butzleri ED-1 | Lipolytic protein
S5PEQ8_9PROT | Arcobacter butzleri 7h1h | Lipolytic enzyme, GDSL domain protein
A8EWS4_ARCB4 | Arcobacter butzleri (strain RM4018) | Lipolytic enzyme, GDSL domain
A0A0L1HYX4_9PLEO | Stemphylium lycopersici | Carbohydrate esterase family 3 protein
E3RJZ5_PYRTT | Pyrenophora teres f. teres (strain 0-1) | Putative uncharacterized protein
G2QGB0_MYCTT | Myceliophthora thermophila (strain ATCC 42464 / BCRC 31852 / DSM 1799) | Carbohydrate esterase family 3 protein
G2QVW9_THITE | Thielavia terrestris (strain ATCC 38088 / NRRL 8126) | Carbohydrate esterase family 3 protein
G0S9F4_CHATD | Chaetomium thermophilum (strain DSM 1495 / CBS 144.50 / IMI 039719) | Putative uncharacterized protein
A0A094AE00_9PEZI | Pseudogymnoascus sp. VKM F-4281 (FW-2241) | Uncharacterized protein
R2P1Z4_9ENTE | Enterococcus raffinosus ATCC 49464 | Uncharacterized protein
R5VMZ1_9FIRM | Firmicutes bacterium CAG:631 | GDSL-like protein
A0A0G1KN57_9BACT | candidate division WWE3 bacterium GW2011_GWC2_44_9 | Secreted protein
K4MTL4_GRYBI | Gryllus bimaculatus | Oskar
E9I8K8_SOLIN | Solenopsis invicta | Putative uncharacterized protein
E2A7I8_CAMFO | Camponotus floridanus | Putative uncharacterized protein
E9IZ46_SOLIN | Solenopsis invicta | Putative uncharacterized protein
A0A026WMY1_CERBI | Cerapachys biroi | Maternal effect protein oskar
F2WJY6_9HYME | Messor pergandei | Oskar
F4WQN7_ACREC | Acromyrmex echinatior | Maternal effect protein oskar
E2BYH0_HARSA | Harpegnathos saltator | Putative uncharacterized protein
A0A0J7KH44_LASNI | Lasius niger | Maternal effect protein oskar
A0A0C9QHR7_9HYME | Fopius arisanus | Osk protein
E1A883_NASVI | Nasonia vitripennis | Oskar
Q2PP79_AEDAE | Aedes aegypti | Oskar
B0WIV7_CULQU | Culex quinquefasciatus | Oskar
Q7PQJ3_ANOGA | Anopheles gambiae | AGAP003545-PA
A0A084WRU4_ANOSI | Anopheles sinensis | AGAP003545-PA-like protein
W5JJ85_ANODA | Anopheles darlingi | Uncharacterized protein
T1DTM7_ANOAQ | Anopheles aquasalis | Uncharacterized protein
U5EFJ8_9DIPT | Corethrella appendiculata | Putative oskar
A0A059PF64_9MUSC | Drosophila pseudoobscura | GA10627
A0A059PGF2_9MUSC | Drosophila pseudoobscura | GA10627
Q295Q4_DROPS | Drosophila pseudoobscura pseudoobscura | Uncharacterized protein, isoform A
B3LZ06_DROAN | Drosophila ananassae | Uncharacterized protein
B3P1W4_DROER | Drosophila erecta | GG13545
B4PTX6_DROYA | Drosophila yakuba | Uncharacterized protein
A0A126GUR4_DROME | Drosophila melanogaster | Oskar, isoform D
E8NH25_DROME | Drosophila melanogaster | RE24380p
B4HKZ1_DROSE | Drosophila sechellia | GM23770
B4N815_DROWI | Drosophila willistoni | Uncharacterized protein
B4JTJ1_DROGR | Drosophila grimshawi | GH23955
B4LXK5_DROVI | Drosophila virilis | Oskar
B4K9E5_DROMO | Drosophila mojavensis | Uncharacterized protein
A0A0M4F3M8_DROBS | Drosophila busckii | Osk
A1Y1T7_DROIM | Drosophila immigrans | Oskar
A0A0L0CP24_LUCCU | Lucilia cuprina | Uncharacterized protein
T1PG45_MUSDO | Musca domestica | GDSL-like Lipase/Acylhydrolase
A0A0A1XRQ4_BACCU | Bactrocera cucurbitae | Maternal effect protein oskar
A0A034WRF5_BACDO | Bactrocera dorsalis | Maternal effect protein oskar
A0A0K8U7J3_BACLA | Bactrocera latifrons | Maternal effect protein oskar
W8CE30_CERCA | Ceratitis capitata | Maternal effect protein oskar
A0A0G3CFD7_METBA | Methanosarcina barkeri CM1 | GDSL family lipase/acylhydrolase

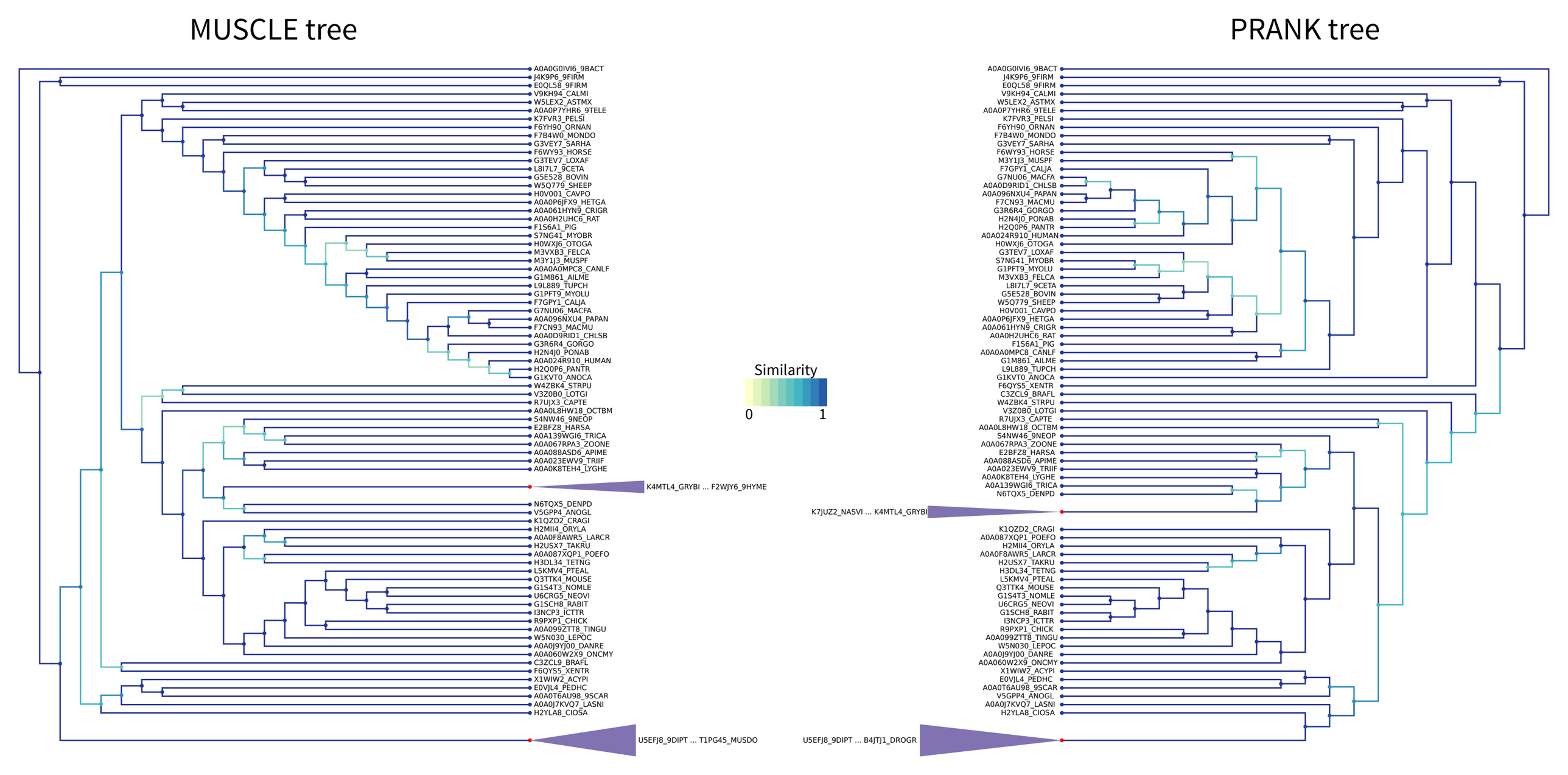

Oskar Protein
Bacteria
Non-Arthropod Eukaryota
Archea
A0A0G3CFD7_METBA | Methanosarcina barkeri CM1 | GDSL family lipase/acylhydrolase
R6T7B3_9BACE | Bacteroides sp. CAG:770 | GDSL-like protein
R5S0B3_9BACE | Bacteroides sp. CAG:545 | GDSL-like protein
A0A069S7Q5_9PORP | Parabacteroides distasonis str. 3776 Po2 i | Uncharacterized protein
A0A0J9FZD4_9PORP | Parabacteroides sp. D26 | Uncharacterized protein
A0A073IAZ3_9PORP | Porphyromonas sp. 31_2 | Uncharacterized protein
E1YW67_9PORP | Parabacteroides sp. 20_3 | GDSL-like protein
A0A0G1KN57_9BACT | candidate division WWE3 bacterium GW2011_GWC2_44_9 | Secreted protein
R5VMZ1_9FIRM | Firmicutes bacterium CAG:631 | GDSL-like protein
R2P1Z4_9ENTE | Enterococcus raffinosus ATCC 49464 | Uncharacterized protein
F2WJY6_9HYME | Messor pergandei | Oskar
F4WQN7_ACREC | Acromyrmex echinatior | Maternal effect protein oskar
A0A026WMY1_CERBI | Cerapachys biroi | Maternal effect protein oskar
E9IZ46_SOLIN | Solenopsis invicta | Putative uncharacterized protein
E2A7I8_CAMFO | Camponotus floridanus | Putative uncharacterized protein
A0A0J7KH44_LASNI | Lasius niger | Maternal effect protein oskar
E2BYH0_HARSA | Harpegnathos saltator | Putative uncharacterized protein
E1A883_NASVI | Nasonia vitripennis | Oskar
A0A0C9QHR7_9HYME | Fopius arisanus | Osk protein
K4MTL4_GRYBI | Gryllus bimaculatus | Oskar
E9I8K8_SOLIN | Solenopsis invicta | Putative uncharacterized protein
Q2PP79_AEDAE | Aedes aegypti | Oskar
B0WIV7_CULQU | Culex quinquefasciatus | Oskar
T1DTM7_ANOAQ | Anopheles aquasalis | Uncharacterized protein
W5JJ85_ANODA | Anopheles darlingi | Uncharacterized protein
A0A084WRU4_ANOSI | Anopheles sinensis | AGAP003545-PA-like protein
Q7PQJ3_ANOGA | Anopheles gambiae | AGAP003545-PA
U5EFJ8_9DIPT | Corethrella appendiculata | Putative oskar
W8CE30_CERCA | Ceratitis capitata | Maternal effect protein oskar
A0A0A1XRQ4_BACCU | Bactrocera cucurbitae | Maternal effect protein oskar
A0A034WRF5_BACDO | Bactrocera dorsalis | Maternal effect protein oskar
A0A0K8U7J3_BACLA | Bactrocera latifrons | Maternal effect protein oskar
A0A0L0CP24_LUCCU | Lucilia cuprina | Uncharacterized protein
T1PG45_MUSDO | Musca domestica | GDSL-like Lipase/Acylhydrolase
Q295Q4_DROPS | Drosophila pseudoobscura pseudoobscura | Uncharacterized protein, isoform A
A0A059PGF2_9MUSC | Drosophila pseudoobscura | GA10627
A0A059PF64_9MUSC | Drosophila pseudoobscura | GA10627
B4PTX6_DROYA | Drosophila yakuba | Uncharacterized protein
B4HKZ1_DROSE | Drosophila sechellia | GM23770
E8NH25_DROME | Drosophila melanogaster | RE24380p
A0A126GUR4_DROME | Drosophila melanogaster | Oskar, isoform D
B3P1W4_DROER | Drosophila erecta | GG13545
B3LZ06_DROAN | Drosophila ananassae | Uncharacterized protein
B4N815_DROWI | Drosophila willistoni | Uncharacterized protein
B4JTJ1_DROGR | Drosophila grimshawi | GH23955
B4K9E5_DROMO | Drosophila mojavensis | Uncharacterized protein
B4LXK5_DROVI | Drosophila virilis | Oskar
A1Y1T7_DROIM | Drosophila immigrans | Oskar
A0A0M4F3M8_DROBS | Drosophila busckii | Osk
A0A0J6BBM7_BREBE | Brevibacillus brevis | Lysophospholipase
A0A0H0SJ00_9BACL | Brevibacillus formosus | Lysophospholipase
J3A568_9BACL | Brevibacillus sp. BC25 | Lysophospholipase L1-like esterase
R5GT16_9FIRM | Eubacterium sp. CAG:786 | GDSL-like protein
R5LL12_9FIRM | Eubacterium sp. CAG:115 | GDSL-like protein
B7KKA6_CYAP7 | Cyanothece sp. (strain PCC 7424) | Lipolytic protein G-D-S-L family
F6DQC0_DESRL | Desulfotomaculum ruminis (strain ATCC 23193 / DSM 2154 / NCIB 8452 / DL) | Lipolytic protein G-D-S-L family
A0A078KJ49_9FIRM | [Clostridium] cellulosi | Uncharacterized protein
U6EVC2_CLOTA | Clostridium tetani 12124569 | Platelet activating factor acetylhydrolase-likeprotein
Q897X6_CLOTE | Clostridium tetani (strain Massachusetts / E88) | Platelet activating factor acetylhydrolase-like protein
R7ADB4_9BACE | Bacteroides pectinophilus CAG:437 | Uncharacterized protein
A0A095ZDI3_9FIRM | Tissierella bacterium S7-1-4 | Uncharacterized protein
A0A0L0WAR6_CLOPU | Clostridium purinilyticum | Lysophospholipase L1
A0A0C1UEU0_9CLOT | Clostridium argentinense CDC 2741 | GDSL-like Lipase/Acylhydrolase family protein
A0A072Y8N5_9CLOT | Clostridium sp. K25 | Acetylhydrolase
R4JC30_9BACT | uncultured bacterium BAC25G1 | Uncharacterized protein
K9WJ28_9CYAN | Microcoleus sp. PCC 7113 | Lysophospholipase L1-like esterase
A0A0M0WNM0_9BACI | Bacillus sp. FJAT-21351 | Lipase
D5DGY9_BACMD | Bacillus megaterium (strain DSM 319) | Lipase/Acylhydrolase (GDSL)
A5N8N5_CLOK5 | Clostridium kluyveri (strain ATCC 8527 / DSM 555 / NCIMB 10680) | Uncharacterized protein
W3AC50_9BACL | Planomicrobium glaciei CHR43 | Uncharacterized protein
A0A098EIL2_9BACL | Planomicrobium sp. ES2 | Multifunctional acyl-CoA thioesterase I and protease I and lysophospholipase L1
K8GFE2_9CYAN | Oscillatoriales cyanobacterium JSC-12 | Lysophospholipase L1-like esterase
G2HS43_9PROT | Arcobacter sp. L | Lipolytic protein
E6L4E3_9PROT | Arcobacter butzleri JV22 | Lipolytic protein
S5PEQ8_9PROT | Arcobacter butzleri 7h1h | Lipolytic enzyme, GDSL domain protein
A0A0G9K3L5_9PROT | Arcobacter butzleri L348 | Lipolytic protein
A0A0M1UPT0_9PROT | Arcobacter butzleri ED-1 | Lipolytic protein
A8EWS4_ARCB4 | Arcobacter butzleri (strain RM4018) | Lipolytic enzyme, GDSL domain
A0A0A6S1U2_STRUB | Streptococcus uberis | Acylneuraminate cytidylyltransferase
A0A0H1UPZ9_STRAG | Streptococcus agalactiae | Acylneuraminate cytidylyltransferase
E3RJZ5_PYRTT | Pyrenophora teres f. teres (strain 0-1) | Putative uncharacterized protein
A0A0L1HYX4_9PLEO | Stemphylium lycopersici | Carbohydrate esterase family 3 protein
G0S9F4_CHATD | Chaetomium thermophilum (strain DSM 1495 / CBS 144.50 / IMI 039719) | Putative uncharacterized protein
G2QVW9_THITE | Thielavia terrestris (strain ATCC 38088 / NRRL 8126) | Carbohydrate esterase family 3 protein
G2QGB0_MYCTT | Myceliophthora thermophila (strain ATCC 42464 / BCRC 31852 / DSM 1799) | Carbohydrate esterase family 3 protein
A0A094AE00_9PEZI | Pseudogymnoascus sp. VKM F-4281 (FW-2241) | Uncharacterized protein
A0A0E3QN36_METBA | Methanosarcina barkeri str. Wiesmoor | Putative tesA-like protease

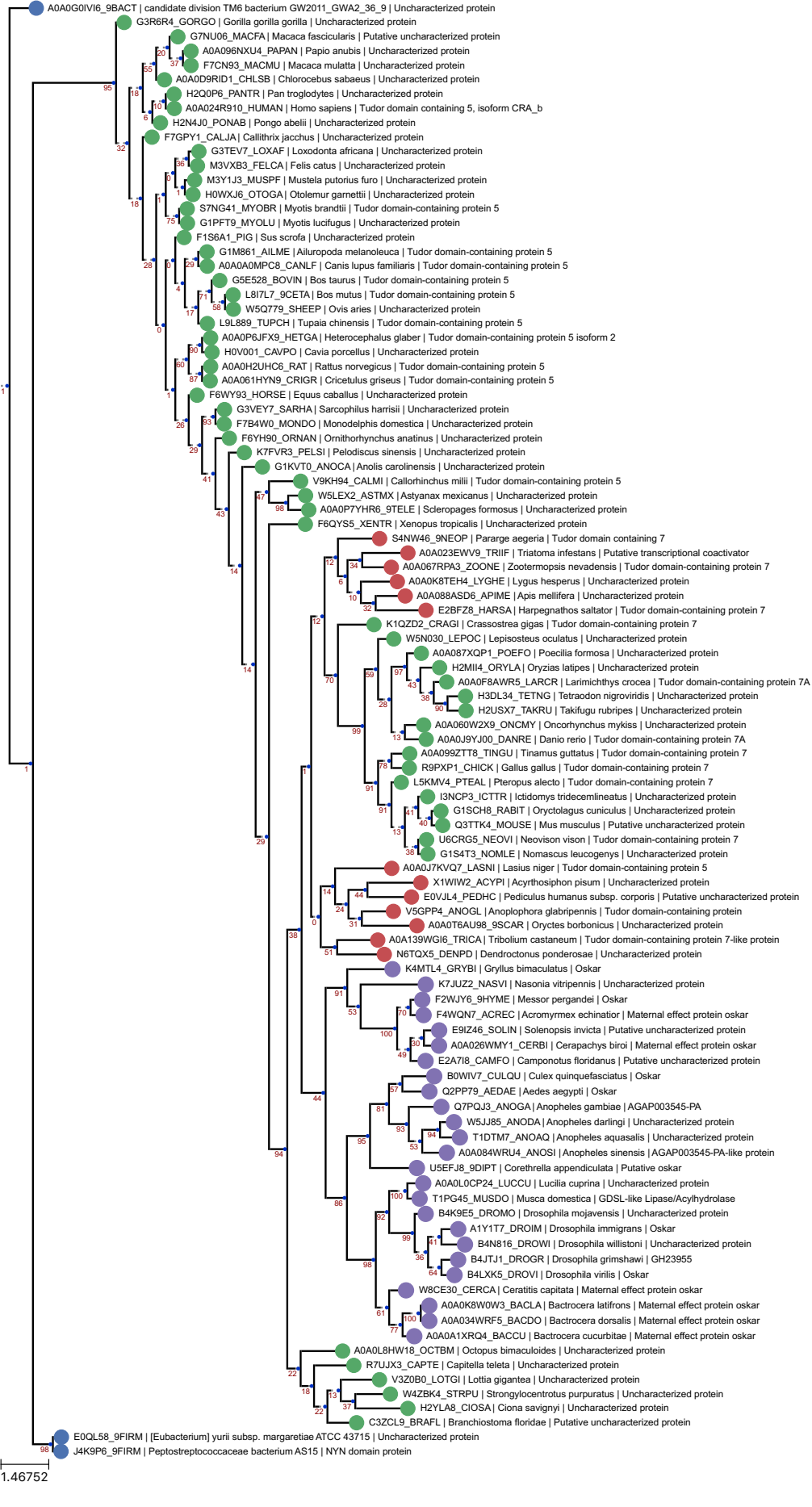

Putative HGT
Constraint
a
Unconstrained tree
Archaea
Bacterias
Bacterias
OSK
Bacterias
CAZ3
VS
Constrained by domain of life
Bacterias
Bacterias
OSK
CAZ3
Archaea
P-value = 0.002
b
Unconstrained tree
Archaea
Bacterias
Bacterias
OSK
Bacterias
CAZ3
VS
Monophyletic Eukaryota
Archaea
Bacterias
Bacterias
OSK
CAZ3
P-value = 0.009
c
Unconstrained tree
Bacterias
LOTUS
Tudor 5
Tudor 7
LOTUS
VS
Monophyletic LOTUS
Bacterias
Tudor 5
Tudor 7
LOTUS
LOTUS
P-value = 0.037

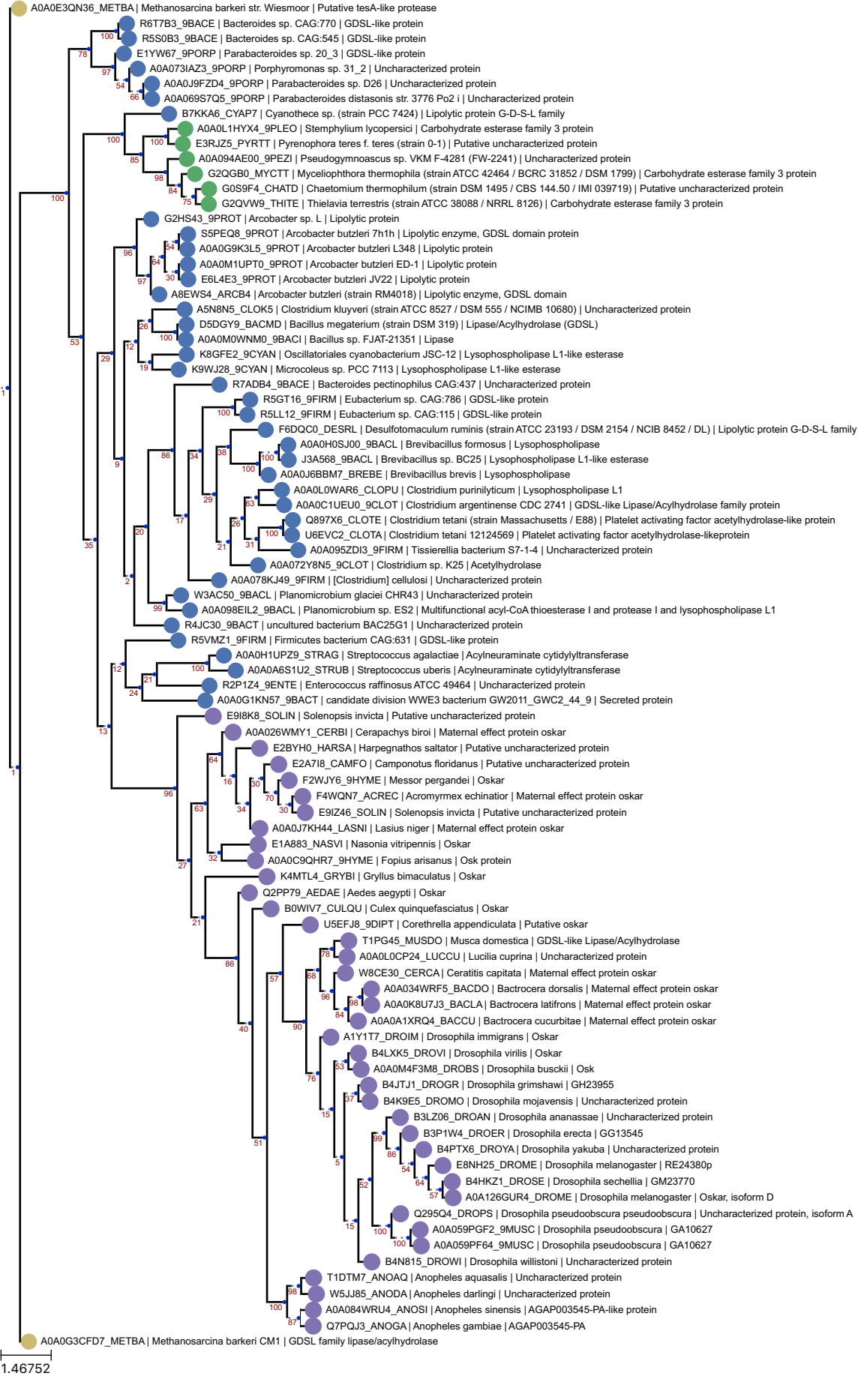

A0A0G0IVI6_9BACT | candidate division TM6 bacterium GW2011_GWA2_36_9 | Uncharacterized protein
V9KH94_CALMI | Callorhinchus milii | Tudor domain-containing protein 5
W5LEX2_ASTMX | Astyanax mexicanus | Uncharacterized protein
A0A0P7YHR6_9TELE | Scleropages formosus | Uncharacterized protein
K7FVR3_PELSI | Pelodiscus sinensis | Uncharacterized protein
F6YH90_ORNAN | Ornithorhynchus anatinus | Uncharacterized protein
F7B4W0_MONDO | Monodelphis domestica | Uncharacterized protein
G3VEY7_SARHA | Sarcophilus harrisii | Uncharacterized protein
A0A061HYN9_CRIGR | Cricetulus griseus | Tudor domain-containing protein 5
A0A0H2UHC6_RAT | Rattus norvegicus | Tudor domain-containing protein 5
M3VXB3_FELCA | Felis catus | Uncharacterized protein
S7NG41_MYOBR | Myotis brandtii | Tudor domain-containing protein 5
G1PFT9_MYOLU | Myotis lucifugus | Uncharacterized protein
G3TEV7_LOXAF | Loxodonta africana | Uncharacterized protein
L8I7L7_9CETA | Bos mutus | Tudor domain-containing protein 5
G5E528_BOVIN | Bos taurus | Tudor domain-containing protein 5
W5Q779_SHEEP | Ovis aries | Uncharacterized protein
H0V001_CAVPO | Cavia porcellus | Uncharacterized protein
A0A0P6JFX9_HETGA | Heterocephalus glaber | Tudor domain-containing protein 5 isoform 2
F6WY93_HORSE | Equus caballus | Uncharacterized protein
M3Y1J3_MUSPF | Mustela putorius furo | Uncharacterized protein
H0WXJ6_OTOGA | Otolemur garnettii | Uncharacterized protein
G7NU06_MACFA | Macaca fascicularis | Putative uncharacterized protein
A0A0D9RID1_CHLSB | Chlorocebus sabaeus | Uncharacterized protein
A0A096NXU4_PAPAN | Papio anubis | Uncharacterized protein
F7CN93_MACMU | Macaca mulatta | Uncharacterized protein
G3R6R4_GORGO | Gorilla gorilla gorilla | Uncharacterized protein
H2N4J0_PONAB | Pongo abelii | Uncharacterized protein
H2Q0P6_PANTR | Pan troglodytes | Uncharacterized protein
A0A024R910_HUMAN | Homo sapiens | Tudor domain containing 5, isoform CRA_b
F7GPY1_CALJA | Callithrix jacchus | Uncharacterized protein
A0A0A0MPC8_CANLF | Canis lupus familiaris | Tudor domain-containing protein 5
G1M861_AILME | Ailuropoda melanoleuca | Tudor domain-containing protein 5
F1S6A1_PIG | Sus scrofa | Uncharacterized protein
L9L889_TUPCH | Tupaia chinensis | Tudor domain-containing protein 5
G1KVT0_ANOCA | Anolis carolinensis | Uncharacterized protein
F6QYS5_XENTR | Xenopus tropicalis | Uncharacterized protein
C3ZCL9_BRAFL | Branchiostoma floridae | Putative uncharacterized protein
W4ZBK4_STRPU | Strongylocentrotus purpuratus | Uncharacterized protein
V3Z0B0_LOTGI | Lottia gigantea | Uncharacterized protein
A0A0L8HW18_OCTBM | Octopus bimaculoides | Uncharacterized protein
R7UJX3_CAPTE | Capitella teleta | Uncharacterized protein
V5GPP4_ANOGL | Anoplophora glabripennis | Tudor domain-containing protein
X1WIW2_ACYPI | Acyrthosiphon pisum | Uncharacterized protein
E0VJL4_PEDHC | Pediculus humanus subsp. corporis | Putative uncharacterized protein
A0A0T6AU98_9SCAR | Oryctes borbonicus | Uncharacterized protein
H2YLA8_CIOSA | Ciona savignyi | Uncharacterized protein
U5EFJ8_9DIPT | Corethrella appendiculata | Putative oskar
A0A084WRU4_ANOSI | Anopheles sinensis | AGAP003545-PA-like protein
Q7PQJ3_ANOGA | Anopheles gambiae | AGAP003545-PA
T1DTM7_ANOAQ | Anopheles aquasalis | Uncharacterized protein
W5JJ85_ANODA | Anopheles darlingi | Uncharacterized protein
Q2PP79_AEDAE | Aedes aegypti | Oskar
B0WIV7_CULQU | Culex quinquefasciatus | Oskar
T1PG45_MUSDO | Musca domestica | GDSL-like Lipase/Acylhydrolase
A0A0L0CP24_LUCCU | Lucilia cuprina | Uncharacterized protein
W8CE30_CERCA | Ceratitis capitata | Maternal effect protein oskar
A0A0A1XRQ4_BACCU | Bactrocera cucurbitae | Maternal effect protein oskar
A0A0K8W0W3_BACLA | Bactrocera latifrons | Maternal effect protein oskar
A0A034WRF5_BACDO | Bactrocera dorsalis | Maternal effect protein oskar
B4LXK5_DROVI | Drosophila virilis | Oskar
A1Y1T7_DROIM | Drosophila immigrans | Oskar
B4N816_DROWI | Drosophila willistoni | Uncharacterized protein
B4K9E5_DROMO | Drosophila mojavensis | Uncharacterized protein
B4JTJ1_DROGR | Drosophila grimshawi | GH23955
A0A0J7KVQ7_LASNI | Lasius niger | Tudor domain-containing protein 5
K1QZD2_CRAGI | Crassostrea gigas | Tudor domain-containing protein 7
A0A087XQP1_POEFO | Poecilia formosa | Uncharacterized protein
H2MII4_ORYLA | Oryzias latipes | Uncharacterized protein
A0A0F8AWR5_LARCR | Larimichthys crocea | Tudor domain-containing protein 7A
H2USX7_TAKRU | Takifugu rubripes | Uncharacterized protein
H3DL34_TETNG | Tetraodon nigroviridis | Uncharacterized protein
L5KMV4_PTEAL | Pteropus alecto | Tudor domain-containing protein 7
Q3TTK4_MOUSE | Mus musculus | Putative uncharacterized protein
G1S4T3_NOMLE | Nomascus leucogenys | Uncharacterized protein
U6CRG5_NEOVI | Neovison vison | Tudor domain-containing protein 7
G1SCH8_RABIT | Oryctolagus cuniculus | Uncharacterized protein
I3NCP3_ICTTR | Ictidomys tridecemlineatus | Uncharacterized protein
R9PXP1_CHICK | Gallus gallus | Tudor domain-containing protein 7
A0A099ZTT8_TINGU | Tinamus guttatus | Tudor domain-containing protein 7
W5N030_LEPOC | Lepisosteus oculatus | Uncharacterized protein
A0A0J9YJ00_DANRE | Danio rerio | Tudor domain-containing protein 7A
A0A060W2X9_ONCMY | Oncorhynchus mykiss | Uncharacterized protein
S4NW46_9NEOP | Pararge aegeria | Tudor domain containing 7
E2BFZ8_HARSA | Harpegnathos saltator | Tudor domain-containing protein 7
A0A088ASD6_APIME | Apis mellifera | Uncharacterized protein
A0A067RPA3_ZOONE | Zootermopsis nevadensis | Tudor domain-containing protein 7
A0A023EWV9_TRIIF | Triatoma infestans | Putative transcriptional coactivator
A0A0K8TEH4_LYGHE | Lygus hesperus | Uncharacterized protein
N6TQX5_DENPD | Dendroctonus ponderosae | Uncharacterized protein
A0A139WGI6_TRICA | Tribolium castaneum | Tudor domain-containing protein 7-like protein
K7JUZ2_NASVI | Nasonia vitripennis | Uncharacterized protein
E9IZ46_SOLIN | Solenopsis invicta | Putative uncharacterized protein
A0A026WMY1_CERBI | Cerapachys biroi | Maternal effect protein oskar
F4WQN7_ACREC | Acromyrmex echinatior | Maternal effect protein oskar
F2WJY6_9HYME | Messor pergandei | Oskar
E2A7I8_CAMFO | Camponotus floridanus | Putative uncharacterized protein
K4MTL4_GRYBI | Gryllus bimaculatus | Oskar
J4K9P6_9FIRM | Peptostreptococcaceae bacterium AS15 | NYN domain protein
E0QL58_9FIRM | [Eubacterium] yurii subsp. margaretiae ATCC 43715 | Uncharacterized protein

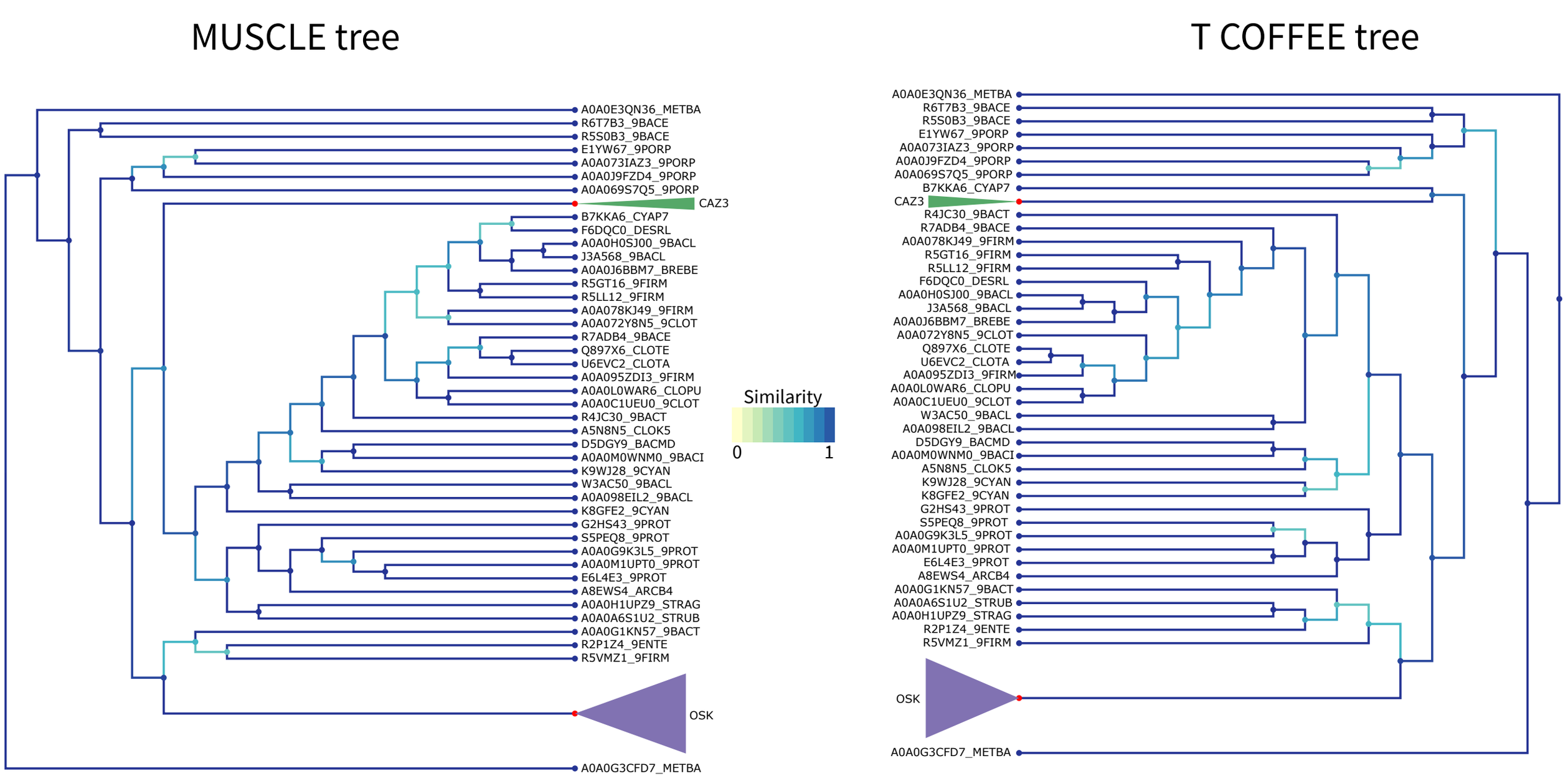

A0A0G3CFD7_METBA | Methanosarcina barkeri CM1 | GDSL family lipase/acylhydrolase
R2P1Z4_9ENTE | Enterococcus raffinosus ATCC 49464 | Uncharacterized protein
A0A0G1KN57_9BACT | candidate division WWE3 bacterium GW2011_GWC2_44_9 | Secreted protein
B4K9E5_DROMO | Drosophila mojavensis | Uncharacterized protein
B4JTJ1_DROGR | Drosophila grimshawi | GH23955
A0A0M4F3M8_DROBS | Drosophila busckii | Osk
B4LXK5_DROVI | Drosophila virilis | Oskar
Q295Q4_DROPS | Drosophila pseudoobscura pseudoobscura | Uncharacterized protein, isoform A
A0A059PGF2_9MUSC | Drosophila pseudoobscura | GA10627
A0A059PF64_9MUSC | Drosophila pseudoobscura | GA10627
B3LZ06_DROAN | Drosophila ananassae | Uncharacterized protein
A0A126GUR4_DROME | Drosophila melanogaster | Oskar, isoform D
E8NH25_DROME | Drosophila melanogaster | RE24380p
B4HKZ1_DROSE | Drosophila sechellia | GM23770
B4PTX6_DROYA | Drosophila yakuba | Uncharacterized protein
B3P1W4_DROER | Drosophila erecta | GG13545
B4N815_DROWI | Drosophila willistoni | Uncharacterized protein
A1Y1T7_DROIM | Drosophila immigrans | Oskar
A0A0L0CP24_LUCCU | Lucilia cuprina | Uncharacterized protein
T1PG45_MUSDO | Musca domestica | GDSL-like Lipase/Acylhydrolase
W8CE30_CERCA | Ceratitis capitata | Maternal effect protein oskar
A0A0K8U7J3_BACLA | Bactrocera latifrons | Maternal effect protein oskar
A0A034WRF5_BACDO | Bactrocera dorsalis | Maternal effect protein oskar
A0A0A1XRQ4_BACCU | Bactrocera cucurbitae | Maternal effect protein oskar
U5EFJ8_9DIPT | Corethrella appendiculata | Putative oskar
Q2PP79_AEDAE | Aedes aegypti | Oskar
B0WIV7_CULQU | Culex quinquefasciatus | Oskar
Q7PQJ3_ANOGA | Anopheles gambiae | AGAP003545-PA
A0A084WRU4_ANOSI | Anopheles sinensis | AGAP003545-PA-like protein
T1DTM7_ANOAQ | Anopheles aquasalis | Uncharacterized protein
W5JJ85_ANODA | Anopheles darlingi | Uncharacterized protein
K4MTL4_GRYBI | Gryllus bimaculatus | Oskar
E2BYH0_HARSA | Harpegnathos saltator | Putative uncharacterized protein
A0A026WMY1_CERBI | Cerapachys biroi | Maternal effect protein oskar
E9IZ46_SOLIN | Solenopsis invicta | Putative uncharacterized protein
F4WQN7_ACREC | Acromyrmex echinatior | Maternal effect protein oskar
F2WJY6_9HYME | Messor pergandei | Oskar
A0A0J7KH44_LASNI | Lasius niger | Maternal effect protein oskar
E2A7I8_CAMFO | Camponotus floridanus | Putative uncharacterized protein
A0A0C9QHR7_9HYME | Fopius arisanus | Osk protein
E1A883_NASVI | Nasonia vitripennis | Oskar
E9I8K8_SOLIN | Solenopsis invicta | Putative uncharacterized protein
R5VMZ1_9FIRM | Firmicutes bacterium CAG:631 | GDSL-like protein
A0A0A6S1U2_STRUB | Streptococcus uberis | Acylneuraminate cytidylyltransferase
A0A0H1UPZ9_STRAG | Streptococcus agalactiae | Acylneuraminate cytidylyltransferase
A0A0G9K3L5_9PROT | Arcobacter butzleri L348 | Lipolytic protein
E6L4E3_9PROT | Arcobacter butzleri JV22 | Lipolytic protein
A0A0M1UPT0_9PROT | Arcobacter butzleri ED-1 | Lipolytic protein
S5PEQ8_9PROT | Arcobacter butzleri 7h1h | Lipolytic enzyme, GDSL domain protein
A8EWS4_ARCB4 | Arcobacter butzleri (strain RM4018) | Lipolytic enzyme, GDSL domain
G2HS43_9PROT | Arcobacter sp. L | Lipolytic protein
A5N8N5_CLOK5 | Clostridium kluyveri (strain ATCC 8527 / DSM 555 / NCIMB 10680) | Uncharacterized protein
K9WJ28_9CYAN | Microcoleus sp. PCC 7113 | Lysophospholipase L1-like esterase
D5DGY9_BACMD | Bacillus megaterium (strain DSM 319) | Lipase/Acylhydrolase (GDSL)
A0A0M0WNM0_9BACI | Bacillus sp. FJAT-21351 | Lipase
A0A0L1HYX4_9PLEO | Stemphylium lycopersici | Carbohydrate esterase family 3 protein
E3RJZ5_PYRTT | Pyrenophora teres f. teres (strain 0-1) | Putative uncharacterized protein
G0S9F4_CHATD | Chaetomium thermophilum (strain DSM 1495 / CBS 144.50 / IMI 039719) | Putative uncharacterized protein
G2QVW9_THITE | Thielavia terrestris (strain ATCC 38088 / NRRL 8126) | Carbohydrate esterase family 3 protein
G2QGB0_MYCTT | Myceliophthora thermophila (strain ATCC 42464 / BCRC 31852 / DSM 1799) | Carbohydrate esterase family 3 protein
A0A094AE00_9PEZI | Pseudogymnoascus sp. VKM F-4281 (FW-2241) | Uncharacterized protein
R5S0B3_9BACE | Bacteroides sp. CAG:545 | GDSL-like protein
R6T7B3_9BACE | Bacteroides sp. CAG:770 | GDSL-like protein
E1YW67_9PORP | Parabacteroides sp. 20_3 | GDSL-like protein
A0A069S7Q5_9PORP | Parabacteroides distasonis str. 3776 Po2 i | Uncharacterized protein
A0A0J9FZD4_9PORP | Parabacteroides sp. D26 | Uncharacterized protein
A0A073IAZ3_9PORP | Porphyromonas sp. 31_2 | Uncharacterized protein
W3AC50_9BACL | Planomicrobium glaciei CHR43 | Uncharacterized protein
A0A098EIL2_9BACL | Planomicrobium sp. ES2 | Multifunctional acyl-CoA thioesterase I and protease I and lysophospholipase L1
K8GFE2_9CYAN | Oscillatoriales cyanobacterium JSC-12 | Lysophospholipase L1-like esterase
R4JC30_9BACT | uncultured bacterium BAC25G1 | Uncharacterized protein
R7ADB4_9BACE | Bacteroides pectinophilus CAG:437 | Uncharacterized protein
A0A0C1UEU0_9CLOT | Clostridium argentinense CDC 2741 | GDSL-like Lipase/Acylhydrolase family protein
Q897X6_CLOTE | Clostridium tetani (strain Massachusetts / E88) | Platelet activating factor acetylhydrolase-like protein
U6EVC2_CLOTA | Clostridium tetani 12124569 | Platelet activating factor acetylhydrolase-likeprotein
A0A0L0WAR6_CLOPU | Clostridium purinilyticum | Lysophospholipase L1
A0A095ZDI3_9FIRM | Tissierellia bacterium S7-1-4 | Uncharacterized protein
A0A078KJ49_9FIRM | [Clostridium] cellulosi | Uncharacterized protein
A0A072Y8N5_9CLOT | Clostridium sp. K25 | Acetylhydrolase
B7KKA6_CYAP7 | Cyanothece sp. (strain PCC 7424) | Lipolytic protein G-D-S-L family
F6DQC0_DESRL | Desulfotomaculum ruminis (strain ATCC 23193 / DSM 2154 / NCIB 8452 / DL) | Lipolytic protein G-D-S-L family
R5GT16_9FIRM | Eubacterium sp. CAG:786 | GDSL-like protein
R5LL12_9FIRM | Eubacterium sp. CAG:115 | GDSL-like protein
A0A0J6BBM7_BREBE | Brevibacillus brevis | Lysophospholipase
A0A0H0SJ00_9BACL | Brevibacillus formosus | Lysophospholipase
J3A568_9BACL | Brevibacillus sp. BC25 | Lysophospholipase L1-like esterase
A0A0E3QN36_METBA | Methanosarcina barkeri str. Wiesmoor | Putative tesA-like protease


MUSCLE tree
A0A0E3QN36_METBA
A0A0G3CFD7_METBA
R5S0B3_9BACE
R6T7B3_9BACE
E1YW67_9PORP
A0A073IAZ3_9PORP
A0A0J9FZD4_9PORP
A0A069S7Q5_9PORP
CAZ3
K8GFE2_9CYAN
W3AC50_9BACL
A0A098EIL2_9BACL
K9WJ28_9CYAN
D5DGY9_BACMD
A0A0M0WNM0_9BACI
A5N8N5_CLOK5
R4JC30_9BACT
R7ADB4_9BACE
Q897X6_CLOTE
U6EVC2_CLOTA
A0A095ZDI3_9FIRM
A0A0C1UEU0_9CLOT
A0A0L0WAR6_CLOPU
R5GT16_9FIRM
R5LL12_9FIRM
B7KKA6_CYAP7
F6DQC0_DESRL
A0A0J6BBM7_BREBE
A0A0H0SJ00_9BACL
J3A568_9BACL
A0A078KJ49_9FIRM
A0A072Y8N5_9CLOT
A0A0G9K3L5_9PROT
E6L4E3_9PROT
A0A0M1UPT0_9PROT
S5PEQ8_9PROT
A8EWS4_ARCB4
G2HS43_9PROT
A0A0A6S1U2_STRUB
A0A0H1UPZ9_STRAG
OSK
A0A0G1KN57_9BACT
R5VMZ1_9FIRM
R2P1Z4_9ENTE
Similarity
0
1
PRANK tree
A0A0E3QN36_METBA
A0A0G3CFD7_METBA
R5S0B3_9BACE
R6T7B3_9BACE
E1YW67_9PORP
A0A073IAZ3_9PORP
A0A0J9FZD4_9PORP
A0A069S7Q5_9PORP
CAZ3
W3AC50_9BACL
A0A098EIL2_9BACL
K8GFE2_9CYAN
K9WJ28_9CYAN
D5DGY9_BACMD
A0A0M0WNM0_9BACI
A5N8N5_CLOK5
R4JC30_9BACT
R7ADB4_9BACE
A0A095ZDI3_9FIRM
A0A0C1UEU0_9CLOT
Q897X6_CLOTE
U6EVC2_CLOTA
A0A0L0WAR6_CLOPU
B7KKA6_CYAP7
F6DQC0_DESRL
R5GT16_9FIRM
R5LL12_9FIRM
A0A0J6BBM7_BREBE
A0A0H0SJ00_9BACL
J3A568_9BACL
A0A078KJ49_9FIRM
A0A072Y8N5_9CLOT
A0A0G9K3L5_9PROT
E6L4E3_9PROT
A0A0M1UPT0_9PROT
S5PEQ8_9PROT
A8EWS4_ARCB4
G2HS43_9PROT
A0A0A6S1U2_STRUB
A0A0H1UPZ9_STRAG
OSK
R5VMZ1_9FIRM
A0A0G1KN57_9BACT
R2P1Z4_9ENTE

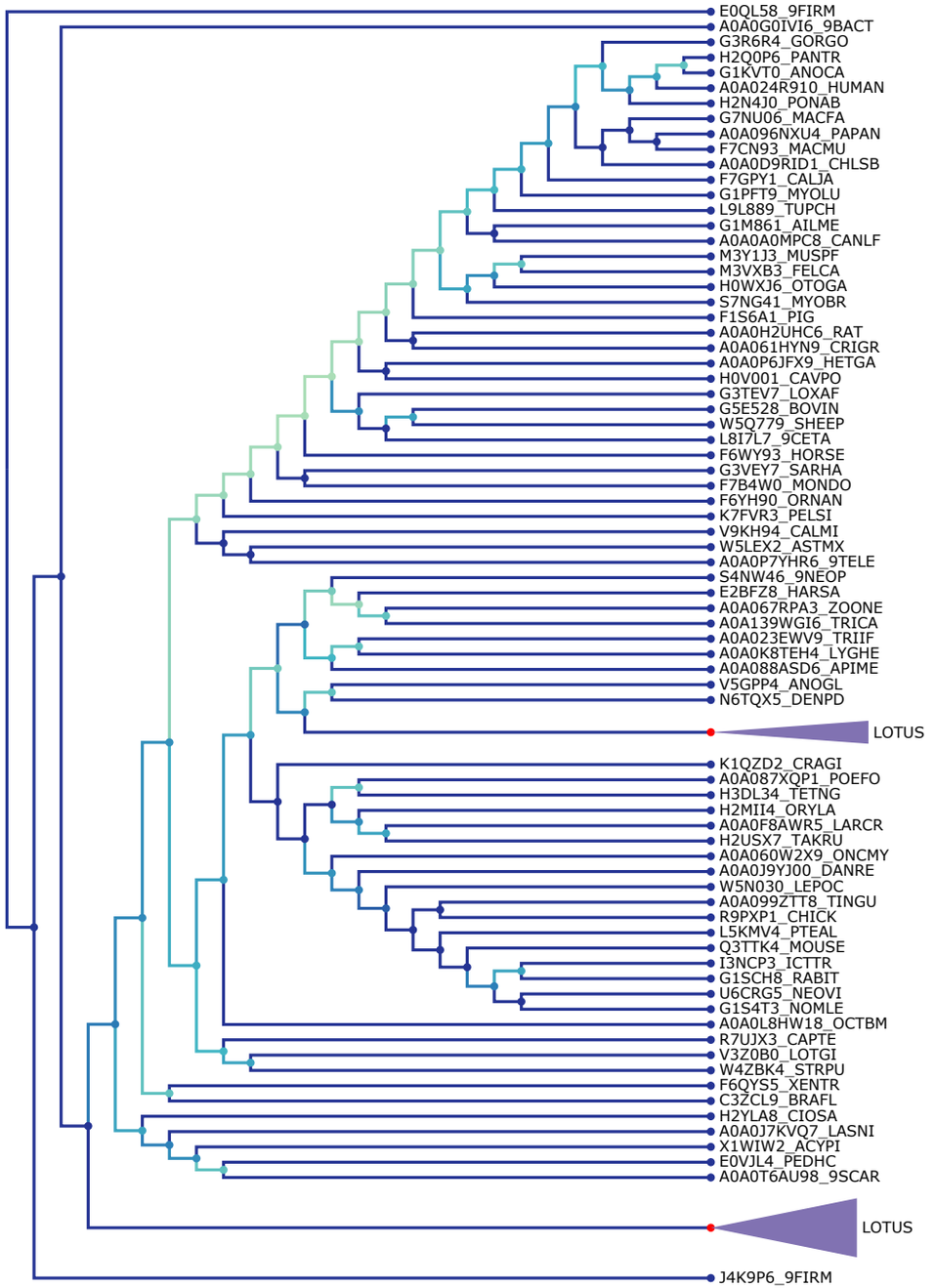

MUSCLE tree
PRANK tree
Similarity
0
1
A0A0G0IVI6_9BACT
J4K9P6_9FIRM
E0QL58_9FIRM
V9KH94_CALMI
W5LEX2_ASTMX
A0A0P7YHR6_9TELE
K7FVR3_PELSI
F6YH90_ORNAN
F7B4W0_MONDO
G3VEY7_SARHA
F6WY93_HORSE
G3TEV7_LOXAF
L8I7L7_9CETA
G5E528_BOVIN
W5Q779_SHEEP
H0V001_CAVPO
A0A0P6JFX9_HETGA
A0A061HYN9_CRIGR
A0A0H2UHC6_RAT
F1S6A1_PIG
S7NG41_MYOBR
H0WXJ6_OTOGA
M3VXB3_FELCA
M3Y1J3_MUSPF
A0A0A0MPC8_CANLF
G1M861_AILME
L9L889_TUPCH
G1PFT9_MYOLU
F7GPY1_CALJA
G7NU06_MACFA
A0A096NXU4_PAPAN
F7CN93_MACMU
A0A0D9RID1_CHLSB
G3R6R4_GORGO
H2N4J0_PONAB
A0A024R910_HUMAN
H2Q0P6_PANTR
G1KVT0_ANOCA
W4ZBK4_STRPU
V3Z0B0_LOTGI
R7UJX3_CAPTE
A0A0L8HW18_OCTBM
S4NW46_9NEOP
E2BFZ8_HARSA
A0A139WGI6_TRICA
A0A067RPA3_ZOONE
A0A088ASD6_APIME
A0A023EWV9_TRIIF
A0A0K8TEH4_LYGHE
K4MTL4_GRYBI ... F2WJY6_9HYME
N6TQX5_DENPD
V5GPP4_ANOGL
K1QZD2_CRAGI
H2MII4_ORYLA
A0A0F8AWR5_LARCR
H2USX7_TAKRU
A0A087XQP1_POEFO
H3DL34_TETNG
L5KMV4_PTEAL
Q3TTK4_MOUSE
G1S4T3_NOMLE
U6CRG5_NEOVI
G1SCH8_RABIT
I3NCP3_ICTTR
R9PXP1_CHICK
A0A099ZTT8_TINGU
W5N030_LEPOC
A0A0J9YJ00_DANRE
A0A060W2X9_ONCMY
C3ZCL9_BRAFL
F6QYS5_XENTR
X1WIW2_ACYPI
E0VJL4_PEDHC
A0A0T6AU98_9SCAR
A0A0J7KVQ7_LASNI
H2YLA8_CIOSA
U5EFJ8_9DIPT ... T1PG45_MUSDO
K7JUZ2_NASVI ... K4MTL4_GRYBI
U5EFJ8_9DIPT ... B4JTJ1_DROGR


A0A0G0IVI6_9BACT | candidate division TM6 bacterium GW2011_GWA2_36_9 | Uncharacterized protein
G3R6R4_GORGO | Gorilla gorilla gorilla | Uncharacterized protein
G7NU06_MACFA | Macaca fascicularis | Putative uncharacterized protein
A0A096NXU4_PAPAN | Papio anubis | Uncharacterized protein
F7CN93_MACMU | Macaca mulatta | Uncharacterized protein
A0A0D9RID1_CHLSB | Chlorocebus sabaeus | Uncharacterized protein
H2Q0P6_PANTR | Pan troglodytes | Uncharacterized protein
A0A024R910_HUMAN | Homo sapiens | Tudor domain containing 5, isoform CRA_b
H2N4J0_PONAB | Pongo abelii | Uncharacterized protein
F7GPY1_CALJA | Callithrix jacchus | Uncharacterized protein
G3TEV7_LOXAF | Loxodonta africana | Uncharacterized protein
M3VXB3_FELCA | Felis catus | Uncharacterized protein
M3Y1J3_MUSPF | Mustela putorius furo | Uncharacterized protein
H0WXJ6_OTOGA | Otolemur garnettii | Uncharacterized protein
S7NG41_MYOBR | Myotis brandtii | Tudor domain-containing protein 5
G1PFT9_MYOLU | Myotis lucifugus | Uncharacterized protein
F1S6A1_PIG | Sus scrofa | Uncharacterized protein
G1M861_AILME | Ailuropoda melanoleuca | Tudor domain-containing protein 5
A0A0A0MPC8_CANLF | Canis lupus familiaris | Tudor domain-containing protein 5
G5E528_BOVIN | Bos taurus | Tudor domain-containing protein 5
L8I7L7_9CETA | Bos mutus | Tudor domain-containing protein 5
W5Q779_SHEEP | Ovis aries | Uncharacterized protein
L9L889_TUPCH | Tupaia chinensis | Tudor domain-containing protein 5
A0A0P6JFX9_HETGA | Heterocephalus glaber | Tudor domain-containing protein 5 isoform 2
H0V001_CAVPO | Cavia porcellus | Uncharacterized protein
A0A0H2UHC6_RAT | Rattus norvegicus | Tudor domain-containing protein 5
A0A061HYN9_CRIGR | Cricetulus griseus | Tudor domain-containing protein 5
F6WY93_HORSE | Equus caballus | Uncharacterized protein
G3VEY7_SARHA | Sarcophilus harrisii | Uncharacterized protein
F7B4W0_MONDO | Monodelphis domestica | Uncharacterized protein
F6YH90_ORNAN | Ornithorhynchus anatinus | Uncharacterized protein
K7FVR3_PELSI | Pelodiscus sinensis | Uncharacterized protein
G1KVT0_ANOCA | Anolis carolinensis | Uncharacterized protein
V9KH94_CALMI | Callorhinchus milii | Tudor domain-containing protein 5
W5LEX2_ASTMX | Astyanax mexicanus | Uncharacterized protein
A0A0P7YHR6_9TELE | Scleropages formosus | Uncharacterized protein
F6QYS5_XENTR | Xenopus tropicalis | Uncharacterized protein
S4NW46_9NEOP | Pararge aegeria | Tudor domain containing 7
A0A023EWV9_TRIIF | Triatoma infestans | Putative transcriptional coactivator
A0A067RPA3_ZOONE | Zootermopsis nevadensis | Tudor domain-containing protein 7
A0A0K8TEH4_LYGHE | Lygus hesperus | Uncharacterized protein
A0A088ASD6_APIME | Apis mellifera | Uncharacterized protein
E2BFZ8_HARSA | Harpegnathos saltator | Tudor domain-containing protein 7
K1QZD2_CRAGI | Crassostrea gigas | Tudor domain-containing protein 7
W5N030_LEPOC | Lepisosteus oculatus | Uncharacterized protein
A0A087XQP1_POEFO | Poecilia formosa | Uncharacterized protein
H2MII4_ORYLA | Oryzias latipes | Uncharacterized protein
A0A0F8AWR5_LARCR | Larimichthys crocea | Tudor domain-containing protein 7A
H3DL34_TETNG | Tetraodon nigroviridis | Uncharacterized protein
H2USX7_TAKRU | Takifugu rubripes | Uncharacterized protein
A0A060W2X9_ONCMY | Oncorhynchus mykiss | Uncharacterized protein
A0A0J9YJ00_DANRE | Danio rerio | Tudor domain-containing protein 7A
A0A099ZTT8_TINGU | Tinamus guttatus | Tudor domain-containing protein 7
R9PXP1_CHICK | Gallus gallus | Tudor domain-containing protein 7
L5KMV4_PTEAL | Pteropus alecto | Tudor domain-containing protein 7
I3NCP3_ICTTR | Ictidomys tridecemlineatus | Uncharacterized protein
G1SCH8_RABIT | Oryctolagus cuniculus | Uncharacterized protein
Q3TTK4_MOUSE | Mus musculus | Putative uncharacterized protein
U6CRG5_NEOVI | Neovison vison | Tudor domain-containing protein 7
G1S4T3_NOMLE | Nomascus leucogenys | Uncharacterized protein
A0A0J7KVQ7_LASNI | Lasius niger | Tudor domain-containing protein 5
X1WIW2_ACYPI | Acyrthosiphon pisum | Uncharacterized protein
E0VJL4_PEDHC | Pediculus humanus subsp. corporis | Putative uncharacterized protein
V5GPP4_ANOGL | Anoplophora glabripennis | Tudor domain-containing protein
A0A0T6AU98_9SCAR | Oryctes borbonicus | Uncharacterized protein
A0A139WGI6_TRICA | Tribolium castaneum | Tudor domain-containing protein 7-like protein
N6TQX5_DENPD | Dendroctonus ponderosae | Uncharacterized protein
K4MTL4_GRYBI | Gryllus bimaculatus | Oskar
K7JUZ2_NASVI | Nasonia vitripennis | Uncharacterized protein
F2WJY6_9HYME | Messor pergandei | Oskar
F4WQN7_ACREC | Acromyrmex echinatior | Maternal effect protein oskar
E9IZ46_SOLIN | Solenopsis invicta | Putative uncharacterized protein
A0A026WMY1_CERBI | Cerapachys biroi | Maternal effect protein oskar
E2A7I8_CAMFO | Camponotus floridanus | Putative uncharacterized protein
B0WIV7_CULQU | Culex quinquefasciatus | Oskar
Q2PP79_AEDAE | Aedes aegypti | Oskar
Q7PQJ3_ANOGA | Anopheles gambiae | AGAP003545-PA
W5JJ85_ANODA | Anopheles darlingi | Uncharacterized protein
T1DTM7_ANOAQ | Anopheles aquasalis | Uncharacterized protein
A0A084WRU4_ANOSI | Anopheles sinensis | AGAP003545-PA-like protein
U5EFJ8_9DIPT | Corethrella appendiculata | Putative oskar
A0A0L0CP24_LUCCU | Lucilia cuprina | Uncharacterized protein
T1PG45_MUSDO | Musca domestica | GDSL-like Lipase/Acylhydrolase
B4K9E5_DROMO | Drosophila mojavensis | Uncharacterized protein
A1Y1T7_DROIM | Drosophila immigrans | Oskar
B4N816_DROWI | Drosophila willistoni | Uncharacterized protein
B4JTJ1_DROGR | Drosophila grimshawi | GH23955
B4LXK5_DROVI | Drosophila virilis | Oskar
W8CE30_CERCA | Ceratitis capitata | Maternal effect protein oskar
A0A0K8W0W3_BACLA | Bactrocera latifrons | Maternal effect protein oskar
A0A034WRF5_BACDO | Bactrocera dorsalis | Maternal effect protein oskar
A0A0A1XRQ4_BACCU | Bactrocera cucurbitae | Maternal effect protein oskar
A0A0L8HW18_OCTBM | Octopus bimaculoides | Uncharacterized protein
R7UJX3_CAPTE | Capitella teleta | Uncharacterized protein
V3Z0B0_LOTGI | Lottia gigantea | Uncharacterized protein
W4ZBK4_STRPU | Strongylocentrotus purpuratus | Uncharacterized protein
H2YLA8_CIOSA | Ciona savignyi | Uncharacterized protein
C3ZCL9_BRAFL | Branchiostoma floridae | Putative uncharacterized protein
E0QL58_9FIRM | [Eubacterium] yurii subsp. margaretiae ATCC 43715 | Uncharacterized protein
J4K9P6_9FIRM | Peptostreptococcaceae bacterium AS15 | NYN domain protein
1.46752


A0A0E3QN36_METBA | Methanosarcina barkeri str. Wiesmoor | Putative tesA-like protease
R6T7B3_9BACE | Bacteroides sp. CAG:770 | GDSL-like protein
R5S0B3_9BACE | Bacteroides sp. CAG:545 | GDSL-like protein
E1YW67_9PORP | Parabacteroides sp. 20_3 | GDSL-like protein
A0A073IAZ3_9PORP | Porphyromonas sp. 31_2 | Uncharacterized protein
A0A0J9FZD4_9PORP | Parabacteroides sp. D26 | Uncharacterized protein
A0A069S7Q5_9PORP | Parabacteroides distasonis str. 3776 Po2 i | Uncharacterized protein
B7KKA6_CYAP7 | Cyanothece sp. (strain PCC 7424) | Lipolytic protein G-D-S-L family
A0A0L1HYX4_9PLEO | Stemphylium lycopersici | Carbohydrate esterase family 3 protein
E3RJZ5_PYRTT | Pyrenophora teres f. teres (strain 0-1) | Putative uncharacterized protein
A0A094AE00_9PEZI | Pseudogymnoascus sp. VKM F-4281 (FW-2241) | Uncharacterized protein
G2QGB0_MYCTT | Myceliophthora thermophila (strain ATCC 42464 / BCRC 31852 / DSM 1799) | Carbohydrate esterase family 3 protein
G0S9F4_CHATD | Chaetomium thermophilum (strain DSM 1495 / CBS 144.50 / IMI 039719) | Putative uncharacterized protein
G2QVW9_THITE | Thielavia terrestris (strain ATCC 38088 / NRRL 8126) | Carbohydrate esterase family 3 protein
G2HS43_9PROT | Arcobacter sp. L | Lipolytic protein
S5PEQ8_9PROT | Arcobacter butzleri 7h1h | Lipolytic enzyme, GDSL domain protein
A0A0G9K3L5_9PROT | Arcobacter butzleri L348 | Lipolytic protein
A0A0M1UPT0_9PROT | Arcobacter butzleri ED-1 | Lipolytic protein
E6L4E3_9PROT | Arcobacter butzleri JV22 | Lipolytic protein
A8EWS4_ARCB4 | Arcobacter butzleri (strain RM4018) | Lipolytic enzyme, GDSL domain
A5N8N5_CLOK5 | Clostridium kluyveri (strain ATCC 8527 / DSM 555 / NCIMB 10680) | Uncharacterized protein
D5DGY9_BACMD | Bacillus megaterium (strain DSM 319) | Lipase/Acylhydrolase (GDSL)
A0A0M0WNM0_9BACI | Bacillus sp. FJAT-21351 | Lipase
K8GFE2_9CYAN | Oscillatoriales cyanobacterium JSC-12 | Lysophospholipase L1-like esterase
K9WJ28_9CYAN | Microcoleus sp. PCC 7113 | Lysophospholipase L1-like esterase
R7ADB4_9BACE | Bacteroides pectinophilus CAG:437 | Uncharacterized protein
R5GT16_9FIRM | Eubacterium sp. CAG:786 | GDSL-like protein
R5LL12_9FIRM | Eubacterium sp. CAG:115 | GDSL-like protein
F6DQC0_DESRL | Desulfotomaculum ruminis (strain ATCC 23193 / DSM 2154 / NCIB 8452 / DL) | Lipolytic protein G-D-S-L family
A0A0H0SJ00_9BACL | Brevibacillus formosus | Lysophospholipase
J3A568_9BACL | Brevibacillus sp. BC25 | Lysophospholipase L1-like esterase
A0A0J6BBM7_BREBE | Brevibacillus brevis | Lysophospholipase
A0A0L0WAR6_CLOPU | Clostridium purinilyticum | Lysophospholipase L1
A0A0C1UEU0_9CLOT | Clostridium argentinense CDC 2741 | GDSL-like Lipase/Acylhydrolase family protein
Q897X6_CLOTE | Clostridium tetani (strain Massachusetts / E88) | Platelet activating factor acetylhydrolase-like protein
U6EVC2_CLOTA | Clostridium tetani 12124569 | Platelet activating factor acetylhydrolase-likeprotein
A0A095ZDI3_9FIRM | Tissierellia bacterium S7-1-4 | Uncharacterized protein
A0A072Y8N5_9CLOT | Clostridium sp. K25 | Acetylhydrolase
A0A078KJ49_9FIRM | [Clostridium] cellulosi | Uncharacterized protein
W3AC50_9BACL | Planomicrobium glaciei CHR43 | Uncharacterized protein
A0A098EIL2_9BACL | Planomicrobium sp. ES2 | Multifunctional acyl-CoA thioesterase I and protease I and lysophospholipase L1
R4JC30_9BACT | uncultured bacterium BAC25G1 | Uncharacterized protein
R5VMZ1_9FIRM | Firmicutes bacterium CAG:631 | GDSL-like protein
A0A0H1UPZ9_STRAG | Streptococcus agalactiae | Acylneuraminate cytidylyltransferase
A0A0A6S1U2_STRUB | Streptococcus uberis | Acylneuraminate cytidylyltransferase
R2P1Z4_9ENTE | Enterococcus raffinosus ATCC 49464 | Uncharacterized protein
A0A0G1KN57_9BACT | candidate division WWE3 bacterium GW2011_GWC2_44_9 | Secreted protein
E9I8K8_SOLIN | Solenopsis invicta | Putative uncharacterized protein
A0A026WMY1_CERBI | Cerapachys biroi | Maternal effect protein oskar
E2BYH0_HARSA | Harpegnathos saltator | Putative uncharacterized protein
E2A7I8_CAMFO | Camponotus floridanus | Putative uncharacterized protein
F2WJY6_9HYME | Messor pergandei | Oskar
F4WQN7_ACREC | Acromyrmex echinatior | Maternal effect protein oskar
E9IZ46_SOLIN | Solenopsis invicta | Putative uncharacterized protein
A0A0J7KH44_LASNI | Lasius niger | Maternal effect protein oskar
E1A883_NASVI | Nasonia vitripennis | Oskar
A0A0C9QHR7_9HYME | Fopius arisanus | Osk protein
K4MTL4_GRYBI | Gryllus bimaculatus | Oskar
Q2PP79_AEDAE | Aedes aegypti | Oskar
B0WIV7_CULQU | Culex quinquefasciatus | Oskar
U5EFJ8_9DIPT | Corethrella appendiculata | Putative oskar
T1PG45_MUSDO | Musca domestica | GDSL-like Lipase/Acylhydrolase
A0A0L0CP24_LUCCU | Lucilia cuprina | Uncharacterized protein
W8CE30_CERCA | Ceratitis capitata | Maternal effect protein oskar
A0A034WRF5_BACDO | Bactrocera dorsalis | Maternal effect protein oskar
A0A0K8U7J3_BACLA | Bactrocera latifrons | Maternal effect protein oskar
A0A0A1XRQ4_BACCU | Bactrocera cucurbitae | Maternal effect protein oskar
A1Y1T7_DROIM | Drosophila immigrans | Oskar
B4LXK5_DROVI | Drosophila virilis | Oskar
A0A0M4F3M8_DROBS | Drosophila busckii | Osk
B4JTJ1_DROGR | Drosophila grimshawi | GH23955
B4K9E5_DROMO | Drosophila mojavensis | Uncharacterized protein
B3LZ06_DROAN | Drosophila ananassae | Uncharacterized protein
B3P1W4_DROER | Drosophila erecta | GG13545
B4PTX6_DROYA | Drosophila yakuba | Uncharacterized protein
E8NH25_DROME | Drosophila melanogaster | RE24380p
B4HKZ1_DROSE | Drosophila sechellia | GM23770
A0A126GUR4_DROME | Drosophila melanogaster | Oskar, isoform D
Q295Q4_DROPS | Drosophila pseudoobscura pseudoobscura | Uncharacterized protein, isoform A
A0A059PGF2_9MUSC | Drosophila pseudoobscura | GA10627
A0A059PF64_9MUSC | Drosophila pseudoobscura | GA10627
B4N815_DROWI | Drosophila willistoni | Uncharacterized protein
T1DTM7_ANOAQ | Anopheles aquasalis | Uncharacterized protein
W5JJ85_ANODA | Anopheles darlingi | Uncharacterized protein
A0A084WRU4_ANOSI | Anopheles sinensis | AGAP003545-PA-like protein
Q7PQJ3_ANOGA | Anopheles gambiae | AGAP003545-PA
A0A0G3CFD7_METBA | Methanosarcina barkeri CM1 | GDSL family lipase/acylhydrolase
1.46752


MUSCLE tree
A0A0E3QN36_METBA
R6T7B3_9BACE
R5S0B3_9BACE
E1YW67_9PORP
A0A073IAZ3_9PORP
A0A0J9FZD4_9PORP
A0A069S7Q5_9PORP
CAZ3
B7KKA6_CYAP7
F6DQC0_DESRL
A0A0H0SJ00_9BACL
J3A568_9BACL
A0A0J6BBM7_BREBE
R5GT16_9FIRM
R5LL12_9FIRM
A0A078KJ49_9FIRM
A0A072Y8N5_9CLOT
R7ADB4_9BACE
Q897X6_CLOTE
U6EVC2_CLOTA
A0A095ZDI3_9FIRM
A0A0L0WAR6_CLOPU
A0A0C1UEU0_9CLOT
R4JC30_9BACT
A5N8N5_CLOK5
D5DGY9_BACMD
A0A0M0WNM0_9BACI
K9WJ28_9CYAN
W3AC50_9BACL
A0A098EIL2_9BACL
K8GFE2_9CYAN
G2HS43_9PROT
S5PEQ8_9PROT
A0A0G9K3L5_9PROT
A0A0M1UPT0_9PROT
E6L4E3_9PROT
A8EWS4_ARCB4
A0A0H1UPZ9_STRAG
A0A0A6S1U2_STRUB
A0A0G1KN57_9BACT
R2P1Z4_9ENTE
R5VMZ1_9FIRM
OSK
A0A0G3CFD7_METBA
Similarity
0
1
T COFFEE tree
A0A0E3QN36_METBA
R6T7B3_9BACE
R5S0B3_9BACE
E1YW67_9PORP
A0A073IAZ3_9PORP
A0A0J9FZD4_9PORP
A0A069S7Q5_9PORP
B7KKA6_CYAP7
CAZ3
R4JC30_9BACT
R7ADB4_9BACE
A0A078KJ49_9FIRM
R5GT16_9FIRM
R5LL12_9FIRM
F6DQC0_DESRL
A0A0H0SJ00_9BACL
J3A568_9BACL
A0A0J6BBM7_BREBE
A0A072Y8N5_9CLOT
Q897X6_CLOTE
U6EVC2_CLOTA
A0A095ZDI3_9FIRM
A0A0L0WAR6_CLOPU
A0A0C1UEU0_9CLOT
W3AC50_9BACL
A0A098EIL2_9BACL
D5DGY9_BACMD
A0A0M0WNM0_9BACI
A5N8N5_CLOK5
K9WJ28_9CYAN
K8GFE2_9CYAN
G2HS43_9PROT
S5PEQ8_9PROT
A0A0G9K3L5_9PROT
A0A0M1UPT0_9PROT
E6L4E3_9PROT
A8EWS4_ARCB4
A0A0G1KN57_9BACT
A0A0A6S1U2_STRUB
A0A0H1UPZ9_STRAG
R2P1Z4_9ENTE
R5VMZ1_9FIRM
OSK
A0A0G3CFD7_METBA

MUSCLE tree



Similarity

0 1

T COFFEE tree

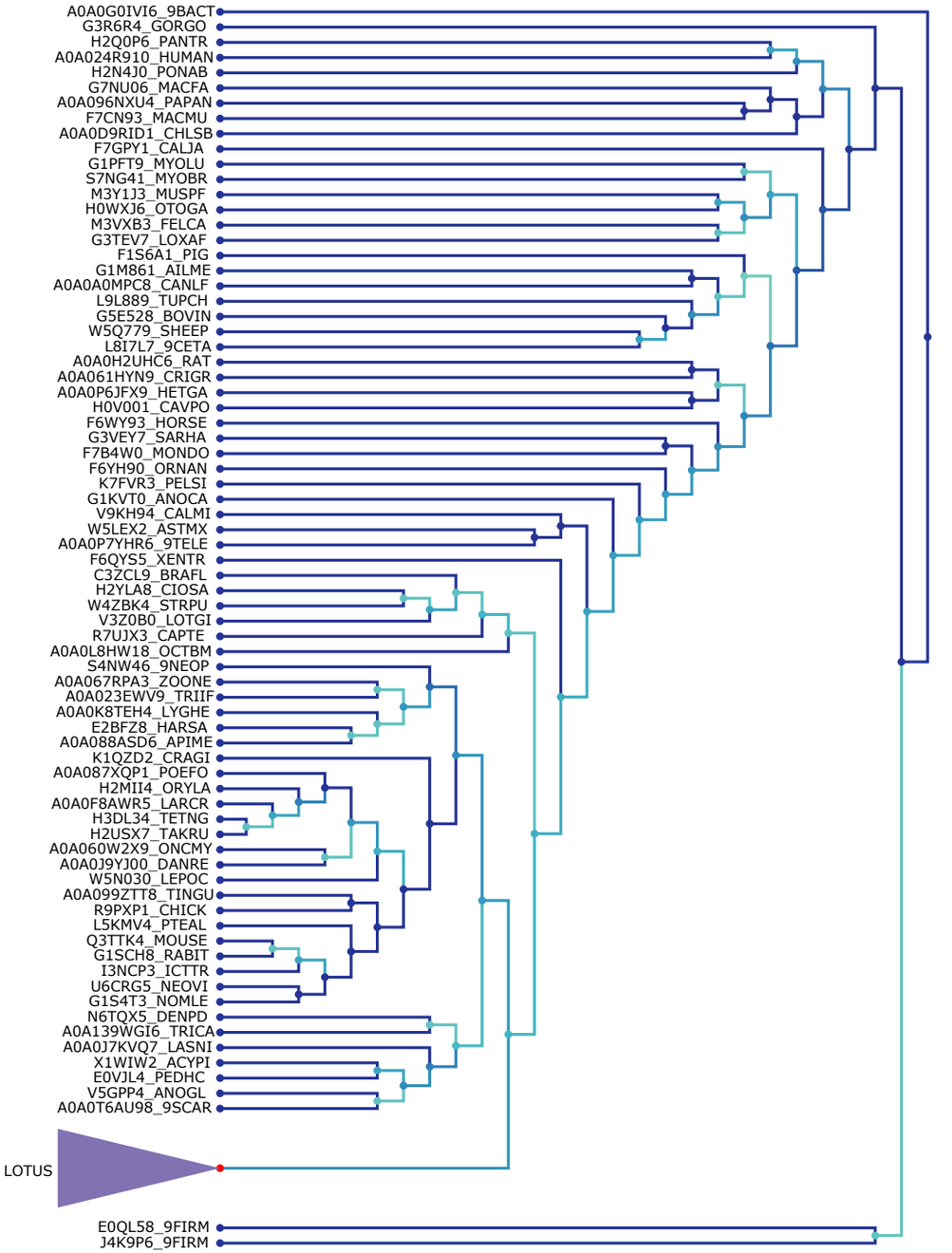
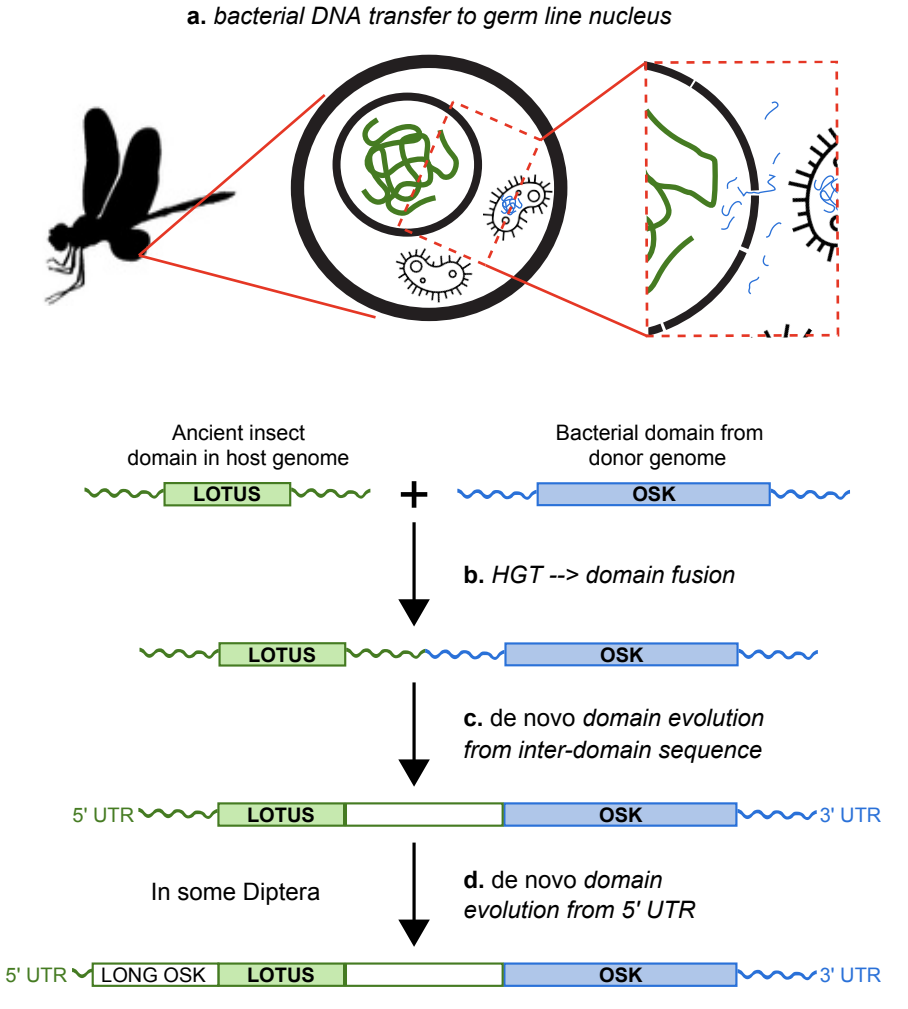



Figure 3
Blondel, Jones & Extavour