

Supplementary File 1 for

Bacterial contribution to genesis of the novel germ line determinant *oskar* *Leo Blondel, Tamsin E. M. Jones and Cassandra G. Extavour*

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

Supplementary File 1

THIS PDF DOCUMENT: FILE NAME BLONDEL_JONES_EXTAVOUR_HGT_SUPPLEMENTARY_FILE_1

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.
- e. Supplementary File 1E: List of genomes analyzed for codon use.

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

44
45
46
47
48
49
50
51

Supplementary File 1A: List of genomes and transcriptomes used for automated *oskar* search.

List of genomes and transcriptomes that were downloaded, annotated, and searched for *oskar* sequences (see “Hidden Markov Model (HMM) generation and alignments of the *OSK* and *LOTUS* domains” in Methods). The table reports the database provenance (NCBI genome or TSA, or 1KITE database) and the accession number. The TSA accession ID can be searched using the NCBI TSA browser here: <https://www.ncbi.nlm.nih.gov/Traces/wgs/?view=TSA>.

Type	Accession ID	Organism Name
Genome	GCA_000001765.2	<i>Drosophila pseudoobscura pseudoobscura</i>
Genome	GCA_000002325.2	<i>Nasonia vitripennis</i>
Genome	GCA_000002335.2	<i>Tribolium castaneum</i>
Genome	GCA_000004775.1	<i>Nasonia giraulti</i>
Genome	GCA_000004795.1	<i>Nasonia longicornis</i>
Genome	GCA_000005115.1	<i>Drosophila ananassae</i>
Genome	GCA_000005135.1	<i>Drosophila erecta</i>
Genome	GCA_000005195.1	<i>Drosophila persimilis</i>
Genome	GCA_000005245.1	<i>Drosophila virilis</i>
Genome	GCA_000005925.1	<i>Drosophila willistoni</i>
Genome	GCA_000005975.1	<i>Drosophila yakuba</i>
Genome	GCA_000006295.1	<i>Pediculus humanus corporis</i>
Genome	GCA_000143395.2	<i>Atta cephalotes</i>
Genome	GCA_000147175.1	<i>Camponotus floridanus</i>
Genome	GCA_000147195.1	<i>Harpegnathos saltator</i>
Genome	GCA_000149185.1	<i>Mayetiola destructor</i>
Genome	GCA_000151625.1	<i>Bombyx mori</i>
Genome	GCA_000151715.1	<i>Bombyx mori</i>
Genome	GCA_000181055.3	<i>Rhodnius prolixus</i>
Genome	GCA_000184785.1	<i>Apis florea</i>
Genome	GCA_000187875.1	<i>Daphnia pulex</i>
Genome	GCA_000187915.1	<i>Pogonomyrmex barbatus</i>
Genome	GCA_000188075.1	<i>Solenopsis invicta</i>
Genome	GCA_000209185.1	<i>Culex quinquefasciatus</i>
Genome	GCA_000211455.3	<i>Anopheles darlingi</i>
Genome	GCA_000214255.1	<i>Bombus terrestris</i>
Genome	GCA_000217595.1	<i>Linepithema humile</i>
Genome	GCA_000220665.2	<i>Drosophila ficusphila</i>
Genome	GCA_000220905.1	<i>Megachile rotundata</i>
Genome	GCA_000224215.2	<i>Drosophila kikkawai</i>
Genome	GCA_000224235.2	<i>Drosophila takahashii</i>
Genome	GCA_000233415.2	<i>Drosophila biarmipes</i>
Genome	GCA_000235995.1	<i>Danaus plexippus plexippus</i>
Genome	GCA_000236305.2	<i>Drosophila rhopaloa</i>

Genome	GCA_000239435.1	<i>Tetranychus urticae</i>
Genome	GCA_000239455.1	<i>Strigamia maritima</i>
Genome	GCA_000259055.1	<i>Drosophila simulans</i>
Genome	GCA_000262585.1	<i>Manduca sexta</i>
Genome	GCA_000262795.1	<i>Phlebotomus papatasi</i>
Genome	GCA_000265325.1	<i>Lutzomyia longipalpis</i>
Genome	GCA_000269505.2	<i>Drosophila miranda</i>
Genome	GCA_000281935.1	<i>Mengenilla moldrzyki</i>
Genome	GCA_000298335.1	<i>Drosophila albomicans</i>
Genome	GCA_000325945.1	<i>Plutella xylostella</i>
Genome	GCA_000330985.1	<i>Plutella xylostella</i>
Genome	GCA_000341935.1	<i>Cephus cinctus</i>
Genome	GCA_000344095.1	<i>Athalia rosae</i>
Genome	GCA_000347755.1	<i>Ceratitis capitata</i>
Genome	GCA_000349025.1	<i>Anopheles minimus</i>
Genome	GCA_000349045.1	<i>Anopheles stephensi</i>
Genome	GCA_000349065.1	<i>Anopheles quadriannulatus</i>
Genome	GCA_000349085.1	<i>Anopheles funestus</i>
Genome	GCA_000349105.1	<i>Anopheles epiroticus</i>
Genome	GCA_000349145.1	<i>Anopheles dirus</i>
Genome	GCA_000349165.1	<i>Anopheles christyi</i>
Genome	GCA_000349185.1	<i>Anopheles arabiensis</i>
Genome	GCA_000371365.1	<i>Musca domestica</i>
Genome	GCA_000439205.1	<i>Anopheles nili</i>
Genome	GCA_000441895.2	<i>Anopheles sinensis</i>
Genome	GCA_000469605.1	<i>Apis dorsata</i>
Genome	GCA_000472105.1	<i>Drosophila suzukii</i>
Genome	GCA_000473185.1	<i>Anopheles maculatus</i>
Genome	GCA_000473375.1	<i>Anopheles culicifacies</i>
Genome	GCA_000473445.2	<i>Anopheles farauti</i>
Genome	GCA_000473505.1	<i>Anopheles atroparvus</i>
Genome	GCA_000473525.2	<i>Anopheles melas</i>
Genome	GCA_000473845.2	<i>Anopheles merus</i>
Genome	GCA_000475195.1	<i>Diaphorina citri</i>
Genome	GCA_000500325.1	<i>Leptinotarsa decemlineata</i>
Genome	GCA_000503995.1	<i>Ceratosolen solmsi marchali</i>
Genome	GCA_000507165.1	<i>Ephemera danica</i>
Genome	GCA_000516895.1	<i>Locusta migratoria</i>
Genome	GCA_000599845.1	<i>Trichogramma pretiosum</i>
Genome	GCA_000611835.1	<i>Ooceraea biroi</i>
Genome	GCA_000648655.1	<i>Copidosoma floridanum</i>
Genome	GCA_000648675.1	<i>Cimex lectularius</i>

Genome	GCA_000648695.1	<i>Onthophagus taurus</i>
Genome	GCA_000648945.1	<i>Limnephilus lunatus</i>
Genome	GCA_000671735.1	<i>Glossina fuscipes fuscipes</i>
Genome	GCA_000688715.1	<i>Glossina pallidipes</i>
Genome	GCA_000688735.1	<i>Glossina austeni</i>
Genome	GCA_000695345.1	<i>Bactrocera tryoni</i>
Genome	GCA_000695645.1	<i>Pachypsylla venusta</i>
Genome	GCA_000696155.1	<i>Zootermopsis nevadensis</i>
Genome	GCA_000696205.1	<i>Oncopeltus fasciatus</i>
Genome	GCA_000696795.1	<i>Halyomorpha halys</i>
Genome	GCA_000696855.1	<i>Homalodisca vitripennis</i>
Genome	GCA_000699065.1	<i>Lucilia cuprina</i>
Genome	GCA_000762945.1	<i>Blattella germanica</i>
Genome	GCA_000775305.1	<i>Belgica antarctica</i>
Genome	GCA_000786065.1	<i>Piezodorus guildinii</i>
Genome	GCA_000786525.1	<i>Chironomus tentans</i>
Genome	GCA_000789215.2	<i>Bactrocera dorsalis</i>
Genome	GCA_000818775.1	<i>Glossina palpalis gambiensis</i>
Genome	GCA_000836235.1	<i>Papilio xuthus</i>
Genome	GCA_000934665.1	<i>Catajapyx aquilonaris</i>
Genome	GCA_000956155.1	<i>Cotesia vestalis</i>
Genome	GCA_000956235.1	<i>Wasmannia auropunctata</i>
Genome	GCA_000956255.1	<i>Anopheles punctulatus</i>
Genome	GCA_000956275.1	<i>Anopheles koliensis</i>
Genome	GCA_001014415.1	<i>Phortica variegata</i>
Genome	GCA_001014435.1	<i>Mayetiola destructor</i>
Genome	GCA_001014445.1	<i>Scaptodrosophila lebanonensis</i>
Genome	GCA_001014505.1	<i>Chironomus riparius</i>
Genome	GCA_001014625.1	<i>Bactrocera oleae</i>
Genome	GCA_001014835.1	<i>Lucilia sericata</i>
Genome	GCA_001014935.1	<i>Liriomyza trifolii</i>
Genome	GCA_001014945.1	<i>Clogmia albipunctata</i>
Genome	GCA_001015115.1	<i>Eutreta diana</i>
Genome	GCA_001015145.1	<i>Eristalis dimidiata</i>
Genome	GCA_001015175.1	<i>Megaselia abdita</i>
Genome	GCA_001015215.1	<i>Holcocephala fusca</i>
Genome	GCA_001015235.1	<i>Sphyracephala brevicornis</i>
Genome	GCA_001015335.1	<i>Stomoxys calcitrans</i>
Genome	GCA_001017275.1	<i>Calliphora vicina</i>
Genome	GCA_001017455.1	<i>Neobellieria bullata</i>
Genome	GCA_001017515.1	<i>Tephritis californica</i>
Genome	GCA_001017525.1	<i>Teleopsis dalmanni</i>

Genome	GCA_001017535.1	<i>Tipula oleracea</i>
Genome	GCA_001077435.1	<i>Glossina morsitans morsitans</i>
Genome	GCA_001186385.1	<i>Diuraphis noxia</i>
Genome	GCA_001188975.2	<i>Bactrocera oleae</i>
Genome	GCA_001272555.1	<i>Dufourea novaeangliae</i>
Genome	GCF_000142985.2	<i>Acyrtosiphon pisum</i>
Genome	GCF_000208615.1	<i>Ixodes scapularis</i>
Genome	GCF_000004015.3	<i>Aedes aegypti</i>
TSA	GBKU	<i>Trichoplusia ni</i>
TSA	GBRL	<i>Sitodiplosis mosellana</i>
TSA	GAAX	<i>Forficula auricularia</i>
TSA	GACV	<i>Philopotamus ludificatus</i>
TSA	GADM	<i>Drosophila malerkotliana malerkotliana</i>
TSA	GAEO	<i>Pissodes strobi</i>
TSA	GAFI	<i>Dendroctonus frontalis</i>
TSA	GAFR	<i>Polistes canadensis</i>
TSA	GAHN	<i>Drosophila serrata</i>
TSA	GAHP	<i>Ostrinia nubilalis</i>
TSA	GAHQ	<i>Ostrinia scapulalis</i>
TSA	GAIW	<i>Ganaspis sp. G1</i>
TSA	GAJA	<i>Leptopilina boulardi</i>
TSA	GAJC	<i>Leptopilina heterotoma</i>
TSA	GAKJ	<i>Sitodiplosis mosellana</i>
TSA	GAMD	<i>Anopheles aquasalis</i>
TSA	GANH	<i>Microplitis demolitor</i>
TSA	GANO	<i>Corethrella appendiculata</i>
TSA	GAPE	<i>Brassicogethes aeneus</i>
TSA	GAPT	<i>Ectopsocus briggsi</i>
TSA	GASE	<i>Prorethra simplex</i>
TSA	GASG	<i>Yponomeuta evonymellus</i>
TSA	GASN	<i>Thermobia domestica</i>
TSA	GASO	<i>Tricholepidion gertschi</i>
TSA	GASQ	<i>Tetrix subulata</i>
TSA	GASS	<i>Platycentropus radiatus</i>
TSA	GAST	<i>Polyommatus icarus</i>
TSA	GASV	<i>Notostira elongata</i>
TSA	GASW	<i>Mantis religiosa</i>
TSA	GASX	<i>Folsomia candida</i>
TSA	GASY	<i>Dyseriocrania subpurpurella</i>
TSA	GATA	<i>Meloe violaceus</i>
TSA	GATB	<i>Metallyticus splendidus</i>
TSA	GATC	<i>Nemophora degeerella</i>

TSA	GATD	<i>Pogonognathellus sp. AD-2013</i>
TSA	GATG	<i>Corydalis cornutus</i>
TSA	GATH	<i>Bittacus pilicornis</i>
TSA	GATI	<i>Bombylius major</i>
TSA	GATJ	<i>Bibio marci</i>
TSA	GATU	<i>Baetis sp. AD-2013</i>
TSA	GATV	<i>Perla marginata</i>
TSA	GATW	<i>Aleochara curtula</i>
TSA	GATY	<i>Chrysis viridula</i>
TSA	GATZ	<i>Sminthurus viridis</i>
TSA	GAUE	<i>Anurida maritima</i>
TSA	GAUF	<i>Leuctra sp. AD-2013</i>
TSA	GAUG	<i>Meinertellus cundinamarcensis</i>
TSA	GAUH	<i>Panorpa vulgaris</i>
TSA	GAUI	<i>Subilla sp. AD-2014</i>
TSA	GAUM	<i>Machilis hrabei</i>
TSA	GAUN	<i>Cercopis vulnerata</i>
TSA	GAUO	<i>Velia caprai</i>
TSA	GAUV	<i>Acanthosoma haemorrhoidale</i>
TSA	GAUW	<i>Apachyus charteceus</i>
TSA	GAUX	<i>Ceuthophilus sp. AD-2013</i>
TSA	GAUY	<i>Gyrinus marinus</i>
TSA	GAUZ	<i>Stenobothrus lineatus</i>
TSA	GAVA	<i>Triarthria setipennis</i>
TSA	GAVB	<i>Triodia sylvina</i>
TSA	GAVM	<i>Hydroptila sp. AD-2013</i>
TSA	GAVV	<i>Pseudomallada prasinus</i>
TSA	GAVW	<i>Epiophlebia superstes</i>
TSA	GAVX	<i>Timema cristinae</i>
TSA	GAWC	<i>Aretaon asperrimus</i>
TSA	GAWD	<i>Medauroidea extradentata</i>
TSA	GAWE	<i>Ramulus artemis</i>
TSA	GAWF	<i>Sipyloidea sipyilus</i>
TSA	GAWG	<i>Extatosoma tiaratum</i>
TSA	GAWK	<i>Ceratophyllus gallinae</i>
TSA	GAWM	<i>Culicoides sonorensis</i>
TSA	GAWP	<i>Grylloblatta bifratrilecta</i>
TSA	GAWQ	<i>Okanagana villosa</i>
TSA	GAWR	<i>Menopon gallinae</i>
TSA	GAWS	<i>Periplaneta americana</i>
TSA	GAWT	<i>Empusa pennata</i>
TSA	GAWU	<i>Aposthonia japonica</i>

TSA	GAWW	<i>Tenthredo koehleri</i>
TSA	GAWX	<i>Trialeurodes vaporariorum</i>
TSA	GAWZ	<i>Gryllotalpa sp. AD-2013</i>
TSA	GAXA	<i>Isonychia bicolor</i>
TSA	GAXB	<i>Tanzaniophasma sp. AD-2013</i>
TSA	GAXC	<i>Thrips palmi</i>
TSA	GAXE	<i>Acerentomon sp. AD-2013</i>
TSA	GAXF	<i>Planococcus citri</i>
TSA	GAXG	<i>Gynaikothrips ficorum</i>
TSA	GAXH	<i>Parides eurimedes</i>
TSA	GAXI	<i>Tetrodontophora bielanensis</i>
TSA	GAXM	<i>Mischocyttarus flavitarsis</i>
TSA	GAXO	<i>Argochrysis armilla</i>
TSA	GAXS	<i>Pepsis grossa</i>
TSA	GAXT	<i>Crioscolia alcione</i>
TSA	GAXW	<i>Euroleon nostras</i>
TSA	GAXX	<i>Rhyacophila fasciata</i>
TSA	GAXZ	<i>Trichocera saltator</i>
TSA	GAYA	<i>Zorotypus caudelli</i>
TSA	GAYB	<i>Zygaena fausta</i>
TSA	GAYC	<i>Osmylus fulvicephalus</i>
TSA	GAYD	<i>Blaberus atropos</i>
TSA	GAYE	<i>Frankliniella cephalica</i>
TSA	GAYH	<i>Conwentzia psociformis</i>
TSA	GAYI	<i>Xenophysella greensladeae</i>
TSA	GAYJ	<i>Atelura formicaria</i>
TSA	GAYK	<i>Boreus hyemalis</i>
TSA	GAYL	<i>Cosmioperla kuna</i>
TSA	GAYM	<i>Calopteryx splendens</i>
TSA	GAYN	<i>Campodea augens</i>
TSA	GAYO	<i>Cordulegaster boltonii</i>
TSA	GAYP	<i>Ctenocephalides felis</i>
TSA	GAYQ	<i>Forficula auricularia</i>
TSA	GAYV	<i>Liposcelis bostrychophila</i>
TSA	GAYY	<i>Acanthocasuarina muellerianae</i>
TSA	GAYZ	<i>Ranatra linearis</i>
TSA	GAZA	<i>Haploembia palaui</i>
TSA	GAZB	<i>Lepicerus sp. AD-2013</i>
TSA	GAZD	<i>Lipara lucens</i>
TSA	GAZE	<i>Mastotermes darwiniensis</i>
TSA	GAZF	<i>Essigella californica</i>
TSA	GAZG	<i>Eurylophella sp. AD-2013</i>

TSA	GAZH	<i>Inocellia crassicornis</i>
TSA	GAZM	<i>Stylops melittae</i>
TSA	GAZN	<i>Cryptocercus wrighti</i>
TSA	GAZQ	<i>Aretaon asperrimus</i>
TSA	GAZT	<i>Prosarthria teretirostris</i>
TSA	GBAB	<i>Musca domestica</i>
TSA	GBBP	<i>Teleopsis dalmanni</i>
TSA	GBCX	<i>Dastarcus helophoroides</i>
TSA	GBDM	<i>Helicoverpa armigera</i>
TSA	GBGV	<i>Polistes metricus</i>
1KITE	INSnfrTAJRAAPEI	<i>Machilis hrabei</i>
1KITE	INSnfrTAFRAAPEI	<i>Meinertellus cundinamarcensis</i>
1KITE	INSfrgTAVRAAPEI	<i>Blaberus atropos</i>
1KITE	INSfrgTAARAPEI	<i>Periplaneta americana</i>
1KITE	INSyvtCDRAAPEI	<i>Cryptocercus wrighti</i>
1KITE	INSbttTCRAAPEI	<i>Arrhenodes minutus</i>
1KITE	INSnfrTBERAAPEI	<i>Gyrinus marinus</i>
1KITE	INSyvtTAJRAAPEI	<i>Lepicerus sp.</i>
1KITE	INShauTAYRAAPEI	<i>Meloe violaceus</i>
1KITE	INShauTBERAAPEI	<i>Aleochara curtula</i>
1KITE	INSbttTIRAAPEI	<i>Folsomia candida</i>
1KITE	INSnfrTAIRAAPEI	<i>Anurida maritima</i>
1KITE	INSjdsTAKRAAPEI	<i>Tetrodontophora bielanensis</i>
1KITE	INShauTAFRAAPEI	<i>Sminthurus viridis/nigromaculatus</i>
1KITE	INShauTAJRAAPEI	<i>Pogonognathellus longicornis/flavescens</i>
1KITE	INSfrgTALRAAPEI	<i>Apachyus charteceus</i>
1KITE	INSjdsTBNRAAPEI	<i>Forficula auricularia</i>
1KITE	INSTmbTABRAAPEI	<i>Campodea augens</i>
1KITE	INSjdsTANRAAPEI	<i>Occasjapyx japonicus</i>
1KITE	INSbusTBKRAAPEI	<i>Bombylius major</i>
1KITE	INSyvtTBWRAAPEI	<i>Lipara lucens</i>
1KITE	INSnfrTBFRAAPEI	<i>Triarthria setipennis</i>
1KITE	INSbusTBCRABPEI	<i>Bibio marci</i>
1KITE	INSjdsTBERAAPEI	<i>Trichocera saltator</i>
1KITE	INSfrgTAZRAAPEI	<i>Aposthonia japonica</i>
1KITE	INSyvtTAHRAAPEI	<i>Haploembia palaui</i>
1KITE	INShauTAKRAAPEI	<i>Baetis sp.</i>
1KITE	INSyvtTCERAAPEI	<i>Eurylophella sp.</i>
1KITE	INSnfrTAKRAAPEI	<i>Ephemera danica</i>
1KITE	INSjdsTAGRAAPEI	<i>Isonychia bicolor</i>
1KITE	INSfrgTAKRAAPEI	<i>Galloisiana yuasai</i>
1KITE	INSnfrTBKRAAPEI	<i>Grylloblatta bifratrilecta</i>

1KITE	INSnfrTANRAAPEI	<i>Cercopis vulnerata</i>
1KITE	INSnfrTBLRAAPEI	<i>Okanagana villosa</i>
1KITE	INSfrgTBCRAAPEI	<i>Nilaparvata lugens</i>
1KITE	INSjdsTARRAAPEI	<i>Xenophysella greensladeae</i>
1KITE	INSnfrTAPRAAPEI	<i>Acanthosoma haemorrhoidale</i>
1KITE	INShauTAPRAAPEI	<i>Notostira elongata</i>
1KITE	INSyvtTANRAAPEI	<i>Ranatra linearis</i>
1KITE	INSnfrTAORAAPEI	<i>Velia caprai</i>
1KITE	INSfrgTAPRAAPEI	<i>Trialeurodes vaporariorum</i>
1KITE	INSyvtTBHRAAPEI	<i>Essigella californica</i>
1KITE	INSjdsTAIRAAPEI	<i>Planococcus citri</i>
1KITE	INSyvtTALRAAPEI	<i>Acanthocasuarina muellerianae</i>
1KITE	INSnfrTAQRAAPEI	<i>Cotesia vestalis</i>
1KITE	INShauTAQRABPEI	<i>Chrysis viridula</i>
1KITE	INSjdsTAURAAPEI	<i>Leptopilina clavipes</i>
1KITE	INSnfrTAARAAPEI	<i>Orussus abietinus</i>
1KITE	INSfrgTATRAAPEI	<i>Tenthredo koehleri</i>
1KITE	INSTmbTBPRAAPEI	<i>Mastotermes darwiniensis</i>
1KITE	INSbusTBMRAAPEI	<i>Prorhinotermes simplex</i>
1KITE	INShauTABRAAPEI	<i>Nemophora degeerella</i>
1KITE	INSbusTBDRAAPEI	<i>Dyseriocrania subpurpurella</i>
1KITE	INSnfrTAVRAAPEI	<i>Triodia sylvina</i>
1KITE	INShauTBGRAAPEI	<i>Polyommatus icarus</i>
1KITE	INSjdsTAJRAAPEI	<i>Parides eurimedes</i>
1KITE	INShauTBFRAAPEI	<i>Yponomeuta evonymellus</i>
1KITE	INSjdsTAWRAAPEI	<i>Zygaena fausta</i>
1KITE	INSfrgTASRAAPEI	<i>Empusa pennata</i>
1KITE	INShauTAARAAPEI	<i>Mantis religiosa</i>
1KITE	INShauTAMRAAPEI	<i>Metallyticus splendidus</i>
1KITE	INSfrgTBBRAAPEI	<i>Tanzaniophasma sp.</i>
1KITE	INSbttTARAPEI	<i>Bittacus pilicornis</i>
1KITE	INSTmbTAWRAAPEI	<i>Boreus hyemalis</i>
1KITE	INShauTACRAAPEI	<i>Panorpa vulgaris</i>
1KITE	INSbttTKRAAPEI	<i>Corydalus cornutus</i>
1KITE	INSnfrTARRAAPEI	<i>Pseudomallada prasinus</i>
1KITE	INSjdsTBQRAAPEI	<i>Conwentzia psociformis</i>
1KITE	INSjdsTATRAAPEI	<i>Euroleon nostras</i>
1KITE	INSjdsTBJRAAPEI	<i>Osmylus fulvicephalus</i>
1KITE	INSjdsTBHRAAPEI	<i>Cordulegaster boltonii</i>
1KITE	INSfrgTAHRAAPEI	<i>Epiophlebia superstes</i>
1KITE	INSTmbTAARAAPEI	<i>Calopteryx splendens</i>
1KITE	INSnfrTAMRAAPEI	<i>Stenobothrus lineatus</i>

1KITE	INSmbTBCRBAPEI	<i>Prosarthria teretirostris</i>
1KITE	INShauTANRAAPEI	<i>Tetrix subulata</i>
1KITE	INSfrgTAXRABPEI	<i>Gryllotalpa sp.</i>
1KITE	INSnfrTBIRAAPEI	<i>Ceuthophilus sp.</i>
1KITE	INSmbTBERAAPEI	<i>Aretaon asperrimus</i>
1KITE	INSfrgTAORAAPEI	<i>Peruphasma schultei</i>
1KITE	INSnfrTBPRAAPEI	<i>Timema cristinae</i>
1KITE	INSmbTBFRAAPEI	<i>Cosmioperla kuna</i>
1KITE	INSnfrTALRAAPEI	<i>Leuctra sp.</i>
1KITE	INShauTALRAAPEI	<i>Perla marginata</i>
1KITE	INSjdsTAHRAAPEI	<i>Acerentomon sp.</i>
1KITE	INSfrgTAFRAAPEI	<i>Menopon gallinae</i>
1KITE	INSyvtTCFRAAPEI	<i>Ectopsocus briggsi</i>
1KITE	INSmbTBGRAAPEI	<i>Liposcelis bostrychophila</i>
1KITE	INSnfrTAGRAAPEI	<i>Xanthostigma xanthostigma</i>
1KITE	INSnfrTBARAAPEI	<i>Ceratophyllus gallinae</i>
1KITE	INSmbTAYRAAPEI	<i>Ctenocephalides felis</i>
1KITE	INSyvtTBKRAAPEI	<i>Stylops melittae</i>
1KITE	INSjdsTADRAAPEI	<i>Gynaikothrips ficorum</i>
1KITE	INSjdsTABRAAPEI	<i>Frankliniella cephalica</i>
1KITE	INSjdsTACRAAPEI	<i>Thrips palmi</i>
1KITE	INSnfrTBJRAAPEI	<i>Hydroptila actia/argosa</i>
1KITE	INSjdsTBSRAAPEI	<i>Rhyacophila fasciata</i>
1KITE	INSjdsTAQRAAPEI	<i>Zorotypus caudelli</i>
1KITE	INSjdsTAVRAAPEI	<i>Atelura formicaria</i>
1KITE	INSbttTJRAAPEI	<i>Tricholepidion gertschi</i>
1KITE	INSbttTSRAAPEI	<i>Thermobia domestica</i>

53
54
55
56
57
58
59

Supplementary File 1B: List of *oskar* sequences used in the final alignment.

List of accession numbers and database provenance of the sequences used in the final alignments of Oskar analysed herein. The table contains the database provenance (*Type*), the database accession number (*ID*), the species, family and order, and extraction notes. In the “Annotation” Column, P = homolog identified by pipeline; DB = homolog identified by database annotation. *Sequence recomposed from two transcripts: GBCX01024638.1 and GBCX01024637.

Type	ID	Species	Family	Order	Annotation	Note
TSA	GAWC01068734.1	<i>Aretaon asperrimus</i>	Heteropterygidae	Phasmatodea	P	HMMER
TSA	GAKJ01010751.1	<i>Sitodiplosis mosellana</i>	Cecidomyiidae	Diptera	P	HMMER
TSA	GATI01010233.1	<i>Bombylius major</i>	Bombyliidae	Diptera	P	HMMER
TSA	GAWW01000144.1	<i>Tenthredo koehleri</i>	Tenthredinidae	Hymenoptera	P	HMMER
TSA	GAIW01009539.1	<i>Ganaspis sp.</i>	Figitidae	Hymenoptera	P	HMMER
TSA	GATY01008637.1	<i>Chrysis viridula</i>	Chrysididae	Hymenoptera	P	HMMER
TSA	GAXM01030263.1	<i>Mischocyttarus flavitarsis</i>	Vespidae	Hymenoptera	P	HMMER
TSA	GAFR01040300.1	<i>Polistes canadensis</i>	Vespidae	Hymenoptera	P	HMMER
TSA	GAZD01106195.1	<i>Lipara lucens</i>	Chloropidae	Diptera	P	HMMER
TSA	GAVA01002196.1	<i>Triarthria setipennis</i>	Tachinidae	Diptera	P	HMMER
TSA	GACV01001831.1	<i>Philopotamus ludificatus</i>	Philopotaminae	Trichoptera	P	HMMER
TSA	GAPE01019095.1	<i>Brassicogethes aeneus</i>	Nitidulidae	Coleoptera	P	HMMER
TSA	GBCX01024638.1_7.1	<i>Dastarcus helophoroides*</i>	Bothrideridae	Coleoptera	P	BLAST 1
TSA	GAKJ01010751.1	<i>Sitodiplosis mosellana</i>	Cecidomyiidae	Diptera	P	HMMER
TSA	GAWM01006639.1	<i>Culicoides sonorensis</i>	Ceratopogonidae	Diptera	P	HMMER
TSA	GAMD01000859.1	<i>Anopheles aquasalis</i>	Culicidae	Diptera	P	HMMER
TSA	GBEO01001325.1	<i>Anopheles sinensis</i>	Culicidae	Diptera	P	HMMER
TSA	GAKP01002609.1	<i>Bactrocera dorsalis</i>	Tephritidae	Diptera	P	HMMER
TSA	GAXM01030263.1	<i>Mischocyttarus flavitarsis</i>	Vespidae	Hymenoptera	P	HMMER
TSA	GAFR01040300.1	<i>Polistes canadensis canadensis</i>	Vespidae	Hymenoptera	P	HMMER
TSA	GBGV01010610.1	<i>Polistes metricus</i>	Vespidae	Hymenoptera	P	HMMER
TSA	GAIW01011550.1	<i>Ganaspis sp. G1</i>	Figitidae	Hymenoptera	P	HMMER
TSA	GAJC01011221.1	<i>Leptopilina heterotoma</i>	Figitidae	Hymenoptera	P	HMMER
TSA	GAIW01009539.1	<i>Ganaspis sp. G1</i>	Figitidae	Hymenoptera	P	HMMER

TSA	GAJA01020544.1	<i>Leptopilina boulandi</i>	Figitidae	Hymenoptera	P	HMMER
TSA	GAJC01009625.1	<i>Leptopilina heterotoma</i>	Figitidae	Hymenoptera	P	HMMER
TSA	GAXO01016630.1	<i>Argochrysis armilla</i>	Chrysididae	Hymenoptera	P	HMMER
TSA	GAEO01004319.1	<i>Pissodes strobi</i>	Curculionidae	Coleoptera	P	HMMER
TSA	GBBP01080309.1	<i>Teleopsis dalmanni</i>	Diopsidae	Diptera	P	HMMER
TSA	GAVM01000124.1	<i>Hydroptila actia/argosa</i>	Hydroptilidae	Trichoptera	P	HMMER
TSA	GAWC01068728	<i>Aretaon asperrimus</i>	Heteropterygidae	Phasmatodea	P	HMMER
TSA	GBEO01001325.1	<i>Anopheles sinensis</i>	Culicidae	Diptera	P	HMMER
Genome	GCA_000349125.1	<i>Anopheles albimanus</i>	Culicidae	Diptera	P	Snap
Genome	GCA_000349065.1	<i>Anopheles quadriannulatus</i>	Culicidae	Diptera	P	Augustus
Genome	GCA_000349185.1	<i>Anopheles arabiensis</i>	Culicidae	Diptera	P	Augustus
Genome	GCA_000473845.2	<i>Anopheles merus</i>	Culicidae	Diptera	P	Augustus
Genome	GCA_000473375.1	<i>Anopheles culicifacies</i>	Culicidae	Diptera	P	Augustus
Genome	GCA_000349085.1	<i>Anopheles funestus</i>	Culicidae	Diptera	P	Augustus
Genome	GCA_000209185.1	<i>Culex quinquefasciatus</i>	Culicidae	Diptera	P	Snap
Genome	GCF_000004015.3	<i>Aedes aegypti</i>	Culicidae	Diptera	P	Augustus
Genome	GCA_001014445.1	<i>Scaptodrosophila lebanonensis</i>	Drosophilidae	Diptera	P	Snap
Genome	GCA_000005925.1	<i>Drosophila willistoni</i>	Drosophilidae	Diptera	P	Augustus
Genome	GCA_000005115.1	<i>Drosophila ananassae</i>	Drosophilidae	Diptera	P	Snap
Genome	GCA_000236285.2	<i>Drosophila bipectinata</i>	Drosophilidae	Diptera	P	Augustus
Genome	GCA_000224215.2	<i>Drosophila kikkawai</i>	Drosophilidae	Diptera	P	Snap
Genome	GCA_000236325.2	<i>Drosophila eugracilis</i>	Drosophilidae	Diptera	P	Augustus
Genome	GCA_000001215.4	<i>Drosophila melanogaster</i>	Drosophilidae	Diptera	P	Augustus
Genome	GCA_000259055.1	<i>Drosophila simulans</i>	Drosophilidae	Diptera	P	Snap
Genome	GCA_000005215.1	<i>Drosophila sechellia</i>	Drosophilidae	Diptera	P	Augustus
Genome	GCA_000005135.1	<i>Drosophila erecta</i>	Drosophilidae	Diptera	P	Augustus
Genome	GCA_000005975.1	<i>Drosophila yakuba</i>	Drosophilidae	Diptera	P	Augustus
Genome	GCA_000220665.2	<i>Drosophila ficusphila</i>	Drosophilidae	Diptera	P	Snap
Genome	GCA_000224195.2	<i>Drosophila elegans</i>	Drosophilidae	Diptera	P	Snap
Genome	GCA_000236305.2	<i>Drosophila rhopaloa</i>	Drosophilidae	Diptera	P	Augustus
Genome	GCA_000233415.2	<i>Drosophila biarmipes</i>	Drosophilidae	Diptera	P	Snap

Genome	GCA_000224235.2	<i>Drosophila takahashii</i>	Drosophilidae	Diptera	P	Augustus
Genome	GCA_000269505.2	<i>Drosophila miranda</i>	Drosophilidae	Diptera	P	Augustus
Genome	GCA_000001765.2	<i>Drosophila pseudoobscura</i>	Drosophilidae	Diptera	P	Snap
Genome	GCA_000005195.1	<i>Drosophila persimilis</i>	Drosophilidae	Diptera	P	Snap
Genome	GCA_000298335.1	<i>Drosophila albomicans</i>	Drosophilidae	Diptera	P	Snap
Genome	GCA_000005155.1	<i>Drosophila grimshawi</i>	Drosophilidae	Diptera	P	Snap
Genome	GCA_000005175.1	<i>Drosophila mojavensis</i>	Drosophilidae	Diptera	P	Snap
Genome	GCA_000005245.1	<i>Drosophila virilis</i>	Drosophilidae	Diptera	P	Snap
Genome	GCA_000371365.1	<i>Musca domestica</i>	Muscidae	Diptera	P	Augustus
Genome	GCA_000648655.1	<i>Copidosoma floridanum</i>	Encyrtidae	Hymenoptera	P	Blast
Protein	ABC54566.1	<i>Anopheles gambiae</i>	Culicidae	Diptera	DB	-
Protein	KYN09041	<i>Trachymyrmex cornetzi</i>	Formicidae	Hymenoptera	DB	-
Protein	KYM88541	<i>Atta colombica</i>	Formicidae	Hymenoptera	DB	-
Protein	XP_012060266	<i>Atta cephalotes</i>	Formicidae	Hymenoptera	DB	-
Protein	XP_011057669	<i>Acromyrmex echinator</i>	Formicidae	Hymenoptera	DB	-
Protein	ADM07366	<i>Messor pergandei</i>	Formicidae	Hymenoptera	DB	-
Protein	XP_008556449	<i>Microplitis demolitor</i>	Microgastrinae	Hymenoptera	DB	-
Protein	XP_012229836	<i>Linepithema humile</i>	Dolichoderinae	Hymenoptera	DB	-
Protein	ABC54566	<i>Anopheles gambiae</i>	Culicidae	Diptera	DB	-
Protein	ACB20969	<i>Culex quinquefasciatus</i>	Culicidae	Diptera	DB	-
Protein	ABC41128	<i>Aedes aegypti</i>	Culicidae	Diptera	DB	-
Protein	XP_004529162	<i>Ceratitis capitata</i>	Tephritidae	Diptera	DB	-
Protein	NP_001234884	<i>Nasonia vitripennis</i>	Pteromalidae	Hymenoptera	DB	-
Protein	XM_011167600	<i>Solenopsis invicta</i>	Formicidae	Hymenoptera	DB	-
Protein	JR477371.1	<i>Rhynchophorus errugineus</i>	Curculionidae	Coleoptera	DB	-
Nucleotide	JO902149	<i>Aedes albopictus</i>	Aedes	Diptera	DB	-
Nucleotide	JO874052	<i>Aedes albopictus</i>	Aedes	Diptera	DB	-
Nucleotide	JO885398	<i>Aedes albopictus</i>	Aedes	Diptera	DB	-
1KITE	INSTmbTBGRAAPEI-33	<i>Liposcelis bostrychophila</i>	Liposcelididae	Psocodea	P	Blast
1KITE	INSjdsTABRAAPEI-20	<i>Frankliniella cephalica</i>	Thripidae	Thysanoptera	P	Blast
1KITE	INSjdsTACRAAPEI-21	<i>Thrips palmi</i>	Thripidae	Thysanoptera	P	Blast

1KITE	INShauTABRAAPEI-93	<i>Nemophora degeerella</i>	Adelidae	Lepidoptera	P	Blast
1KITE	INSbttTARAAPEI-9	<i>Platycentropus radiatus</i>	Limnephilidae	Trichoptera	P	Blast
1KITE	INSnfrTALRAAPEI-31	<i>Leuctra sp.</i>	Leuctridae	Plecoptera	P	Blast
1KITE	INShauTAKRAAPEI-90	<i>Baetis pumilus</i>	Baetidae	Ephemeroptera	P	Blast
1KITE	INSjdsTADRAAPEI-22	<i>Gynaikothrips ficorum</i>	Phlaeothripidae	Thysanoptera	P	Blast
1KITE	INStmbTAWRAAPEI-13	<i>Boreus hyemalis</i>	Boreidae	Mecoptera	P	Blast
1KITE	INSyvtCDRAAPEI-35	<i>Cryptocercus wrighti</i>	Cryptocercidae	Blattodea	P	Blast
1KITE	INSnfrTAQRAAPEI-37	<i>Cotesia vestalis</i>	Braconidae	Hymenoptera	P	Blast
1KITE	INSjdsTAURAAPEI-62	<i>Leptopilina clavipes</i>	Figitidae	Hymenoptera	P	Blast

60

61 **Supplementary File 1C: List of sequences and their BLAST results used for phylogenetic analysis of the LOTUS domain.**

62 The sequences were obtained by searching the TrEMBL database using hmmsearch and the final HMM generated for LOTUS
 63 (Supplementary files: HMM>LOTUS.hmm). Reported are the UniProtID (*Accession Number*), the Domain and Phylum origin of the
 64 sequence, the E-value, score and bias given by hmmsearch, and the description of the target from UniProt. To obtain sequences for each
 65 entry, either search UniProt directly (<https://www.uniprot.org/>) or consult the final alignment in Supplementary Files:
 66 Alignments>LOTUS_TREE.fasta. Phylum abbreviations: A = Arthropoda; An = Annelida; E = Echinodermata; F = Firmicutes; M =
 67 Mollusca; T = Tunicata; V = Vertebrata; ? = unclassified
 68

Accession ID	Domain	Phylum	E-value	score	bias	Description of Target
V3Z0B0_LOTGI	Eukarya	M	8.30E-32	120.8	0.1	Uncharacterized protein OS=Lottia gigantea GN=LOTGIDRAFT_236389 PE=4 SV=1
R7UJX3_CAPTE	Eukarya	An	4.00E-30	115.3	0	Uncharacterized protein OS=Capitella teleta GN=CAPTEDRAFT_218952 PE=4 SV=1
E9IZ46_SOLIN	Eukarya	A	2.80E-27	106.2	0.1	Putative uncharacterized protein (Fragment) OS=Solenopsis invicta GN=SINV_01516 PE=4 SV=1
K7JU22_NASVI	Eukarya	A	9.00E-27	104.6	0.2	Uncharacterized protein OS=Nasonia vitripennis GN=oskar PE=4 SV=1
F4WQN7_ACREC	Eukarya	A	1.30E-25	100.9	0.1	Maternal effect protein oskar OS=Acromyrmex echinator GN=G5I_08127 PE=4 SV=1
W4ZBK4_STRPU	Eukarya	E	1.70E-25	100.5	0.1	Uncharacterized protein OS=Strongylocentrotus purpuratus GN=Sp-Tdrd5 PE=4 SV=1
E2A7I8_CAMFO	Eukarya	A	1.90E-25	100.4	2.1	Putative uncharacterized protein OS=Camponotus floridanus GN=EAG_03874 PE=4 SV=1
F2WJY6_9HYME	Eukarya	A	3.40E-25	99.6	0	Oskar (Fragment) OS=Messor pergandei PE=2 SV=1
B4N816_DROWI	Eukarya	A	8.50E-25	98.3	9.6	Uncharacterized protein OS=Drosophila willistoni GN=Dwil\GK11116 PE=4 SV=2
B4K9E5_DROMO	Eukarya	A	8.70E-25	98.2	0.6	Uncharacterized protein OS=Drosophila mojavensis GN=Dmoj\GI10055 PE=4 SV=2
A0A026WMY1_CERBI	Eukarya	A	2.20E-24	97	0.1	Maternal effect protein oskar OS=Cerapachys biroi GN=X777_01612 PE=4 SV=1
C3ZCL9_BRAFL	Eukarya	A	5.40E-24	95.7	0	Putative uncharacterized protein OS=Branchiostoma floridae GN=BRAFLDRAFT_64001 PE=4 SV=1
B4LXK5_DROVI	Eukarya	A	8.00E-24	95.2	0.9	Oskar OS=Drosophila virilis GN=osk PE=4 SV=1
K4MTL4_GRYBI	Eukarya	A	6.70E-23	92.2	0	Oskar OS=Gryllus bimaculatus PE=2 SV=1 GH23955 OS=Drosophila grimshawi GN=Dgri\GH23955 PE=4 SV=1
B4JTJ1_DROGR	Eukarya	A	8.30E-23	91.9	1.4	
A1Y1T7_DROIM	Eukarya	A	8.60E-23	91.9	0.8	Oskar OS=Drosophila immigrans GN=osk PE=4 SV=1
Q2PP79_AEDAE	Eukarya	A	1.70E-22	90.9	0	Oskar OS=Aedes aegypti PE=4 SV=1

W8CE30_CERCA	Eukarya	A	2.00E-22	90.7	0.3	Maternal effect protein oskar OS=Ceratitidis capitata GN=OSKA PE=2 SV=1
T1PG45_MUSDO	Eukarya	A	1.60E-21	87.8	1.2	GDSL-like Lipase/Acylhydrolase OS=Musca domestica PE=2 SV=1
A0A0L8HW18_OCTBM	Eukarya	M	1.70E-21	87.7	7.5	Uncharacterized protein OS=Octopus bimaculoides GN=OCBIM_22005378mg PE=4 SV=1
F6QYS5_XENTR	Eukarya	V	6.10E-20	82.7	0.2	Uncharacterized protein (Fragment) OS=Xenopus tropicalis PE=4 SV=1
B0WIV7_CULQU	Eukarya	A	8.90E-20	82.2	0.1	Oskar OS=Culex quinquefasciatus GN=CpipJ_CPIJ007471 PE=2 SV=1
A0A034WRF5_BACDO	Eukarya	A	1.30E-19	81.6	13.1	Maternal effect protein oskar OS=Bactrocera dorsalis GN=OSKA PE=4 SV=1
Q7PQJ3_ANOGA	Eukarya	A	1.40E-19	81.5	0	AGAP003545-PA OS=Anopheles gambiae GN=osk PE=4 SV=3
X1WIW2_ACYPI	Eukarya	A	1.70E-19	81.3	12.3	Uncharacterized protein OS=Acyrtosiphon pisum GN=LOC100162069 PE=4 SV=1
A0A023EWW9_TRIIF	Eukarya	A	1.80E-19	81.2	0.2	Putative transcriptional coactivator (Fragment) OS=Triatoma infestans PE=2 SV=1
A0A0L0CP24_LUCCU	Eukarya	A	2.40E-19	80.8	0.5	Uncharacterized protein OS=Lucilia cuprina GN=FF38_12727 PE=4 SV=1
A0A0K8W0W3_BACLA	Eukarya	A	5.00E-19	79.8	0	Maternal effect protein oskar OS=Bactrocera latifrons GN=osk_1 PE=4 SV=1
A0A0A1XRQ4_BACCU	Eukarya	A	7.20E-19	79.3	6.4	Maternal effect protein oskar OS=Bactrocera cucurbitae GN=osk PE=4 SV=1
A0A088ASD6_APIME	Eukarya	A	2.90E-18	77.3	0.2	Uncharacterized protein OS=Apis mellifera GN=LOC726241 PE=4 SV=1
K1QZD2_CRAGI	Eukarya	M	1.60E-17	75	0	Tudor domain-containing protein 7 OS=Crassostrea gigas GN=CGI_10018436 PE=4 SV=1
A0A139WGI6_TRICA	Eukarya	A	1.70E-17	74.9	0.1	Tudor domain-containing protein 7-like protein OS=Tribolium castaneum GN=TcasGA2_TC034722 PE=4 SV=1
F6YH90_ORNAN	Eukarya	V	1.80E-17	74.8	0	Uncharacterized protein OS=Ornithorhynchus anatinus GN=TDRD5 PE=4 SV=1
W5JJ85_ANODA	Eukarya	A	2.10E-17	74.6	0	Uncharacterized protein OS=Anopheles darlingi GN=AND_005442 PE=4 SV=1
A0A084WRU4_ANOSI	Eukarya	A	2.40E-17	74.4	0.2	AGAP003545-PA-like protein OS=Anopheles sinensis GN=ZHAS_00021239 PE=4 SV=1
H2YLA8_CIOSA	Eukarya	T	2.60E-17	74.3	0.1	Uncharacterized protein OS=Ciona savignyi GN=Csa.10307 PE=4 SV=1
G1KVT0_ANOCA	Eukarya	V	2.70E-17	74.2	0.1	Uncharacterized protein OS=Anolis carolinensis GN=TDRD5 PE=4 SV=1
G3TEV7_LOXAF	Eukarya	V	4.50E-17	73.5	0	Uncharacterized protein OS=Loxodonta africana GN=TDRD5 PE=4 SV=1
A0A0K8TEH4_LYGHE	Eukarya	A	6.40E-17	73	0	Uncharacterized protein OS=Lygus hesperus PE=4 SV=1

A0A067RPA3_ZOONE	Eukarya	A	7.10E-17	72.9	0.1	Tudor domain-containing protein 7 OS=Zootermopsis nevadensis GN=L798_01728 PE=4 SV=1
G3VEY7_SARHA	Eukarya	V	1.30E-16	72.1	0	Uncharacterized protein OS=Sarcophilus harrisii GN=TDRD5 PE=4 SV=1
L8I7L7_9CETA	Eukarya	V	1.40E-16	72	0.1	Tudor domain-containing protein 5 (Fragment) OS=Bos mutus GN=M91_03486 PE=4 SV=1
W5Q779_SHEEP	Eukarya	V	1.40E-16	71.9	0.5	Uncharacterized protein OS=Ovis aries GN=TDRD5 PE=4 SV=1
G5E528_BOVIN	Eukarya	V	2.30E-16	71.3	0	Tudor domain-containing protein 5 OS=Bos taurus GN=TDRD5 PE=4 SV=1
H2MII4_ORYLA	Eukarya	V	2.90E-16	71	0	Uncharacterized protein OS=Oryzias latipes GN=TDRD7 (1 to many) PE=4 SV=1
N6TQX5_DENPD	Eukarya	A	3.20E-16	70.8	0.1	Uncharacterized protein (Fragment) OS=Dendroctonus ponderosae GN=YQE_11709 PE=4 SV=1
F6WY93_HORSE	Eukarya	V	5.80E-16	70	0	Uncharacterized protein OS=Equus caballus GN=TDRD5 PE=4 SV=1
A0A0J9YJ00_DANRE	Eukarya	V	5.90E-16	69.9	0	Tudor domain-containing protein 7A (Fragment) OS=Danio rerio GN=tdrd7a PE=1 SV=2
U5EFJ8_9DIPT	Eukarya	A	9.50E-16	69.3	0.9	Putative oskar (Fragment) OS=Corethrella appendiculata PE=2 SV=1
H2USX7_TAKRU	Eukarya	V	9.60E-16	69.3	0	Uncharacterized protein OS=Takifugu rubripes PE=4 SV=1
F1S6A1_PIG	Eukarya	V	1.10E-15	69.1	0	Uncharacterized protein OS=Sus scrofa GN=TDRD5 PE=4 SV=2
T1DTM7_ANOAAQ	Eukarya	A	1.10E-15	69	0	Uncharacterized protein (Fragment) OS=Anopheles aquasalis PE=2 SV=1
A0A060W2X9_ONCMY	Eukarya	V	2.20E-15	68.1	0	Uncharacterized protein OS=Oncorhynchus mykiss GN=GSONMT00078733001 PE=4 SV=1
S7NG41_MYOBR	Eukarya	V	2.60E-15	67.9	0.1	Tudor domain-containing protein 5 OS=Myotis brandtii GN=D623_10022817 PE=4 SV=1
F7B4W0_MONDO	Eukarya	V	2.70E-15	67.8	0	Uncharacterized protein OS=Monodelphis domestica GN=TDRD5 PE=4 SV=2
V9KH94_CALMI	Eukarya	V	3.50E-15	67.5	0	Tudor domain-containing protein 5 OS=Callorhynchus milii PE=2 SV=1
A0A0P7YHR6_9TELE	Eukarya	V	4.70E-15	67	0	Uncharacterized protein OS=Scleropages formosus GN=Z043_114704 PE=4 SV=1
G1PFT9_MYOLU	Eukarya	V	5.70E-15	66.8	0.1	Uncharacterized protein OS=Myotis lucifugus GN=TDRD5 PE=4 SV=1
H3DL34_TETNG	Eukarya	V	7.70E-15	66.4	0	Uncharacterized protein OS=Tetraodon nigroviridis PE=4 SV=1
A0A096NXU4_PAPAN	Eukarya	V	8.70E-15	66.2	0	Uncharacterized protein OS=Papio anubis GN=TDRD5 PE=4 SV=1
F7GPY1_CALJA	Eukarya	V	1.10E-14	65.9	0	Uncharacterized protein OS=Callithrix jacchus GN=TDRD5 PE=4 SV=1

W5N030_LEPOC	Eukarya	V	1.20E-14	65.8	0.1	Uncharacterized protein OS=Lepisosteus oculatus PE=4 SV=1
M3VXB3_FELCA	Eukarya	V	1.30E-14	65.6	0	Uncharacterized protein OS=Felis catus GN=TDRD5 PE=4 SV=1
A0A099ZTT8_TINGU	Eukarya	V	1.40E-14	65.6	0	Tudor domain-containing protein 7 OS=Tinamus guttatus GN=N309_12928 PE=4 SV=1
G3R6R4_GORGO	Eukarya	V	1.50E-14	65.5	0	Uncharacterized protein OS=Gorilla gorilla gorilla GN=TDRD5 PE=4 SV=1
L9L889_TUPCH	Eukarya	V	1.60E-14	65.4	0.3	Tudor domain-containing protein 5 OS=Tupaia chinensis GN=TREES_T100015801 PE=4 SV=1
A0A0J7KVQ7_LASNI	Eukarya	A	1.70E-14	65.3	0	Tudor domain-containing protein 5 (Fragment) OS=Lasius niger GN=RF55_5458 PE=4 SV=1
E0VJL4_PEDHC	Eukarya	A	1.70E-14	65.3	0.9	Putative uncharacterized protein OS=Pediculus humanus subsp. corporis GN=Phum_PHUM247930 PE=4 SV=1
A0A0D9RID1_CHLSB	Eukarya	V	1.70E-14	65.2	0	Uncharacterized protein OS=Chlorocebus sabaeus GN=TDRD5 PE=4 SV=1
E2BFZ8_HARSA	Eukarya	A	2.00E-14	65	0.9	Tudor domain-containing protein 7 OS=Harpegnathos saltator GN=EAI_14615 PE=4 SV=1
A0A0T6AU98_9SCAR	Eukarya	A	2.10E-14	65	3.9	Uncharacterized protein (Fragment) OS=Oryctes borbonicus GN=AMK59_7658 PE=4 SV=1
R9PXP1_CHICK	Eukarya	V	2.30E-14	64.8	0	Tudor domain-containing protein 7 OS=Gallus gallus GN=TDRD7 PE=4 SV=1
F7CN93_MACMU	Eukarya	V	2.70E-14	64.6	0	Uncharacterized protein OS=Macaca mulatta GN=TDRD5 PE=4 SV=1
A0A0A0MPC8_CANLF	Eukarya	V	2.70E-14	64.6	0	Tudor domain-containing protein 5 OS=Canis lupus familiaris GN=TDRD5 PE=4 SV=1
M3Y1J3_MUSPF	Eukarya	V	2.90E-14	64.5	0	Uncharacterized protein OS=Mustela putorius furo GN=TDRD5 PE=4 SV=1
H2N4J0_PONAB	Eukarya	V	3.00E-14	64.5	0	Uncharacterized protein OS=Pongo abelii GN=TDRD5 PE=4 SV=1
H2Q0P6_PANTR	Eukarya	V	3.20E-14	64.4	0	Uncharacterized protein OS=Pan troglodytes GN=TDRD5 PE=4 SV=1
W5LEX2_ASTMX	Eukarya	V	3.50E-14	64.3	0	Uncharacterized protein OS=Astyanax mexicanus PE=4 SV=1
A0A024R910_HUMAN	Eukarya	V	4.00E-14	64.1	0	Tudor domain containing 5, isoform CRA_b OS=Homo sapiens GN=TDRD5 PE=4 SV=1
A0A0P6JFX9_HETGA	Eukarya	V	4.10E-14	64.1	0	Tudor domain-containing protein 5 isoform 2 OS=Heterocephalus glaber GN=TDRD5 PE=4 SV=1
H0WXJ6_OTOGA	Eukarya	V	4.70E-14	63.8	0	Uncharacterized protein OS>Otolemur garnettii GN=TDRD5 PE=4 SV=1
H0V001_CAVPO	Eukarya	V	5.20E-14	63.7	0	Uncharacterized protein OS=Cavia porcellus GN=TDRD5 PE=4 SV=1
G7NU06_MACFA	Eukarya	V	5.80E-14	63.6	0	Putative uncharacterized protein OS=Macaca fascicularis GN=EGM_01633 PE=4 SV=1

G1S4T3_NOMLE	Eukarya	V	6.40E-14	63.4	0	Uncharacterized protein OS=Nomascus leucogenys GN=TDRD7 PE=4 SV=1
G1M861_AILME	Eukarya	V	6.80E-14	63.3	0	Tudor domain-containing protein 5 OS=Ailuropoda melanoleuca GN=TDRD5 PE=4 SV=1
A0A061HYN9_CRIGR	Eukarya	V	7.10E-14	63.3	0	Tudor domain-containing protein 5 OS=Cricetulus griseus GN=H671_5g14992 PE=4 SV=1
A0A0F8AWR5_LARCR	Eukarya	V	1.00E-13	62.7	0	Tudor domain-containing protein 7A OS=Larimichthys crocea GN=EH28_10800 PE=4 SV=1
V5GPP4_ANOGL	Eukarya	A	1.10E-13	62.6	1.4	Tudor domain-containing protein (Fragment) OS=Anoplophora glabripennis GN=TDRD7 PE=4 SV=1
A0A087XQP1_POEFO	Eukarya	V	1.10E-13	62.6	0	Uncharacterized protein OS=Poecilia formosa PE=4 SV=2
S4NW46_9NEOP	Eukarya	A	1.20E-13	62.5	0	Tudor domain containing 7 (Fragment) OS=Pararge aegeria PE=4 SV=1
L5KMV4_PTEAL	Eukarya	V	2.10E-13	61.8	0	Tudor domain-containing protein 7 OS=Pteropus alecto GN=PAL_GLEAN10008027 PE=4 SV=1
I3NCP3 ICTTR	Eukarya	V	2.40E-13	61.6	0	Uncharacterized protein OS=Ictidomys tridecemlineatus GN=TDRD7 PE=4 SV=1
G1SCH8_RABIT	Eukarya	V	2.50E-13	61.5	0	Uncharacterized protein OS=Oryctolagus cuniculus GN=TDRD7 PE=4 SV=1
Q3TTK4_MOUSE	Eukarya	V	2.50E-13	61.5	0	Putative uncharacterized protein (Fragment) OS=Mus musculus GN=Tdrd7 PE=2 SV=1
A0A0H2UHC6_RAT	Eukarya	V	3.90E-13	60.9	0	Tudor domain-containing protein 5 OS=Rattus norvegicus GN=Tdrd5 PE=4 SV=1
U6CRG5_NEOVI	Eukarya	V	3.90E-13	60.9	0	Tudor domain-containing protein 7 OS=Neovison vison GN=TDRD7 PE=2 SV=1
K7FVR3_PELSI	Eukarya	V	5.10E-13	60.5	0	Uncharacterized protein OS=Pelodiscus sinensis GN=TDRD5 PE=4 SV=1
J4K9P6_9FIRM	Bacteria	F	2.90E-07	42.1	31.8	NYN domain protein OS=Peptostreptococcaceae bacterium AS15 GN=HMPREF1142_1162 PE=4 SV=1
E0QL58_9FIRM	Bacteria	F	0.00014	33.5	29.1	Uncharacterized protein OS=[Eubacterium] yurii subsp. margaretiae ATCC 43715 GN=HMPREF0379_1756 PE=4 SV=1
A0A0G0IVI6_9BACT	Bacteria	?	0.0025	29.5	1.9	Uncharacterized protein OS=candidate division TM6 bacterium GW2011_GWA2_36_9 GN=US32_C0002G0021 PE=4 SV=1

69
70

71 **Supplementary File 1D: List of sequences and their BLAST results used for phylogenetic analysis of the OSK domain.**

72 The sequences were obtained by searching the TrEMBL database using hmmsearch and the final HMM generated for OSK (Supplementary
 73 files: HMM>OSK.hmm). Reported parameters are as described for Supplementary Table S3. To obtain sequences for each entry, either
 74 search UniProt directly (<https://www.uniprot.org/>) or consult the final alignment in Supplementary Files: Alignments>OSK_TREE.fasta.
 75 Phylum Abbreviations: A = Arthropoda; Ar = Archaea; As = Ascomycota; B = Bacteroidetes; C = Cyanobacteria; Eu = Euryarchaeota; F =
 76 Firmicutes; Fu = Fungi; P = Proteobacteria

77

Accession Number	Domain	Phylum	E-value	Score	Bias	Description of target
A1Y1T7_DROIM	Eukarya	A	2.90E-34	128.4	0.6	Oskar OS=Drosophila immigrans GN=osk PE=4 SV=1
F2WJY6_9HYME	Eukarya	A	3.50E-34	128.2	0.9	Oskar (Fragment) OS=Messor pergandei PE=2 SV=1
B4JTJ1_DROGR	Eukarya	A	5.00E-34	127.6	0.1	GH23955 OS=Drosophila grimshawi GN=Dgr\GH23955 PE=4 SV=1 Maternal effect protein oskar OS=Acromyrmex echinator
F4WQN7_ACREC	Eukarya	A	8.10E-34	127	0.9	GN=G5I_08127 PE=4 SV=1 Putative uncharacterized protein OS=Camponotus floridanus
E2A7I8_CAMFO	Eukarya	A	2.80E-33	125.2	0.4	GN=EAG_03874 PE=4 SV=1 Maternal effect protein oskar OS=Ceratitis capitata GN=OSKA PE=2
W8CE30_CERCA	Eukarya	A	5.80E-33	124.2	0.3	SV=1 Maternal effect protein oskar OS=Bactrocera cucurbitae GN=osk
A0A0A1XRQ4_BACCU	Eukarya	A	7.50E-33	123.9	0.4	PE=4 SV=1 Uncharacterized protein OS=Drosophila willistoni GN=Dwi\GK11117
B4N815_DROWI	Eukarya	A	8.10E-33	123.8	0.3	PE=4 SV=2 Maternal effect protein oskar OS=Cerapachys biroii GN=X777_01612
A0A026WMY1_CERBI	Eukarya	A	2.40E-32	122.3	0.7	PE=4 SV=1 Putative uncharacterized protein (Fragment) OS=Solenopsis invicta
E9IZ46_SOLIN	Eukarya	A	2.50E-32	122.2	0.1	GN=SINV_01516 PE=4 SV=1 Maternal effect protein oskar OS=Bactrocera dorsalis GN=OSKA
A0A034WRF5_BACDO	Eukarya	A	2.70E-32	122.1	0.5	PE=4 SV=1 Maternal effect protein oskar OS=Bactrocera latifrons GN=osk_2
A0A0K8U7J3_BACLA	Eukarya	A	5.70E-32	121	0.5	PE=4 SV=1
Q2PP79_AEDAE	Eukarya	A	6.30E-32	120.9	0.1	Oskar OS=Aedes aegypti PE=4 SV=1 Uncharacterized protein OS=Drosophila mojavensis
B4K9E5_DROMO	Eukarya	A	8.60E-32	120.5	0.1	GN=Dmoj\GI10055 PE=4 SV=2
B4LXK5_DROVI	Eukarya	A	1.20E-31	120	0.1	Oskar OS=Drosophila virilis GN=osk PE=4 SV=1
A0A0M4F3M8_DROBS	Eukarya	A	1.80E-31	119.4	0.4	Osk OS=Drosophila busckii GN=Dbus_chr3Rg607 PE=4 SV=1 Uncharacterized protein OS=Lucilia cuprina GN=FF38_12727 PE=4
A0A0L0CP24_LUCCU	Eukarya	A	2.90E-31	118.8	1.5	SV=1 Uncharacterized protein OS=Drosophila yakuba GN=Dyak\GE25914
B4PTX6_DROYA	Eukarya	A	1.00E-30	117	0.5	PE=4 SV=1

B3P1W4_DROER	Eukarya	A	2.70E-30	115.6	0.5	GG13545 OS=Drosophila erecta GN=Dere\GG13545 PE=4 SV=1
T1PG45_MUSDO	Eukarya	A	4.20E-30	115.1	0.7	GDSL-like Lipase/Acylhydrolase OS=Musca domestica PE=2 SV=1
B4HKZ1_DROSE	Eukarya	A	5.20E-30	114.8	0.2	GM23770 OS=Drosophila sechellia GN=Dsec\GM23770 PE=4 SV=1
Q295Q4_DROPS	Eukarya	A	6.40E-30	114.5	0.2	Uncharacterized protein, isoform A OS=Drosophila pseudoobscura pseudoobscura GN=Dpse\GA10627 PE=4 SV=2
E8NH25_DROME	Eukarya	A	7.30E-30	114.3	0.2	RE24380p (Fragment) OS=Drosophila melanogaster GN=osk-RA PE=2 SV=1
T1DTM7_ANO AQ	Eukarya	A	1.00E-29	113.8	0	Uncharacterized protein (Fragment) OS=Anopheles aquasalis PE=2 SV=1
B3LZ06_DROAN	Eukarya	A	2.20E-28	109.5	0.1	Uncharacterized protein OS=Drosophila ananassae GN=Dana\GF17692 PE=4 SV=1
E1A883_NASVI	Eukarya	A	4.20E-28	108.6	0.2	Oskar OS=Nasonia vitripennis PE=2 SV=1
A0A084WRU4_ANOSI	Eukarya	A	1.30E-27	107.1	0.1	AGAP003545-PA-like protein OS=Anopheles sinensis GN=ZHAS_00021239 PE=4 SV=1
W5JJ85_ANODA	Eukarya	A	4.50E-27	105.3	0	Uncharacterized protein OS=Anopheles darlingi GN=AND_005442 PE=4 SV=1
Q7PQJ3_ANOGA	Eukarya	A	6.00E-27	104.9	0	AGAP003545-PA OS=Anopheles gambiae GN=osk PE=4 SV=3
B0WIV7_CULQU	Eukarya	A	1.30E-26	103.9	0.4	Oskar OS=Culex quinquefasciatus GN=CpipJ_CPIJ007471 PE=2 SV=1
U5EFJ8_9DIPT	Eukarya	A	4.00E-25	99.1	0	Putative oskar (Fragment) OS=Corethrella appendiculata PE=2 SV=1
A0A126GUR4_DROME	Eukarya	A	3.10E-24	96.2	0.3	Oskar, isoform D OS=Drosophila melanogaster GN=osk PE=4 SV=1
A0A059PGF2_9MUSC	Eukarya	A	3.10E-24	96.2	0.5	GA10627 (Fragment) OS=Drosophila pseudoobscura GN=GA10627 PE=4 SV=1
K4MTL4_GRYBI	Eukarya	A	1.20E-23	94.2	0.2	Oskar OS=Gryllus bimaculatus PE=2 SV=1
A0A0C9QHR7_9HYME	Eukarya	A	5.40E-22	89	0	Osk protein OS=Fopius arisanus GN=osk PE=4 SV=1
A0A059PF64_9MUSC	Eukarya	A	1.90E-19	80.8	0.7	GA10627 (Fragment) OS=Drosophila pseudoobscura GN=GA10627 PE=4 SV=1
A0A0J7KH44_LASNI	Eukarya	A	1.40E-13	61.9	0	Maternal effect protein oskar (Fragment) OS=Lasius niger GN=RF55_10783 PE=4 SV=1
E2BYH0_HARSA	Eukarya	A	4.00E-12	57.3	0	Putative uncharacterized protein OS=Harpegnathos saltator GN=EAI_08923 PE=4 SV=1
A0A0J6BBM7_BREBE	Bacteria	F	3.10E-10	51.2	0	Lysophospholipase OS=Brevibacillus brevis GN=AB432_04505 PE=4 SV=1
A0A0H0SJ00_9BACL	Bacteria	F	9.60E-10	49.6	0	Lysophospholipase OS=Brevibacillus formosus GN=AA984_15375 PE=4 SV=1
E9I8K8_SOLIN	Bacteria	A	5.90E-09	47.1	1	Putative uncharacterized protein (Fragment) OS=Solenopsis invicta GN=SINV_16199 PE=4 SV=1
J3A568_9BACL	Bacteria	F	5.80E-08	43.9	0	Lysophospholipase L1-like esterase OS=Brevibacillus sp. BC25 GN=PMI05_03395 PE=4 SV=1

G2HS43_9PROT	Bacteria	P	1.10E-07	43	1	Lipolytic protein OS=Arcobacter sp. L GN=ABLL_2651 PE=4 SV=1
E6L4E3_9PROT	Bacteria	P	1.70E-07	42.4	0.7	Lipolytic protein OS=Arcobacter butzleri JV22 GN=HMPREF9401_1319 PE=4 SV=1
R5GT16_9FIRM	Bacteria	F	2.60E-07	41.8	0	GDSL-like protein OS=Eubacterium sp. CAG:786 GN=BN782_00012 PE=4 SV=1
A0A078KJ49_9FIRM	Bacteria	F	3.40E-07	41.4	0	Uncharacterized protein OS=[Clostridium] cellulosi GN=CCDG5_0508 PE=4 SV=1
A8EWS4_ARCB4	Bacteria	P	3.80E-07	41.3	0.7	Lipolytic enzyme, GDSL domain OS=Arcobacter butzleri (strain RM4018) GN=Abu_2183 PE=4 SV=1
S5PEQ8_9PROT	Bacteria	P	4.00E-07	41.2	0.4	Lipolytic enzyme, GDSL domain protein OS=Arcobacter butzleri 7h1h GN=A7H1H_2115 PE=4 SV=1
A0A0G9K3L5_9PROT	Bacteria	P	4.00E-07	41.2	0.6	Lipolytic protein OS=Arcobacter butzleri L348 GN=AA20_04280 PE=4 SV=1
R6T7B3_9BACE	Bacteria	B	4.80E-07	41	0	GDSL-like protein OS=Bacteroides sp. CAG:770 GN=BN777_00744 PE=4 SV=1
A0A0A6S1U2_STRUB	Bacteria	F	4.90E-07	40.9	0.2	Acylneuraminate cytidyltransferase OS=Streptococcus uberis GN=NC01_08240 PE=4 SV=1
W3AC50_9BACL	Bacteria	F	5.50E-07	40.8	0.3	Uncharacterized protein OS=Planomicrobium glaciei CHR43 GN=G159_16940 PE=4 SV=1
A0A0M1UPT0_9PROT	Bacteria	P	5.50E-07	40.8	0.4	Lipolytic protein OS=Arcobacter butzleri ED-1 GN=ABED_1978 PE=4 SV=1
U6EVC2_CLOTA	Bacteria	F	6.30E-07	40.6	0.7	Platelet activating factor acetylhydrolase-like protein OS=Clostridium tetani 12124569 GN=BN906_00617 PE=4 SV=1
A0A098EIL2_9BACL	Bacteria	F	7.90E-07	40.3	0.1	Multifunctional acyl-CoA thioesterase I and protease I and lysophospholipase L1 OS=Planomicrobium sp. ES2 GN=BN1080_01055 PE=4 SV=1
A0A069S7Q5_9PORP	Bacteria	B	1.10E-06	39.9	0	Uncharacterized protein OS=Parabacteroides distasonis str. 3776 Po2 i GN=M090_4091 PE=4 SV=1
Q897X6_CLOTE	Bacteria	F	1.20E-06	39.7	0.8	Platelet activating factor acetylhydrolase-like protein OS=Clostridium tetani (strain Massachusetts / E88) GN=CTC_00594 PE=4 SV=1
A0A0J9FZD4_9PORP	Bacteria	B	1.30E-06	39.6	0	Uncharacterized protein OS=Parabacteroides sp. D26 GN=HMPREF1000_00856 PE=4 SV=1
K9WJ28_9CYAN	Bacteria	C	1.60E-06	39.3	1.2	Lysophospholipase L1-like esterase OS=Microcoleus sp. PCC 7113 GN=Mic7113_4114 PE=4 SV=1
A0A0G1KN57_9BACT	Bacteria	?	3.70E-06	38.1	0	Secreted protein OS=candidate division WWE3 bacterium GW2011_GWC2_44_9 GN=UW82_C0006G0015 PE=4 SV=1
R5VMZ1_9FIRM	Bacteria	F	5.00E-06	37.7	0.1	GDSL-like protein OS=Firmicutes bacterium CAG:631 GN=BN742_01282 PE=4 SV=1
A0A0L0WAR6_CLOPU	Bacteria	F	7.20E-06	37.2	0.4	Lysophospholipase L1 OS=Clostridium purinilyticum GN=CLPU_6c00720 PE=4 SV=1
R5S0B3_9BACE	Bacteria	B	7.30E-06	37.2	0	GDSL-like protein OS=Bacteroides sp. CAG:545 GN=BN702_00435 PE=4 SV=1

A0A073IAZ3_9PORP	Bacteria	B	7.60E-06	37.1	0	Uncharacterized protein OS=Porphyromonas sp. 31_2 GN=HMPREF1002_01104 PE=4 SV=1
R4JC30_9BACT	Bacteria	?	7.90E-06	37.1	0.1	Uncharacterized protein OS=uncultured bacterium BAC25G1 GN=metaSSY_00600 PE=4 SV=1
A0A0C1UEU0_9CLOT	Bacteria	F	7.90E-06	37	0.8	GDSL-like Lipase/Acylhydrolase family protein OS=Clostridium argentinense CDC 2741 GN=U732_2423 PE=4 SV=1
R5LL12_9FIRM	Bacteria	F	1.00E-05	36.7	0	GDSL-like protein OS=Eubacterium sp. CAG:115 GN=BN470_02036 PE=4 SV=1
A0A0E3QN36_METBA	Archaea	Eu	1.20E-05	36.5	0.2	Putative tesA-like protease OS=Methanosarcina barkeri str. Wiesmoor GN=MSBRW_2234 PE=4 SV=1
B7KKA6_CYAP7	Bacteria	C	1.60E-05	36.1	0.3	Lipolytic protein G-D-S-L family OS=Cyanotheca sp. (strain PCC 7424) GN=PCC7424_2577 PE=4 SV=1
R2P1Z4_9ENTE	Bacteria	F	1.60E-05	36	0.1	Uncharacterized protein OS=Enterococcus raffinosus ATCC 49464 GN=UAK_02837 PE=4 SV=1
A0A072Y8N5_9CLOT	Bacteria	F	1.80E-05	35.9	1.5	Acetylhydrolase OS=Clostridium sp. K25 GN=Z957_08245 PE=4 SV=1
K8GFE2_9CYAN	Bacteria	C	2.00E-05	35.8	0.1	Lysophospholipase L1-like esterase OS=Oscillatoriales cyanobacterium JSC-12 GN=OscopyDRAFT_3941 PE=4 SV=1
A0A095ZDI3_9FIRM	Bacteria	F	2.00E-05	35.8	0.3	Uncharacterized protein OS=Tissierellia bacterium S7-1-4 GN=HMPREF1634_08565 PE=4 SV=1
E1YW67_9PORP	Bacteria	B	2.20E-05	35.6	0	GDSL-like protein OS=Parabacteroides sp. 20_3 GN=HMPREF9008_00759 PE=4 SV=1
A0A0M0WNM0_9BACI	Bacteria	F	2.60E-05	35.4	0.4	Lipase OS=Bacillus sp. FJAT-21351 GN=AMS61_13120 PE=4 SV=1
A0A0G3CFD7_METBA	Archaea	Eu	2.80E-05	35.3	0.1	GDSL family lipase/acylhydrolase OS=Methanosarcina barkeri CM1 GN=MCM1_0752 PE=4 SV=1
F6DQC0_DESRL	Bacteria	F	3.40E-05	35	0	Lipolytic protein G-D-S-L family OS=Desulfotomaculum ruminis (strain ATCC 23193 / DSM 2154 / NCIB 8452 / DL) GN=Desru_1430 PE=4 SV=1
D5DGY9_BACMD	Bacteria	F	3.50E-05	35	0.5	Lipase/Acylhydrolase (GDSL) OS=Bacillus megaterium (strain DSM 319) GN=BMD_3140 PE=4 SV=1
A5N8N5_CLOK5	Bacteria	F	3.70E-05	34.9	1.6	Uncharacterized protein OS=Clostridium kluyveri (strain ATCC 8527 / DSM 555 / NCIMB 10680) GN=CKL_1624 PE=4 SV=1
R7ADB4_9BACE	Bacteria	F	3.90E-05	34.8	0.2	Uncharacterized protein OS=Bacteroides pectinophilus CAG:437 GN=BN656_00903 PE=4 SV=1
A0A0H1UPZ9_STRAG	Bacteria	F	4.30E-05	34.7	0.2	Acylneuraminate cytidyltransferase OS=Streptococcus agalactiae GN=WA03_09270 PE=4 SV=1
A0A094AE00_9PEZI	Eukarya	As	0.00036	31.7	0.1	Uncharacterized protein (Fragment) OS=Pseudogymnoascus sp. VKM F-4281 (FW-2241) GN=V493_03380 PE=4 SV=1
A0A0L1HYX4_9PLEO	Eukarya	As	0.00066	30.9	0	Carbohydrate esterase family 3 protein OS=Stemphylium lycopersici GN=TW65_91054 PE=4 SV=1

G2QGB0_MYCTT	Eukarya	As	0.00074	30.7	0.1	Carbohydrate esterase family 3 protein OS=Myceliophthora thermophila (strain ATCC 42464 / BCRC 31852 / DSM 1799) GN=MYCTH_53698 PE=4 SV=1
E3RJZ5_PYRTT	Eukarya	As	0.0014	29.8	0	Putative uncharacterized protein OS=Pyrenophora teres f. teres (strain 0-1) GN=PTT_08513 PE=4 SV=1
G2QVW9_THITE	Eukarya	As	0.0066	27.7	0.1	Carbohydrate esterase family 3 protein OS=Thielavia terrestris (strain ATCC 38088 / NRRL 8126) GN=THITE_2042744 PE=4 SV=1
G0S9F4_CHATD	Eukarya	As	0.01	27.1	0	Putative uncharacterized protein OS=Chaetomium thermophilum (strain DSM 1495 / CBS 144.50 / IMI 039719) GN=CTHT_0045680 PE=4 SV=1

78

79 **Supplementary File 1E: List of genomes analyzed for codon use.**

80 This table lists the 17 genomes that were downloaded and analyzed for codon use as described in
 81 “*Selection of sequences for codon use analysis*” in Methods. All genomes can be downloaded from
 82 <https://www.ncbi.nlm.nih.gov/genome/browse#!/overview/>. The table lists the species name (*Species*),
 83 family (*Family*) and Order (*Order*), NCBI genome accession number (*Genome ID*), and the *oskar*
 84 NCBI Nucleotide accession number (*oskar Nucleotide ID*).
 85

Species	Family	Order	Genome ID	Oskar Nucleotide ID††
<i>Drosophila melanogaster</i>	Drosophilidae	Diptera	GCA_001014345.1	NM_169248.4
<i>Nasonia vitripennis</i>	Pteromalidae	Hymenoptera	GCA_000002325.2	HM535628.1
<i>Culex quinquefasciatus</i>	Culicidae	Diptera	GCA_000209185.1	EU517695.1
<i>Drosophila virilis</i>	Drosophilidae	Diptera	GCA_000005245.1	L22556.1
<i>Ceratitis capitata</i>	Tephritidae	Diptera	GCA_000347755.4	LOC101450245
<i>Musca domestica</i>	Muscidae	Diptera	GCA_000371365.1	LOC101890691
<i>Acromyrmex echinatior</i>	Formicidae	Hymenoptera	GCA_000204515.1	LOC105147973
<i>Harpegnathos saltator</i>	Formicidae	Hymenoptera	GCA_000147195.1	LOC105187957
<i>Bactrocera dorsalis</i>	Tephritidae	Diptera	GCA_000789215.2	LOC105232054
<i>Fopius arisanus</i>	Braconidae	Hymenoptera	GCA_000806365.1	LOC105267990
<i>Athalia rosae</i>	Tenthredinidae	Hymenoptera	GCA_000344095.2	LOC105692731
<i>Orussus abietinus</i>	Orussidae	Hymenoptera	GCA_000612105.2	LOC105696794
<i>Stomoxys calcitrans</i>	Muscidae	Diptera	GCA_001015335.1	LOC106086381
<i>Bactrocera oleae</i>	Tephritidae	Diptera	GCA_001188975.2	LOC106622417
<i>Copidosoma floridanum</i>	Encyrtidae	Hymenoptera	GCA_000648655.2	LOC106642594
<i>Polistes canadensis</i>	Vespidae	Hymenoptera	GCA_001313835.1	LOC106790143
<i>Neodiprion lecontei</i>	Diprionidae	Hymenoptera	GCF_001263575.1	LOC107223453

86