

Patterns of molecular evolution of the germ line specification gene *oskar* suggest that a novel domain may contribute to functional divergence in *Drosophila*

Abha Ahuja · Cassandra G. Extavour

Received: 10 September 2013 / Accepted: 7 December 2013
© Springer-Verlag Berlin Heidelberg 2014

Abstract In several metazoans including flies of the genus *Drosophila*, germ line specification occurs through the inheritance of maternally deposited cytoplasmic determinants, collectively called germ plasm. The novel insect gene *oskar* is at the top of the *Drosophila* germ line specification pathway, and also plays an important role in posterior patterning. A novel N-terminal domain of *oskar* (the Long Oskar domain) evolved in Drosophilids, but the role of this domain in *oskar* functional evolution is unknown. Trans-species transgenesis experiments have shown that *oskar* orthologs from different *Drosophila* species have functionally diverged, but the underlying selective pressures and molecular changes have not been investigated. As a first step toward understanding how Oskar function could have evolved, we applied molecular evolution analysis to *oskar* sequences from the completely sequenced genomes of 16 *Drosophila* species from the *Sophophora* subgenus, *Drosophila virilis* and *Drosophila immigrans*. We show that overall, this gene is subject to purifying selection, but that individual predicted structural and functional domains are subject to heterogeneous selection pressures. Specifically, two domains, the *Drosophila*-specific Long Osk domain and the region that interacts with the germ plasm protein Lasp, are

evolving at a faster rate than other regions of *oskar*. Further, we provide evidence that positive selection may have acted on specific sites within these two domains on the *D. virilis* branch. Our domain-based analysis suggests that changes in the Long Osk and Lasp-binding domains are strong candidates for the molecular basis of functional divergence between the Oskar proteins of *D. melanogaster* and *D. virilis*. This molecular evolutionary analysis thus represents an important step towards understanding the role of an evolutionarily and developmentally critical gene in germ plasm evolution and assembly.

Keywords *Drosophila* · Positive selection · Oskar · Germline specification · Germ plasm · Novelty

Introduction

The advent of a dedicated germ line is a major evolutionary transition associated with the origin of multicellularity (Michod 2005). In all sexually reproducing animals, the specification of the germ line early in embryogenesis is a critical developmental event. Two modes of germ line specification have been identified in metazoans: inheritance of maternally synthesized cytoplasmic germ line determinants (germ plasm), and the induction of germ cell fate by signals from neighboring somatic cells (Extavour and Akam 2003). Phylogenetic analyses of these developmental patterns suggest that the inductive mode may be ancestral in metazoans, with germ plasm-driven mechanisms having evolved independently in multiple lineages (Extavour and Akam 2003; Blackstone and Jasker 2003). In *Drosophila melanogaster*, germ cells are specified by the inheritance mode, and germ plasm is assembled during oogenesis by the products of the *oskar* gene. Oskar protein physically interacts with and recruits germ plasm components, including Valois, Lasp, and Vasa proteins

Communicated by: Claude Desplan

Electronic supplementary material The online version of this article (doi:10.1007/s00427-013-0463-7) contains supplementary material, which is available to authorized users.

A. Ahuja (✉) · C. G. Extavour (✉)
Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, MA 02138, USA
e-mail: abha_ahuja@hms.harvard.edu
e-mail: extavour@oeb.harvard.edu

Present Address:
A. Ahuja
Curriculum Fellows Program, Department of Cell Biology, Harvard Medical School, Boston, MA, USA

and *nanos* mRNA (Ephrussi et al. 1991; Breitwieser et al. 1996; Suyama et al. 2009; Cavey et al. 2005). *oskar* is necessary and sufficient for germ plasm assembly (Ephrussi and Lehmann 1992), and because it localizes the posterior determinant *nanos*, it is also required for posterior body patterning (Ephrussi et al. 1991).

Surprisingly, in contrast to most other critical metazoan germ line genes, *oskar* is not highly conserved across animals. It is instead a novel gene that evolved in the lineage leading to insects (Ewen-Campen et al. 2012), and may have facilitated the evolution of germ plasm in holometabolous insects (those that undergo true metamorphosis) (Lynch et al. 2011). *oskar* orthologs have been identified to date only from flies and mosquitoes (Diptera, the furthest derived clade of holometabolous insects) (Goltsev et al. 2004; Juhn and James 2006; Juhn et al. 2008), ants and wasps (Hymenoptera, the most basally branching clade of the Holometabola) (Lynch et al. 2011), and a basally branching hemimetabolous insect, the cricket *Gryllus bimaculatus* (Ewen-Campen et al. 2012). *oskar* is absent from the genomes of multiple holometabolous insects that lack germ plasm, including the silk moth *Bombyx mori*, the beetle *Tribolium castaneum*, and the honeybee *Apis mellifera*, indicating that this gene has been secondarily lost several times in holometabolous evolution (Lynch et al. 2011).

The domain organization of *oskar* (Fig. 1a) further underscores the dynamic evolutionary history of this gene. Three domains are common to all known *oskar* orthologs. The first is a predicted N-terminal RNA-binding LOTUS domain, which is also present in the highly conserved Tudor domain family members Tdrd5 and Tdrd7 (Callebaut and Mornon 2010; Anantharaman et al. 2010). However, *oskar* lacks Tudor domains, and is thus not a subclass of Tudor family genes (Lynch et al. 2011). The second is a C-terminal domain that shares greatest sequence and physicochemical similarity to SGNH class hydrolases, an ancient group of lipid-interacting proteins present in all kingdoms of life (Juhn et al. 2008). Finally, part of the region between the LOTUS and SGNH domains has been shown to interact with the actin-binding protein Lasp in *D. melanogaster* (Suyama et al. 2009). In contrast to these three relatively conserved domains, the N-terminal-most domain, called Long Osk, is completely absent in mosquito, hymenopteran, and cricket *oskar* orthologs, and is thus likely an innovation that arose at some point after the divergence of the lineages leading to *Drosophila* and mosquitoes (approximately 260 Mya) (Gaunt and Miles 2002).

Even within Drosophilids, *oskar* has diverged functionally (Fig. 2). In some cases, *oskar* function appears highly conserved. For example, the *oskar* ortholog from *D. immigrans* (*immosk*), which diverged from the *D. melanogaster* lineage 30–60 Mya (Remsen and O’Grady 2002; Obbard et al. 2012), can rescue germ cell and body patterning defects of *D. melanogaster* *oskar* null mutants (Jones and Macdonald 2007). *D. melanogaster* *oskar* null flies carrying an *immosk*

transgene display pole plasm morphology more similar to that of *D. immigrans* than that of *D. melanogaster*, but this pole plasm still contains the conserved germ plasm component Aubergine and is sufficient to form germ cells (Jones and Macdonald 2007). This indicates that although *immosk* may function differently from *D. melanogaster* *oskar* with respect to the specific morphology of the germ plasm that it confers, *immosk*-like germ plasm is still sufficient to induce germ cell formation in *D. melanogaster*, indicating essential functional conservation between *immosk* and *D. melanogaster* *oskar* with respect to a role in germ cell formation.

In contrast, *oskar* from the equally distantly related species *Drosophila virilis* (*virosk*) does not show functional conservation of its germ plasm role in a *D. melanogaster* context. In *D. virilis*, *virosk* transcript is localized to the posterior pole of oocytes and embryos like its homolog in *D. melanogaster* (Webster et al. 1994). *D. virilis* embryos also form posterior germ plasm and subsequently pole cells (Webster et al. 1994), suggesting that the germ plasm function of *oskar* is conserved in *D. virilis*, as in *D. immigrans*. In transgenic *D. melanogaster* *oskar* loss of function mutants carrying a *virosk* transgene, Virosk appears to recruit sufficient *D. melanogaster* *nanos* mRNA to rescue posterior patterning in these mutants (Webster et al. 1994). However, in *D. melanogaster*, Virosk is unable to assemble functional germ plasm, and thus unable to direct germ cell formation (Webster et al. 1994). This suggests that although Virosk may retain some ability to interact with *D. melanogaster* *nanos* mRNA, its interactions with other *D. melanogaster* germ plasm components may be too divergent to permit assembly of functional pole plasm in a *D. melanogaster* context, indicating essential functional divergence. Given that *oskar*’s role in functional germ plasm assembly appears conserved even in Hymenoptera (Lynch et al. 2011), it is likely that *virosk* and *immosk* direct functional germ plasm assembly in *D. virilis* and *D. immigrans* respectively, via interactions with the germ plasm component orthologs in these species. In this paper, we focus on the functional divergence within the genus *Drosophila* that prevent Virosk’s fruitful interaction with *D. melanogaster* germ plasm gene products, despite the high level of conservation of most other germ plasm genes (Ewen-Campen et al. 2010).

Although *oskar* plays an indispensable role in Drosophilid germ cell specification, the nature of the selective pressures and molecular changes responsible for its functional divergence within the genus *Drosophila* are unknown. To gain insight into the molecular evolution of this novel and critical gene, we took advantage of the completely sequenced genomes of 16 *Drosophila* species from the *Sophophora* subgenus, the *D. virilis* genome sequence and the sequenced *oskar* locus from *D. immigrans*. The goal of this study is to assess patterns of change in the *oskar* nucleotide sequence to evaluate potential variation in the evolutionary rate of distinct functional protein domains. We test the hypothesis that

Fig. 1 **a** Domain organization of *Drosophila* Oskar. Amino acid residues corresponding to the *Drosophila*-specific Long Osk domain (green), conserved predicted structural domains (LOTUS: yellow; SGNH hydrolase: blue), and regions shown to interact with conserved germ line specification genes in *D. melanogaster* (gray shades) are indicated. **b** ω value estimates for *oskar* predicted structural and interaction domains from combined maximum likelihood analysis based on two different MSA methods using fixed sites model E. Valois-interacting domains were concatenated for analysis

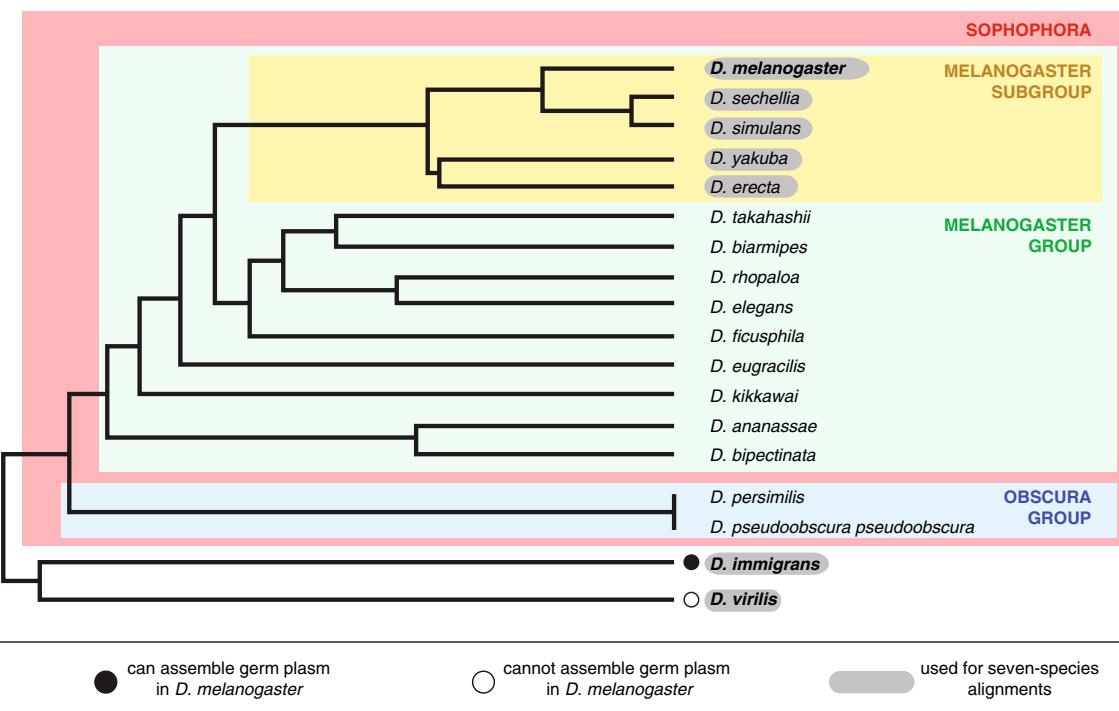
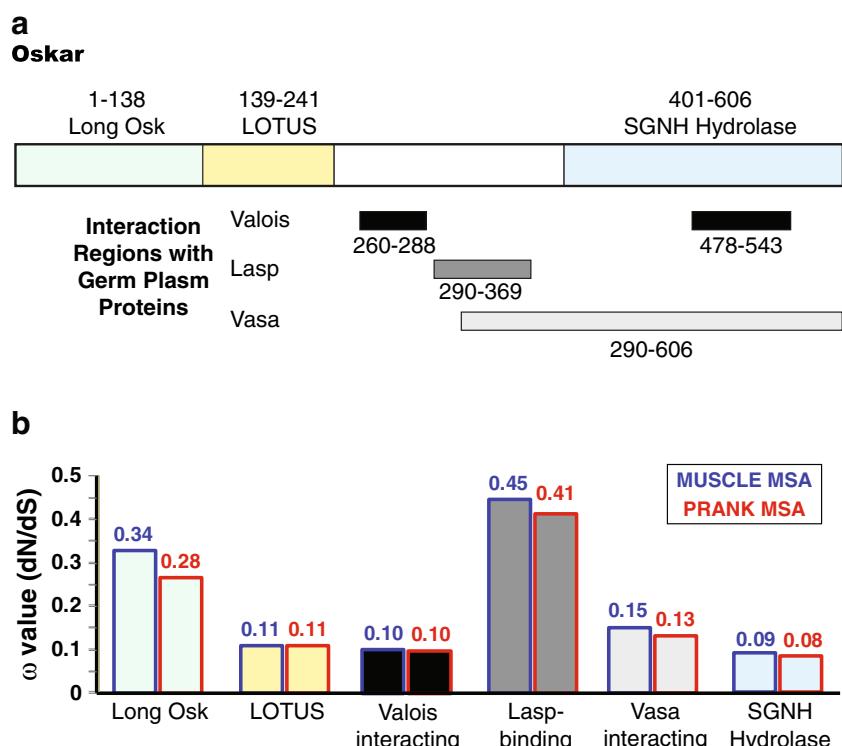


Fig. 2 Phylogenetic tree of *Drosophila* species showing topology used for PAML analysis, which is essentially consistent with current phylogenetic hypotheses (Yang et al. 2012; Kopp 2006). Species where functional analysis of *oskar* has been performed (Webster et al. 1994; Jones and Macdonald 2007) are indicated in bold. *D. immigrans* *oskar* shows functional conservation with respect to germ plasm assembly in

D. melanogaster (black circle), whereas this *oskar* function has diverged in *D. virilis* (white circle). The 18 species analyses used *oskar* sequences from all species shown; the seven species analyses used sequences from those species outlined in gray; and the five species analyses used sequences only from the *melanogaster* subgroup species (yellow box)

positive selection drives the evolution of *Drosophila oskar*, and identify regions that are likely to underlie functional divergence between *D. virilis* and *D. melanogaster oskar*, providing candidates for future study of the evolutionary changes that prevent *virosk* from specifying germ plasm in a *D. melanogaster* background.

Methods

Annotated *oskar* orthologs from *D. melanogaster* [GenBank NM_169248.1; FlyBase CG10901-PA], *Drosophila simulans* [GenBank XM_002104160.1; FlyBase GD18580-PA], *Drosophila sechellia* [GenBank XM_002031933.1; FlyBase GM23770-PA], *Drosophila yakuba* [GenBank XM_002096839.1; FlyBase GE25914-PA], *Drosophila erecta* [GenBank XM_001980858.1; FlyBase GG13545-PA], *Drosophila ananassae* [GenBank XM_001953262.1; FlyBase GF17692-PA], *Drosophila pseudoobscura* [GenBank XM_001359471.2; FlyBase GA10627], *Drosophila persimilis* [GenBank XM_002017349.1; FlyBase GL21554-PA], *Drosophila willistoni* [GenBank XM_002070244.1; FlyBase GK11117-PA], and *D. virilis* [GenBank XM_002053233.1; FlyBase GJ23790-PA] were obtained from FlyBase (www.flybase.org). Coding sequence of *D. immigrans oskar* was obtained from GenBank [DQ823084.1]. We also identified *oskar* orthologs from the following recently sequenced species whose genomes have not been annotated: *Drosophila eugracilis* [genomic scaffold: GenBank JH402624.1], *Drosophila ficusphila* [genomic scaffold: GenBank GL987928.1], *Drosophila biarmipes* [genomic scaffold: GenBank JH400370.1], *Drosophila takahashii* [genomic scaffold: GenBank JH112313.1], *Drosophila elegans* [genomic scaffold: GenBank JH110107.1], *Drosophila rhopaloa* [genomic scaffold: GenBank JH406433.1], *Drosophila kikkawai* [genomic scaffold: GenBank JH111367.1], *Drosophila bipectinata* [genomic scaffold: GenBank JH401929.1]. Genome sequences for these species were accessed via the Drosophila Species Stock Center at <https://stockcenter.ucsd.edu/info/welcome.php>. We conducted a tBLASTN search of each Drosophilid genome with *D. melanogaster Oskar* protein as the query to uncover the most similar coding sequence with respect to amino acid conservation. As *oskar* is a single-copy gene in all genomes examined to date, and shares no significant overall sequence similarity with non-*oskar* genes (Lynch et al. 2011), the top tBLASTN hit was considered to be the ortholog. To extract the coding sequence from genomic scaffold sequences we used the Augustus gene predictor (Keller et al. 2011), and manually curated the resulting sequences to mask stop codons and frame shift mutations. Reciprocal BLAST of the top predicted coding sequences to *D. melanogaster* was performed to confirm orthology.

Nucleotide sequences of all *oskar* orthologs analysed in this study are provided in Online Resource 1.

As the results of PAML analyses can be sensitive to the alignment methods used (Blackburne and Whelan 2013), we generated multiple sequence alignments using two different methods, and performed analyses on the results of both MSAs. The sequences of the 18 *Drosophila* species were multiply aligned using MUSCLE implemented in TranslatorX (Abascal et al. 2010), or using PRANK (Löytynoja and Goldman 2008). We did not include the *D. willistoni* sequence as its predicted length was less than 60 % that of *D. melanogaster oskar* (not shown). There is no evidence for unusual divergence of *D. willistoni oskar*; this is therefore likely due to sequencing or annotation error.

Predicted structural and interaction domains were manually extracted from the whole gene alignment using the known acid residues corresponding to each domain in *D. melanogaster* (Fig. 1) (Breitwieser et al. 1996; Suyama et al. 2009; Anne 2010). Sequences from Valois-interacting regions were concatenated for analysis. The PRANK alignment was subjected to GBlocks analysis using the default options (minimum number of sequences for a conserved position: 10; minimum number of sequences for a flanking position: 10; maximum number of contiguous positions: 8; minimum length of a block: 5).

We repeated the branch site test for domains that showed a consistent signature for positive selection using the reported cDNA sequence of *D. virilis oskar* (Genbank_L22556.1) that was used in the experiments that suggested functional divergence of *D. virilis* and *D. melanogaster oskar* with respect to germ plasm assembly (Webster et al. 1994) (Online Resource 1). We also used Sanger sequencing to confirm the entire nucleotide sequence of this *D. virilis* allele (see Online Resource 1).

Results and Discussion

The complete *oskar* coding region is under purifying selection

We obtained *oskar* coding sequences from the Drosophilid genome sequences, as well as the sequenced *D. immigrans oskar* coding region (Fig. 2; Online Resource 1), and generated multiple sequence alignments using two different multiple sequence alignment (MSA) tools, the similarity-based MSA MUSCLE (Edgar 2004) (Online Resource 2, Figure S1) and the evolutionarily informed MSA PRANK (Löytynoja and Goldman 2008) (Online Resource 3, Fig S2). Because the choice of MSA can influence the outcome of such analyses (Blackburne and Whelan 2013), we performed evolutionary rate analyses using both MSA outputs. We conducted maximum likelihood analyses with codeml implemented in PAML v4 (Yang 2007) to estimate non-synonymous (dN) and synonymous (dS) substitution

rates, and their ratio ($\omega=dN/dS$). We then used the Likelihood Ratio Test to compare the fit of different evolutionary models to the data (Yang 2007; Yang et al. 2000b).

We first applied the simplest model M0, which sets all branches and sites to evolve at the same rate, to obtain a single global ω estimate for the entire *oskar* coding region alignment (Yang et al. 2000a). The log likelihood for this model was $-4,397.07$ using the MUSCLE alignment, and $-16,563.12$ using the PRANK alignment, with ω estimates of 0.32 and 0.16, respectively, indicating overall purifying selection of full-length *oskar*.

Distinct *oskar* domains are evolving at different rates

Next, to estimate ω for each domain separately and test if these domains were under different selective constraints, we applied different fixed site models (Yang and Swanson 2002). The highest log likelihood was obtained using model E (MG4), which assumes different ω , κ (transition/transversion ratio), π (equilibrium codon frequencies), and rs (proportional branch lengths) between *oskar* domains. Using either the MUSCLE MSA (Online Resource 2, Table S1) or the PRANK MSA (Online Resource 3, Table S4), the fit of this model was significantly better than that of model B, which assumes identical κ , ω , and π but different rs ($P<0.001$). This shows that the strength of selection varies among distinct *oskar* domains. Estimates of ω were less than 1 for all domains, indicating overall purifying selection on each domain (Fig. 1b). However, the SGNH, LOTUS, and Valois-interacting domains are the most constrained, while the Lasp-binding and Long Osk domains are evolving approximately four to five and two to three times faster, respectively, than the former three domains. This suggests that evolutionary changes in *oskar* function may not be due to changes within the SGNH, LOTUS, or Valois-interacting domains. Moreover, it is consistent with the relatively poor conservation of the Lasp-binding region across insects (Lynch et al. 2011) and the recent evolution of Long Osk in the *Drosophila* lineage. As selective pressures on individual *oskar* domains are heterogeneous, further analyses were conducted separately for each domain.

Evidence for positive selection on specific *oskar* domains

Given the functional divergence of *Drosophila oskar*, we asked whether positive selection might be acting on specific amino acids of any Oskar domain. To test this hypothesis, we applied different site models that set the ω ratio to vary among sites (Yang et al. 2000a, 2005). Comparison of models M1 (nearly neutral; assumes two site classes in the sequence with $\omega_0<1$ and $\omega_1=1$) and model M2 (adds site class ω_2 allowing for positive selection) was not significant for any domain

using either of the MUSCLE (Online Resource 2, Table S2) or PRANK (Online Resource 3, Table S5) alignments (Table 1). Comparison of models M7 (assumes a beta distribution of ω over sites) and M8 (adds an extra site class with a free ω ratio estimated from the data, allowing ω values greater than 1) also showed no significant differences for any single domain for either of the MUSCLE or PRANK alignments (Table 1).

The results of the analyses thus far show that when using site models to consider the evolution of Oskar orthologs from these 18 *Drosophila* lineages, no specific domain appears to be under positive selection. However, if a gene evolves under purifying selection most of the time, but is occasionally subject to episodes of adaptive change, site models may not yield a significant dN/dS value. Given the inability of *D. virilis* *oskar* to substitute for *D. melanogaster* *oskar*, we wished to test the hypothesis that positive selection might be acting on the *D. virilis* branch alone. We therefore assessed differences in ω on that specific branch using branch site models (Zhang et al. 2005). We used the modified model A, or test 2, because it is more conservative, yields fewer false positives, and is better able to distinguish relaxed selective constraint from positive selection (Zhang et al. 2005). Setting the *D. virilis* branch as the foreground branch, we tested whether model MA (which assumes three site classes, $0>\omega_0<1$ between 0 and 1, and $\omega_1=1$ and $\omega_2>1$ only for the foreground branch) fit the data better than the alternative model MA_{fix} (which fixes ω_2 at 1 in foreground branches). Using either of the MUSCLE (Online Resource 2, Table S3) or PRANK (Online Resource 3, Table S6) alignments, we found that comparisons between MA and MA_{fix} were not significant for the LOTUS or SGNH predicted structural domains, or for the Valois interaction domain (Table 1). Under the MUSCLE MSA analysis of the Lasp-binding domain, the log likelihood of the MA model ($-2,350.39$) was greater than that of the MA_{fix} model ($-2,352.01$) ($\chi^2=3.24$, $P=0.07$) (Table 1; Online Resource 2, Table S3), but the difference was not significant. We note here that PAML documentation recommends the use of a critical value of χ^2 (3.84 at 5 % and 5.99 at 1 %) to guide against violation of model assumptions for branch site models (Yang 2007).

In contrast, the PRANK MSA analysis did reveal a statistically significant signature of positive selection for the Lasp-binding domain ($\chi^2=7.9$; $P<0.01$) on the *D. virilis* branch (Table 1; Online Resource 3, Table S6). The PRANK MSA analysis also identified significant signatures of positive selection for the Vasa-interacting domain ($\chi^2=5.1$; $P<0.05$) (Table 1; Online Resource 3, Table S6). This suggests some residues within these domains could be evolving under positive selection and may contribute to *oskar* functional evolution. Strikingly, we found that under both the MUSCLE ($\chi^2=14.7$; $P<0.001$) and PRANK ($\chi^2=5.43$; $P<0.05$) MSA analyses, the MA

Table 1 Likelihood Ratio Test statistics for Oskar predicted structural and interaction domains: 18 species analyses

	MUSCLE MSA			PRANK MSA		
	M1 vs M2 ¹	M7 vs M8	MA vs MA _{fix}	M1 vs M2 ¹	M7 vs M8	MA vs MA _{fix}
Long Osk	0 (1.0000)	0 (1.0000)	14.7 (0.0001)***	0 (1.0000)	1.47 (0.4795)	5.43 (0.0198)*
LOTUS	0 (1.0000)	1.44 (0.4858)	1.05 (0.3055)	0 (1.0000)	0.08 (0.9608)	1.05 (0.3055)
Lasp-binding	0.04 (0.9802)	2.91 (0.2334)	3.24 (0.0719)	0.09 (0.9560)	3.71 (0.1565)	7.9 (0.0049)**
SGNH Hydrolase	0 (1.0000)	5.73 (0.0561)	0 (1.0000)	0 (1.0000)	5.94 (0.0515)	0 (1.0000)
Vasa-interacting	0 (1.0000)	0.35 (0.8395)	0.66 (0.4166)	0 (1.0000)	3.24 (0.1979)	5.1 (0.0239)*
Valois-interacting	0 (1.0000)	3.72 (0.1557)	0 (1.0000)	0 (1.0000)	4.57 (0.1080)	0 (1.0000)

Values in each column correspond to twice the difference of log likelihood of the two nested models, with *P* values indicated in parentheses

P*<0.05; *P*<0.01; ****P*<0.0001

model fit the data significantly better than MA_{fix} for the Long Osk domain (Table 1), suggesting that non-synonymous changes at some Long Osk codons may additionally have been subject to positive selection on the *D. virilis* branch.

While the branch site analyses indicate positive selection for Long Osk and Lasp-binding domains, they may be affected by the large divergence time between *D. immigrans*, *D. virilis*, and *D. melanogaster*, the three species for which functional data are available. Highly divergent sequences could suffer from alignment errors, which can lead to false positives with the branch site model (Markova-Raina and Petrov 2011; Anisimova and Yang 2007). Removing unreliable alignment regions is expected to increase the accuracy of positive selection inference, although such filtering may also significantly decrease the power of the test, as positively selected regions are fast evolving, and those same regions are often those that are most difficult to align (Privman et al. 2012). In order to gain insight into whether Oskar sequence divergence across the 18 Drosophilids may have affected our inference of positive selection, we repeated the branch site test for positive selection on the Lasp-binding, Long Osk and Vasa-interacting domain using a new set of edited alignments, that we call here “conserved sequence block” alignments. To make these alignments, we removed poorly aligned sites in the original full-length Oskar PRANK alignment using Gblocks (Castresana 2000) and then conducted the branch site test analysis on the remaining conserved sequence blocks of each domain (Online Resource 4, Figure S3). Even with this conservative approach, we continued to see a significant signature for positive selection on the *D. virilis* branch for both the Lasp-binding ($\chi^2=4.21$; *P*<0.05) and Long Osk domains ($\chi^2=4.76$; *P*<0.05) (Table 2; Online Resource 4, Table S7). Comparisons between MA and MA_{fix} were not significant for the Vasa-interacting region (*P*=0.34)

(Online Resource 4, Table S7). Thus, even with a conservative approach the Long Osk and Lasp-binding domains are indicated as being under positive selection on the *D. virilis* branch in our analysis of 18 *Drosophila* species.

Candidate sites for functional divergence between *D. virilis* and *D. melanogaster* oskar

To identify candidates for specific sites that could be evolving under positive selection in our 18 species analysis, we applied the Bayes Empirical Bayes (BEB) approach to all alignments that showed a statistically significant signature for adaptive evolution on the *D. virilis* branch, namely (1) the MUSCLE alignment of the Long Osk domain extracted from the original alignment of full-length Oskar (Online Resource 2, Figure S1); (2) PRANK alignments of the entire Long Osk, Lasp-binding, and Vasa-interacting domains extracted from the original alignment of full-length Oskar (Online Resource 3, Figure S2); and (3) the conserved sequence blocks of the Long Osk and Lasp-binding domains obtained by trimming poorly aligned regions from the original full-length PRANK MSA alignment (Online Resource 4, Figure S3). To be conservative we focused only on sites with a BEB posterior probability of being under positive selection greater than 0.9.

Table 2 Likelihood ratio test statistics for MA vs MA_{fix} model for Long Osk, Lasp-binding and Vasa-interacting domains aligned with PRANK MSA

	Conserved sequence blocks	Seven species alignment
Long Osk	4.76 (0.0291)*	0 (1.0000)
Lasp-binding	4.21 (0.0402)*	0 (1.0000)
Vasa-interacting	0.92 (0.3375)	0.43 (0.5120)

Values in each column correspond to twice the difference of log likelihood of the two nested models, with *P* values indicated in parentheses

**P*<0.05

Using these criteria, we identified five residues, two in the Long Osk domain and three in the Lasp-binding domain that may be evolving under positive selection on the *D. virilis* branch.

Analysis of the Long Osk domain extracted from the MUSCLE MSA of full-length Oskar identified one residue with signatures of positive selection (*D. melanogaster* F51 (residue numbers throughout refer to the position in the primary amino acid sequence of *D. melanogaster* Long Oskar); Online Resource 2, Figure S1). In addition, analyses of the Long Osk domain extracted from the PRANK alignment of full-length Oskar, and of the conserved sequence blocks of Oskar, both supported the hypothesis of positive selection at a second residue (*D. melanogaster* R65; Online Resource 3, Figure S2; Online Resource 4, Figure S3).

In the Lasp-binding domain, analysis of this domain extracted from the PRANK MSA of full-length Oskar (Online Resource 3, Figure S2) identified two residues (*D. melanogaster* E306 and P353) as being under positive selection. These two residues were also identified as being under positive selection by analysis of the Vasa-interaction domain extracted from the full-length PRANK MSA (Online Resource 3). Finally, a third residue (*D. melanogaster* Y339) was identified by analysis of

the Lasp-binding domain derived from the conserved sequence block PRANK MSA (Online Resource 4, Figure S3).

If changes of amino acid identity at these sites contributed to the functional divergence between *D. melanogaster* and *D. virilis* Oskar, we might expect these residues to have different physicochemical characteristics in the two species. Consistent with this idea, we found that at three of the five candidate sites, one in the Long Osk domain and two in the Lasp-binding domain, amino acid differences between *D. virilis* and *D. melanogaster* result in changes in one or both of polarity and acid–base properties (Table 3, Fig. 3). This supports the hypothesis that changes at these residues may be linked to the functional divergence between *D. melanogaster* and *D. virilis* Oskar, making these residues promising candidates for future functional verification in vivo.

Analysis of *D. melanogaster* subgroup species

A second issue related to the high divergence time between our species of interest is that the saturation of dS in highly divergent sequences is expected to inflate false positive rates of likelihood ratio tests for positive selection (Gharib and Robinson-Rechavi 2013). Determining the optimal analytical

Table 3 Chemical properties of amino acids under positive selection in the branch leading to *D. virilis*

PP value of positive selection	AA position in <i>D. melanogaster</i> ^a	<i>D. virilis</i> AA			<i>D. melanogaster</i> AA			Polarity change	Acidity change	
		AA	Polarity	pH	Hydro	AA	Polarity	pH	Hydro	
Long Oskar domain—MUSCLE MSA (supported by full-length genomic and Macdonald allele)										
0.99 ^b	51	<i>K</i>	Polar	Basic	Philic	<i>F</i>	Nonpolar	Phobic	<i>Y</i>	<i>Y</i>
Long Oskar domain—PRANK MSA (supported by analysis of full-length and conserved sequence blocks alignments of genomic allele and full-length Macdonald allele)										
0.92/0.91/0.95 ^c	65	<i>R</i>	Polar	Basic	Philic	<i>R</i>	Polar	Basic	Philic	<i>N</i>
Lasp-binding domain—PRANK MSA (supported by full-length alignment)										
0.99 ^d	353	<i>A</i>	Nonpolar	Phobic	<i>P</i>	Nonpolar	Phobic	N	<i>N</i>	
0.98/0.95 ^e	306	<i>M</i>	Nonpolar	Phobic	<i>E</i>	Polar	Acidic	Philic	<i>Y</i>	
Lasp-binding domain—PRANK MSA (supported by genomic allele conserved sequence blocks alignment and full-length Macdonald allele)										
0.95/0.98 ^f	339	<i>H</i>	Polar	Basic	Philic	<i>Y</i>	Polar	Philic	<i>N</i>	

^a Amino acid position refers to the *D. melanogaster* residue position in the unaligned primary amino acid sequence

^b Rows in italics indicate residues under positive selection that have different chemical properties between *D. virilis* and *D. melanogaster*. The BEB posterior probability values from MUSCLE MSA alignment analyses using the genomic *D. virilis* allele or the Macdonald allele were the same for both alignments (see Online Resources 2, 8)

^c Indicates the posterior probability values from three different PRANK MSA analyses for one of two residues in the Long Osk domain identified as being under positive selection. BEB values are from analysis of full-length Oskar alignment and conserved sequence blocks alignment with genomic *D. virilis* allele, and full-length alignment generated using *D. virilis* Macdonald allele, respectively (see Online Resources 3, 4, 9)

^d This residue differs between the genomic and Macdonald alleles, and shows signatures of positive selection only under the analysis of a PRANK alignment using the genomic allele, but not the Macdonald allele (see Online Resources 3, 7)

^e Indicates the posterior probability values from two different PRANK MSA analyses for the first of three residues in the Lasp-binding domain identified as being under positive selection. BEB values are from analysis of full-length Oskar alignments generated with the genomic and Macdonald *D. virilis* alleles, respectively (see Online Resources 3, 9)

^f Indicates the posterior probability values for the second of three residues in the Lasp-binding domain identified by two different PRANK MSA analyses as being under positive selection (see Online Resources 4, 9). BEB values are from analysis of conserved sequence blocks alignment and full-length alignment generated using *D. virilis* allele from Webster et al (1994), respectively

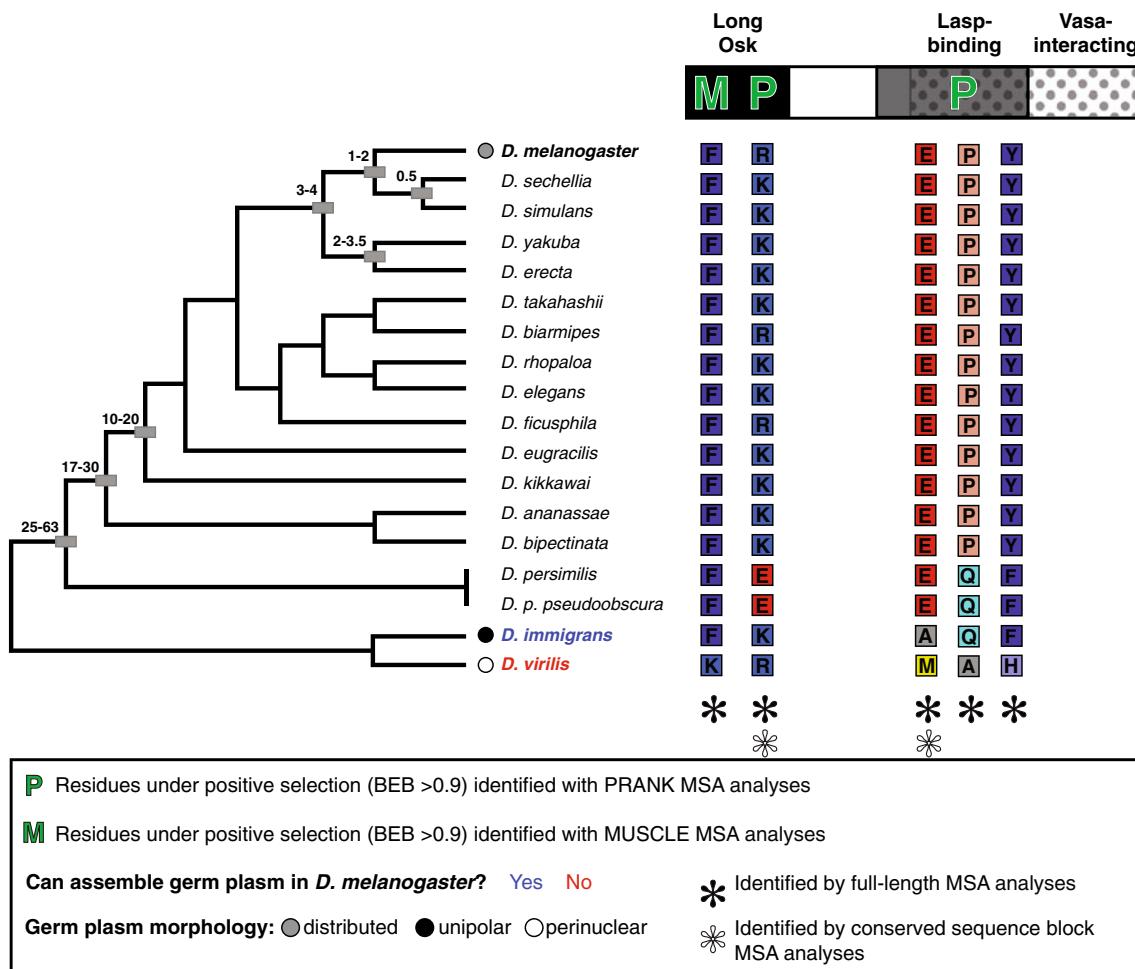


Fig. 3 Phylogenetic relationships and physicochemical properties of positively selected *oskar* residues of 18 *Drosophila* species. Left side shows phylogenetic relationships between the 18 species whose *oskar* sequences were analyzed in this study. Species names in colored text (red/blue) in tree at left indicate functional conservation of *oskar* with respect to potential for germ plasm assembly in *D. melanogaster* (Webster et al. 1994; Jones and Macdonald 2007). Shaded circles indicate germ plasm morphology as determined by histology and electron microscopy (black unipolar; gray distributed; white perinuclear; see Additional File 12 for details) (Counce 1963; Mahowald 1968). Among the species shown, functional data on *oskar*'s germ plasm assembly capabilities in *D. melanogaster*, and morphological data on germ plasm morphology, are only currently available for *D. melanogaster*, *D. immigrans* and

D. virilis. Phylogenetic relationships according to topology used for PAML analysis (Fig. 2), which is consistent with current hypotheses (Yang et al. 2012; Kopp 2006). Numbers at selected nodes indicated estimated divergence times in Mya; estimates from (Obbard et al. 2012; Remsen and O'Grady 2002). Right side shows amino acid identities of those Oskar residues that appeared positively selected in the branch leading to *D. virilis* by each of the analyses performed with two different MSA, MUSCLE (in the Long Osk domain; leftmost residue) and PRANK (one in the Long Osk domain and three in the Lasp-binding and Vasa-interacting domains; remaining four residues). Residues are shown in the N-terminal to C-terminal order in which they appear in the indicated protein domains, colored according to the RasMol amino acid color scheme (Sayle and Milner-White 1995) (see Table 3 for details)

design for these studies is challenged, however, by the fact that the phylogenetic boundaries for points of saturation, and the extent to which the branch site test is affected by divergence time, are not clear (Gharib and Robinson-Rechavi 2013). Nonetheless, we undertook two additional analyses as a further step towards gaining some insight into the effect of divergence on our inferences of positive selection.

First, we assessed the relative rates of evolution of predicted structural and functional domains using a reduced set of sequences with relatively low divergence. We generated MUSCLE alignments of the five species of the *D. melanogaster*

subgroup (Online Resource 5, Figure S4), which are often considered to be at an appropriate level of divergence for PAML analyses (Clark et al. 2007; Larracuente et al. 2008), as they diverged from a last common ancestor only approximately 3–4 Mya (Tamura et al. 2004). Analyses of only these sequences should thus avoid the potential problem of saturation of dS. The global rate of evolution of *oskar* in these five species using model MO was 0.25 with a log likelihood of −4,086.41, indicating overall purifying selection as observed in the analysis of 18 species. Next, we used this alignment to assess rates of evolution of the component subdomain models using the MG4

model, and again obtained the same results as those obtained with the alignment containing 18 species (Fig. 1b). ω values for each domain showed that SGNH was the most conserved domain, while the Long Osk and Lasp-binding domains were the two most rapidly evolving (Online Resource 5, Table S8). However, in contrast to our original 18 species analysis, neither partition nor site tests detected significant positive selection for any domain except LOTUS, which was significant with $P=0.04$ (Online Resource 5, Table S9, Table S10).

Second, to test for positive selection on the *D. virilis* branch, we generated another set of PRANK alignments using a subset of sequenced *Drosophila* species comprising those in *D. melanogaster* subgroup, plus *D. immigrans* and *D. virilis*, the only two non-*melanogaster* species for which data on *oskar* function are available (Webster et al. 1994; Jones and Macdonald 2007) (Online Resource 6, Figure S5). These alignments, like those of the *melanogaster* subgroup sequences, should be less likely than the full 18 species analysis to suffer from problems with saturated synonymous sites and inaccurate alignments. We used these seven species alignments to assess rates of evolution of the component subdomain using the MG4 model, and obtained the same results as those obtained with the alignment containing 18 species (Fig. 1b). ω values for each domain showed that SGNH was the most conserved domain, while the Long Osk and Lasp-binding domains were the two most rapidly evolving (Online Resource 6, Table S11). However, branch site test comparison of MA with MA_{fix} was not significant for any of the Lasp-binding ($\chi^2=0$; $P=1$), Long Osk ($\chi^2=0$; $P=1$) or Vasa-interacting domains ($P=0.51$) (Table 2; Online Resource 6, Table S12).

Assessment of putative polymorphisms in *D. virilis oskar*

All analyses described above used the *oskar* nucleotide sequence annotated from the *D. virilis* genome (Clark et al. 2007) (NCBI accession XM_002053233.1; henceforth designated “genomic allele”). However, we noticed that the translation of this sequence differs from the reported *D. virilis oskar* cDNA translation (Webster et al. 1994) (NCBI accession L22556.1; henceforth designated “Macdonald allele”) by two amino acids in the Long Osk region (QQ66-67 indel), and three amino acids in the Lasp-binding region (A341P, H348 indel, P383R) (Online Resource 7, Figure S6). We used Sanger sequencing to confirm that the sequence of the Macdonald allele was identical to that reported in NCBI accession L22556.1, except for one residue (A341P) that matched the genomic allele rather than the Macdonald allele (Online Resource 7). Because the Macdonald allele was used for the transgenic experiments that suggested functional divergence between *D. virilis* and *D. melanogaster oskar* (Webster et al. 1994), we repeated the branch site test for 18 species using the Macdonald allele sequence, and focused our analysis on those

protein domains and residues that showed a consistent signal for positive selection. Specifically, these were the Long Oskar domain aligned with the MUSCLE MSA, and the Long Oskar and Lasp-binding domains aligned with the PRANK MSA. We found that the same domains and residues were predicted to be under positive selection regardless of whether we used the genomic or Macdonald alleles for analysis. Specifically, we found a significant signal of positive selection on the Long Osk domain using the MUSCLE alignment of the Macdonald allele ($\chi^2=8.44$; $P=0.0037$) (Online Resource 8, Table S13). Within the Long Osk domain, the same residue (*D. melanogaster* F51) was identified as being under positive selection in analyses of both alleles. Similarly, we found a significant signal of positive selection on the Long Osk domain using the PRANK alignment of the Macdonald allele ($\chi^2=5.76$; $P=0.016$) (Online Resource 9, Table S14), and within this domain, the same residue (*D. melanogaster* R65) was identified as being under positive selection in analyses of both alleles. We also detected a significant signature of positive selection for the Lasp-binding region ($\chi^2=5.5$; $P=0.019$) on the *D. virilis* branch using the PRANK alignment and the Macdonald allele. This analysis identified two residues: *D. melanogaster* E306, which was also identified in the analysis of the full-length PRANK alignment of the genomic allele (Online Resource 3), and *D. melanogaster* Y339, which was also identified in the PRANK analysis of conserved sequence blocks of the genomic allele (Online Resource 4).

Thus, the minor sequence differences between the Macdonald and genomic alleles do not change our detection of signatures of positive selection on the same domains and residues of Oskar. The specific residues thus identified differ as a function of the alignment method used, but not due to the *D. virilis oskar* sequence used. The only exception is a single residue, *D. melanogaster* P353, which was identified as being under positive selection in the PRANK analysis of the genomic allele but not of the Macdonald allele. This is likely due to the fact that this residue is different in the Macdonald allele (A341P). We therefore cannot conclude that this particular residue is a candidate for functional divergence between the two species. Overall, we conclude that while there are a small number of polymorphisms in sequenced *D. virilis oskar* alleles, our results are generally robust to these minor sequence differences.

In summary, all analyses clearly support heterogeneous rates of evolution on distinct Oskar domains, with Long Osk and Lasp-binding being the fastest evolving. However, we cannot exclude the possibility that signatures of positive selection on these domains, which were detected only with the 18 species analyses, may be false positives. Alternatively, the lack of a significant signal for positive selection in the seven species analysis could be due to the decrease in power that results from using fewer sequences. In support of the positive selection result obtained with the 18 species analyses,

simulations have shown that the branch site model is more robust to high sequence divergence than might have been expected (Gharib and Robinson-Rechavi 2013; Studer et al. 2008), and that high dS values may be more of a concern for false negatives than false positives (Gharib and Robinson-Rechavi 2013).

Overall, the effects of sequence divergence and dS saturation are complex and difficult to fully resolve. Nonetheless, the consistent pattern of higher rates of evolution of Long Osk and Lasp-binding domains seen in all analyses of five species, seven species and 18 species alignments make changes in these domains strong candidates for functional divergence between *Drosophila* Oskar proteins. We further suggest that the specific sites within these domains exhibiting changes in physicochemical properties between *D. virilis* and *D. melanogaster* are strong candidates for changes underlying the functional divergence of Oskar between *D. melanogaster* and *D. virilis*. However, we note that caution is warranted in the inference of positive selection on these sites along the *D. virilis* branch.

The roles of *oskar* in the evolution of germ plasm morphology

To date the functional characteristics of *oskar* with respect to germ plasm assembly have been tested for the orthologs of only three *Drosophila* species: *D. melanogaster*, *D. immigrans* (*immosk*), and *D. virilis* (*virosk*). All of these species possess germ plasm, but the germ plasm morphology is distinct in each species (Mahowald 1962, 1968; Counce 1963). Mapping germ plasm morphology on to a phylogeny of Drosophilids (Online Resource 10) shows that germ plasm morphology displays some clade-specific patterns. We hypothesize that changes in germ plasm morphology may be the result of evolutionary changes in *oskar* function. In support of this hypothesis, as noted above *immosk* can assemble a functional germ plasm that has a *D. immigrans* morphology in a *D. melanogaster* context (Jones and Macdonald 2007). In contrast, *virosk* cannot assemble functional germ plasm in a *D. melanogaster* context (Webster et al. 1994). This indicates that the characteristics of germ plasm morphology and the assembly of germ plasm itself are both processes directed by *oskar*. At present, *oskar* orthologs have not been identified for most species with well-studied germ plasm morphology (Online Resource 10), and germ plasm morphology has not been characterized for most of the species whose *oskar* sequence has been determined. Further studies that fill these knowledge gaps will be useful in determining the specific changes in *oskar* that accompanied and/or led to evolutionary changes in germ plasm morphology and function within Drosophilids.

Possible significance of the evolution of the *Drosophila*-specific Long Oskar domain

The Long Oskar domain was identified by both MUSCLE and PRANK MSA analyses as possibly being under positive selection in the lineage leading to *D. virilis*. Known *oskar* loss of function alleles in *D. melanogaster* do not contain lesions in the Long Oskar domain, and therefore do not shed light on the potential functions of the positively selected amino acids identified in this study. However, evidence from studies of *D. melanogaster* suggests that this domain may play a role in stabilizing assembled germ plasm. During oogenesis, *oskar* mRNA is synthesized and transported to the posterior pole of the oocyte (Rongo et al. 1995), where it is translated into the Long Osk and Short Osk isoforms from two alternative start codons (Markussen et al. 1995). The Long Osk isoform includes all four protein domains described above, whereas the Short Osk isoform excludes the Long Oskar domain and thus consists only of the LOTUS, Lasp-binding and SGNH domains. The two isoforms exhibit distinct molecular affinities for germ plasm components (Breitwieser et al. 1996; Suyama et al. 2009; Babu et al. 2004) and localize to distinct subcellular compartments (Vanzo et al. 2007), but both are required for effective germ plasm assembly. Short Osk alone is able to assemble enough germ plasm to yield a low frequency of germ cell formation (Markussen et al. 1995), but cannot efficiently maintain *oskar* mRNA or protein at the posterior pole (Vanzo and Ephrussi 2002). In contrast, Long Osk can maintain both *oskar* mRNA and Oskar protein at the posterior pole, but no germ cells form when only Long Osk is expressed, suggesting that it cannot efficiently assemble stable germ plasm. The current model for Osk-mediated germ plasm assembly in *D. melanogaster* is thus one in which Long Osk anchors Short Osk to the posterior oocyte cortex, and Short Osk in turn localizes multiple germ cell determinants. The fact that *Nasonia vitripennis* (Hymenoptera) *oskar* lacks the Long Osk domain but nonetheless appears to act as a functional germ plasm determinant (Lynch et al. 2011), raises the question of what the functional significance of the evolution of the Long Osk domain might be.

One possibility is that the evolution of the Long Oskar domain was linked to the emergence of additional mRNA localization mechanisms that increased the stability of *oskar* mRNA at the oocyte posterior. This could have been beneficial in improving the stability of germ plasm and increasing robustness of *oskar*-directed germ cell specification. Consistent with this hypothesis, although *oskar* mRNA is localized to embryonic germ plasm in all studied holometabolous insects, its localization to the oocyte posterior cortical cytoplasm prior to germ cell formation is variable, and correlates with the presence or absence of the Long Oskar domain. In *D. melanogaster*, *oskar* mRNA (as visualized by *in situ* hybridization) is very tightly localized in a thin crescent

(Ephrussi et al. 1991), but in *D. virilis*, comparable gene expression methods show that only some *oskar* mRNA is localized to a small spot at the posterior cortex, while additional transcripts are localized in a diffuse cloud in an apparent posterior to anterior gradient (Webster et al. 1994). In holometabolous insects without a Long Oskar isoform (the wasp *Nasonia vitripennis*, the ant *Messor pergandi* (Lynch et al. 2011), and the mosquitoes *Aedes aegypti* (Juhn and James 2006) and *Culex quinquefasciatus* (Juhn et al. 2008)), posterior *oskar* mRNA localization also appears more diffuse than in *D. melanogaster*. Finally, the only *oskar* ortholog identified to date in a hemimetabolous insect (the cricket *Gryllus bimaculatus*) also lacks a Long Oskar domain, and its mRNA is distributed ubiquitously in oocytes and embryos, showing no asymmetric localization at all (Ewen-Campen et al. 2012).

Conclusions

While *D. melanogaster* *oskar* function has been well characterized at the genetic level, the specific molecular mechanisms by which it functions remain unknown. Previous comparative approaches using the entire gene have shown functional divergence of *Drosophila oskar* (Jones and Macdonald 2007; Webster et al. 1994) but have not identified the specific regions or selective pressures involved. Our molecular evolution analysis shows that the two fastest evolving domains, Long Osk and Lasp-binding, also show a statistically significant signature of positive selection on the *D. virilis* branch in our analysis of 18 species orthologs. Specific putatively positively selected sites within these domains also exhibit major differences in the physicochemical properties of amino acids between *D. melanogaster* and *D. virilis*. However, the high divergence time between *D. melanogaster*, *D. immigrans*, and *D. virilis* means that we can only cautiously infer positive selection at these sites. Further polymorphism-based tests of positive selection will be required to elucidate the selection pressures involved. Nonetheless, based on their faster rate of evolution, we suggest that changes in the Long Oskar and Lasp-binding domains underlie functional differences between the Oskar proteins. Functional verification of the roles of these domains and specific sites will be required to evaluate the contributions of each of these candidates to the evolution of *oskar* function. The Long Osk and Lasp-binding domains are the first candidates for *oskar* functional evolution identified using an evolutionary approach, and provide specific hypotheses that can be tested for functional verification in future *in vivo* studies. This analysis thus represents an important step towards understanding the role of Oskar in germ plasm evolution and assembly.

Acknowledgments Thanks to Paul Macdonald for the plasmid containing the *D. virilis* *oskar* cDNA, to John Srouji for Sanger sequencing

and discussion of the results, to Victor Zeng and Amit Indap for assistance with preliminary analyses, and to Extavour lab members for discussion of the manuscript. This work was partly supported by NSF grant IOS-0817678 to CGE and funds from Harvard University.

Competing interests The authors declare that they have no competing interests.

Author contributions AA conceived of the study, created alignments, and performed evolutionary rate analyses. CGE assisted with study design and performed analyses of amino acid physicochemical properties and phylogenetic distribution of germ plasm morphology. Both authors wrote and approved the final manuscript.

References

- Abascal F, Zardoya R, Telford MJ (2010) TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. doi:[10.1093/nar/gkq291](https://doi.org/10.1093/nar/gkq291), 38 (Web Server issue): W7-13
- Anantharaman V, Zhang D, Aravind L (2010) OST-HTH: a novel predicted RNA-binding domain. Biol Direct 5:13. doi:[10.1186/1745-6150-5-13](https://doi.org/10.1186/1745-6150-5-13)
- Anisimova M, Yang Z (2007) Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. Mol Biol Evol 24(5):1219–1228. doi:[10.1093/molbev/msm042](https://doi.org/10.1093/molbev/msm042)
- Anne J (2010) Targeting and anchoring Tudor in the pole plasm of the *Drosophila* Oocyte. PLoS ONE 5(12):e14362. doi:[10.1371/journal.pone.0014362](https://doi.org/10.1371/journal.pone.0014362)
- Babu K, Cai Y, Bahri S, Yang X, Chia W (2004) Roles of Bifocal, Homer, and F-actin in anchoring Oskar to the posterior cortex of *Drosophila* oocytes. Genes Dev 18(2):138–143. doi:[10.1101/gad.282604](https://doi.org/10.1101/gad.282604)
- Blackburne BP, Whelan S (2013) Class of multiple sequence alignment algorithm affects genomic analysis. Mol Biol Evol 30(3):642–653. doi:[10.1093/molbev/mss256](https://doi.org/10.1093/molbev/mss256)
- Blackstone NW, Jasker BD (2003) Phylogenetic considerations of clonality, coloniality, and mode of germline development in animals. J Exp Zool B Mol Dev Evol 297B(1):35–47
- Breitwieser W, Markussen F-H, Horstmann H, Ephrussi A (1996) Oskar protein interaction with Vasa represents an essential step in polar granule assembly. Genes Dev 10:2179–2188
- Callebaut I, Mornon J-P (2010) LOTUS, a new domain associated with small RNA pathways in the germline. Bioinformatics 26(9):1140–1144. doi:[10.1093/bioinformatics/btq122](https://doi.org/10.1093/bioinformatics/btq122)
- Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17(4): 540–552
- Cavey M, Hijal S, Zhang X, Suter B (2005) *Drosophila valois* encodes a divergent WD protein that is required for Vasa localization and Oskar protein accumulation. Development 132(3):459–468
- Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA, Kaufman TC, Kellis M, Gelbart W, Iyer VN, Pollard DA, Sackton TB, Larracuente AM, Singh ND, Abad JP, Abt DN, Adryan B, Aguade M, Akashi H, Anderson WW, Aquadro CF, Ardell DH, Arguello R, Artieri CG, Barbash DA, Barker D, Barsanti P, Batterham P, Batzoglou S, Begun D, Bhutkar A, Blanco E, Bosak SA, Bradley RK, Brand AD, Brent MR, Brooks AN, Brown RH, Butlin RK, Caggese C, Calvi BR, Bernardo de Carvalho A, Caspi A, Castrezana S, Celniker SE, Chang JL, Chapple C, Chatterji S, Chinwalla A, Civetta A, Clifton SW, Comerón JM, Costello JC, Coyne JA, Daub J, David RG, Delcher AL, Delehaunty K, Do CB, Ebling H, Edwards K, Eickbush T, Evans JD, Filipski A, Findeiss S,

- Freyhult E, Fulton L, Fulton R, Garcia ACL, Gardiner A, Garfield DA, Garvin BE, Gibson G, Gilbert D, Gnerre S, Godfrey J, Good R, Gotea V, Gravely B, Greenberg AJ, Griffiths-Jones S, Gross S, Guigo R, Gustafson EA, Haerty W, Hahn MW, Halligan DL, Halpern AL, Halter GM, Han MV, Heger A, Hillier L, Hinrichs AS, Holmes I, Hoskins RA, Hubisz MJ, Hultmark D, Huntley MA, Jagadeeshan S, Jeck WR, Johnson J, Jones CD, Jordan WC, Karpen GH, Kataoka E, Keightley PD, Kheradpour P, Kirkness EF, Koerich LB, Kristiansen K, Kudrna D, Kulathinal RJ, Kumar S, Kwok R, Lander E, Langley CH, Lapoint R, Lazzaro BP, Lee S-J, Levesque L, Li R, Lin C-F, Lin MF, Lindblad-Toh K, Llopart A, Long M, Low L, Lozovsky E, Lu J, Luo M, Machado CA, Makalowski W, Marzo M, Matsuda M, Matzkin L, McAllister B, McBride CS, McKernan B, McKernan K, Mendez-Lago M, Minx P, Mollenhauer MU, Montooth K, Mount SM, Mu X, Myers E, Negre B, Newfeld S, Nielsen R, Noor MAF, O'Grady P, Pachter L, Papaceit M, Parisi MJ, Parisi M, Parts L, Pedersen JS, Pesole G, Phillippe AM, Ponting CP, Pop M, Porcelli D, Powell JR, Prohaska S, Pruitt K, Puig M, Quesneville H, Ram KR, Rand D, Rasmussen MD, Reed LK, Reenan R, Reily A, Remington KA, Rieger TT, Ritchie MG, Robin C, Rogers Y-H, Rohde C, Rozas J, Rubenfield MJ, Ruiz A, Russo S, Salzberg SL, Sanchez-Gracia A, Saranga DJ, Sato H, Schaeffer SW, Schatz MC, Schlenke T, Schwartz R, Segarra C, Singh RS, Sirot L, Sirota M, Sisneros NB, Smith CD, Smith TF, Spieth J, Stage DE, Stark A, Stephan W, Strausberg RL, Strempl S, Sturgill D, Sutton G, Sutton GG, Tao W, Teichmann S, Tobari YN, Tomimura Y, Tsolas JM, Valente VLS, Venter E, Venter JC, Vicario S, Vieira FG, Villegas AJ, Villasante A, Walenz B, Wang J, Wasserman M, Watts T, Wilson D, Wilson RK, Wing RA, Wolfner MF, Wong A, Wong GK-S, Wu C-I, Wu G, Yamamoto D, Yang H-P, Yang S-P, Yorke JA, Yoshida K, Zdobnov E, Zhang P, Zhang Y, Zimin AV, Baldwin J, Abdouelleil A, Abdulkadir J, Abebe A, Abere B, Abreu J, Acer SC, Aftuck L, Alexander A, An P, Anderson E, Anderson S, Arachi H, Azar M, Bachantsang P, Barry A, Bayul T, Berlin A, Bessette D, Bloom T, Blye J, Boguslavskiy L, Bonnet C, Boukhalter B, Bourzgui I, Brown A, Cahill P, Channer S, Cheshatsang Y, Chuda L, Citroen M, Collymore A, Cooke P, Costello M, D'Aco K, Daza R, De Haan G, DeGray S, DeMaso C, Dhargay N, Dooley K, Dooley E, Doricent M, Dorje P, Dorjee K, Dupes A, Elong R, Falk J, Farina A, Faro S, Ferguson D, Fisher S, Foley CD, Franke A, Friedrich D, Gadbois L, Gearin G, Gearin CR, Giannoukos G, Goode T, Graham J, Grandbois E, Grewal S, Gyaltset K, Hafez N, Hagos B, Hall J, Henson C, Hollinger A, Honan T, Huard MD, Hughes L, Hurlhula B, Husby ME, Kamat A, Kanga B, Kashin S, Khazanovich D, Kisner P, Lance K, Lara M, Lee W, Lennon N, Letendre F, LeVine R, Lipovsky A, Liu X, Liu J, Liu S, Lokyitsang T, Lokyitsang Y, Lubonja R, Lui A, MacDonald P, Magnisalis V, Maru K, Matthews C, McCusker W, McDonough S, Mehta T, Meldrim J, Meneus L, Mihai O, Mihalev A, Mihova T, Mittelman R, Mlenga V, Montmayeur A, Mulrain L, Navidi A, Naylor J, Negash T, Nguyen T, Nguyen N, Nicol R, Norbu C, Norbu N, Novod N, O'Neill B, Osman S, Markiewicz E, Oyono OL, Patti C, Phunkhang P, Pierre F, Priest M, Raghuraman S, Rege F, Reyes R, Rise C, Rogov P, Ross K, Ryan E, Settipalli S, Shea T, Sherpa N, Shi L, Shih D, Sparrow T, Spaulding J, Stalker J, Stange-Thomann N, Stavropoulos S, Stone C, Strader C, Tesfaye S, Thomson T, Thoulutsang Y, Thoulutsang D, Topham K, Topping I, Tsamla T, Vassiliev H, Vo A, Wangchuk T, Wangdi T, Weiand M, Wilkinson J, Wilson A, Yadav S, Young G, Yu Q, Zembek L, Zhong D, Zimmer A, Zwirko Z, Jaffe DB, Alvarez P, Brockman W, Butler J, Chin C, Grabherr M, Mauceli E, MacCallum I (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450(7167):203–218. doi:[10.1038/nature06341](https://doi.org/10.1038/nature06341)
- Counce SJ (1963) Developmental morphology of polar granules in *Drosophila* including observations on pole cell behavior and distribution during embryogenesis. *J Morphol* 112(2):129–145
- Da Lage JL, Kerfoot GJ, Maczkowiak F, Silvain FF, Cariou ML, Lachaise D (2006) A phylogeny of Drosophilidae using the *Amyrel* gene: questioning the *Drosophila melanogaster* species group boundaries. *J Zool Syst Evol Res* 45(1):47–63
- Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinforma* 5:113. doi:[10.1186/1471-2105-5-113](https://doi.org/10.1186/1471-2105-5-113)
- Ephrussi A, Lehmann R (1992) Induction of germ cell formation by *oskar*. *Nature* 358(6385):387–392
- Ephrussi A, Dickinson LK, Lehmann R (1991) Oskar organizes the germ plasm and directs localization of the posterior determinant *nanos*. *Cell* 66(1):37–50
- Ewen-Campen B, Schwager EE, Extavour CG (2010) The molecular machinery of germ line specification. *Mol Reprod Dev* 77(1):3–18
- Ewen-Campen B, Srouji JR, Schwager EE, Extavour CG (2012) *oskar* predates the evolution of germ plasm in insects. *Curr Biol* 22(23):2278–2283
- Extavour CG, Akam ME (2003) Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. *Development* 130(24):5869–5884
- Gaunt MW, Miles MA (2002) An insect molecular clock dates the origin of the insects and accords with palaeontological and biogeographic landmarks. *Mol Biol Evol* 19(5):748–761
- Gharib WH, Robinson-Rechavi M (2013) The branch-site test of positive selection is surprisingly robust but lacks power under synonymous substitution saturation and variation in GC. *Mol Biol Evol* 30(7):1675–1686. doi:[10.1093/molbev/mst062](https://doi.org/10.1093/molbev/mst062)
- Goltsev Y, Hsiung W, Lanzaro G, Levine M (2004) Different combinations of gap repressors for common stripes in *Anopheles* and *Drosophila* embryos. *Dev Biol* 275(2):435–446
- Jones JR, Macdonald PM (2007) Oskar controls morphology of polar granules and nuclear bodies in *Drosophila*. *Development* 134(2):233–236
- Juhn J, James AA (2006) *oskar* gene expression in the vector mosquitoes, *Anopheles gambiae* and *Aedes aegypti*. *Insect Mol Biol* 15(3):363–372
- Juhn J, Marinotti O, Calvo E, James AA (2008) Gene structure and expression of *nanos* (*nos*) and *oskar* (*osk*) orthologues of the vector mosquito, *Culex quinquefasciatus*. *Insect Mol Biol* 17(5):545–552
- Keller O, Kollmar M, Stanke M, Waack S (2011) A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics* 27(6):757–763. doi:[10.1093/bioinformatics/btr010](https://doi.org/10.1093/bioinformatics/btr010)
- Kopp A (2006) Basal relationships in the *Drosophila melanogaster* species group. *Mol Phylogen Evol* 39(3):787–798. doi:[10.1016/j.ympev.2006.01.029](https://doi.org/10.1016/j.ympev.2006.01.029)
- Larracueta AM, Sackton TB, Greenberg AJ, Wong A, Singh ND, Sturgill D, Zhang Y, Oliver B, Clark AG (2008) Evolution of protein-coding genes in *Drosophila*. *Trends Genet* 24(3):114–123. doi:[10.1016/j.tig.2007.12.001](https://doi.org/10.1016/j.tig.2007.12.001)
- Löytynoja A, Goldman N (2008) Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. *Science (New York, NY)* 320(5883):1632–1635. doi:[10.1126/science.1158395](https://doi.org/10.1126/science.1158395)
- Lynch JA, Özüak O, Khila A, Abouheif E, Desplan C, Roth S (2011) The phylogenetic origin of *oskar* coincided with the origin of maternally provisioned germ plasm and pole cells at the base of the holometabola. *PLoS Genet* 7(4):e1002029. doi:[10.1371/journal.pgen.1002029](https://doi.org/10.1371/journal.pgen.1002029)

- Mahowald AP (1962) Fine structure of pole cells and polar granules in *Drosophila melanogaster*. *J Exp Zool* 151:201–215
- Mahowald AP (1968) Polar granules of *Drosophila*. II. Ultrastructural changes during early embryogenesis. *J Exp Zool* 167(2):237–261. doi:[10.1002/jez.1401670211](https://doi.org/10.1002/jez.1401670211)
- Markova-Raina P, Petrov D (2011) High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 *Drosophila* genomes. *Genome Res* 21(6):863–874. doi:[10.1101/gr.115949.110](https://doi.org/10.1101/gr.115949.110)
- Markussen FH, Michon AM, Breitwieser W, Ephrussi A (1995) Translational control of *oskar* generates short OSK, the isoform that induces pole plasm assembly. *Development* 121(11):3723–3732
- Michod RE (2005) On the transfer of fitness from the cell to the multicellular organism. *Biol Philos* 20(5):967–987
- Obbard DJ, MacLennan J, Kim KW, Rambaut A, O’Grady PM, Jiggins FM (2012) Estimating divergence dates and substitution rates in the *Drosophila* phylogeny. *Mol Biol Evol* 29(11):3459–3473. doi:[10.1093/molbev/mss150](https://doi.org/10.1093/molbev/mss150)
- Oliveira DCSG, Almeida FC, O’Grady PM, Armella MA, DeSalle R, Etges WJ (2012) Monophyly, divergence times, and evolution of host plant use inferred from a revised phylogeny of the *Drosophila repleta* species group. *Mol Phylogenet Evol* 64(3):533–544. doi:[10.1016/j.ympev.2012.05.012](https://doi.org/10.1016/j.ympev.2012.05.012)
- Privman E, Penn O, Pupko T (2012) Improving the performance of positive selection inference by filtering unreliable alignment regions. *Mol Biol Evol* 29(1):1–5. doi:[10.1093/molbev/msr177](https://doi.org/10.1093/molbev/msr177)
- Remsen J, O’Grady P (2002) Phylogeny of Drosophilinae (Diptera: Drosophilidae), with comments on combined analysis and character support. *Mol Phylogenet Evol* 24(2):249–264
- Robe LJ, Valente VLS, Budnik M, Loreto ELS (2005) Molecular phylogeny of the subgenus *Drosophila* (Diptera, Drosophilidae) with an emphasis on Neotropical species and groups: a nuclear versus mitochondrial gene approach. *Mol Phylogenet Evol* 36(3):623–640. doi:[10.1016/j.ympev.2005.05.005](https://doi.org/10.1016/j.ympev.2005.05.005)
- Rongo C, Gavis ER, Lehmann R (1995) Localization of *oskar* RNA regulates *oskar* translation and requires Oskar protein. *Development* 121(9):2737–2746
- Sayle RA, Milner-White EJ (1995) RASMOL: biomolecular graphics for all. *Trends Biochem Sci* 20(9):374
- Studer RA, Penel S, Duret L, Robinson-Rechavi M (2008) Pervasive positive selection on duplicated and nonduplicated vertebrate protein coding genes. *Genome Res* 18(9):1393–1402. doi:[10.1101/gr.076992.108](https://doi.org/10.1101/gr.076992.108)
- Suyama R, Jenny A, Curado S, Pellis-van Berkel W, Ephrussi A (2009) The actin-binding protein Lasp promotes Oskar accumulation at the posterior pole of the *Drosophila* embryo. *Development* 136(1):95–105. doi:[10.1242/dev.027698](https://doi.org/10.1242/dev.027698)
- Tamura K, Subramanian S, Kumar S (2004) Temporal patterns of fruit fly (*Drosophila*) evolution revealed by mutation clocks. *Mol Biol Evol* 21(1):36–44. doi:[10.1093/molbev/msg236](https://doi.org/10.1093/molbev/msg236)
- van der Linde K, Houle D, Spicer GS, Steppan SJ (2010) A supermatrix-based molecular phylogeny of the family Drosophilidae. *Genet Res* 92(1):25–38
- Vanzo NF, Ephrussi A (2002) Oskar anchoring restricts pole plasm formation to the posterior of the *Drosophila* oocyte. *Development* 129(15):3705–3714
- Vanzo N, Oprins A, Xanthakis D, Ephrussi A, Rabouille C (2007) Stimulation of endocytosis and actin dynamics by Oskar polarizes the *Drosophila* oocyte. *Dev Cell* 12(4):543–555
- Webster PJ, Suen J, Macdonald PM (1994) *Drosophila virilis oskar* transgenes direct body patterning but not pole cell formation or maintenance of mRNA localization in *D. melanogaster*. *Development* 120(7):2027–2037
- Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24(8):1586–1591. doi:[10.1093/molbev/msm088](https://doi.org/10.1093/molbev/msm088)
- Yang Z, Swanson WJ (2002) Codon-substitution models to detect adaptive evolution that account for heterogeneous selective pressures among site classes. *Mol Biol Evol* 19(1):49–57
- Yang Z, Nielsen R, Goldman N, Pedersen AM (2000a) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155(1):431–449
- Yang Z, Swanson WJ, Vacquier VD (2000b) Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. *Mol Biol Evol* 17(10):1446–1455
- Yang Z, Wong WS, Nielsen R (2005) Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol* 22(4):1107–1118. doi:[10.1093/molbev/msi097](https://doi.org/10.1093/molbev/msi097)
- Yang Y, Hou Z-C, Y-h Q, Kang H, Zeng Q-t (2012) Increasing the data size to accurately reconstruct the phylogenetic relationships between nine subgroups of the *Drosophila melanogaster* species group (Drosophilidae, Diptera). *Mol Phylogenet Evol* 62(1):214–223. doi:[10.1016/j.ympev.2011.09.018](https://doi.org/10.1016/j.ympev.2011.09.018)
- Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 22(12):2472–2479. doi:[10.1093/molbev/msi237](https://doi.org/10.1093/molbev/msi237)

Additional File 1. Nucleotide sequences of *oskar* orthologues used for analysis. Sequence names indicate species name and GenBank accessions for cDNA sequences; “scaffold” in the sequence name indicates that the sequence was identified in scaffolds of unannotated *Drosophila* genomes.

```
>Drosophila_melanogaster_NM_169248.1
ATGGCCGCAGTCACAAGTGAATTCCCCAGCAAACCGATCAGTTATACCAGCACCAATACTTCCGCCAAAACCTATTATCTTAAGTCCGTGAAAAAGCAGGGTGACCACGTGTTCCAGCAGTGCCTCAGTCAGATCTCGTAAGGTCCGATTCTCCGCATCGGGGAACGCTTCGGAAAGATCTTCAAATCGGCGGAAAAACCGAATTGCCGGAGTTGTGGAAGGTGCCATTGGTGGCCCATGAACTCACCAGCCGGCAGAGCAGCAGTGCAAGTCTGAGCCAGACTCTTCTCGTCCACTCAGATTTCACCAAGGAAATCACCTACAAACAGCAACAGCAACACCAGCGAGAACAAATGACCACATCGAGAGCAACTATATATCCGTGCCGAGGAATATCCCATAAGATA GTGAGGTGCGGCCATATTGCTGAGCCACGCCAGAATGGAATCAGGATATCGAGCATTCAAGTGAACCGCTGAGTGCAAGCGAGTCCGGTAAGCGGATCTTCAACCTGAAAGCAGCCTGAAGAACGGTACCTCCTGGATATGGTGTCAACCAGAAAAGCGCACCAGCGACTACAGCAGCGAGCTCCGTCCCTGGAGAACATACCACGAGCACCCTCCACGCTACTGGAAGAACCTTCAAACGGAGGCTCTGCCCAGCTGAACACCAGCCCGAGGACCGTGCCTAAGATAACGGATGAAAAGACCAAGGATATGCCACCGGCCGGTTTCGCTGCATCAAATGCCAATGAGGCAAGCAGCTGAACTGGTGTACCAAGGATAATTGGAAGCATCTCAACAATTCTACCAGCAAGCCAGCGTAAATGCGCCAAAATGCCAGTACCCATCAACATCTACAGCCCCGATGCCCGAGGAACCAATCAATTGGCTCCACCTGGGCATCAGCCAAGCTGCAGAACCCAAAGCCAGAAAACCGAACCGACTGAAAACCGCCATTGGGCATCTTGTGCATCCATTAAACGGCATGAACATAATGAAGAGACGCCACGAAATGACGCCACGCAACGATTAACTCGGATTACGATGCCTATCTGCTGGACTTCCGTTATGGCGATGATTTATGCTATACTCGCCCGAATGGAGCTAAATGCCATTAGGCGTCACGAACGCGCTCTGCAGTCAGGACTTGTGTATCCGGACTGACGATCAATGGGCCCGAAATCGTTAAAAGAGTCCAAATTACCCAGGGTACTCAGATCATCGTCAATATCGATGGTGGACATTATGCGCGGAAAGCCTTGGTCAGATCGAGCACGATTTCGGCTACTGATCAAGGAGATGCACAATATGCGATTGGTGCCGATTCTAACAAATCTTGCAACCGCTGGCAACTATTGTCACGATAAGGTTATATGTGACAAATCTACCGATTCAACAAGTAACTTATCCGAAGCGAATGCTGTCACCTAAAGGTGATCGACATACACTCCTGTCTAACAGAAAGGGCGTGGTGCAGTTGATTGCTTTCAGGCCTCACCACGCCAAGTCACCGGTTCCAAGGAACCTATCTGTTCTGGAACAAAATCGGTCGGCAGCGCTACTGCAAGTTATTGAAACGAGTCTGGAGTATTAA
```

```
>Drosophila_ananassae_XM_001953262.1
ATGGCCGCAGTCACAAGTGAATTACCCAGAGAGGCCACCTACCAACACCAAGAGCGATCGCCTGAGGGCAGTGAACAAACCGTGTGACCACGTGTTGCAGCAGTGGTGCAGAACCCGACTATACGGCATTGGCGATAAATTCAAATCCGTCGCCAATTTCACCGGATCTCAAGGGCGGAAGGAAAGGCAAAGAACCGCCGGAGTCGGTCGATGAGCGCCAGCAGTGCAATTGTCAGACTTTCTCCTCCACCTTGATATCCACGGAGGAGATATACGGAAACAGCAGCCGGAAAAGTCAGAGTTCAACAGCAACAGCATGACCATTATCGAGAGCAACTATATGCCGTAAGGGAGGAGAACCGATATAGACGCCAGAGGTCCGCCATTGTCTACAGAGTGAATATCGGAA
```

CAAACGGGCAACACCTCCGCTGCACGACAATGTCACCGACTTCTGCTCACCAATTCC
ATATGTCACCGCCGAATGCGAGCAGGCAAGAGGGATCTCAATATGAAGCCCAACA
AGGACAATCGTCATCTGTACGACATGGTCTGAATCAAAGCAGCGAACGACTTCAGC
AGCGGTCCGCCCTCCACCGTCGACTACAATAGCAAGCCGGGGCAGCGTGACCGCC
CCGCCACTGGCGGTATCAGTACAAGCGCAGGGCCAATCCCCGCTGGAAAATCTAAATC
AGATGAAACACCTGAACACCGAAAGGCAAGTCCAGGACAAGGCGATTAACGCCAATTG
GCTACCTGAGAAACTGGCAAGGCTGCAGGGAGTCGAATTGGTGTACCAAGGATAA
TTGGAATCATCTCAACAACCTTACCAACAAAGCAACGTCAATGGACACCAAAATCAAG
TGCCACCCAACATTACGTCGATACTCCGGATCGCAGACCCAGGTACAGATCCAGAG
CAACATTGCCAGAACCCACACCCCAGCCAGAACCAAAACTCTCAACAGGATATACAA
TAACCAGAACAGAGAACGTTGAATTAAAGAGACCTCGTACGCTAACACCCACTCCC
CCATGACCAGTGCCACCCACAACGATTGCGATGCCAGACGATCAGCTCCGATTTGACGCC
TTTCTGCTGATTTCCGCTCATGGCGATGATTCTTCTATATCTGGCCCGATGG
GCTAAACTGTCGCTTCAAGCGGTACGAGCGGGCTCTGCAGTCAGGACTCTGTGTCTCGG
GACAAACGATCATTGGGCCAGGAACAGGTTCCGAAAGTCTATTGCCGGAGGGTACA
CAGATCATAGTCAACATCGGATCGTGGACATAATGCCGGCAAGCCGCTGGTCAAAT
CGAACACGATTTCGACTCCTGATCAAGGAGATCCACAGCCGGAGGTGCATCCGATCT
TAACAAATCTGGCACCTCTGGCTAACTACTGTCACGACAAGGTCCTGTGCGATAAGATC
ACAAAGTTCAATAGGTTGTCGAAACGAGGAGTCATCTGAAGTTCATCGACAT
TAATTCTTGTCTCATCAACGAGAAGAAAATGTCCTATTGATTGTTCAAAGTGCAC
CGCGTACCGTCACGGTTCCAAGGAACCGTATCTGTTGAAACAAGATCGGACGGCAA
AGAGTCCTTCAAACAATCGAATCCTGCCTGGAGTACTAA

>*Drosophila erecta*_XM_001980858.1

ATGGCCGCAGTCACAAGTGAATTCCCCAGCAAGCCGATCAGTTGTACCAAGCACCAATAC
ATCCGCCAAATCCAATTATCTAAAGTCCGTAAAAAACGGGTGACCACGTGTTCCAGC
AGTTGCCGAAAAACTCCAGGCATCCGGTCTTCGCAAGAGTTCCCCAGCCGCTC
TACAAGATCTCGTAAAGGCCATTTCGCCCTGTGGTAACGGTTTCAGAGGATCTT
CAAATCGCGCGAAAGACCCAATGCCGGAGTTGTGGAAGGTGCCATTGCCCGCCATG
ACCTCACCAACGCCGGCAGACCAGCCAGCAGTTGCAATCGTAGCCAGACTCTCTCA
TCCACTCAGATTCCACCAAGGAAATCACCTACAGCACAGCAACAAACACCAGCAG
CTATAGCTTCAGGAGCCGGAAAGCGAGAGCTGCAGGAGCAACATGACCATTATCGAGA
GCAACTATATAGCCGTGCGCGAGGATTATCCGACATAGATGTCGAGGTGCGGCCATA
TTGCTGAGCCACGCCAGAATGGAATCAGATATCGAACATCAAGAGTGAATATCGGAA
ACTGACGGGCAATCGTTCCACTGCACGACAACGTGACGGACTTCTGCTGACCATAC
CCAATGTGACCGCCGAGTCAGCGAGTCCGGAAAGCGGATCTCAACCTGAAGGCGAAC
CTGAAGAACCATCACCTCTGGAGATGGTCTCAACCAGAAACAGCCACCGCAGCAG
CGACTACAGCAGCGGAGCTCCATCCGTGGAGAGCATGCCCGTGCACCAACGCTACT
GGAAGAATCCCTCAAGCGGAGGGCTCTGCCCAGCTGAACAAACCTGAACACCGAGTCC
AAGATTACAGATGAGCAGACCCAGGAGACCTCCACCAAGACCCGGCTTCGCTGCACCAAAT
GGCCAATGAGGCAGCGGAGTCGAACCTGGTCTACCAAGGATAATTGGAAGCATCTCAACA
ACTTCTACCAGCAAGCCAGCGTAAATGCCAAAATGCCAGTACCCATCAATATCTAC
AGCGATGCCCGAGGAACGCATCCAAGTGGATGCACCTGGCATCAGCCAGCTGCAG
AACCCAAAACCAGAAAACCAACACCCGACTGAAAACGCCACTTGGCATCTTGTGC
ATCCATTCAAGCGATATGAACGTAATGAAGAGACGCCACGAAATGACGCCACGCCACG
CCCACGCCTACGACCTTATCGAGTGGAACCTACGCCATTCCCTGCTGACGATCAACTC
GGATTACGATGCCATTCTGTTGACTTCACTGATGGCGATGATTGATGCTGTATC
TCGCCGAATGGAGCTTAAGTGGCATTAGCGGCCACGAGCGAGTCCTCAATCAGGA

CTTTGTGCATCCGGACTGACGATCAATGGCGCCGAAATCGTTGAGGAGAGTCCAAC
ACCGAGGGCACCCAGATCATCGTCAATATCGGATCGGTGGACATTATGCCGGGCAAGC
CATTGGTCAGATCGAGCACGACTTCGGCTGCTGATCAAGGAGATGCACAGTATGCGT
TATGTGCCGATCCTCACAAATCTGGCGCCGCTGGCAAACATTGCCACGATAAGGTTT
GTGTGACAAGGTCCACCGATTCAACAAGTTCATCCGAAGCGAGTGCTGTCACCTGAAGG
TCATCGACATACACTCCTGCCTGATCAACGAACGGGGCGTTGTGCCTTGATTGCTTT
CAGAGCTGCCACGCCAAGTCAGCGGTTCAAGGAACCCATTCTGTTCTGGAACAAAAT
CGGTGGCAGCGCTGCTGCAAGTCATTGAAACGAGTCTGGAGTATTAA

>Drosophila_persimilis_XM_002017349.1

ATGGCCACGTATAGAAGTGCCAACAGCGAGCCGTACACGAGTAAGAGCAACAGCAACAG
CAACAGCAGCAGCAGTAGTCGCCTGACGGCATTGAGAAAACGCCTGACCACGTGTTGC
AGCAGTGGTGCCAACATCTCCAGAACGGGCTTCCTCCAAAAGTGCACAAAGTGGCCCC
AACACCAGTTTACAAGATCTTGAGGAGGTATTGACAGCAAGGTATTGGCGACAA
GACCAAATATTCGGCAGTTAAGAAGATTTCAAAGGAACAAGGACTTAGCTAAGA
AAAGTGAGCCAGGGAGGGTAGTGCTAGTGCTAGTGCCCAGAAACCGAGTGACGACACA
GAGGAGAGCCAGAACAGCTGCAGCTGTGGCCAGACTCTTCATCCACACTGAT
ATCCGCGAAGGAGGAGATCTCCTGTAGCTACAACACAGCAGCAGCATGACTATCATCGACA
GCAATTACATCGAAGTGCAGGAGTACCCGGAGATTGACAGCGAGATACGGTCCATA
CTGCTGGCCAATGCCAAAGTGGCATCAGATATCGACCATCAAAATGAATATCGAAA
ATTGACGGGCAATCCGTTCCGCTGCACGAAAATATCACAGACTTTCTGCTACCATCC
CCCACGTGACCGCCGAATGCTGCGTGTCCGGGAAGCGAATTAACTGAAGGCCACGC
CCGGAGACTGCCACCTGTACGATATGGTGTCCATCAGAAGCGAACCAACGACCTGGG
GACCAACGACCATCCGGTGGCTGCACCACCCCCACGCCTCTGGCGCTGTCAGTACAAGC
GTCGGGCCTGTACTCACTGAACACCAACAACAACACTACAATCTCAACCAGAAAT
ATCAATCACATCGAGGAGAACGGTACCGGCTGTTGGCAAAGAAGTCGCCATTGAAGA
CCAGGGGGTCACTGCAGCAACTGCCAGGGCTGCCGGGAGTCACACTGGTGTATC
AGGACAACCTGGAATCATCTCAAGAACTGCTATCAGCGAGCTAACATCTTGCCATCCT
CCAGAGAACCCACGCCGTTGCCGGCCAAGAAAACCATCTCGTAATTCAACAGCCA
GAACCATCGGGATAACCTAAAGAACCAACCGAACGGATTGCTAGTGAACGAACATCATAGTA
TGCAGCAGGAGTATCAGTATCAGTATCCCACCACCCCCACACCCACCATATCG
AGTGGGACTCAGCATGACTCCATGTTCACCATCAACTCGGATTACGATGCCTACTGTT
GGACTTCCCCTACTGGGAGATGATTCCTTGTATCTGGCTCGGATGGAGCTGAGGT
GTCGTTCAAGCGGACCGAGAGGGCCTGCAGTCGGCTGTGCGTATCTGGCAGACT
ATAAGTGGGCCGGTCTGCGTCTGCATATTGCTAGTGAACAAGGGTACGCAAATAAT
TGTCAATATTGGATCCGTGGACATAATGCCGGGAGACCGATTGTCAAATCCAACATG
ACTTCGTCAGTTGGTCAAGGACATGCACAATCGGGACTGGTCCGATCCTAACCA
TTGGCCCCGCTGCCAATTATTGCCAGACAAAGCGATGTGCGATAAGGTGGTCAAAGTT
CAACCAGTTCAATTGGAAAGAGTGCAGCTGATTACCTGAAGGTGATCGATACACTCGT
GCCTGGTCAACGAAATGGCGTCCGGTTCGATTGTTCCAGTACTCATCGCTAAC
GTGACGGGCTCCAAGGAATCCTATGTGTTCTGGAACAAAATTGGTGGCAACGTGTCTT
GCAAATGATTGAAGCGAGTCTGGAGTATTGA

>Drosophila_pseudoobscura_pseudoobscura_XM_001359471.2

ATGGCCACGTATAGAAGTGCCAACAGCGAGCCGTACACGAGTAAGAGCAACAGCAACAG
CAACAGCAGCAGCAGTAGTCGCCTGACGGCATTGAGAAAACGCCTGACCACGTGTTGC
AGCAGTGGTGCCAACATCTCCAGAACGGGCTTCCTCCAAAAGTGCACAAAGTGGCCCC
AACACCAGTTTACAAGATCTTGAGGAGGTATTGACAGCAAGGTATTGGCGACAA

GACCAAATATTCGGCAGTTAAGAAGATTTCAAAGGAAACAAGGACTTAGCTAAGA
AAAGTGAGCCAGGGAGGGTAGTGCTAGTGCTAGTGCCCAGAAAACCAGTGGACGACACA
GAGGAGAGCCAGAAGAACAGCTGCAGCTGTGGCCAGACTCTTTCATCCACACTGAT
ATCCGCGAAGGAGGAGATCTCCTGTAGCTACAACAGCAGCAGCATGACTATCATCGACA
GCAATTACATCGAAGTGGGGAGGTACCCGGAGATTGACAGCGAGATACTGGTCCATA
CTGCTGGCCAATGCCAAAGTGGCATCACGATATCGACCATAAAAATGAATATCGAAA
ATTGACGGGCAATCCGTTCCGCTGCACGAAAATATCACAGACTTCTGCTCACCATCC
CCCACGTGACCGCCGAATGCTGCGTGTCCGGGAAGCGAATTAAACATGAAGCCACGC
CCGGAGACTCGCCACCTGTACGATATGGTCGTCCATCAGAAGCGAACCAACGACCTGGG
GACCAACGACCATCCGGTGGCTGCACCACCCCCACGCCTCTGGCGCTGTCAGTACAAGC
GTCGGCCCTGTACTCACTGAACACCAACAACAACACTACAATCTCAACCAGAAAT
ATCAATCACATCGAGGAGAAGGTACCGGCTGTTGGCAAAGAAGTCACCATTGAAGA
CCAGGGGGTCCACTGCAGCACTGGCCAGGGCTGCGCGAGTCAAACTGGTGTATC
AGGACAACCTGGAATCATCTCAAGAACTGCTATCAGCGAGCTAACATCTTGCCGATCCT
CCAGAGAAGCCACGCCGTTGCCGCCAAGAAAACCATCTCGTAATTCAAACAGCCA
GAACCATCGGGATAACCTAAAGAACCAACCGAATTGCTAGTGAACGAACATCATAGTA
TGCAGCAGGAGTATCAGTATCAGTATCCCACCAACCCCCACACCCACCCATATCG
AGTGGGACTCAGCATGACTCCATGTTCACCATTAACCTGGATTACGATGCCTACTTGT
GGACTTCCCCTACTGGGAGATGATTCCCTTGATCTGGCTCGGATGGAGCTGAGGT
GTCGTTCAAGCGGACCGAGAGGGCTGCAGTCGGGCTGTGCGTATCTGGCAGACT
ATAAGTGGGCCCGGTCGCTGCATCATTTGCTAGTGAACAAGGGTACGCAAATAAT
TGTCAATATTGGATCCGTGGACATAATGCCGGGAGACCGATTGTCCAATCCAACATG
ACTTCGTCAGTTGGTCAAGGACATGCACAATGGGACTGGTCCGATCCTAACACACA
TTGGCCCCGCTGCCAATTATTGCCACGACAAAGCGATGTGCGATAAGGTGGTCAAGTT
CAACCAGTTCATGGAAAGAGTGCAGCTTACCTGAAGGTGATCGATATACACTCGT
GCCTGGTCAACGAAATGGCGTGTCCGGTTGATTGTTCCAGTACTCATCGCGTAAC
GTGACGGGCTCAAGGAATCCTATGTGTTCTGGAACAAAATTGGTGGCAACGTGTCTT
GCAAATGATTGAAGCGAGTCTGGAGTATTGA

>*Drosophila_sechellia_XM_002031933.1*

ATGGCCACAGTCACAAGTGAATTCCCCAGCAAACCGATCAGTTATACCAGCACCAATAC
TTCCGCCAAAACCTATTATCTTAAGTCCGTAAAAAGCGGCTGACCACGTGTTCCAGC
AGTTCGCGAAAAACTCCAGCCATTGGTCTTCGCAAGAGTTCCCTCCAACCGCCTC
TACAAGATCTCGTAAAGTCCGATTCTCCGCATCGGGTAACGCTTCAGAAGATCTT
CAAATCGCGCGAAAACCGAATGCCGGAGTTGGAAGGTGCCATTGGTGGCTCATG
AACTCACCAGCCGGCAGAGCAGCCAGCAGTTGCAAATCGTAGCCAGACTTTCTCGTCC
ACTCAGATTCCACCAAGGAATCACCTACAGCAGCAACAACAACACCAGCGAGACCAA
CATGACCATTATCGAGAGCAACTATATATCCGTGCGCGAGGATTATCCGATATAGATG
CCGAGGTGCGGCCATTGCTGAGCCACGCCAGAATGGAATCACGATATCGAACATC
AAGAGTGAATATCGAAAACGTACGGCAATCCCTTCCACTGCACGACAACGTGACGGA
CTTCCCTGCTGACCATTCCAATGTGACGCCAGTGCAGCGAGACGGGAAACGGATCT
TCAACCTGAAAGCGAACCTGAAGAACGGTCACCTCTGGATATGGTGTCAACCCAGAAA
CAGCGCACCACCAACAGCACCAGCGACTACAGCAGCGAGCTCCGCTCCCTGGAGAACAT
ACCACGGAGCACCACCGCTACTGGAAGAATCCCTTAAACGGAGGGCTCTGCCCCAGC
TGAACACCAGCCAAGGACCGTGGCCAAGATAACGAATGAGAGGACCAAGGAGATCGCC
ACCAGACGGCTCGCTGCACCAAATGGCAATGAGGCAGCGGAGTCGAACCTGGTGCTA
CCAGGATAATTGGAAGCATCTCAACAATTCTACCAGCAAGCCAGCGTAAATCGGCC
AAATGCCAGTACCCATCAACATCTACAGCGATGCCCGGAGGAACCAACCCATTGGCT

GCACCTGGCAACAGCCAAGCTGCTGCAGAACCCAGAGCAAGAAAATGAACCGACTGA
AAACGCCATTGGCATCTTGTGCATCCATTAAACGACATGAACGTATTGAAGAGAC
GCCACGAAATGACGCCACGCCACGATTAAACCACTGAAACCTACAACGATTCTCTG
CTGACGATTAACCTGGATTACGATGCCTATCTGCTGGACTTCCGTTATGGCGATGA
TTTATGCTATATCTGCCGAATGGAGCTGAAATGCCATTAGCGGCACGAACGCG
TCCTGCAGTCAGGACTTGTATCCGGACTGACGATCAATGGCAGGGAAATCGTTG
AAAAGAGTGCATTACCCGAGGGTACCCAGATCATCGTAATATCGGATCGTGGACAT
TATGCGTGGCAGGCCTTGGTTAGATCGAGCACGATTCCGGCTGCTGATCAAGGAGA
TGCACAATATGCGATTGGTGCCATTCTCACAAATCTGCACCGCTGGCAACTATTGC
CACGACAAGGTATTGTGTGACAAGATCTACCGATTCAACAAGTTATCCGAAGCGAGTG
CTGTCACCTAAAGGTATCGACATACTCATGTCAGTCAACGAAAGGGCGTTGTGC
GATTTGATTGTTTCAGACATCACCAGCCAAGTCACCGGTTCCAAGGAACCTATCTG
TTCTGGAACAAAATCGGGCGCAGCGCTGCAAATTATTGAAACGAGTCTGGAGTA
TTAA

>Drosophila_simulans_XM_002104160.1
ATGGCCACAGTCACAAGTGAATTCCCCAGCAAACCGATCAGTTATACCAGCACCAATAC
TTCCGCCAAAACCTATTATCTTAAGTCCGTAAAAAGCGGCTGACCACGTGTTCCAGC
AGTTGCGGAAAAACTCCAGCCATTGGTCTTCGCAAGAGTTCTCATCCAGCCGC
CTCTACAAGATCTCGTGAAGTCCGATTCTCCGATCGGGTAACGCTTCAGAAGAT
CTTCAAATCGGCGGAAAAACCGAATGCCGGAGTTGTGGAAGGTGCCATTGGTGGCTC
ATGAACTCACCAGCCGGCAGAGCAGCCAGCAGTTGCACATCGTAGCCAGACTCTTCTCG
TCCACTCAGATTCCACCAAGGAAATCACCTACAGCAGCAACAGAACACCAGCAGAG
CAACATGACCATTATCGAGAGCAACTATATATCCGTGCGCAGGATTATCCGATATAG
ATGCCGAGGTGCGCGCCATTGCTGAGGCCACGCCAGAATGGAATCAGATATCGAAC
ATCAAGAGTGAATATCGAAAATGACGGCAATCCATTCCACTTCACGACAACGTGAC
GGACTTCCTGCTGACCATTCCAAATGTGACCGCCAGTGCAGCGAGACCGGGAAAGCGGA
TCTTCAACCTGAAAGCGAACCTGAAGAACGGTCACCTCCTGGATATGGTGTCAACCAG
AAACAGCGCACCACCAACAGCACCAGCAGCAGTACAGCAGCGAGCTCCGTCCTGGAGAA
CATACCACGAGCACCACCGACTGGAAAGAATGCCTTAACGGAGGGCTCTGTCCC
AGCTGAACACCAAGCCCAGGGACCGTGCCAAAGATAACGGATGAGCAGACTAAGGAGATC
CCCACCAAGCCGCTTCGCTGCATCAAATGCCAAATGAGGAGCGGAGTCGAACGGTG
CTACCAGGATAATTGGAAGCATCTCAACAATTCTACCAGCAAGCCAGCGTAAATGCGC
CCCAAATGCCAGTACCCATCAACATCTACAGCGATGCCCGAGGAACCAACCCATTG
GCTGCACCTGGCAACAGCCAAGCTGCTGCAGAACCCAGAGCAAGAACCGAACCCAC
TGAAAACGCCATTGGCATCTTGTGCATCCATTAAACGACATGAACGTATTGAAGA
GACGCCACGAAATGACGCCACGCCACGATTAAACCACTGGAACCTACAACGATTCT
CTGCTGACGATTAACCTGGATTACGATGCCTATCTGCTGGACTTCCGTTATGGCGA
TGATTATGCTATATCTGCCGAATGGAGCTAAATGCCGATTAGCGGCCACGAAC
GCGCTCTGCAGTCAGGACTTGTATCCGGACTGACGATCAATGGCGCCGAAATCGT
TTGAAAAGAGTGCATTACCCGAGGGTACCCAGATCATCGTAATATCGGATCGGTGGA
CATTATGCGTGGCAGGCCTTGGTTAGATCGAGCACGATTTCCGGTTGCTGATCAAGG
AGATGCACAATATCGGATTGGTGCCTTCAACAAATCTGCACCGCTGGCAACTAT
TGCCACGACAAGGTATTGTGTGACAAGATCTACCGATTCAACAAGTTCATCCGAAGCGA
GTGCTGTACCTAAAGGTATCGACATACACTCATGTCAGTCAACGAAAGGGCGTTG
TGCGATTGATTGCTTCAGACCTCACCAACGCCAAGTCACCGGTTCAAGGAACCTAT
CTGTTCTGGAACAAAATCGGGCGCAGCGCTGCAAATTATTGAAACGAGTCTGGGA
GTATTAA

>Drosophila_yakuba_XM_002096839.1

ATGGCCGCAGTCACAAGTGAATTCCCCAGCAAACCGATCAGTTATACCAGCACCAATACTTCTGCCAAACCTATTATCTTAAGTCCGTGAAAAAGCGGGTGACCACGTGTTCCAGCAGCTGCGCAGAAAACCTCCAGGCATCCGGTCTTCGCAAGAGCTCTACCAGCCGCCTCTACAAGATCTCGTAAAGGCCATTTCGCCGCCTCGGGTGAACGCTTCAAAGGATCTTCAAATCAGCGCAGAAAACCGAATGCCGGAGTTGGAAGGTGCCATTGCCCATGACCTCACCACTCAGATTCCACCAGGGAAATCACCTACACCAACAGCAACAAACAGCAGCTATAGCTTCAGGAGGCGGAAAAGCGAGAGCCACCGGAGCAATATGACCATCATCGAGAGCAACTATATAAGCGTGCAGGATTATCCGATATAGATAACGAGGTGCGGCCATAATTGCTAGCCACGCCAGAACATTTCCACTGCACGACAACGTGACGGACTTCTGCTGACCATTCCAATGTGACGGCAATAACATTTCCACTGCACGACAACGTGACGGACTTCTGCTGACCATTCCAATGTGACGGCCAGACTGCAGCGAGTCAGTCCGGAGCGGATCTTCAACCTGAAGGCGAACCTGAAGAACCGTCACCTCTGGACATGGTCTCAACCAGAAGCAGGCACCGGCACAGCCAACGGCACCAGCGACTACAGCAGCGAGCTCAGTCCGTGGAGAACATGCCCGGGACCAACACCAGCTACTGGAAGAATTCTTCAAGCGTAGGGCTCTGTCAGCTGAACGACAACCTGAACACCGTGCCTAACGCTTACGGATGAACAGACCAAGGAGAGCGCCACCAACACGGCTTCGCTGACCAAAATGGCAATGAGGCAGCGAGTCGAACCTGGTCTACCAGGATAATTGGAAGCATCTAAACAACCTCTACCAGCAATCCAGCGTAAATGCCCAAAATGCCAGTACCCGTCAACATCTACAGCGATGCCCGAGGAACGAATCCATTGGATGCACTGGCATCAGCCAAGCTGCAGAACCCAAAACCAACAGAAAACCAACAGACGACTGAGAACCGCCACTTGGGCATCTTTGTGCATCCATTCAACGATATGAACGTAATGAAGAGACGCCACGAAATGACGCCACGCCACGATTTAACCGACTACAGCGATTCTCTGCTGACGATTAACCTCGATTACGATGCCATTCTGGACTTCCGCTTATGGCGATGATTATGCTATATCTCGCCGAATGGAGCTAAAGTCCGATTAGACGGCACGAGCGAGTCCTTCAGTCACTTTGTGATCCGGACTGACGATCAATGGTCCGAAATCGTTGAAAAGAGTGCAACTACCGGAGGGTACGCAGATCATCGCAATATCGGATCGGTGGACATTATGCCGGCAAGCCTTGGTCAGATCGAACACGATTTCGGCTGCTGATCAAGGAAATGCACAATATGCGAATTGTCGCGATCCTCACAAATCTGGCACCGCTGCCAACTATTGCCACGATAAGGTATTGTGTGATAAGATCCACCGATTCAACAAGTTATCCGAAAGCGAGTGCTGCCACCTAAAGTCATCGACATACACTCCTGCCTGATTAACGAACGGGGTGTGCGATTGATTGTTTCAAGAGCTCACCACGCCAGTCAGCGGTTCAAGGAACCCATTCTTCTGGAAACAAACCGTGGCAGCGTGCAGTTATTGAAACGAGTCTGGAGTATTAA

>Drosophila_biarmipes_osk (scaffold_JH400370.1)

ATGGCCGCAGTTGGAAAGTGAATTCCAAGCGAACCGATCAGTTGAGGAACACCAACAC TCCC GCCAAACCGATCGCCTGAGGTCACTGAAAAAACGTGTGACCACGTGCTTGCAGCAGTTGCGCAGAAAACCTCCGGCAGCCGGATCCTTCGCAAGAGCTCCACCGGCCGTTCTACCAAGGTCTCGTCAGGCCGACTTTCCGCCTTGGGAGAGGTTTCGGAAGATCTTAAAGTCGGCGCGAAAACCGAATCGCCGGAGTTGGAAGGTGCTCTGCCCATGACGTCACTGGGCCAGCAGCCAGCATTGCACTGAGCCAGACTCTTCTCATCAACTCTGATTTCACCAAGGAGATCTGTACAGCAACCACAGATATACTGCAGGAGCGGGAAAAGCGAGAGCTACAGCAGCAACATGACCAATTATCGAGAGCAACTATAGCGGTGCGCGAGGAGCACCCGATATCGATAAGGAGATTGCGCCATATTGCTGAGCCACGCCAGAACATGGAAATCAGGATATCGAGCATCAAGAGTGAATATCGGAAACAGACGCCAACCCCTTCCACTGCACGACAACATCACGGACTTCCGCTGACCACTCCAAACGTGACCGCCAGTGCAGCGAGTCGGGCAAGCGGATCTCAACCTGAAGCCCAGCCTGAAGAACCGCCACCTGCTCGAA

ATGGTGCTCAACCAGAAGCAGCGCTCCAGCGACTGCAGCAGCGGAGCTCCCTCCGTGGA
GGACCAGCCACGTGCCCAACCACGCTACTGGAAGAACATCCCTACAAGCGGCGGGCCCTCT
CCCAGCTGGACAACAACCTGAACACCCTGCCCCAGCTCACGGCTGACCAGACCAAGGCG
GTCGACACCAGGGCGGCTCCGCTGCAGCAACTGGCCAAGGAGGCAGCGGAGTCGAACCTG
GTGCTACCAGGATAATTGGAATCACCTGAACAAACATTACCGATCCAGCGTAAGCG
AGCCCGAGATAACAGCTGCCGGTCCCCATCAACATCTACAGCGATGCCCGAGGAACCG
AACCAGATGAATGTTCCCCAGCACCACCCACTCTGCCATAACCGCAACCAGAACATCCGAG
CCAGAACCGAGCAGGCCAACACAGGACGCAAACACCAGCACCTGGGAATCTTGTGCAGCCGT
TCAGCGACATGAACGTGCTGAAGAGACGCCAAGAGATGACGCCACGCCACGCCCTA
TCGAGTGGCACCTACAACGATTGATGCTGACGATCAACTCGGACTACGACGCCCTACCT
GCTGGACTTCCGCTGATGGCGACGACTTCATGCTCTATCTGCCCGCATGGAGCTCA
AGTGCCGGTTCAGGCGCCACGAGCGAGTCCTGCAGTCGGCTCTGCGTGTCCGGACAG
ACGATCAACGGGGCCCGGATGCCGTGAAGAAGTCCATCTGCCCGAGGGCACCCAGAT
CATAGTCAACATAGGGCGGTGGACATATTGAGGGGAAGGCCCTGGTCCAGATCGAAC
ACGACTTCGACTGCTGATCAAGGAGATGCACAGCCAGCGATTGCTGCCATCCTCACG
AACCTGGCGCCGCTGCCAACTACTGCCACGACAAGGAGCTGTGCGACAAGGTCAACCG
ATTCAACAAGTCATCCGGAGCGAGGGCAGACACCTCAAGGTGGTGGACATACACTCCT
GTCTGATCAACGAGAGGGCGTGTGCGATTGATTGTTCCAGCAGGAGGCCCTACCTG
TTCTGGAACAAAATCGGGCGACAGCGGTGCTGCAAATCATCGAATCGAGTCTGGAGTA
CTAA

>*Drosophila_bipectinata_osk* (scaffold_JH401929.1)

ATGGCCGAGTCACAAGTGAATTACCTAGAGAGGCCACAACACCACCAAGAG
CGATCGCCTGAGGGCAGTAAAAACGTGTGACCACGTGTTGCAGCAGTGGTGCAGA
AGCTCCAGGCCGCCCTTCGCAAGAGTTCTCCAGCCGTTCTACAGGATCTC
GTGAAGGCCGATTATAACGGCAATTGGCGATAAAATTCAAATCCGTTGCCAATTCCACAA
GATCTCAAGGCCGGAAAGGAAAAAAAGCAAAGAACCGGCCAGTCGGTCAAG
AGCAGTTGCAGTTGTGGCGAGACTCTTCTCCTCCACATTAATTCCACGGAGGAGATT
TACGGTAACGGAAAAACCGACTACAGCAGCCGAAAGTCAGAGTTACAACAGAACAG
CATGACCATTATTGAGAGCAACTATATAGCCTAAGGGAGGAGAACATCCGATATAGACG
CAGAGGTGCGGCCATTGCTATCTCATGCCAGAATGGCATTACGATATCGAGCATT
AAGAGTGAATATCGGAAACAAACGGCAATACCTCCGCTGCACGACAATGTCACCGA
TTTCCTGCTAACCATTCATATGTCACCGCGAATGCGAGACGGCAAGAGGATT
TCAACATGAAGCCCAACAAGGATAACCGTCATCTGTACGATATGGCTCTTAATCAAAG
CAGCGCAACGACTGCAGCAGCGGCCCTCCGTGGAGTACAATAGACCTGTGGG
GCAGCGTGGACGCCCGTCACTGGCGGTATCAGTACAAGCGCCGGCCAAATCCCCGC
TGGAAAACACAAACCGATCAAACAACTGAACACCGAACAGAATGGTATGGAGAAGCTC
CAGGACAATGTCAAATCAGAGCGATTAAGACCAATTGGCTCCCTGCAGAAACTGGC
CAAGGCTGCAGGGAGTCGAATTGGTGTACCAAGGATAATTGGAATCATCTAACAAATT
TCTACCAACAAAACAACGTGAATGGACCACAAATCCAGTGCACCCACATTTACATC
GATACTCCGGATCAGCCCACCCAGGTGACAGATCCAGAGCTACTCATTGCCAGAACCG
AAATCCCAGCCAGAACCAAAACTACCAACCCAGGATTATATAATAACCCATAACAGAGATA
TGCTTGAAATTAAAGAGACCTCGTACTTAACACCCACTCCACCGACAGTGGCACC
CAAATGACTCGATGATGACTATCAGCTCGATTGATGCCCTTCTGCTTGTGATT
GCTCATGGCGATGACTTTCTATATCTGCCCGCATGGAGCTTAACGCGCTTCA
AGCGATAACGAGCGGGTCCGTGACTCAGGACTTTGTGTCTCCGGACAGACAAATCATCGGT
GCCAGAAAACAGGTTAGGAAAGTTACTTACCGGAGAACATCCAGATCATAGTCAACAT
CGGATCGGTGGATATAATCGGGGGCAAGGCCCTCGTCCAGATCGAACATGATTAGAT

TGCTAATCAAGGAGATTACAGCCGGAGATGCATCCGATCTAACAAATCTGGCACCT
CTGGCCAACACTGTACGATAAGGTCTGTGCGATAAGATCTCAAATTAAAGTT
TGTACGCAACGAGGCAAGAAGTCATCTGAAGTTATCGATATTAAATTCTGTCTCATAA
ACGAAAAGAAAAATGT CCTATTGATTGTTCAAAGTGCACCACGTACCGTCACGGGT
TCCAAGGAACCGTATCTCTTTGGAACAAATAGGACGGAACGTGTCCTCCAAACAAT
CGAATCCTGCCTGGAGTATTGA

>*Drosophila elegans*_osk (scaffold_JH110107.1)
ATGGCCGCAGTCAGAAGTGAATTCCCCAGCGAACCGAACAGTAATACCAACACCAATAC
CTCCACCAAAACTGATCGCCTGAGGTCCGTGAAAATCGGGTGACCACGTGTTGAAGC
AGTTGTGCGAAAACCTCCAGGCAACTGGATCTTCGCAAGAGTTCTCCAGCCATT
TACAAGATCTCGTAAGGCTGATTATTCCGGTTACGGGAGAAATCCAAATTGCG
CCAGTTACAGAAGATCTATAATCGACCGAACCGAACATGCCGGAGTTGTGGAAAG
TGCCGTTGGCCCATGACGTACCCGGCAGAGCAGCCAGCAGTGCACATCGTAACCAGA
CTTTCTCATCCACTCTGATTCCACCGAGGAAATCACCTACAACACCAACAAAGAA
CAACAACAGCTGAAAAGTCGAAAAGCCATAAAAGCTACAAAGCAGCATGACCA
TCGAGAGCAACTATAGCTGTGCGGAGGAGTTCCGATATCGACACCGAGGTGCGC
GCCATATTGCTCAGCCACGCCAGAATGGGATCACGATATCGAACATCAAAGTGAATA
TCGAAAGCAGACGGCTGCCAGTTCCACTGCACGACAACGTGACGGACTTCCTGCTGA
CCATTCCGAACGTGACCGCCGAGTGCAGCGAGTCCGGCAAGAGGATCTAACCTGAAG
CCGAACGTGAAGAACCGCCACCTGCTCGACATGGCCTCAACCAGAACAGCAGCGCACCAG
CGACTACAGCAGCGGAGCTCCCTCCGTGGAGGACAAACCTCGTCCGCCACCACGATACT
GGAAGTTCCCCTACAAACGGCGGGCTTGTCCCAGCTGGAAAACAACTGAACACCGGT
CCAAAACGTGAAGGAACAGGAACACCAAGTCCAGGGCCACCAAGGCGCTCCACT
GCAGCAACTGGCAAGGAGGCAGCGGAGTCGAACGGTGTACCCAGGACAATTGGAAAC
ATCTCAACAACCTCTACCAGCAATCCAGCGCAAACGAACCGCAGATGCCAGTGC
AACTTCTGCTACAGAGGAACGAACTCAGTTGCTGTTCCGAGCCGGGAC
ATGCGGCAAGAACCAACCAGAACATAACAAACCAATGTAATCAGCAACTGGGATCAT
TTGTGAACCCATTAAATGATATGAATATACTGAAGAGACGCCACGATATGACGCC
CCCACGACTTATCCAGTGGAACGTATAACGATTCAATGCTGACCATCAGTCGGATTA
CGATGCCCTACCTGCTGGACTTCCGCTCATGGCGATGATTGCTATATCTGCC
GCATGGAGCTAAATGCCGATTCCAGCGACACGAGCGTATCCTCAGTCCGGACTTGC
GTCTCCGGCAAACGATCAATGGGCCGGCGGGTTGAAGAAAGTTCAACTGCC
AGGCACCCAGATCATCGTAACATCGGCTCCGTGGATATTGTGCGGAAGGCCTTG
TTCAAATCGAACACGATTCCGGCTGCTGATCAAGGAGATGCATAGCCAGCGATTG
CCCATTCTCACAATCTGGCCCCGCTGCCAACTATTGTCATGACAAGACTATGTG
TAAGATCTATCGGTTCAACAAATTCACTCCGAGCGAGTGTGTCACCTGAAATACATCG
ACATACACTCCTGTCTGATCAACGAAAGGGCATCGTGCAGGTTGATTGTTCAAAGC
TCACCAAGCCAAGTCACTGGTTCCAATCAGCCATATCTGTTCTGGAACAAATCGGGCG
GCAACGTGTGCTGCAAACATTGAATCAAATCTGGAGTATTGA

>*Drosophila eugracilis*_osk (scaffold_JH402624.1)
ATGGCCACAGTCACAAGTGAATTCCCCAGCAAACCGATAAGTTATACGAATACCAATAC
CAATACTCCACAAACCTATTGCTGAGGTCACTGAAAAGGCGGGTGACCACGTGTT
TGCAGCAGTTGCGCGAAAACCTCCGGCAACCGGATCATTGCAAGAGTTCCACCAGC
CGTTTCACAGGTTTCTATAAAAGTCAGATTCCGACTTAGTGAAAAGCTGAGAA
GATCTTAAATCGGCGCGAAAATCTGAATGCCAGAGTTGCGGAGAGTTCCATTGAGCA
CCCATGACCAGCAGACCAGCGGGCAGAGCAGCCAGCAGTCAGGCAAATCGTAGCCAGACTT

TTCTCATCCACATTGATTCCACCAAAGAAATCACCTTAACAACCACACTAAAAATAG
CTTAGCTATAAGTCAGGAAAAGCAGGGTTACAGCAGCAATATGACCATCATCGAGG
GCAACTATATAGACGTGCGCGAGGAGAATCCCATTGATGCCAAGTGCAGGCCATA
TTGCTTACCCACGCCAGAATGGAATCACGATATCGAGCATCAAGAGTGAATATCGAA
ACTGACGGGCAATCCCTTCCACTGCACGACAACGTGACGGACTTCCTGCTGACCATT
CCAATGTGACCGCCGAGTGCAGCGAGACTGGCAAACGGATCTCAACTTGAAGCAAAC
CTGAAGAACCGACACCTCTCGACATGGTGCTCAACCAGAAGCAGAAGCCCAGCGACAA
CAGCAGCGAAGCTCCAACGGTGGAGGATAAATCCCGTGTACCGCCGCGCTACTGGAAGA
ATCCCTACAAGCGTGTGCACTTCCCAGCTAGACAACAAACCTAAACACCCTCACAAAG
CTAACGGATGACCAGGCCAAGGCAATGTCCATGCATACCAGAGCGGCTCCACTGCAGCA
ATTGCCAAGGTGGCAGGGAGTCGAACGGTGTACCCAGGACAATTGGAAGCATCTCA
ACAATTCTACCAGCAATCCAGCGTAAACGAGGCCAGAGGCCTGTTCCCATTACATC
TACACCGATGCCTGGAGGATTGCCAATTCACTGGATAATCCCGAGCATCACCCAAA
CTGAAAAACAAACACCAAAATATCAAACAAAACGTTAACAGCATCTGGAAATTTG
TTAACCCATTAAACGATGTTAACCCATTAAAAGACGCCAAGATATAACGCCACACCA
ACAATTTTTACCACTGGAGGTTCTAGCAATAATTCCCTGATGACCATCAATTGGATT
CGATGCCCTTCACTGGACTTCCCTGATGGCGATGATTATGTTATATCTGCC
GCATGGAGTTGAAATGTAGATTCAAGAGGCACGAACGTGTCCTCAGTCAGGACTTTGT
GCATCTGGACAGACGATAATGGTGCCTGAAATCGGTTGAGGAAAGTCCAACCTCCGA
GGGCACACAGATTATTGTTAATATAGGATCGGTTGATATAATGAAAGGAAACCTTAG
TTCAGATAGAACACGATTCCGATTGCTGATCAAGGAGATGCATAATCAGAGATATATA
CCGATCCTTACAAATCTAGCACCTTGGCCAATATTGCCATGATAAGGTTGTGTGA
TAAGATCACGCGATTTAATAAGTTATAGCAAGTGAAGGCAGGCACCTGAAGGTATTA
ATATACACTCCTGTTGATCAACGAAAGGGGCATCGTGAGATTGATTGTTCAAAGG
AACCTTATCTATTCTGGAACAAATCGGGCGCAGCGCTCTGCAAGTTATTGAATCG
AGTTGGAGTACTAAGTCGGGTGATTGA

>*Drosophila_ficusphila_osc* (scaffold_GL987928.1)
ATGCCGCAGTCAGAAGTGANCTGGAGGTCGGTAAAAAAACGAGTTACCAACGTTGAAAGAAATTGT
GCGAAAAACTCCAGGCAGCCGGATCTTCGCAAGAGTTCCAGCCGATTCTATAAG
ATTTCTGAGGGCCGACTATTCCGATTGCTCCAAGTTGCGCCGGTTTCAAGAAAAT
CTTAAATCGGCAGAAAACGAATGCCGGAGTCCTGGAAAGTGCCGTTGGCTCATG
ACGTACCAGGCAGAGCAGCATTCACTCAGTTGCAAATCGTAGCCAGACTTTCTCATCC
ACCCAGATTCCACCGAGGAATCATCTACAACACATCAGCTCAACAGCAGAAAAAG
CCAGAGCCACAGCAGCAACATGACCATTATTGAAAGCAATTATAGCTGTGCGGGAGG
AGTACCCGATATTGATGCCAGGGTGCAGCCATATTGCTGAGCCACGCCAGAATGGG
ATTACGATATCGAACATCAAGAGTGAATATCGAACACAGACGGCAACCCCTTCCAGT
GCAAGAAAACGTGACGGACTTCCTGCTGACCATTCCAATGTGACGCCAGTCAGTG
AGTCCGGCAAGAGGATCTCAACCTGAAGCCAACACGAAGAACCGTCACCTGCTGGAC
TTGGTTCTCAACCAGAACAGCAGCGACCAGCGACACCAGCAGTGGCGCTCCGTGG
GAGCCATTCACTGGCACCTCCCTACTACTGGAGGAATTGCTACAAGCGGAAGGCTTT
CCCAGCTGGACAGCAACATGAAACATCAGGCCGAAAACGGCGGACTACCAGGCCAAGGAT
ATAGCCACCAAGGCGGCTCCGCTGCCAACTGGCAAGGAGGCAGCGGAGTCGAATTG
GTGCTACCAGGACAATTGAAACATCTCAACAACTACTACCAGCAATCCAGTGC
AACCGCAGATTCCAGTGCCATCAACATCTATACCGATGCAACAGAGGAACCAACCAA
TTTGCTGTTCCGAGCACCTGCCACCCCTGCAAGAACCCACAAGCAGAACACCAA
CGATAACCAACATTGGCATTGTTATGCGTCCCTATAATGATATGAACGTACTCAAGA

AACGACGCGATTGACGCCAACACCCACGACTTATCCAGCGAACCTTAACGATTG
ATTCTGACGATGAGTCGGATTACGATGCCATCTACTGGACTTCCGCTCATGGCGA
TGACTCTGCTATATCTGCCCGCATGGAGCTCAACTGTCGATTCAAGCGATACGAGC
GGGTCCCTCAGTCAGGACTTGCCTCCGGCAAACGATCAATGGCGCCGAAACCGG
TTGAAGAAAGCAAACCTCCGAGGGACGCAGATCATTGCAATATTGGTTCTGTGGA
CATAAATGAGGGCAAGCCCTGGTTAGATCGAACACGATTCCGGCTGCTCATCAAGG
AGATGCACAACCGCGCTTGTGCCATCTGACCAACCTGGCACCGCTGGCGAACTAC
TGCCACGATACTGCATGTGATAAGATGTCCCATTCAACAAGTCATCCGAGCGA
GGGGTGTCACTGAAGATCATTGACATCCACTCCTGCCTGGTCAACGAAAGGGCATAG
TGCATTGATTGTTCAAAGTTCACCACGGCAGGTCACTGGTTCTAAGGAGCCATAT
CTGTTCTGGAACAAAATCGGGCGGCAGCGCTGCAAACAATCGAAACGAATCTGGA
GTATTAA

>Drosophila_kikkawai_osk (scaffold_JH111367.1)
ATGGCCGCAGTCAGAAGTGAATTCCCCAGCGAGTCGAACAGTAATCGAGTACCAAAGC
CGACAGTTGAGGTCGTTGAAAAAACACGTGGCCACGTGTTGGCACAATTGCAAAA
AACTCCAGCAGCCGCCTCTTCGCAAGAGTTCCATAGCCGTTCTACAAGATCTT
TTGAAGGCCACTGCTCTGCAGGTTCCGGAAAAGTCGAAATTTCAGCCAATTCA
CCGGATCTCAAGGCAGGCCGGAAAAGCGAAGGCCGGTGGTAACCGGACCCAGGCCGG
AGTTGTGAAAGAAGTGCCGTTGGCGATGAGCAAAGAAGCCAAGAGCAATTGCAAATC
GTAGCGAGACTATTTCCTACTCTAACTCCACCAAGGAGGAGGATCATCACCAC
AAGCAGGAGTAGCAATTGTAACACCAAGCTACAGGAGAAGAAAAGCCAGAGCTACTGCA
GCACCAGCACGATGACCATAATCGAAGGCAACTATAGGAAGTGCCTGACGAGTATCCC
GACATTGATGCCAGGGTGCCTGCATATTGCTAAGCCATGCCAGAATGGAATCACGAT
ATCGAGCATCAAAGATGAATATCGCAAACAGACGGCAACACCTTCCCTGCACGACA
ACGTGACAGACTTCTGCTTACCAATTCCGTATGTGACCGCTGAGTGCAGTCATACTGGC
AAGCGGATCTCAACCTGAAACCCAACAAGGAGAATAGCCACCTGCTGGAGATGGTACT
TAACCAGAAGCAGCGCAGTGATTACAGCAGCGGAGCACCCTGCATAGAGGACAAATACC
GAATGGGTCCACCGCAGGTGCCACGCTACTGGCTAACGGATTCAAGCGCGGGCA
CTGTCCTGGTCAAGGCAACACCTCAATGTCAAGGTCAACTCGGAGAAGCCGGGATGA
CCAGAGCAAGGCATCAACACCAAGATGGCAGCGCTGCCATATGCCAAGGCCAG
CGGAGTCGAACTGGTCTACCAAGGACAATTGGAATCATCTAAAAATATCTACCAAGCAA
TCCAGGGAAATGCACCACACATGCATGAGCCAACCAGATTATACTGATCCCCGG
GCTGTCAAACCACTGGCTAACCGGAGCTGATCAAGCCACCCCTGCAAGAACATAACA
GCCCGAATCTCAACCAGATACTCAACCCAAACCTGGACATCTTAGCGTCCACGTTCC
TCCTGCGACTTGAAGCGGCCACGAGAGAACGCCACACCCACGACCATGTCCTCTGC
AACTCATAACGATTGATGCTGACCATAGCTGGACTATGATGCCATCTTCTGGACT
TTCCTCTAATGGCGATGATCTACTTTATATCTGCCGAAATGGAGCTCAAGTGCCGA
TTCAGGCCTTACGAACGGGCTGCAGTCGGACTTGCCTGCCAGACGATCAA
CGCTGCCGGAATCGGCTGAAGAGGGTCAACCTGCCAGGGCACCCAGATAATTGTGA
ACATTGGTTGGATATACTACGGGCAAGCCGCTGTGAGATCGAGCACGACTTT
CGGCTGCTCATCAAGGAGATGCATAACCGGCGATTGTCCGATACTCACCAATCTTGC
CCCGCTGGCAAACCTATTGCCACGACAAGGGCTGTGAGCAAGGTCTGCGATTAAACA
ACTTTATTGCTCCGAGGGCAGGAGTCACCTGAAGGTGATTGATATACTCGTGCCTG
ATCAACGAGCAGGGCGTCGTTGCTTCGATTGTTCCAGAACGCTCCCGCTGTGTC
GGGTTCAAAGGAGCCCTATTGTTCTGGAACAAAATCGGCAGGCAGCGCTGCTGCAGA
TGATTGAATCGAGTCTGGAGTACTGA

>Drosophila_rhopaloa_osc (scaffold_JH406433.1)
ATGGCCGCAGTCAGAAGTGAATTCCCCAGCGAACCGTACAGTAATACCAACAACAATACTC
CTCCACCAAAACCGATTGCCTGAGGTCCGTGAAAAAACGAGTGACCACGTGTTGAAGC
AGTTGTGCGAAAAACTCCAGGCACCCGGATCTTCGCAAGAGTCTCCAGCCGATTCT
TACAAGATATTGCGAAGGCCATTATTCCGGATTAGGGAGAAATCGAAATTGTCGCG
CCAATTTCAGAAGATCTCAAGTCGGCGAAAACCGAGTCGCCGGAGTTGGAAAG
TGCCATTGGCCCATGACGTCACTGGGCAGAACAGCCAGCAGTGCAGATCGAAATCGTAGA
CTTTCTCATCCACTCTGTTCCACCGAGGGAGATCACCTATCACACCAACATCAACAG
CTACAAGAGCCGGAAAAGTCAGAGCTACGACTGCGACATGACCATTATCGAGAGCAACT
ATATAGCTGTGCGGAGGAGTATCCGATATCGACACCGAGGTTCGCGCCATTGCTC
AGCCACGCCAGAATGGAATCACGATATCGAACATCAAAGTGAATATCGAAAACAGAC
GGGCATTCACTGTTCCACTGCACGACAACGTGACGGACTTCCTGTTGACCATTCCGAATG
TGACCGCCGAGTGCAGCGAGTCCGGCAAGCGGATCTCAACCTGAAGCAAATGCGAAG
AACCGCCACTTGCTGGACATGGCTCTCAACCAGAACAGCAGCGACCAGCGACTACAGCAG
TGGAACTCCTCCGTGGAGGACAATCCTGTCGCCACCACGCTACTGGAAGTCCCCT
ACAAACGTCGAGCTATTCCAGCTGGACAACAAACCTGAACACCAGGCCGAAACTGGAG
GATCACCAGTTCAAGGCATAACCGCAAGACGACTCCACTGCACCAATTGCCAAGGA
GGCAGCGGAGTCGAACTGGTGCTATCAGGACAATTGAAACATCTCAACAACTCTACC
AGCAATCCAGCGCAAATGAACCGCAGAGGCCGTGCCATCAACATCTACAGCGATCCT
CCGGAGGAGCGGACTCAGTCGAGTCCCGAGCTCCACACACTGCAGGAACCCCAA
CCAGAATACCAAACAGAACGTAACCGAGCATTGGAAATCTTGTGAATCCATTCAACG
ATATGAATGTACTGAAGAACGCCACGATATGACCCCCACACCCAGCATTCCAGT
GGAATCTACAGCGATTGATACTGACGATCAGTTCGGATTACGATGCCCTACCTACTGGA
CTTCCACTCATGGCGATGACTTATGCTATATCTGCCGTATGGAGCTCAAATGCC
GATTCAAGGCCATGAGCGGGTCCTCAGTCAGGACTCTCGGTCTCCGACAAACGATC
AATGGCGCCCGAAGCGGTGAAGAACGTCATCTGACCGAACGGACCCAAATCATCGT
GAACATTGGATCGGTGGATATATTGTGCGGGAGGCCGTTGGTCAAGATGAAACACGATT
TCCGGCTGCTGATCAAGGAGATGCACAGCCAGCAGTGTGCCATCCTCACAAATCTG
GCACCGCTGCCAACTACTGCCACGATAAGATCCTGCGACAAGGTCATCGTTCAA
CAAATTCACTCCGCAGCGAGGGAAAGTCACCTGAAATACATCGACATTCACTCCTGCCTGA
TCAACGAAAGGGCATTGTGCGATTGATTGTTCCAGAGCTCACCCGCCAAGTCAC
GGCTCCAAGGAGCCATATCTGTTCTGGAACAAAATCGGGCGGCAGCGCGTGCTGCAAGT
TATTGAATCGAGTCTGGAGTACTAG

>Drosophila_takahashii_osk (scaffold_JH112313.1)
ATGGCCGCAGTCAGAAGTGAATTCCAAGCGAACCGATCAGTTATAACGAACACCAATAC
TTCCAACAAATCCAATTGCCTGAGGTCCGTGAAAAAACGTGTGACCACGTGTTGCAGC
AGTTGCGTAAAAACTACCGGCATCCGGATCTTCGCAAGAGTCTCCAGCCGTTTT
TACAAGATCTCGTAAAGGCTGATTTTCCGCCTTGGAGAGCGCTTCAGAAGATCTT
TAAATCGGCGCGAAAAACCGAATGCCGGAGTTGTGGAAAGTGCCGTTGGCCATGACG
TCACCGGGCAGAGCGGCAGCAATTGCAAATCGTAGCCAGACTTTCTCATCCACGCTG
ATTTCCACCAAGGAAATCCACTGCAACAAACAGCTACAGCTACAAGAGCAGGAAAAG
CGAGAGCTACAGCAGAACATGACCATCATCGAGAGCAACTATATAACCGTGCAGGAGG
AATTTCGGACATCGATGCCGAGGTCGCGCCATATTGCTGACCCATGCCAGAATGGAA
ATCACGATATCGAACATCAAAAGTGAATATCGCAAACAGACGGGCAACGCATTCCACT
GCACGACAACATAACGGACTTCCTGCTGACCATTCCCCATGTGACGGCGAGTGCAGCG
AGTCCGGCAAGCGGATCTCAACCTCAAGCGAACCTGAAGAACCGCCACCTGCTGGAG
ATGGTTCTCAACCAGAAACAGCGCCAGCGACTACAGCAGCGGAGCTCCGTCTGGAA

GGACAAACCCGAGCACC CGCGCTACTGGAAGAATCCCTACAAACGGCGGGCTCTT
CACAGCTGGACAACAA CCTGAACACCGTGACAA ACTAACGGATGAGCAGACCAAGGC
ATCAACACCAAGGCGGCTCGCTGCAGCAACTGGCAAGGAGGCAGCGGAGTCGA ACTG
GTGCTACCAGGATAATTGGAATCATCTCAACAACTTCTACCA CAGCAATCCAGAGTAAATG
AACCA CAGATGCCGGTGC CCATCAACATTACAGCGATGCAACAGAGGAGCCAATCCAG
TTGGATGCTCCCAGGCATCACCCACACTGCCAGAACCAACGAGAATCAGAATCCGAA
TCAGAATCAGAATACCAAACAGAACGTAACCAGCACCTGGCATTTCGTGCAGCCTT
TCAATGATATGAACATTCTGAAGAGACGCCACGA ACTGACGCCACACCCACGACATT
TCGAGTGGAACCTACAACGATTGATGCTGACGATCAATTGGACTACGATGCCTTC
GCTGGACTTTCCCCTCATGGCGATGATT TATGCTATATCTGCCCGATGGAGCTCA
AGTGTGTTTCAGGCGCTACGAGCGATCCTCAGTCGGCCTTGC GTGTCCGGACAG
ACGATCAACGGCGCCCGGAATCGGTTGAAGAAATTCCACCTCCGAGGGCACCCAGAT
CATCGTGAATATCGGATCGGTGGATATATTGCGGGGCAGGCCTTGGTTCAGATCGAAC
ATGATTCCGGCTGCTGATCAAGGAGATGCACAACCAACGATATGTGCCCATCCTCACG
AATCTAGCACCGCTGGCAA ACTATTGCCAGCAGAACGGATCTGTGCAGAACAGTCAGCCG
ATTCAACAAGTCATCCGGGCCGAGGGCAGACACCTGAAGGTCGTTGACATACACTC
GCCTGATCAACGAAAGTGGCATCGTGCATTGCTTCAGAGCTACCACGCCAA
GTCACCGGTTCCAGGGAACCATATCTGTTCTGGAACAAAATCGGCCGTAGCGCGTGCT
TCGAATTATTGAATCGAGTCTGGAGTATTAA

>Drosophila immigrans DQ823084.1

ATGGCAGACATTCAAGAAGTGAATTAAACAGCGTGCCGAACACTCACCCCCGATCGCCTGAC
GGCTTCCGAAAAAAAGCTAACCAACGTGTTCCATCAGTGGGCCAGAAACTGCAGCATT
TTAACGACGACTACTCCACGTTGCGAGAAAAAGTATTGCTGGCAGAAATTCAAAAAG
ATTACACAGCAACAAAGCTGATTAGCCAGGAAGAGGTCGAAGAGCAGCAGCTACAAAT
TGTGGCCACGCTATTTCATCGACACTGATATCGAAATACACTAGCTCAAAGAATTCCA
GCAGCTGCAACATGACCATCATGGATGATTGCTATTGGGTGCGGGACGAGTACCCC
GATATAGATGCTGAAATACGCCATATTGCTGCCAATGCCAGAAAGGAATCACGAT
ATCGAGCATCAAAAGTGAATATCGAAATATGACGGGCAATGCATTCCACTGCGCGAGA
ACATCACAGACTCCTGCTGACCATAACCGTACGTCACGGCCGAGTGCTGCATGTCTGGC
AAGCGCATCTCAACATCAAGCCCACGGAGGACACCCGTCACTGACGACATGGTCCT
GAATCAACGGCAGCGTAAAATGGTAACAACCACAACAGCAACAAACCATCACGTCCTTGG
ACGCACAGGAACCGCCACGCTTGTGGCCTTCCAAATACAAGCGTCGGTGTCTGCAGAAT
GAGAATTACAATTCAACTCAACATCTGCGAGAAACCGCAGCTGCCAAAATAACGCC
AGTGCAGCACCTGGCACGGCTGCAATGGGCCATTGGTGTCTATCAAGACAACCTGGG
AGCATCTGAACAACTCAGTATCAGCTCTCACAGTTAGCCACTCAAACAGCAATCAATT
TCCTGCATTGTTGGATCCTGTACAGCTACAAATAGCACCACATCCGCCAGAACAGCA
ACACCAACAAACACGTCAAGCAACACAGCAGCAAATTGGAGAGTATTACATAAGCGAC
GTCATGAATTACGCCACGCCCACAACACTGTCTGACATGACTCCATG
TTTACCATCAACTCGGACTATGATGCATATCTCTTAGATTTCGTTGCTGGCGACGA
TTTCTCTGTACTTGGCGCGATGGAGCTAAATGTCGCTTCAAGAAGGATGAGAAGG
TGCTACAGTCTGGTCTGTGCATCTGGGCAAACCATAAATGCCGACGACAGCGGGTG
CAGCACGTAGAAGTGCAGGAGATGACACAAATCATGTTAACATCGGTTCCGTGACAT
AATGCCGGGCAAGCCGTGGTCAAATCGAGCATGATTCCGTCAGCTAACAGGAGA
TGCACAATCGCAGATTGTGCCTGTGCTGACCACGCTAGCACCGCTGGCCAACATTGC
CACGACAAGCAGACATGCGAGAAAGTGTGAGATTAAACAAATTACGCAACGAGGG
ACGACATCTCACTGTTATTGATATACACACGTGTGATTAATGAGAATGGCGTTGTAC
GCTTCGATTGCTTCCAGAAGGGGCCACGTAGTGTGACGGGCTCAGTAGAGCCATATGTA

TTCTGGAACAAAATCGGACGACAGCGTGTGTTACAGATGATCGAACAAATTGGAATA
TCACTGA

>*Drosophila virilis*_L22556.1 (reported cDNA sequence from Webster et al. (1994))

ATGGCGACATTCAAAGTGAATTCAACAGCGTGCCAACACTTACAACCAGCGCCTGAC
GGCATTGAAGAAAAAGCTGACCACGTGTTTCAGCAGTGGCGACAGCAGCTGCTCAAAC
TTCACCAACAGCAGCAGGCCAAGCAACAGCAACAGCAACAGCAACATCAATCCCAA
CACCAACACCAACAGCAGAAGCAACAGCAACAGCAACATCAATCCCAA
TGTGCGCGCGGACTGCTCTATTGAGCGAAAAAGTTACCTATTGCGCAAGTTCAAGA
GGATTTCACAGCGGCCAGAATTGGCGAGTGCTCGGACAGCAGAACAGCGGCATCA
GCCAAACAGCCAGTGCTTGGCAGGTGCATTGCCAGGAGGAGCAGCAAATGCAAATAGT
GGCAAGCCTCTCTCATCAACATTGATATCGAACAGCACAGCTCCAGTGCAAGTATCACCA
GCTCCAGCATGACCATCATGGACAACACCTATATTGGGGTGCAGCAGTACCCGAT
ATTGATACTGAAATACGCTCCATACTGCTGGCTAATGCCAGAATGGTATCACGATATC
GAGCATCAAAAGGAATATCGACAATTGACGGGCACGGCATTCCGCTGCACGACAACA
TCACGGACTTTCTGCTTACCATACCGCATGTCACGGCAGAGTGCTGCAGTCTGGGAAG
CGCATTTAACATCAAGCCCACCGGAGCATAACCGCCATCTGCACGAGATGATACTGCA
GCAGCGACAACCGCAGTGAGCAACCCAATCCAAGCACAGGAGGCCACGCTTGT
GGCGCGCCAGTACAAGCGCGAATCCCACAGCACTTCAACTTCAATCTAAATACAAGT
GAGAAGCCTCCAGCGTCAAGATATCAAACACTGCAGGCCCTGGCACGGCTGCAGCGAT
GTCTAACGATGTCTATCAGGACAACCTGGAAACATCTCAACAATCAGTATCAGCTCCCAC
AGCTCAATGCTCCAAAGAACAAACATCCACTCGCACATCGCTGAATCCTGCACAGTTA
CAGGCAGCACTGCCGGCAGAGCATCATCCATCCAACACAGTGGAGGAATATGCGCACAA
GCGGCGTCAATACGCGCACGCAACACGTTGTCGTGCATCAACGCAACATG
ACTCCATGTTACCCTCAACTCGGACTATGATGCATATCTATTGGACTTCCACTGCTG
GGCGATGACTTCTTCTGTATTGGCGCGATGGAGCTGAAATGCCGTTCAAGAAGTT
CGAAAAGGTGCTGCAGTCTGGTCTGTGCATCTGGACAAACCATAATGCTGCACGGC
AACGGTTGCGCCTGGTGGAGCTGCCAGATGACACAAATTATTGTTAACATCGGCTCC
GAGGACATAATGCGCGGCAGGTGCTGGTGCAGATCGAGCACGATTCCGTGCTAGT
GAAGGAGATGCACAACCGTAGATTGCCCCAGTGTGACGACATGCTGACCGCTGGCCA
ACTGTCGCCACGACAAGCAGACATGCGACAAGGTGCTCCGATTCAACAAGTTTACGC
AGCGAGGGGCACATCTGAAAGTGTGATTGATATACACTCATGCTGATCAATGAGAATGG
CATCGTTGATTGCTTCCAGAACGGGCCACGTAGTGTGACGGCTCATCTGAGC
CGTATGTATTCTGGAACAAAATCGGACGGCAGCGTGTGCACATGATTGAAGAAAAC
TTGGAGTATTATTAG

>*Drosophila virilis*_XM_002053233.1 (annotated from genome sequence)

ATGGCGACATTCAAAGTGAATTCAACAGCGTGCCAACACTTACAACCAGCGCCTGAC
GGCATTGAAGAAAAAGCTGACCACGTGTTTCAGCAGTGGCGACAGCAGCTGCTCAAAC
TTCACCAACAGCAGCAGGCCAAGCAACAGCAACAGCAACAGCAACATCAATCCCAA
CACCAACACCAACAGCAACAGCAGAAGCAAGCAACAGCAACAGCAACATCAATCCCAA
CATTTTGCGCGCGGACTGCTCTATTGAGCGAAAAAGTTACCTATTGCGCAAGT
TCAAGAGGATTTCACAGCGGCCAGAATTGGCGAGTGCTCGGACAGCAGCAAAGCG
GCATCAGCCAACAGCCAGTGCTGGCAGGTGCATTGCCAGGAGGAGCAGCAAATGCA
AATAGTGGCAAGCCTCTCATCAACATTGATATCGAACAGCACAGCTCCAGTGCAAGTA
TCACCAGCTCCAGCATGACCATCATGGACAACACCTATATTGGGGTGCAGCAGTAC

CCCGATATTGATACTGAAATACGCTCCACTGCTGGCTAATGCCAGAACGGTATCAC
GATATCGAGCATCAAAAGGAATATCGACAATTGACGGGCACGGCATTCCGCTGCACG
ACAACATCACGGACTTCTGCTTACCATACCGCATGTCACGGCAGAGTGCTGCGAGTCT
GGGAAGCGCATCTTAACATCAAGCCCACCGAGCATAACCGCATCTGCACGAGATGAT
ACTGCAGCAGCGACAACCGACAGTGTGAGCAACCCAATCCAAGCACAGGAGGCCAC
GCTTGTGGCGCGCCAGTACAAGCGCGAATCCCACAGCACTCAACTCAATCTAAAT
ACAAGTGAGAAGCCTCCAGCGTCAAGATATCAAACACTGCAGCCCCGGCCACGGCTGC
AGCGATGTCTAACGATGTCTATCAGGACAACGGAAACATCTCAACAATCAGTATCAGC
TCCCTCAGCTCAATGCTGCCAAGAATAACATCCACTCCATCGCTCGAATCCTGCACAG
TTACAGGCAGCACTGCCGGCAGAGCATCATCCATCAAACACGTGGAGGAATATGCGCA
CAAGCGCGTCATGAATATACGCCAACCGAACACGTTGTCGTGCCATCAACGCAAC
ATGACTCCATGTTACCACATCAACTCGGACTATGATGCATATCTATTGGATTTCCACTG
CTGGCGATGACTTCTTGTATTGGCGCGATGGAGCTGAAATGCGCTCAAGAA
GTTGAAAAGGTGCTGAGTCTGGTCTGTGCATCTGGACAAACCATAAATGCTGCAC
GGCAACGGTTGCGCCTGGTGGAGCTGCCAGAGATGACACAAATTATTGTTAACATCGGC
TCCGAGGACATAATGCGCGCAGGTGCTGGTGCAGATCGAGCACGATTCCGTCTGCT
AGTGAAGGAGATGCACAACCGTAGATTGTGCCAGTGTGACCACCTGGCACCGCTGG
CCAACGTGCCACGACAAGCAGACATGCGACAAGGTGTCCCATTCAACAAGTTATA
CGCAGCGAGGGCGACATCTGAAAGTGATTGATATACACTATGTCATGATCAATGAGAA
TGGCATCGTCGATTGCTTCCAGAACGGGCCACGTAGTGTGACGGCTCATCTG
AGCGTATGTATTCTGGAACAAATCGGACGGCAGCGTGTGCTGCACATGATTGAAGAA
AACTGGAGTATTATTAG

>*Drosophila virilis* sequenced (Sanger sequence verified)
ATGGCGACATTCAAGAAGTGAATTCAACAGCGTGCCGAACACTTACAACCAGCGCCTGAC
GGCATTGAAGAAAAAGCTGACCACTGGACTTTCAGCAGTGGCGACAGCAGCTGCTCAAAC
TTCACCAACAGCAGCAGCAGGCCAAGCAACAGCAACAGCAACAGCAACATCAATCCAA
CACCAACACCAACAGCAGCAGAAGCAGAAGCAACAGTGGCAGTGAACACATTCTACAACATT
TGTGCGCGGACTGCTCTATTGAGCAGGAAAGTACCTATTGCGCAAGTTCAAGA
GGATTTCACAGCGGCCAGAATTGGCGAGTGCTCGGACAGCAGCAACAGCGGCATCA
GCCAAACAGCCAGTGCTGGCAGGTGCATTGCCAGGAGGAGCAGCAACTGCAAATAGT
GGCAAGCCTCTCATCAACATTGATATCGAACAGCACAGCTCCAGTGCAGTATCACCA
GCTCCAGCATGACCACATGGACAACACCTATTGGGGTGCAGCAGGAGTACCCGAT
ATTGATACTGAAATACGCTCCATACTGCTGGCTAATGCCAGAACGGTATCACGATATC
GAGCATCAAAAGGAATATCGACAATTGACGGGCACGGCATTCCGCTGCACGACAACA
TCACGGACTTCTGCTTACCATACCGCATGTCACGGCAGAGTGCTGCGAGTCTGGGAAG
CGCATTTAACATCAAGCCCACCGAGCATAACCGCATCTGCACGAGATGATACTGCA
GCAGCGACAACCGACAGTGTGAGCAACCCAATCCAAGCACAGGAGGCCACGCTTGT
GGCGGCCAGTACAAGCGCGAATCCCACAGCACTCAACTCAATCTAAATACAAGT
GAGAAGCCTCCAGCGTCAGATATCCAACACTGCAGCCCCGGCCACGGCTGCAGCGAT
GTCTAACGATGTCTATCAGGACAACACTGGAAACATCTCAACAACTCAGTATCAGCTCCAC
AGCTCAATGCTCCCAAGAACACATCCACTCGCACATCGCTCGAATCCTGCACAGTTA
CAGGCAGCACTGCCGGCAGAGCATCATCCAAACACGTGGAGGAATATGCGCACAA
GCGCGCTCATGAATATACGCCACGCCAACACGTTGTCGTGCCATCAACGCAACATG
ACTCCATGTTACCATCAACTCGGACTATGATGCATATCTATTGGACTTCCACTGCTG
GGCGATGACTTCTTGTATTGGCGCGATGGAGCTGAAATGCCGCTCAAGAAGTT
CGAAAAGGTGCTGAGTCTGGTCTGTGCATCTGGACAAACCATAAATGCTGCACGGC
AACGGTTGCGCCTGGTGGAGCTGCCGAGATGACACAAATTATTGTTAACATCGGCTCC

GAGGACATAATGCGCGGCAGGTGTTGGTGCAGATCGAGCACGATTCCGTCTGCTAGT
GAAGGAGATGCACAACCGTAGATTGTGCCAGTGCTGACCACCTGGCACCCTGGCCA
ACTGTCGCCACGACAAGCAGACATGCGACAAGGTGTCCTGATTCAACAAGTTATACGC
AGCGAGGGGCACATCTGAAAGTGATTGATATACTCATGTCTGATCAATGAGAATGG
CATCGTTGATTGCTTCCAGAACCGGGCACGTAGTGTGACGGGCTCATCTGAGC
CGTATGTATTCTGGAACAAAATCGGACGGCAGCGTGTGCTGCACATGATTGAAGAAAAC
TTGGAGTATTAT

Online Resource 2

This file contains the Oskar amino acid alignment generated with sequences from 18 Drosophilids using the MUSCLE MSA, and the results of multiple PAML analyses of this alignment. The following elements comprise this Additional File:

Figure S1: Amino acid alignment.

Table S1: Log likelihood values and parameter estimates under partition models of Oskar domains.

Table S2: Log likelihood values and parameter estimates under site models of Oskar domains.

Table S3. Log likelihood values and parameter estimates under branch site models of Oskar domains.

Figure S1. Amino acid alignment of Oskar generated by MUSCLE MSA used for PAML analysis. Residue with BEB probability of positive selection = 0.99 is indicated in red.

dper	MATYRSAH-SEPYTSKSNNSNSNSSSSRLTALRKRLTTCLQWCQHL-----
dpse	MATYRSAH-SEPYTSKSNNSNSNSSSSRLTALRKRLTTCLQWCQHL-----
dimm	MATFRSEFNSVPNTH-----PDRLTAFRKLLTCFHQWRQKL-----
Dvir	MATFRSEFNSVPNTY-----NQLTALKKKLTTCFQOWRQQLKLHQQQQQPKQO
dana	MAAVTSELPPREP-----TTTTNTKSDRLRAVKRVTTCLQOWCQKL-----
dbip	MAAVTSELPPREP-----TTTTNTKSDRLRAVKRVTTCLQOWCQKL-----
dkik	MAAVRSEFPSE-----SNSNPSTKADSLRSLKHKVATCWHLCKKL-----
deug	MATVTSEFPSKPISYTNTNTSTKTYCLRSVKRRVTTCLQQLREKL-----
dmel	MAAVTSEFPSKPISY--TSTNTSAKTYYLKSVKVRVTTCFQQLRDKL-----
dsec	MATVTSEFPSKPISY--TSTNTSAKTYYLKSVKVRVTTCFQQLRDKL-----
dsim	MATVTSEFPSKPISY--TSTNTSAKTYYLKSVKVRVTTCFQQLRDKL-----
dere	MAAVTSEFPSKPIS--TSTNTSAKSNYLKSVKVRVTTCFQQLRDKL-----
dyak	MAAVTSEFPSKPISY--TSTNTSAKTYYLKSVKVRVTTCFQQLRDKL-----
dfic	MAAVRSEFPSPKN---SNTYTNKFDDLRSVKVRVTTCLKKLCEKL-----
dele	MAAVRSEFPSEPNSN--TNTNTSTKTDRLRSVKNRVTTCLKQLCEKL-----
drho	MAAVRSEFPSEPYSN--TNNNTSTKTDCLRSVKVRVTTCLKQLCEKL-----
dbia	MAAVGSEFPSEPISC--RNTNTPAKPDRLRSVKVRVTTCLQQLREKL-----
dtak	MAAVRSEFPSEPISY--TNTNTSNKSNCRSVKVRVTTCLQQLREKL-----
dper	-----QKPASFQKCKSGPNHQFYKIFEEVFDKV-FGDKTKYIROFKKIFKGN
dpse	-----QKPASFQKCKSGPNHQFYKIFEEVFDKV-FGDKTKYIROFKKIFKGN
dimm	-----QHFKH-----DYST-LRAKSICWQKFKKIYTAT
Dvir	-----QQQQOHOSOHOHQQQQQOKQKORG--SEHFYNIFVRA-DCSI-LSEKVTYLRKFKRIFTAR
dana	-----QOPASFRK-SS--SSRFYRLFVKA-DYTA-FGDKFKSVRQFHKIFKAG
dbip	-----QOPASFRK-SS--SSRFYRIFVKA-DYTA-IGDKFKSVRQFHKIFKAG
dkik	-----QOPASFRK-SS--HSRFYKIFLKA-DCSASFPEKSKFLRQFHRIFKAG
deug	-----PATGSFRK-SS--TSRFHRFFIKS-DFSD-FSEK-----LQKIFKSA
dmel	-----QSSGSFRK-SS--SSCLNQIFVRS-DFSA-CGER-----FRKIFKSA
dsec	-----QPGFSFRK-SS--SNRLYKIFVKS-DFSA-CGER-----FQKIFKSA
dsim	-----QPGFSFRKSSSS--SSRLYKIFVKS-DFSA-CGER-----FQKIFKSA
dere	-----QASGSFRK-SS--PSRLYKIFVKA-DFAA-CGER-----FQRIFKSA
dyak	-----QASGSFRK-SS--TSRLYKIFVKA-DFAA-CGER-----FQRIFKSA
dfic	-----QAPGSFRK-SS--SSRFHKIFVRA-DYSD-CSK--FVRRFQKIFKSA
dele	-----QATGSFRK-SS--SSRFYKIFVKA-DYSG-LREKSKFLRQLQKIKYST
drho	-----QAPGSFRK-SS--SSRFYKIFVKA-DYSG-FREKSKFVRQFQKIFKSA
dbia	-----PAAGSFRK-SS--TGRFYQVFVRA-DFSA-FGER-----FRKIFKSA
dtak	-----PASGSFRK-SS--SSRFYKIFVKA-DFSA-FGER-----FQKIFKSA
dper	KDLAK--KSEPGRVVLVLVPEKPVDL-EESQKKQLOLVARLFSSTLISAKEEISCsyn-
dpse	KDLAK--KSEPGRVVLVLVPEKPVDL-EESQKKQLOLVARLFSSTLISAKEEISCsyn-
dimm	KLISQ-----EEVEEEQQLQIVATLFSSTLIS-----KYTS
Dvir	QNLAS----ARTAATAASAKQPVLGRCIGQEEQQLQIVASLFSSTLISKHS-----
dana	RKGKE-----PAESVDER-----QQLQFVARLFSSTLISTEE-----IYGN
dbip	RKSK-----EPAESVDER-----QQLQFVARLFSSTLISTEE-----IYGN
dkik	RKSEGGGEPDPEPELWKEVPLAD----EQRSQEQLQIVARLFSSTLISTKEEEIITTSR
deug	RKSES-----PELRR-VPLSTHDQQTSGQSSQQLQIVARLFSSTLISTKE---ITFNN
dmel	RKTEL-----PELWK-VPLVAHELT-SRQSSQQLQIVARLFSSTOISTKE---ITYNS
dsec	RKTES-----PELWK-VPLVAHELT-SRQSSQQLQIVARLFSSTOISTKE---ITYSS
dsim	RKTES-----PELWK-VPLVAHELT-SRQSSQQLHIVARLFSSTOISTKE---ITYSS
dere	RKTQS-----PELWK-VPLAAHDLTNGRQTSQQLRIVARLFSSTQISTKE---ITYST
dyak	RKTES-----PELWK-VPLAHDLLTTSRQSSQQLRIVARLFSSTQISTRE---ITYTN
dfic	RKTES-----PESWK-VPL-AHDVT-RQSSIHQLQIVARLFSSTQISTEE---IIYNN
dele	RKTES-----PELWK-VPL-AHDVT--GQSSQQLQIVTRLFSSTLISTEE---ITYNT
drho	RKPES-----PELWK-VPL-AHDVT--GQNSQQLQIVARLFSSTLVSTEE---ITYHT
dbia	RKTES-----PELWK-VPL-AHDVT--GPSSQQLQIVARLFSSTLISTKE---ISYSN
dtak	RKTES-----PELWK-VPL-AHDVT--GQSGQQLQIVARLFSSTLISTKE---IHCNN
dper	-----SSSMTIIDSNYIEVREEYPEIDSEIRSILLANAQSGITI

dpse -----SSSMTIIDSNYIEVREEYPEIDSEIRSILLANAQSGITI
 dimm -----SKNSS-----SCNMTIMDDSYIGVRDEYPDIDAEIRAILLANAQKGITI
 Dvir -----SSASIT-----SSSMTIMDNTYIGVRDEYPDIDTEIRSILLANAQNGITI
 dana -----SSRKS-QSF-NSNSMTIIESNYIAVREENPDIDAEVRAILLSHAQNGITI
 dbip -----GKTSYSSRK-SQY-NSNSMTIIESNYIAVREENPDIDAEVRAILLSHAQNGITI
 dkik -----SSN-CNTSYTRRK-SQYCSTSTMIIIEGNYREVRDEYPDIDAEVRAILLSHAQNGITI
 deug -----HTK--NSFSYKCRKS-EGY--SSNMTIIEGNYIDVREENPDIDAEVRAILLTHAQNGITI
 dmel -----NSNTS-----ENNMTIIESNYISVREEYPDIDSEVRAILLSHAQNGITI
 dsec -----NNNTS-----ETNMTIIESNYISVREDYPDIDAEVRAILLSHAQNGITI
 dsim -----NSNTS-----ESNMTIIESNYISVREDYPDIDAEVRAILLSHAQNGITI
 dere -----SNNNTSSYSFRSRKS-ESC--RSNMTIIESNYIAVREDYPDIDTEVRAILLSHAQNGITI
 dyak -----SNN-SSYSFRRK-SESH--RSNMTIIESNYIGVREDYPDIDTEVRAILLSHAQNGITI
 dfic -----ISFNSRK-SQSH--SSNMTIIESNYIAVREEYPDIDAEVRAILLSHAQNGITI
 dele -----NNNSCKSRKSHKSY--KSSMTIIESNYIAVREEFPDIDTEVRAILLSHAQNGITI
 drho -----NINSYKSRKS-QSY--DCDMTIIESNYIAVREEYPDIDTEVRAILLSHAQNGITI
 dbia -----HYRSCRGKS-ESY--SSNMTIIESNYIAVREEHPDIDKEIRAILLSHAQNGITI
 dtak -----NSYSYKSRKS-ESY--SSNMTIIESNYITVREEFPDIDAEVRAILLTHAQNGITI

dper STIKNEYRKLGTGNPFPLHENITDFLLTIPHVTAECCVSGKRIFNMKPRPETRHLYDMVVH
 dpse STIKNEYRKLGTGNPFPLHENITDFLLTIPHVTAECCVSGKRIFNMKPRPETRHLYDMVVH
 dimm SSIKSEYRNMTGNAFPLRENITDFLLTIPVTAECMSGKRIFNIKPTEDETRHLHDMVLN
 Dvir SSIKKEYRQLTGTAFPLHDNITDFLLTIPVTAECSETGKRIFNMKPNKDNRHLYDMVLN
 dana SSIKSEYRKQTGNTFFPLHDNVTDFLLTIPVTAECSETGKRIFNMKPNKDNRHLYDMVLN
 dbip SSIKSEYRKQTGNTFFPLHDNVTDFLLTIPVTAECSETGKRIFNLKPNKENSHLLEMVLN
 dkik SSIKDEYRKQTGNTFFPLHDNVTDFLLTIPVTAECSETGKRIFNLKPNKENSHLLEMVLN
 deug SSIKSEYRKLTGNNPFPPLHDNVTDFLLTIPVTAECSETGKRIFNLKANLKNRHLDDMVLN
 dmel SSIKSEYRKLTGNNPFPPLHDNVTDFLLTIPVTAECSETGKRIFNLKASLKNRHLDDMVLN
 dsec SNIKSEYRKLTGNNPFPPLHDNVTDFLLTIPVTAECSETGKRIFNLKANLKNRHLDDMVLN
 dsim SNIKSEYRKLTGNNPFPPLHDNVTDFLLTIPVTAECSETGKRIFNLKANLKNRHLDDMVLN
 dere SNIKSEYRKLTGNNPFPPLHDNVTDFLLTIPVTAECSETGKRIFNLKANLKNRHLDDMVLN
 dyak SNIKSEYRKLTGNNPFPPLHDNVTDFLLTIPVTAECSETGKRIFNLKANLKNRHLDDMVLN
 dfic SNIKSEYRTQGTGQFPVQENVTDFLLTIPVTAECSESQKRIFNLKPNVKNRHLDDMVLN
 dele SNIKSEYRKQTGQFPVQENVTDFLLTIPVTAECSESQKRIFNLKPNVKNRHLDDMVLN
 drho SSIKSEYRKQTGNAFPLHDNITDFLLTIPVTAECSESQKRIFNLKPSLKNRHLLEMVLN
 dbia SNIKSEYRKQTGNAFPLHDNITDFLLTIPVTAECSESQKRIFNLKPNLKNRHLLEMVLN
 dtak

dper QKRT-----NDLGTNDHP-V-----AAPPPRLWRCQYKRRALYSLNTNNNNNNYNL
 dpse QKRT-----NDLGTNDHP-V-----AAPPPRLWRCQYKRRALYSLNTNNNNNNYNL
 dimm QRQE---NGNNHNSNNHHVLD-----AQEPPRLWRSQYKRRCLQNEFY----NFNL
 Dvir QRQR-----DSVSNPQI-----AQEPPRLWRAQYKRRIPQHFNF----NLNT
 dana QKQR-----NDFSSGPPSTVDNSKPGQRGPPRHWRYQYKRRQAQSPL-----NLNQ
 dbip QKQR-----NDCSSGPPSSVEYN-RPVGQRGPPRHWRYQYKRRQAQSPL-----NTNQ
 dkik QKQR-----SDYSSGAPS-IEDKYRMGPPQVPPRYWRNGFKRRALSPLDN----NLNV
 deug QKOK-----PSDNSSEAPT-VEDK-----SRVPPRYWKNPYKRRALSQLDN----NLNT
 dmel QKER-----TSDYSSGAPS-LENI-----PRAPPRYWKNPFKRRALSQLNT----SPRT
 dsec QKORT--TNSTSDYSSGAPS-LENI-----PRAPPRYWKNPFKRRALSQLNT----SPRT
 dsim QKORT--TNSTSDYSSGAPS-LENI-----PRAPPRYWKNAFKRRALSQLNT----SPRT
 dere QKORT---GSSDYSSGAPS-VESM-----PRAPPRYWKNPFKRRALSQLN----NLNT
 dyak QKORTGTANGTSDYSSGAQS-VENM-----PRAPPRYWKNSFKRRALSQLND----NLNT
 dfic QKQR-----TSDTSSGAPS-VESH-----SRPPPYYWRNSYKRKALSQLDS----NLNI
 dele QKQR-----TSDYSSGAPS-VEDK-----PRPPPRYWKFPYKRRALSQLEN----NLNT
 drho QKQR-----TSDYSSGTPS-VEDN-----PRPPPRYWKFPYKRRALSQLDN----NLNT
 dbia QKQR-----SSDCSSGAPS-VEDQ-----PRAPPRYWKNPYKRRALSQLDN----NLNT
 dtak QKQR-----ASDYSSGAPS-LEDK-----PRAPPRYWKNPYKRRALSQLDN----NLNT

dper N---QNINHIE-----EKPAVGGKEVAIEDQ-AGPLQQLARAAAESNWCYQDNWNH
 dpse N---QNINHIE-----EKPAVGGKEVTIEDQ-AGPLQQLARAAAESNWCYQDNWNH
 dimm N-----IC-----EKP-----PAAK-----ITPVQHLLATAAMAPFGVYQDNWKH
 Dvir S-----EKP-----PAVK-----ISKLOPLATAAAMSNDVYQDNWKH
 dana M---KHLN-TE-----RQV-----QDKAINAN-SATLQQLAKAAGESNWCYQDNWNH
 dbip I---KQLN-TEQNGMEKLQDNV-----QIRAIKTN-SASLQQLAKAAGESNWCYQDNWNH
 dkik KVNSEKP--AD-----DQS-----KAIN--TR-SAALPHMAAKAAESNWCYQDNWNH
 deug V---HKL--TD-----DQA-----KAMSMHTR-AAPLQQLAKVAAESNWCYQDNWKH
 dmel V---PKI--TD-----EKT-----KDIA--TR-PVSLHQMANEEAESNWCYQDNWKH

dsec	V---PKI--TN-----ERT-----KEIA--TR-PASLHQMANEAAESNWCYQDNWKH
dsim	V---PKI--TD-----EQT-----KEIP--TR-PASLHQMANEAAESNWCYQDNWKH
dere	E---SKI--TD-----EQT-----QETS--TR-PASLHQMANEAAESNWCYQDNWKH
dyak	V---PKL--TD-----EQT-----KESA--TRQPASLHQMANEAAESNWCYQDNWKH
dfic	R---PCT--AD-----YQA-----KDIA--TK-AAPLPQLAKEAAESNWCYQDNWKH
dele	G---PKLKEQE-----HQF-----QAAT--TK-AAPLQLAKEAAESNWCYQDNWKH
drho	R---PKL--ED-----HQF-----KAIT--AK-TTPLHOLAKEAAESNWCYQDNWKH
dbia	V---PKL--TA-----DQT-----KAVD--TR-AAPLQLAKEAAESNWCYQDNWNH
dtak	V---HKL--TD-----EQT-----KAIN--TK-AAPLQLAKEAAESNWCYQDNWNH

dper	LKNCYQRA-----NIFA-DPPEK-PPPLPA---QENHL-----RNS
dpse	LKNCYQRA-----NIFA-DPPEK-PPPLPA---QENHL-----RNS
dimm	LNNQYQLSQLATO-----NSNQFS-CIRSD-PVQLQI---APHPPE-----QQ
Dvir	LNNQYQLPQLNAA-----KNNIH-SIASN-PAQLQAALPAEHHPS-----
dana	LNNFYQOOSNVNGH---QNQVPPNIYV-DTPDR-QTOVTD---PEQH---C-----O
dbip	LNNFYQOQNNVNGP---QNPVPPNIYI-DTPDQ-PTQVTD---PELLH---C-----Q
dkik	LKNIYQOSSGNAP--HMHEPNQIYT-DPPEL-SNHLAN---PELIKPPC-----KNH
deug	LNNFYQOSSVNEA--QRPVPINIYT-DALEDLPIQLDN---PEHHPN-C-----KNN
dmel	LNNFYQQASVNAP--KMPVPINIYS-DAPEE-PINLAP---PGHQPS-C-----RTQ
dsec	LNNFYQQASVNAP--QMPVPINIYS-DAPEE-PTHLAA---PGQOPSCC-----RTQ
dsim	LNNFYQQASVNAP--QMPVPINIYS-DAPEE-PTHLAA---PGQOPSCC-----RTQ
dere	LNNFYQQASVNAP--KMPVPINIYS-DAPEE-RIQVDA---PGHQPS-C-----RTQ
dyak	LNNFYQQASVNAP--KMPVPVNIS-DAPEE-RIHLDA---PGHQPS-C-----RTQ
dfic	LNNYYQOSSANE-----QIPVPINIYT-DATEE-PNQFAV---PEHLPP-C-----KNH
dele	LNNFYQOSSANE-----QMPVPINIYS-DATEE-RTQFAV---PEPRP--C-----GKNH
drho	LNNFYQOSSANE-----QRPVPINIYS-DPPEE-RTQFAV---PELPPH-C-----RNP
dbia	LNNIYQRSSVSEPOQLPVPINIYS-DAPEE-PNQMNV---PQHHPL-C--HNRNQNPQ
dtak	LNNFYQQSRVNEP--QMPVPINIYS-DATEE-PIQLDA---PEHHPH-CQNHNEQNPQ

dper	NSQNHRDNLKNQRNLLVNEHHSMOO-EYQYQY-PTTTPTPTTISS-GTOHDSMFTINS DY
dpse	NSQNHRDNLKNQRNLLVNEHHSMOO-EYQYQY-PTTTPTPTTISS-GTOHDSMFTINS DY
dimm	HOOOR-OOQQQQIGEYSH-----KRRHEF---TPTPTTLSC-PSQHDSMFTINS DY
Dvir	-----KHVEEYAH-----KRRHEY---TPTPTTLSCPSTQHDSMFTINS DY
dana	NPHPS-QNONSQODIYNNQNRELFEFKRPRDL---TPTPTMTS--ATHNDSMRTI SDF
dbip	NRNPS-QNQNYHQDLYNNHNRMNLEFKRPRDL---TPTPTTS--GTQNDMMTI SDF
dkik	NSPNLNQILNPNLDIFRRPRSSCD-LKRRHER---TPTPTTMSS-ATHNDMSLT I SDY
deug	NONIK-QNVNQHLGNFVNPLNDVNPFKRRQDI---TPTPTIFTSGGSSNNSLMT I SDY
dmel	SQKTE-PTENRHLGI FVHPFNGMNIMKRRHEM---TPTPTILTS-GTYND SLLT I SDY
dsec	SKTE-PTENRHLGI FVHPFNDMVNLKRRHEM---TPTPTILTS-GTYND SLLT I SDY
dsim	SKPTE-PTENRHLGI FVHPFNDMVNLKRRHEM---TPTPTILTS-GTYND SLLT I SDY
dere	NOKTQHPTENRHLGI FVHPFSDMNVMKRRHEMPTPTPTPTTLSS-GTYADSL LT I SDY
dyak	NOKTQQTTEENRHLGI FVHPFNDMVNMVKRRHEM---TPTPTILTS-GTYSDS LLT I SDY
dfic	KONTK-PNDNQHLGI FMRPYNDMVNLKRRRDL---TPTPTTLSS-GTFNDS ILT MSS DY
dele	NONNK-PNVNQQLGSFVNPFNDMVNLKRRHD---TPTPTTLSS-GTYND SMLT I SDY
drho	NONTK-QNVNQHLGI FVNPFPNDMVNLKRRHD---TPTPTTLSS-GIYSDS ILT I SDY
dbia	NOQAK-QDANQHLGI FVQPFSDMVNLKRRQEM---TPTPTTLSS-GTYND SMLT I SDY
dtak	NONTK-QNVNQHLGI FVQPFSDMVNLKRRHEL---TPTPTTLSS-GTYND SMLT I SDY

dper	DAYLLDFPLLGDDFLLYLARMELRCRKTERVLQSGLCVSGOTISGARSRLHHLLVNKG
dpse	DAYLLDFPLLGDDFLLYLARMELRCRKTERVLQSGLCVSGOTISGARSRLHHLLVNKG
dimm	DAYLLDFPLLGDDFLLYLARMELKCRFKDEKV LQSGLCISGOTINAARQRLRLVELPEM
Dvir	DAYLLDFPLLGDDFLLYLARMELKCRFKFEKV LQSGLCISGOTINAARQRLRLVELPEM
dana	DAFLDFPLMGDDFLY LARMELNCRFKRYERVLQSGLCVSGOTIIIGARNRFRKV YLPEG
dbip	DAFLDFPLMGDDFLY LARMELNCRFKRYERVLQSGLCVSGOTIIIGARNRFRKV YLPEG
dkik	DAYLLDFPLMGDDFLY LARMELKCRFRRYERVLQSGLCVSGOTINAARNRLKRV NLP EG
deug	DAYLLDFPLMGDDFLY LARMELKCRFRRYERVLQSGLCASGOTINGARNRLKRV QLPEG
dmel	DAYLLDFPLMGDDFLY LARMELKCRFRRYERVLQSGLCASGOTINGARNRLKRV QLPEG
dsec	DAYLLDFPLMGDDFLY LARMELKCRFRRYERVLQSGLCASGOTINGARNRLKRV QLPEG
dsim	DAYLLDFPLMGDDFLY LARMELKCRFRRYERVLQSGLCASGOTINGARNRLKRV QLPEG
dere	DAYLLDFPLMGDDFLY LARMELKCRFRRYERVLQSGLCASGOTINGARNRLR RV QLPEG
dyak	DAYLLDFPLMGDDFLY LARMELKCRFRRYERVLQSGLCASGOTINGARNRLK RV QLPEG
dfic	DAYLLDFPLMGDDFLY LARMELNCRFKRYERVLQSGLCVSGOTTINGARNRLKK ANLPEG
dele	DAYLLDFPLMGDDFLY LARMELKCRFRRYERVLQSGLCVSGOTTINGARR RLKKV QLPEG
drho	DAYLLDFPLMGDDFLY LARMELKCRFRRYERVLQSGLCVSGOTTINGARKRL KK VHLTEG
dbia	DAYLLDFPLMGDDFLY LARMELKCRFRRYERVLQSGLCVSGOTTINGARM RLKKFHLPEG

dtak

DAFLLDFPLMGDDFMLYLARMELKCRFRRYERILQSGLCVSGOTINGARNRLKKFHLPEG

dper
dpse
dimm
Dvir
dana
dbip
dkik
deug
dmel
dsec
dsim
dere
dyak
dfic
dele
drho
dbia
dtak

TQIIVNIGSVDIMRGRPIVQIQHDFRQLVKDMHNRGLVPILTTPLAPLANYCHDKAMCDKV
TQIIVNIGSVDIMRGRPIVQIQHDFRQLVKDMHNRGLVPILTTPLAPLANYCHDKAMCDKV
TQIIVNIGSVDIMRGRPIVQIEHDFRQLIKEHMHNRRFVPLTTPLAPLANYCHDKQTCVK
TQIIVNIGSEDIMRGRSILVQIEHDFRLLVKEHMHNRRFVPLTTPLAPLANCRRHKQTCDKV
TQIIVNIGSVDIMRGKPLVQIEHDFRLLIKEIHSSRCIPILTNLAPLANYCHDKVLCDKI
TQIIVNIGSVDIMRGKPLVQIEHDFRLLIKEIHSSRCIPILTNLAPLANYCHDKVLCDKI
TQIIVNIGSVDIMRGKPLVQIEHDFRLLIKEHMHNRRFVPLTTPLAPLANYCHDKGLCDKV
TQIIVNIGSVDIMRGKPLVQIEHDFRLLIKEHMHNQRYPIILTNLAPLANYCHDKVLCDKI
TQIIVNIGSVDIMRGKPLVQIEHDFRLLIKEHMHNMRVPLTTPLAPLGNYCHDKVLCDKI
TQIIVNIGSVDIMRGKPLVQIEHDFRLLIKEHMHSQRFVPLTTPLAPLANYCHDKTMCDKI
TQIIVNIGSVDIMRGKPLVQIEHDFRLLIKEHMHSQRVPLTTPLAPLANYCHDKTMCDKI
TQIIVNIGSVDIMRGKPLVQIEHDFRLLIKEHMHSQRVPLTTPLAPLANYCHDKILCDKV
TQIIVNIGSVDIMRGKPLVQIEHDFRLLIKEHMHNQRYVPLTTPLAPLANYCHDKELCDKV
TQIIVNIGSVDIMRGKPLVQIEHDFRLLIKEHMHNQRYVPLTTPLAPLANYCHDKDLCRKV

dper
dpse
dimm
Dvir
dana
dbip
dkik
deug
dmel
dsec
dsim
dere
dyak
dfic
dele
drho
dbia
dtak

VKFNOFIWKECASYLKVIDIHSCLVNENGVVRFDCFQYS---SRNVTGSKESYVFWNKIG
VKFNOFIWKECASYLKVIDIHSCLVNENGVVRFDCFQYS---SRNVTGSKESYVFWNKIG
LRFNKFIRNEGR-HLTVIDIHTCLINENGVVRFDCFQKG---PRSVTGSVEPYVFWNKIG
SRFNKFIRSEGR-HLKVIDIHSCLINENGIVRFDCFQNG---PRSVTGSSEPYVFWNKIG
TKFNRFVRNEARSHLKFIDINSCLINEKKNVLFDCFQSA---PRTVTGSKEPYLFWNKIG
SKFNKFVRNEARSHLKFIDINSCLINEKKNVLFDCFQSA---PRTVTGSKEPYLFWNKIG
LRFNNFIRSEGRSHLKFIDINSCLINEKKNVLFDCFQNA---PRSVSGSKEPYLFWNKIG
TRFNKFIASEGR-HLKVINIHSCLINERGVVRFDCFQNS---PRQVTGSKEPYLFWNKIG
YRFNKFIRSECC-HLKVIDIHSCLINERGVVRFDCFQAS---PRQVTGSKEPYLFWNKIG
YRFNKFIRSECC-HLKVIDIHSCLVNTERGVVRFDCFQTS---PRQVTGSKEPYLFWNKIG
YRFNKFIRSECC-HLKVIDIHSCLVNTERGVVRFDCFQTS---PRQVTGSKEPYLFWNKIG
HRFNKFIRSECC-HLKVIDIHSCLINERGVVRFDCFQSS---PRQVTGSKEPYLFWNKIG
HRFNKFIRSECC-HLKVIDIHSCLINERGVVRFDCFQSS---PRQVTGSKEPYLFWNKIG
SRFNKFIRSEGC-HLKVIDIHSCLVNTERGVVRFDCFQSS---PRQVTGSKEPYLFWNKIG
YRFNKFIRSECC-HLYKIDIHSCLINERGVVRFDCFQSS---PRQVTGSNQPYLFWNKIG
HRFNKFIRSEGS-HLYKIDIHSCLINERGVVRFDCFQSS---PRQVTGSKEPYLFWNKIG
NRFNKFIRSEGR-HLKVVDIHSCLINERGVVRFDCFQ-----QEPYLFWNKIG
SRFNKFIRAEGR-HLKVVDIHSCLINESGIVRFDCFQSS---PRQVTGSREPYLFWNKIG

dper
dpse
dimm
Dvir
dana
dbip
dkik
deug
dmel
dsec
dsim
dere
dyak
dfic
dele
drho
dbia
dtak

RORVLQMIASLEYX-
RORVLQMIASLEYX-
RORVLQMIEQNLEYHX
RORVLHMIEENLEYX-
RORVLQTIESCLEYX-
RORVLQTIESCLEYX-
RORVLQMIESSLEYX-
STKSGDX-----
RORVLQVIETSLEYX-
RORVLQIIETSLEYX-
RORVLQIIETSLEYX-
RORVLQVIETSLEYX-
RORVLQVIETSLEYX-
RORVLQTIETNLEYX-
RORVLOTIESNLEYX-
RORVLQVIESSLEYX-
RORVLQIIESSLEYX-
RORVLRIIESSLEYX-

Table S1. Log likelihood values and parameter estimates under partition models for Oskar structural and interaction domains using MUSCLE MSA.

	Model	np ¹	<i>l</i> ²	<i>rs</i> ₂	<i>rs</i> ₃	<i>rs</i> ₄	<i>rs</i> ₅	<i>rs</i> ₆	κ	ω
Mgene=0	B (different <i>rs</i> ³)	50	-25256.89	0.46	1.06	0.71	0.84	0.8	1.58	0.15
Mgene=2	C (different <i>rs</i> and π ⁴)	95	-25247.84	0.49	1.13	0.73	0.87	0.86	1.63	0.16
Mgene=3	D (different <i>rs</i> , κ ⁵ and ω ⁶)	60	-25120.72	0.53	1.03	0.89	0.95	0.98	$\kappa_1=1.62$	$\omega_1=0.33$
									$\kappa_2=1.5$	$\omega_2=0.12$
									$\kappa_3=1.39$	$\omega_3=0.37$
									$\kappa_4=1.73$	$\omega_4=0.08$
									$\kappa_5=1.57$	$\omega_5=0.14$
									$\kappa_6=1.89$	$\omega_6=0.11$
Mgene=4	E (different <i>rs</i> , κ , ω and π)	105	-25104.23	0.56	1.05	0.91	0.97	1.09	$\kappa_1=1.67$	$\omega_1=0.34$
									$\kappa_2=1.57$	$\omega_2=0.11$
									$\kappa_3=1.56$	$\omega_3=0.45$
									$\kappa_4=1.75$	$\omega_4=0.09$
									$\kappa_5=1.57$	$\omega_5=0.15$
									$\kappa_6=1.98$	$\omega_6=0.10$

¹ np: number parameters

² *l*: log likelihood value

³ *rs*: substitution rate. *rs*₁=1 for all models (not shown). Subscript corresponds to six partitions: (1) Long Osk domain; (2) LOTUS domain; (3) Lasp-binding domain; (4) SGNH Hydrolase domain; (5) Vasa-interacting domain; and (6) Valois-interacting domain.

⁴ π : equilibrium codon frequencies

⁵ κ : transition/transversion ratio

⁶ ω : dN/dS ratio

Table S2. Log likelihood and parameter estimates from site models for Oskar structural and interaction domains using MUSCLE MSA.

	I	κ	Parameter estimates
Long Osk			
M1	-4367.34	1.78	$p_o=0.75, p_1=0.25, \omega_0=0.23, \omega_1=1$
M2	-4367.34	1.78	$p_o=0.75, p_1=0.15, p_2=0.1; \omega_0=0.23, \omega_1=1, \omega_2=1$
M7	-4355.18	1.74	$p=1.16, q=2.05$
M8	-4355.18	1.73	$p_o=0.99, p_1=0.001, p=1.16, q=2.05; \omega=1$
LOTUS			
M1	-1925.78	1.65	$p_o=0.9, p_1=0.1; \omega_0=0.07, \omega_1=1$
M2	-1925.78	1.65	$p_o=0.9, p_1=0.07, p_2=0.02; \omega_0=0.07, \omega_1=1, \omega_2=1$
M7	-1903.99	1.59	$p=0.28, q=1.82$
M8	-1903.27	1.58	$p_o=0.97, p_1=0.03, p=0.35, q=3.13; \omega=1$
Lasp-binding			
M1	-2353.15	1.56	$p_o=0.5, p_1=0.5, \omega_0=0.16, \omega_1=1$
M2	-2353.13	1.57	$p_o=0.51, p_1=0.0, p_2=0.49; \omega_0=0.1, \omega_1=1, \omega_2=1.03$
M7	-2350.60	1.54	$p=0.5, q=1.03$
M8	-2349.14	1.59	$p_o=0.6, p_1=0.4, p=0.88; q=2.52, \omega=1.10$
SGNH Hydrolase			
M1	-4916.54	1.95	$p_o=0.9, p_1=0.1, \omega_0=0.07, \omega_1=1$
M2	-4916.54	1.95	$p_o=0.9, p_1=0.1, p_2=0; \omega_0=0.1, \omega_1=1, \omega_2=1.03$
M7	-4867.25	1.80	$p=0.37, q=2.99$
M8	-4864.38	1.80	$p_o=0.994, p_1=0.006; p=0.39, q=3.5; \omega=1.79$
Vasa-interacting			
M1	-8424.96	1.84	$p_o=0.75, p_1=0.25, \omega_0=0.09, \omega_1=1$
M2	-8424.95	1.84	$p_o=0.75, p_1=0.24, p_2=0; \omega_0=0.09, \omega_1=1, \omega_2=34$
M7	-8335.16	1.63	$p=0.42, q=1.77$
M8	-8334.98	1.63	$p_o=0.99, p_1=0.003; p=0.42, q=1.85; \omega=1.65$
Valois-interacting			
M1	-2559.41	2.37	$p_o=0.8, p_1=0.2, \omega_0=0.06, \omega_1=1$
M2	-2559.41	2.37	$p_o=0.8, p_1=0.16, p_2=0.04; \omega_0=0.06, \omega_1=1, \omega_2=1$
M7	-2528.22	2.02	$p=0.36, q=2.24$
M8	-2526.36	2.03	$p_o=0.99, p_1=0.01; p=0.39, q=2.9; \omega=1.72$

Table S3. Log likelihood and parameter estimates from branch site models for Oskar structural and interaction domains using MUSCLE MSA.

		ω values under Model: MA					ω values under Model: MA _{fix}			
	Site Class	Proportion	Background	<i>D. virilis</i>	<i>I</i>	Proportion	Background	<i>D. virilis</i>	<i>I</i>	
Long Osk	0	0.65	0.24	0.24	-4358.84	0.64	0.23	0.23	-4366.19	
	1	0.19	1.00	1.00		0.22	1.00	1.00		
	2a	0.12	0.24	803.22		0.10	0.23	1.00		
	2b	0.04	1.00	803.22		0.04	1.00	1.00		
LOTUS	0	0.84	0.07	0.07	-1923.41	0.79	0.07	0.07	-1923.94	
	1	0.08	1.00	1.00		0.08	1.00	1.00		
	2a	0.07	0.07	3.46		0.11	0.07	1.00		
	2b	0.01	1.00	3.46		0.01	1.00	1.00		
Lasp-binding	0	0.42	0.15	0.15	-2350.39	0.20	0.15	0.15	-2352.01	
	1	0.40	1.00	1.00		0.19	1.00	1.00		
	2a	0.09	0.15	999.00		0.31	0.15	1.00		
	2b	0.09	1.00	999.00		0.30	1.00	1.00		
SGNH Hydrolase	0	0.85	0.06	0.06	-4916.29	0.85	0.06	0.06	-4916.29	
	1	0.10	1.00	1.00		0.10	1.00	1.00		
	2a	0.05	0.06	1.00		0.05	0.06	1.00		
	2b	0.01	1.00	1.00		0.01	1.00	1.00		
Vasa-interacting	0	0.72	0.09	0.09	-8423.51	0.68	0.09	0.09	-8423.84	
	1	0.23	1.00	1.00		0.22	1.00	1.00		
	2a	0.04	0.09	3.63		0.07	0.09	1.00		
	2b	0.01	1.00	3.63		0.02	1.00	1.00		
Valois-interacting	0	0.70	0.06	0.06	-2558.43	0.70	0.06	0.06	-2558.43	
	1	0.17	1.00	1.00		0.17	1.00	1.00		
	2a	0.10	0.06	1.00		0.10	0.06	1.00		
	2b	0.03	1.00	1.00		0.03	1.00	1.00		

Online Resource 3

This file contains the Oskar amino acid alignment generated with sequences from 18 Drosophilids using the PRANK MSA, and the results of multiple PAML analyses of this alignment. The following elements comprise this Additional File:

Figure S2: Amino acid alignment.

Table S4: Log likelihood values and parameter estimates under partition models of Oskar domains.

Table S5: Log likelihood values and parameter estimates under site models of Oskar domains.

Table S6. Log likelihood values and parameter estimates under branch site models of Oskar domains.

Figure S2. Amino acid alignment of Oskar generated by PRANK MSA used for PAML analysis. Residues with BEB probability of positive selection = 0.92 indicated in red, p = 0.98 in blue, and p = 0.99 in green.

Dmel	MAAVTSE----FPSKPISYTST--NTSAK-TYYLKSVKRVTTCFQQLRDKL-----
Dsec	MATVTSE----FPSKPISYTST--NTSAK-TYYLKSVKRRLTTCFQQLREKL-----
Dsim	MATVTSE----FPSKPISYTST--NTSAK-TYYLKSVKRVTTCFQQLREKL-----
Dere	MAAVTSE----FPSKPISCTST--NTSAK-SNYLKSVKRVTTCFQQLREKL-----
Dyak	MAAVTSE----FPSKPISYTST--NTSAK-TYYLKSVKRVTTCFQQLREKL-----
Dfic	MAAVRSE----FPSKPNSNTY--N--SK-FDDLRSVKKRVTTCFLKKLCEKL-----
Dele	MAAVRSE----FPSEPNNSNTN--NTSTK-TDRLRSVKNRVTTCFLKQLCEKL-----
Drho	MAAVRSE----FPSEPYNSNTNN--NTSTK-TDCLRSVKKRVTTCFLKQLCEKL-----
Dbia	MAAVGSE----FPSEPISCRNT--NTPAK-PDRLRSVKKRVTTCFLQQLREKL-----
Dtak	MAAVRSE----FPSEPISYTNT--NTSNK-SNCLRSVKKRVTTCFLQQLREKL-----
Deug	MATVTSE----FPSKPISYTNTNTNTSTK-TYCLRSVKKRVTTCFLQQLREKL-----
Dkik	MAAVRSE----FPSE---SNS--NPSTK-ADSLRSLKKHVATCWHNLCKKL-----
Dbip	MAAVTSE---LPRE---PTT--TTNTK-SDRLRAVKKRVTTCFLQQCQKL-----
Dana	MAAVTSE---LPRE---PTT--TTNTK-SDRLRAVKKRVTTCFLQQCQKL-----
Dper	MATYRSAHSEPYTSK---SNS--NSNSSSSSRLTALRKRLTTCLQQCQHL-----
Dpse	MATYRSAHSEPYTSK---SNS--NSNSSSSSRLTALRKRLTTCLQQCQHL-----
Dvir	MATFRSE---F-----NS--VPNTY-NQRLTALKKKLTTCFQQWRQQLLKLHQQQQ
Dimm	MATFRSE---F-----NS--VPNTH-PDRLTAFRKKLTTCFHQWRQKL-----
Dmel	-----QSSGSFRKS-SSS---CLNQIFV RS -DFSA
Dsec	-----QPFGSFRKS-SSN---RLYKIFV KS -DFSA
Dsim	-----QPFGSFRKSSSS---RLYKIFV KS -DFSA
Dere	-----QASGSFRKS-SPS---RLYKIFV KA -DFAA
Dyak	-----QASGSFRKS-STS---RLYKIFV KA -DFAA
Dfic	-----QAPGSFRKS-SSS---RFHKIFV RA -DYS--
Dele	-----QATGSFRKS-SSS---RFYKIFV KA -DYSG
Drho	-----QAPGSFRKS-SSS---RFYKIFV KA -DYSG
Dbia	-----PAAGSFRKS-STG---RFYQVFV RA -DFSA
Dtak	-----PASGSFRKS-SSS---RFYKIFV KA -DFSA
Deug	-----PATGSFRKS-STS---RFHRFFI KS -DFSD
Dkik	-----QQPASFRKS-SHS---RFYKIFL KA -DCSAS
Dbip	-----QQPASFRKS-SSS---RFYRIFV KA -DYTA
Dana	-----QQPASFRKS-SSS---RFYRLFV KA -DYTA
Dper	-----QKPASFQKC-KSGPNHQFYKIFE EV FDSKV
Dpse	-----QKPASFQKC-KSGPNHQFYKIFE EV FDSKV
Dvir	QPKQQQQQQQHQSQHQHQQQQQQKQKQRGSE-----HFYNIFV RA -DCSI
Dimm	-----Q-----HF----- KH -DYST
Dmel	CGE----R----FRKIFKS-----ARK-TEL-----
Dsec	CGE----R----FQKIFKS-----ARK-TES-----
Dsim	CGE----R----FQKIFKS-----ARK-TES-----
Dere	CGE----R----FQRIFKS-----ARK-TQS-----
Dyak	CGE----R----FQRIFKS-----ARK-TES-----
Dfic	--DCSKFVRR----FQKIFKS-----ARK-TES-----
Dele	LREKSKFLRQ----LQKIFYKS-----TRK-TES-----
Drho	FREKSKFVRQ----FQKIFKS-----ARK-PES-----
Dbia	FGER-----FRKIFKS-----ARK-TES-----
Dtak	FGER-----FQKIFKS-----ARK-TES-----
Deug	FSE----K----LQKIFKS-----ARK-SES-----
Dkik	FPE----KSKFLRQFHIFKA-----GRK-SEGGEPD
Dbip	IGD----KFKSVRQFHIFKA-----GRKKS-----
Dana	FGD----KFKSVRQFHIFKA-----GRK-G-----
Dper	FGD----KTKYIRQFKKIFKGN-----KDLAKK-S-----
Dpse	FGD----KTKYIRQFKKIFKGN-----KDLAKK-S-----
Dvir	LSE----KVTYLRKFKRIFTARQNLASARTAATAASAKQPVLGRCIG-----
Dimm	LRA----KSICWQKFKKIYTAT-----KLIS-----

Dmel --PELWK-----VPLVAHEL-T-SRQSS---Q--QLQVVARLFSSTQIST-KE--
 Dsec --PELWK-----VPLVAHEL-T-SRQSS---Q--QLQIVARLFSSTQIST-KE--
 Dsim --PELWK-----VPLVAHEL-T-SRQSS---Q--QLHIVARLFSSTQIST-KE--
 Dere --PELWK-----VPLAAHDL-TNGRQTS---Q--QLRIVARLFSSTQIST-KE--
 Dyak --PELWK-----VPLA-HDLTTTSRQSS---Q--QLRIVARLFSSTQIST-RE--
 Dfic --PESWK-----VPL-AHDV-T--RQSS---IH--QLQIVARLFSSTQIST-EE--
 Dele --PELWK-----VPL-AHDV-T--GQSS---Q--QLQIVTRLFSSTLIST-EE--
 Drho --PELWK-----VPL-AHDV-T--GQNS---Q--QLQIVARLFSSTLVST-EE--
 Dbia --PELWK-----VPL-AHDV-T--GPSS---Q--HLQIVARLFSSTLIST-KE--
 Dtak --PELWK-----VPL-AHDV-T--GQSG---Q--QLQIVARLFSSTLIST-KE--
 Deug --PELRR-----VPLSTHDQ-QTSGQSS---Q--QLQIVARLFSSTLIST-KE--
 Dkik PEPELWKE-----VPLA-DE-Q--RSQ---E--QLQIVARLFSSTLIST-KEEE
 Dbip -----K-----EPAESVDE-R-----Q--QLQFVARLFSSTLIST-EE--
 Dana -----K-----EPAESVDE-R-----Q--QLQFVARLFSSTLIST-EE--
 Dper -----E-PGRVVLVLVPEKPVDD-T--EES---QKKQQLQVARLFSSTLISAKEE--
 Dpse -----E-PGRVVLVLVPEKPVDD-T--EES---QKKQQLQVARLFSSTLISAKEE--
 Dvir -----QEE---Q--QLQIVASLFSSTLIS-----
 Dimm -----QEEVEE-Q--QLQIVATLFSSTLIS-----

Dmel -IT---Y-NS---NS---NTSE-----NN-MTIIESNYISVREEYPDIDS
 Dsec -IT---Y-SS---NN---NTSE-----TN-MTIIESNYISVREDYPDIDA
 Dsim -IT---Y-SS---NS---NTSE-----SN-MTIIESNYISVREDYPDIDA
 Dere -IT---YSTS---NN---NTSYSFRSRKSE-SC--RSN-MTIIESNYIAVREDYPDIDV
 Dyak -IT---YTNS---NN---SSYSFRRKSE-SH--RSN-MTIIESNYIGVREDYPDIDT
 Dfic -II---Y-NN-----ISFNSRKSQ-SH--SSN-MTIIESNYIAVREEYPDIDA
 Dele -IT---Y-NT---NNKNN---NSCKSRKSHKSY--KSS-MTIIESNYIAVREEFPDIDT
 Drho -IT---Y-HT---NI-----NSYKSRKSQ-SY--DCD-MTIIESNYIAVREEYPDIDT
 Dbia -IS---Y-SN---HR-----YSCRSGKSE-SY--SSN-MTIIESNYIAVREEHPDIDK
 Dtak -IH---C-NN---NS-----YSYKSRKSE-SY--SSN-MTIIESNYITVREEFPDIDA
 Deug -IT---F-NNHTKNS-----FSYKCRKSE-GY--SSN-MTIIEGNYIDVREENPDIDA
 Dkik IITTSRS-SN---CN-----TSYTRRKSO-SYCSTST-MTIIEGNYREVRDEYPDIDA
 Dbip -I---Y-GN---GK-----TSYSSRKSQ-SY--NSNSMTIIESNYIAVREENPDIDA
 Dana -I---Y-GN-----SSRKSQ-SF--NSNSMTIIESNYIAVREENPDIDA
 Dper -I-----SC-SY--NSSSMTIIDSNYIEVREEYPEIDS
 Dpse -I-----SC-SY--NSSSMTIIDSNYIEVREEYPEIDS
 Dvir -----KHSSSASI-T---SSSSMTIMDNTYIGVRDEYPDIDT
 Dimm -----KYTSSKNS-S---SCNMTIMDDSYIGVRDEYPDIDA

Dmel EVRAILLSHAQNGITISSIKSEYRKLGNPFPFLHDNVTDFLLTIPNVTAECSESGKRIFN
 Dsec EVRAILLSHAQNGITISNIKSEYRKLGNPFPFLHDNVTDFLLTIPNVTAECSETGKRIFN
 Dsim EVRAILLSHAQNGITISNIKSEYRKLGNPFPFLHDNVTDFLLTIPNVTAECSETGKRIFN
 Dere EVRAILLSHAQNGITISNIKSEYRKLGNPFPFLHDNVTDFLLTIPNVTAECSESGKRIFN
 Dyak EVRAILLSHAQNGITISNIKSEYRKLGNTFPLHDNVTDFLLTIPNVTAECSESGKRIFN
 Dfic EVRAILLSHAQNGITISNIKSEYRTQTGNPFPVQENVTDFLLTIPNVTAECSESGKRIFN
 Dele EVRAILLSHAQNGITISNIKSEYRKQTGCQFPFLHDNVTDFLLTIPNVTAECSESGKRIFN
 Drho EVRAILLSHAQNGITISNIKSEYRKQTGIQFPFLHDNVTDFLLTIPNVTAECSESGKRIFN
 Dbia EIRAILLSHAQNGITISSIKSEYRKQTGNPFPFLHDNITDFLLTIPNVTAECSESGKRIFN
 Dtak EVRAILLTHAQNGITISNIKSEYRKQTGNAFPFLHDNITDFLLTIPHVTAECSESGKRIFN
 Deug EVRAILLTHAQNGITISSIKSEYRKLTGNPFPFLHDNVTDFLLTIPNVTAECSETGKRIFN
 Dkik EVRAILLSHAQNGITISSIKDEYRKQTGNTFPLHDNVTDFLLTIPVTAECSEHTGKRIFN
 Dbip EVRAILLSHAQNGITISSIKSEYRKQTGNTFPLHDNVTDFLLTIPVTAECSETGKRIFN
 Dana EVRAILLSHAQNGITISSIKSEYRKQTGNTFPLHDNVTDFLLTIPVTAECSETGKRIFN
 Dper EIRSILLANAQSGITISTIKNEYRKLGNPFPFLHENITDFLLTIPHVTAECCVSGKRIFN
 Dpse EIRSILLANAQSGITISTIKNEYRKLGNPFPFLHENITDFLLTIPHVTAECCVSGKRIFN
 Dvir EIRSILLANAQNGITISSIKKEYRQLTGTAFPFLHDNITDFLLTIPHVTAECCCESGKRIFN
 Dimm EIRAILLANAQKGITISSIKSEYRNMTGNAFPFLRENITDFLLTIPVTAECCMSGKRIFN

Dmel LKASLKNGHLLDMVLNQKER-----TSDYSSGAPSLEN---I----P---RAPPRYWK
 Dsec LKANLKNHGHLLEMVLNQKORT--TNSTSDYSSGAPSLEN---I----P---RAPPRYWK
 Dsim LKANLKNHGHLLEMVLNQKORT--TNSTSDYSSGAPSLEN---I----P---RAPPRYWK
 Dere LKANLKNHGHLLEMVLNQKORTG---SSDYSSGAPSVES---M----P---RAPPRYWK
 Dyak LKANLKNRHLLEMVLNQKORTGTANGTSDYSSGAQSVEN---M----P---RAPPRYWK
 Dfic LKPNTKNRHLLEMLVNLNQKORT-----SDTSSGAPSVES---H----S---RPPPYYWR
 Dele LKPNVKNRHLLEMVLNQKORT-----SDYSSGAPSVED---K----P---RPPPYYWK
 Drho LKPNAKNRHLLEMVLNQKORT-----SDYSSGTPSVED---N----P---RPPPYYWK
 Dbia LKPSLKNRHLLEMVLNQKORS-----SDCSSGAPSVED---Q----P---RAPPRYWK
 Dtak LKPNLKNRHLLEMVLNQKORA-----SDYSSGAPSLED---K----P---RAPPRYWK
 Deug LKANLKNRHLLEMVLNQKQP-----SDNSSEAPTVED---K----S---RVPPRYWK
 Dkik LKPNKENSHLLEMVLNQKQR-----SDYSSGAPSIED---KYRMGPP---QVPPRYWR
 Dbip MKPNKDNRHLYDMVLNQKQR-----NDCSSGPPSSVEYN-R---PVGQRGPPRHWR
 Dana MKPNKDNRHLYDMVLNQKQR-----NDFSSGPPSTVDYNSK---PVGQRGPPRHWR
 Dper MKPRPETRHLHYDMVVHQKRT-----NDLG---TND---H---PVAAP-PPRLWR
 Dpse MKPRPETRHLHYDMVVHQKRT-----NDLG---TND---H---PVAAP-PPRLWR
 Dvir IKPTEHTRHLHEMILQQROR-----DSVSNP-----IQAQEPPRLWR
 Dimm IKPTEDTRHLHDMVLNQQROR-----ENGNNHNSNNH---H---VLDAQEPPRLWR

Dmel NPFKRRALSQLN-----T---SPRTVP-----KI--TDE-KTKDIA--TR-PV
 Dsec NPFKRRALSQLN-----T---SPRTVP-----KI--TNE-RTKEIA--TR-PA
 Dsim NAFKRRALSQLN-----T---SPRTVP-----KI--TDE-QTKEIP--TR-PA
 Dere NPFKRRALSQLN-----NLNTES-----KI--TDE-QTQETS--TR-PA
 Dyak NSFKRRALSQLN-----D---NLNTVP-----KL--TDE-QTKESA--TRQPA
 Dfic NSYKRKALSQLD-----S---NLNIRP-----KT--ADY-QAKDIA--TK-AA
 Dele FPYKRRALSQLE-----N---NLNTGP-----KLKEQEH-QFQAAT--TK-AA
 Drho FPYKRRALSQLD-----N---NLNTRP-----KL--EDH-QFKAIT--AK-TT
 Dbia NPYKRRALSQLD-----N---NLNTVP-----KL--TAD-QTKAVD--TR-AA
 Dtak NPYKRRALSQLD-----N---NLNTVH-----KL--TDE-QTKAIN--TK-AA
 Deug NPYKRRALSQLD-----N---NLNTVH-----KL--TDD-QAKAMSMHTR-AA
 Dkik NGFKRRALSPLD-----N---NLN-VKV---NSEKP--ADD-QSKAIN--TR-SA
 Dbip YQYKRRAQSPLE-----NTNQIKQLNTEQN---GMEKL--QDNVQIRAIK--TN-SA
 Dana YQYKRRAQSPLE-----NLNQMKHLNTERQ-----VQDKAIN--AN-SA
 Dper CQYKRRALYSLNTNNNNNNYLNQNINHIEEKVPAVGGKEVA--IED-QA-----G
 Dpse CQYKRRALYSLNTNNNNNNYLNQNINHIEEKVPAVGGKEVT--IED-QA-----G
 Dvir AQYKRRIP---Q--HFNFNLN-----TSEKPPAV---KI--S-----
 Dimm SQYKRRCL---QNEYNFNLN-----ICEKPPAA---KI--T-----

Dmel SLHOMANEAAESNWCYQDNWKLHNNFYQQASVNAPK--MPVPINIYSPDAPEE-PINLAP
 Dsec SLHOMANEAAESNWCYQDNWKLHNNFYQQASVNAPQ--MPVPINIYS-DAPEE-PTHLAA
 Dsim SLHOMANEAAESNWCYQDNWKLHNNFYQQASVNAPQ--MPVPINIYS-DAPEE-PTHLAA
 Dere SLHOMANEAAESNWCYQDNWKLHNNFYQQASVNAPK--MPVPINIYS-DAPEE-RIQVDA
 Dyak SLHOMANEAAESNWCYQDNWKLHNNFYQQSSVNAPK--MPVPVNIIYS-DAPEE-RIHLDA
 Dfic PLPQLAKEAAESNWCYQDNWKLHNNYYQQSSANEPO--IPVPINIYT-DATEE-PNQFAV
 Dele PLQQLAKEAAESNWCYQDNWKLHNNFYQQSSANEPO--MPVPINIYS-DATEE-RTQFAV
 Drho PLHQLAKEAAESNWCYQDNWKLHNNFYQQSSANEPO--RPVPINIYS-DPPEE-RTQFAV
 Dbia PLQQLAKEAAESNWCYQDNWNHLNNIYQRSSVSEPOIQLPVPINIYS-DAPEE-PNQMV
 Dtak PLQQLAKEAAESNWCYQDNWNHLNNFYQQSRVNEPO--MPVPINIYS-DATEE-PIQLDA
 Deug PLQQLAKVAAESNWCYQDNWKLHNNFYQQSSVNEAO--RPVPINIYT-DALEDLPIQLDN
 Dkik ALPHMAAKAAESNWCYQDNWNHLKNIYQQSSGNAPH--MHEPNQIYT-DPPEL-SNHLAN
 Dbip SLQQLAKAAGEESNWCYQDNWNHLNNFYQQNNVNGPO--NPVPPNIYI-DTPDQ-PTQVTD
 Dana TLQQLAKAAGEESNWCYQDNWNHLNNFYQQSVNGHQ--NQVPPNIYV-DTPDR-QTOVTD
 Dper PLQQLARAAAESNWCYQDNWNHLKNCYQRA-----NIFA-DPPEK-PPPLPA
 Dpse PLQQLARAAAESNWCYQDNWNHLKNCYQRA-----NIFA-DPPEK-PPPLPA
 Dvir KLQPLATAAMMSNDVYQDNWKLHNNQYQL-----PQLNA
 Dimm PVQHLATAAMAPFGVYQDNWKLHNNQYQL-----SQLAT

Dmel **PGH**--QPS-CRTQS-----Q----KTE-PTEN-----
 Dsec **PGQ**--QPSCCRTQS-----K----KTE-PTEN-----
 Dsim **PGQ**--QPSCCRTQS-----K----KPE-PTEN-----
 Dere **PGH**--QPS-CRTQN-----Q----KTQHPTEN-----
 Dyak **PGH**--QPS-CRTQN-----Q----KTQQTEN-----
 Dfic **PEH**--LPP-CKNHK-----Q----NTK-PNDN-----
 Dele **PEP**--RPC-GKNHN-----Q----NNK-PNVN-----
 Drho **PEL**--PPH-CRNP-----Q----NTK-QNVN-----
 Dbia **PQH**--HPL-C--HNRNQNPSQN-----Q----QAK-QDAN-----
 Dtak **PEH**--HPH-CQNHNEQNQPNQN-----Q----NTK-QNVN-----
 Deug **PEH**--HPN-CKNNN-----Q----NIK-QNVN-----
 Dkik **PELI**--KPP-CKNHN-----S----PNLN-----
 Dbip **PEL**--LH--CQN-----RNPS-----
 Dana **PEQ**--H--CQN-----PHPS-----
 Dper **QEN**--H--LRNSN-----S----QN-----
 Dpse **QEN**--H--LRNSN-----S----QN-----
 Dvir **A**-----KNN-----IHS-IASNPAQLQAALPAE---HHPS-----
 Dimm **Q**-----NSN-----QFSCIRSDPVQLOIA-----PHPPEQQHQQRQOO

Dmel -RH--LGIFVHPFNGMNI-----MKRRHEM-----
 Dsec -RH--LGIFVHPFNDMV-----LKRRHEM-----
 Dsim -RH--LGIFVHPFNDMV-----LKRRHEM-----
 Dere -RH--LGIFVHPFSDMV-----MKRRHEMPTP-----
 Dyak -RH--LGIFVHPFNDMV-----MKRRHEM-----
 Dfic -QH--LGIFMRPYNDMV-----LKKRDL-----
 Dele -QQ--LGSFVNPFNDMNI-----LKRRHDM-----
 Drho -QH--LGIFVNPFNDMV-----LKKRHD-----
 Dbia -QH--LGIFVQPFSDMV-----LKRRQEM-----
 Dtak -QH--LGIFVQPFNDMNI-----LKRRHEL-----
 Deug -QH--LGNFVNPLNDVNP-----FKRRODI-----
 Dkik -QI--L-----NPNLDI-----FRRPRSSCDLKRRHER-----
 Dbip -QN--Q-----NYHQDLYNNHNNDMLEFKRPR-----DL-----
 Dana -QN--Q-----NSQQDIYNNQNRELFEKRPR-----DL-----
 Dper -----HRDN-----LKNQR-----NL---LVNEHHSM-----
 Dpse -----HRDN-----LKNQR-----NL---LVNEHHSM-----
 Dvir -KHVEE-----YAH-----KRRH-----EY-----
 Dimm QQQIGE-----YSH-----KRRH-----EF-----

Dmel -----TPTPTILTS-GT-YNDSLLTINSDYDAYLLDFPLMGDDFMLYLARMELK
 Dsec -----TPTPTILTS-GT-YNDSLLTINSDYDAYLLDFPLMGDDFMLYLARMELK
 Dsim -----TPTPTILTS-GT-YNDSLLTINSDYDAYLLDFPLMGDDFMLYLARMELK
 Dere -----TPTPTTLSS-GT-YADSLLTINSDYDAYLLDFPLMGDDFMLYLARMELK
 Dyak -----TPTPTILTS-GT-YSDSLLTINSDYDAYLLDFPLMGDDFMLYLARMELK
 Dfic -----TPTPTTLSS-GT-FNDSILTMSSDYDAYLLDFPLMGDDFLYLARMELN
 Dele -----TPTPTTLSS-GT-YNDSMLTISSDYDAYLLDFPLMGDDFMLYLARMELK
 Drho -----TPTPTTLSS-GI-YSDSILTISSDYDAYLLDFPLMGDDFMLYLARMELK
 Dbia -----TPTPTTLSS-GT-YNDSMLTINSDYDAYLLDFPLMGDDFMLYLARMELK
 Dtak -----TPTPTTLSS-GT-YNDSMLTINSDYDAFLDFPLMGDDFMLYLARMELK
 Deug -----TPTPTIFTSGGS-SNNSLMTINSDYDAFLDFPLMGDDFMLYLARMELK
 Dkik -----TPTPTTMSS-AT-HNDSMLTISSDYDAYLLDFPLMGDDFLYLARMELK
 Dbip -----TPTPT-TTS-GT-QNDSSMMTISSDFDAFLDFPLMGDDFFFLYLARMELN
 Dana -----TPTPT-MTS-AT-HNDSMRTIISDFDAFLDFPLMGDDFFFLYLARMELN
 Dper QQEYQYQYPTTPTPTTISS-GT-QHDSMFTINSDYDAYLLDFPLLGDDFLYLARMELR
 Dpse QQEYQYQYPTTPTPTTISS-GT-QHDSMFTINSDYDAYLLDFPLLGDDFLYLARMELR
 Dvir -----TPTPTTLSC-PSTQHDSMFTINSDYDAYLLDFPLLGDDFFFLYLARMELK
 Dimm -----TPTPTTLSC-PS-QHDSMFTINSDYDAYLLDFPLLGDDFFFLYLARMELK

Dmel CRFRRHERVLQSGLCVSGLTINGARNRLKRVQLPEGTQIIVNIGSVDIMRGKPLVQIEHD
Dsec CRFRRHERVLQSGLCVSGLTINGARNRLKRVQLPEGTQIIVNIGSVDIMRGKPLVQIEHD
Dsim CRFRRHERVLQSGLCVSGLTINGARNRLKRVQLPEGTQIIVNIGSVDIMRGKPLVQIEHD
Dere CRFRRHERVLQSGLCASGLTINGARNRLRRLQVQLPEGTQIIVNIGSVDIMRGKPLVQIEHD
Dyak CRFRRHERVLQSGLCVSGLTINGARNRLKRVQLPEGTQIIVNIGSVDIMRGKPLVQIEHD
Dfic CRFKRYERVLQSGLCVSGQTINGARNRLKKANLPEGTQIIVNIGSVDIMRGKPLVQIEHD
Dele CRFRRHERILQSGLCVSGQTINGARRRLKKVQLPAGTQIIVNIGSVDILCGRPLVQIEHD
Drho CRFKRHERVLQSGLCVSGQTINGARKRLKKVHLTEGTQIIVNIGSVDILCGRPLVQIEHD
Dbia CRFRRHERVLQSGLCVSGQTINGARMRLKKFHLPEGTQIIVNIGSVDILRGRPLVQIEHD
Dtak CRFRRYERILQSGLCVSGQTINGARNRLKKFHLPEGTQIIVNIGSVDILRGRPLVQIEHD
Deug CRFKRHERVLQSGLCASGQTINGARNRLRKVQLPEGTQIIVNIGSVDIMKGKPLVQIEHD
Dkik CRFRRYERVLQSGLCVSGQTINAARNRLKRVNLPEGTQIIVNIGSVDILRGKPLVQIEHD
Dbip CRFKRYERVLQSGLCVSGQTIIGARNRFRKVYLPENTQIIVNIGSVDIMRGKPLVQIEHD
Dana CRFKRYERVLQSGLCVSGQTIIGARNRFRKVYLPETQIIVNIGSVDIMRGKPLVQIEHD
Dper CRFKRTERVLQSGLCVSGQTISGARSLHLLVNKGTOIIVNIGSVDIMRGRPIVQIQHD
Dpse CRFKRTERVLQSGLCVSGQTISGARSLHLLVNKGTOIIVNIGSVDIMRGRPIVQIQHD
Dvir CRFKFEKVLQSGLCISGQTINAARQRLRLVELPEMTQIIVNIGSEDIMGRGRSLVQIEHD
Dimm CRFKDEKVLQSGLCISGQTINAARQRVQHVELQEMTQIIVNIGSVDIMRGKPLVQIEHD

Dmel FRLLIKEHMNRMLVPILTNLAPLGNYCHDKVLCDKIYRFNKFIRSECC-HLKVIDIHSL
Dsec FRLLIKEHMNRMLVPILTNLAPLGNYCHDKVLCDKIYRFNKFIRSECC-HLKVIDIHSL
Dsim FRLLIKEHMNRMLVPILTNLAPLGNYCHDKVLCDKIYRFNKFIRSECC-HLKVIDIHSL
Dere FRLLIKEHSMRYVPILTNLAPLANYCHDKVLCDKVHRFNKFIRSECC-HLKVIDIHSL
Dyak FRLLIKEHMNRMLVPILTNLAPLANYCHDKVLCDKIHRFNKFIRSECC-HLKVIDIHSL
Dfic FRLLIKEHMNRRFVPILTNLAPLANYCHDKMCDKMSRFNKFIRSEGC-HLKIIDIHSCL
Dele FRLLIKEMHSORFVPILTNLAPLANYCHDKTMCDKIYRFNKFIRSECC-HLKYIDIHSCL
Drho FRLLIKEMHSORLVPILTNLAPLANYCHDKILCDKVHRFNKFIRSEGS-HLKYIDIHSCL
Dbia FRLLIKEMHSORFVPILTNLAPLANYCHDKELCDKVNRFNKFIRSEGR-HLKVVDIHSCL
Dtak FRLLIKEHNQRYVPILTNLAPLANYCHDKDLCDKVSRFNKFIRAEGR-HLKVVDIHSCL
Deug FRLLIKEHNQRYIPILTNLAPLANYCHDKVLCDKITRFNKFIASEG-R-HLKVINIHSCL
Dkik FRLLIKEHNRRFVPILTNLAPLANYCHDKGLCDKVLRFNNFIRSEGRSHLKVIDIHSL
Dbip FRLLIKEIHSRCIPILTNLAPLANYCHDKVLCDKISKFNKFVRNEARSHLKFDINSCL
Dana FRLLIKEIHSRCIPILTNLAPLANYCHDKVLCDKITKFNRFVRNEARSHLKFDINSCL
Dper FRQLVKDMHNRLVPILTTLAPLANYCHDKAMCDKVVFKNFQFIWKECASYLKVIDIHSL
Dpse FRQLVKDMHNRLVPILTTLAPLANYCHDKAMCDKVVFKNFQFIWKECASYLKVIDIHSL
Dvir FRLLVKEMHNRRFVPVLTTLAPLANCRAHDKQTCDFKVSRFNKFIRSEGR-HLKVIDIHSL
Dimm FRQLIKEHMNRRFVPVLTTLAPLANYCHDKQTCDFKVSRFNKFIRNEGR-HLTVIDIHTCL

Dmel INERGVVRFDCFOASPRQVTGSKEPYLFWNKIGRQRLQVIETSLEY-----
Dsec VNERGVVRFDCFOQSPRQVTGSKEPYLFWNKIGRQRLQIIETSLEY-----
Dsim VNERGVVRFDCFOQSPRQVTGSKEPYLFWNKIGRQRLQIIETSLEY-----
Dere INERGVVRFDCFOSSPRQVSGSKEPYLFWNKIGRQRLQVIETSLEY-----
Dyak INERGVVRFDCFOSSPRQVSGSKEPYLFWNKIGRQRLQVIETSLEY-----
Dfic VNERGIVRFDCFOSSPRQVTGSKEPYLFWNKIGRQRLQTIETNLEY-----
Dele INERGIVRFDCFOSSPRQVTGSNQPYLFWNKIGRQRLQTIETNLEY-----
Drho INERGIVRFDCFOSSPRQVTGSKEPYLFWNKIGRQRLQVIESSLEY-----
Dbia INERGVVRFDCFO-----QEPYLFWNKIGRQRLQIIESSLEY-----
Dtak INESGIVRFDCFOSSPRQVTGSREPYLFWNKIGRQRLRIIESSLEY-----
Deug INERGIVRFDCFO-----NLIYSGTKSGGSASCKL
Dkik INERGVVRFDCFOQAPRSVSGSKEPYLFWNKIGRQRLQMIESSLEY-----
Dbip INEKKNVLFDCFOQAPRTVTGSKEPYLFWNKIGRQRLQTIETSCLEY-----
Dana INEKKNVLFDCFOQAPRTVTGSKEPYLFWNKIGRQRLQTIETSCLEY-----
Dper VNENGVVRFDCFOYSSRNTGSKESYVFWNKIGRQRLQMIASLEY-----
Dpse VNENGVVRFDCFOYSSRNTGSKESYVFWNKIGRQRLQMIASLEY-----
Dvir INENGIVRFDCFOQNGPRSVTGSSEPYVFWNKIGRQRLHMIENLEY-----
Dimm INENGVVRFDCFOQKGPRSVTGSVEPYVFWNKIGRQRLQMIQNLEY-----

Dmel -----X
Dsec -----X
Dsim -----X
Dere -----X
Dyak -----X
Dfic -----X
Dele -----X
Drho -----X
Dbia -----X
Dtak -----X
Deug LNRVWSTKSGDX
Dkik -----X
Dbip -----X
Dana -----X
Dper -----X
Dpse -----X
Dvir -----YX
Dimm -----HX

Table S4. Log likelihood values and parameter estimates under partition models for Oskar structural and interaction domains using PRANK MSA.

	Model	np¹	<i>I</i>²	<i>rs</i>₂	<i>rs</i>₃	<i>rs</i>₄	<i>rs</i>₅	<i>rs</i>₆	<i>K</i>	<i>ω</i>
Mgene=0	B (different <i>rs</i> ³)	50	-24498.39	0.93	0.50	0.75	0.84	0.88	1.65	0.14
Mgene=2	C (different <i>rs</i> and π ⁴)	95	-24497.72	0.88	0.51	0.74	0.83	0.90	1.70	0.14
Mgene=3	D (different <i>rs</i> , κ ⁵ and ω ⁶)	60	-24387.97	0.96	0.58	0.94	0.98	1.08	$\kappa_1=1.56$	$\omega_1=0.34$
									$\kappa_2=1.60$	$\omega_2=0.27$
									$\kappa_3=1.49$	$\omega_3=0.12$
									$\kappa_4=1.79$	$\omega_4=0.08$
									$\kappa_5=1.66$	$\omega_5=0.12$
									$\kappa_6=1.97$	$\omega_6=0.10$
Mgene=4	E (different <i>rs</i> , κ , ω and π)	105	-24380.18	0.93	0.61	0.95	0.98436	1.17	$\kappa_1=1.76$	$\omega_1=0.41$
									$\kappa_2=1.65$	$\omega_2=0.28$
									$\kappa_3=1.56$	$\omega_3=0.11$
									$\kappa_4=1.83$	$\omega_4=0.08$
									$\kappa_5=1.67$	$\omega_5=0.13$
									$\kappa_6=2.06$	$\omega_6=0.10$

¹ np: number parameters

² *I*: log likelihood value

³ *rs*: substitution rate. *rs*₁=1 for all models (not shown). Subscript corresponds to six partitions: (1) Lasp-binding domain (2) Long Osk domain; (3) LOTUS domain; (4) SGNH Hydrolase domain; (5) Vasa-interacting domain; and (6) Valois-interacting domain .

⁴ π : equilibrium codon frequencies

⁵ *K*: transition/transversion ratio

⁶ *ω*: dN/dS ratio

Table S5. Log likelihood and parameter estimates from site models for Oskar structural and interaction domains using PRANK MSA.

	I	κ	Parameter estimates
Long Osk			
M1	-4195.42	1.73	$p_0 = 0.82, p_1 = 0.17, \omega_0 = 0.21, \omega_1 = 1$
M2	-4195.42	1.73	$p_0 = 0.82, p_1 = 0.02, p_2 = 0.15; \omega_0 = 0.21, \omega_1 = 1, \omega_2 = 1$
M7	-4186.96	1.71	$p = 1.08, q = 2.46$
M8	-4186.22	1.71	$p_0 = 0.93, p_1 = 0.06, p = 1.41; q = 3.91, \omega = 1.06$
LOTUS			
M1	-1925.78	1.65	$p_0 = 0.89, p_1 = 0.10, \omega_0 = 0.07, \omega_1 = 1$
M2	-1925.78	1.65	$p_0 = 0.89, p_1 = 0.10, p_2 = 0; \omega_0 = 0.07, \omega_1 = 1, \omega_2 = 1$
M7	-1903.99	1.58	$p = 0.27, q = 1.82$
M8	-1904.03	1.58	$p_0 = 0.99, p_1 = 0.00, p = 0.27; q = 1.83, \omega = 4.39$
Lasp-binding			
M1	-2144.17	1.71	$p_0 = 0.52, p_1 = 0.47, \omega_0 = 0.14, \omega_1 = 1$
M2	-2144.12	1.73	$p_0 = 0.53, p_1 = 0, p_2 = 0.46; \omega_0 = 0.14, \omega_1 = 1, \omega_2 = 1.04$
M7	-2142.52	1.66	$p = 0.44, q = 0.45$
M8	-2140.66	1.72	$p_0 = 0.63, p_1 = 0.36, p = 0.80; q = 2.55, \omega = 1.16$
SGNH Hydrolase			
M1	-4860.40	1.94	$p_0 = 0.89, p_1 = 0.10, \omega_0 = 0.05, \omega_1 = 1$
M2	-4860.40	1.94	$p_0 = 0.89, p_1 = 0.07, p_2 = 0.03; \omega_0 = 0.05, \omega_1 = 1, \omega_2 = 1$
M7	-4803.77	1.79	$p = 0.29, q = 2.57$
M8	-4800.80	1.80	$p_0 = 0.99, p_1 = 0.005; p = 0.31, q = 3.03; \omega = 1.97$
Vasa-interacting			
M1	-8190.91	1.93	$p_0 = 0.78, p_1 = 0.21, \omega_0 = 0.08, \omega_1 = 1$
M2	-8190.91	1.93	$p_0 = 0.78, p_1 = 0.20, p_2 = 0.007; \omega_0 = 0.08, \omega_1 = 1, \omega_2 = 1$
M7	-8107.90	1.87	$p = 0.37, q = 1.72$
M8	-8106.28	1.73	$p_0 = 0.98, p_1 = 0.01; p = 0.4, q = 2.01; \omega = 1.86$
Valois-interacting			
M1	-2592.65	2.20	$p_0 = 0.84, p_1 = 0.15, \omega_0 = 0.06, \omega_1 = 1$
M2	-2592.65	2.20	$p_0 = 0.84, p_1 = 0.11, p_2 = 0.03; \omega_0 = 0.06, \omega_1 = 1, \omega_2 = 1$
M7	-2568.01	2.04	$p = 0.36, q = 2.21$
M8	-2565.72	2.06	$p_0 = 0.98, p_1 = 0.0; p = 0.42, q = 3.14; \omega = 1.64$

Table S6. Log likelihood and parameter estimates from branch site models for Oskar structural and interaction domains using PRANK MSA.

		ω values under Model: MA				ω values under Model: MA _{fix}			
	Site Class	Proportion	Background	<i>D. virilis</i>	<i>I</i>	Proportion	Background	<i>D. virilis</i>	<i>I</i>
Long Osk	0	0.76	0.20	0.20	-4191.98	0.73	0.20	0.20	-4194.7
	1	0.15	1	1		0.15	1	1	
	2a	0.05	0.2	16.5		0.09	0.20	1	
	2b	0.01	1	16.5		0.02	1	1	
LOTUS	0	0.83	0.07	0.07	-1923.4	0.79	0.07	0.07	-1923.93
	1	0.08	1	1		0.08	1	1	
	2a	0.07	0.07	3.46		0.11	0.07	1	
	2b	0.00	1	3.46		0.01	1	1	
Lasp-binding	0	0.41	0.13	0.13	-2139.20	0.44	0.13	0.13	-2143.18
	1	0.35	1	1		0.39	1	1	
	2a	0.12	0.13	35.09		0.08	0.13	1	
	2b	0.10	1	35.09		0.07	1	1	
SGNH Hydrolase	0	0.83	0.05	0.05	-4859.96	0.83	0.05	0.05	-4859.96
	1	0.09	1	1		0.09	1	1	
	2a	0.06	0.05	1		0.06	0.05	1	
	2b	0.07	1	1		0.07=	1	1	
Vasa-interacting	0	0.76	0.08	0.08	-8186.49	0.72	0.07	0.07	-8189.04
	1	0.20	1	1		0.20	1	1	
	2a	0.02	0.08	13.26		0.05	0.07	1	
	2b	0.00	1	13.26		0.01	1	1	
Valois-interacting	0	0.74	0.06	0.06	-2591.50	0.74	0.06	0.063	-2591.50
	1	0.13	1.00	1		0.13	1	1	
	2a	0.10	0.06	1		0.10	0.06	1	
	2b	0.018	1.00	1		0.018	1	1	

Online Resource 4

This file contains amino acid alignments of conserved sequence blocks of specific predicted structural and interaction Oskar domains from 18 Drosophilids using the PRANK MSA, and the results of the PAML analysis of these alignments. The following elements comprise this Additional File:

Figure S3: Amino acid alignments.

Table S7: Log likelihood values and parameter estimates under branch site models of trimmed alignments of individual Oskar domains from 18 Drosophilids.

Figure S3: Conserved sequence blocks of the Long Osk, Lasp-binding, and Vasa-interacting domains from Oskar sequences of 18 Drosophilids aligned using the PRANK MSA. Sites with BEB >0.9, indicating sites under positive selection, are indicated with highlighted in red and marked with an asterisk.

Long Osk domain

	0.91 *
Dper	MATYRSAYTSKSNS----N-LTALRKRLTTCLQQWCQHLQKPASFQKCKSGQFYKIFE E V
Dpse	MATYRSAYTSKSNS----N-LTALRKRLTTCLQQWCQHLQKPASFQKCKSGQFYKIFE E V
Dimm	MATFRSEFNSVPNT----H-LTAFRKKLTTCFHQWRQKLQ-----HF----- K H
Dvir	MATFRSEFNSVPNT----Y-LTALKKKLTTCFQOWRQGLE-----HFYNIFV R A
Deug	MATVTSEFPSPKPISYTNTNTLRSVKRRVTTCLQQLREKLPATGSFRKSSTSFRHRF I K S
Dtak	MAAVRSEFPSEPISYTNTNTLRSVKRKVTTCLQQLREKLPASGSFRKSSSSRFYKIFV K A
Dbia	MAAVGSEFPSEPISCRNTNTLRSVKRKVTTCLQQLREKLPAAAGSFRKSSTGRFYQVFV R A
Dmel	MAAVTSEFPSPKPISYTNTLKVKKRVTTCFQQLRDKLQSSGSFRKSSSSCLNQIFV R S
Dsec	MATVTSEFPSPKPISYTNTLKVKKRVLTTCFQQLREKLOPFGSFRKSSSNRLYKIFV K S
Dsim	MATVTSEFPSPKPISYTNTLKVKKRVLTTCFQQLREKLOPFGSFRKSSSSRLYKIFV K S
Dyak	MAAVTSEFPSPKPISYTNTLKVKKRVTTCFOOLREKLOASGSFRKSSTSRLYKIFV K A
Dere	MAAVTSEFPSPKPICTNTLKVKKRVTTCFQQLREKLOQASGSFRKSSPSRLYKIFV K A
Drho	MAAVRSEFPSEPYNSNTNNNTLRSVKRKVTTCLQQLCEKLOQAPGSFRKSSSSRFYKIFV K A
Dele	MAAVRSEFPSEPNNTNTLRSVKRNVTTCFQQLCEKLOQATGSFRKSSSSRFYKIFV K A
Dfic	MAAVRSEFPSPKNSNTYTN-LRSVKRKVTTCLKKLCEKLOQAPGSFRKSSSSRFHKIFV R A
Dkik	MAAVRSEFPSESNS---NPLRSLKKHVATCWHNLCKLQOPASFRKSSHRSRFYKIFL K A
Dana	MAAVTSELPREPTT---TTLRAVKKRVTTCQOWCQKLQOPASFRKSSSSRFYRLFV K A
Dbip	MAAVTSELPREPTT---TTLRAVKKRVTTCQOWCQKLQOPASFRKSSSSRFYRIFV K A
Dper	GDKTKYIRQFKKIFKG--L----KQLQLVARLFSSSTLIS
Dpse	GDKTKYIRQFKKIFKG--L----KQLQLVARLFSSSTLIS
Dimm	RAKSICWQKFKKIYTA--I----QQLQIVATLFSSSTLIS
Dvir	SEKVTYLRKFKRIFTA--I----QQLQIVASLFSSTLIS
Deug	SE-----KLQKIFKSARPELRRQQLQIVARLFSSSTLIS
Dtak	GE-----RFQKIFKSARPELWKQQLQIVARLFSSSTLIS
Dbia	GE-----RFRKIFKSARPELWKQHLQIVARLFSSSTLIS
Dmel	GE-----RFRKIFKSARPELWKQQLQVVARLFSSSTQIS
Dsec	GE-----RFQKIFKSARPELWKQQLQIVARLFSSSTQIS
Dsim	GE-----RFQKIFKSARPELWKQQLHIVARLFSSSTQIS
Dyak	GE-----RFQRIFKSARPELWKQQLRIVARLFSSSTQIS
Dere	GE-----RFQRIFKSARPELWKQQLRIVARLFSSSTQIS
Drho	REKSKFVRQFQKIFKSARPELWKQQLQIVARLFSSSTLVS
Dele	REKSKFLRQLQKIYKSTRPELWKQQLQIVTRLFSSTLIS
Dfic	-DCSKFVRRFQKIFKSARPESWKHQLQIVARLFSSSTQIS
Dkik	PEKSKFLRQFHRIFKAGRPELWKEQLQIVARLFSSSTLIS
Dana	GDKFKSVRQFHRIFKAGR-----QQLQFVARLFSSSTLIS
Dbip	GDKFKSVRQFHRIFKAGR-----QQLQFVARLFSSSTLIS

Lasp-binding domain

0.95

*

Dper	IGPLQQLARAAAESNWCYQDNWNHLKNCYQRA-----	NIFDPPEKLPAQELRNSN
Dpse	IGPLQQLARAAAESNWCYQDNWNHLKNCYQRA-----	NIFDPPEKLPAQELRNSN
Dimm	-TPVQHLATAAMAPFGVYQDNWKHLNNQYQNS-----	NQFDPVQLHPPEQQQORQ
Dvir	-SKLQPLATAAAMSNDVYQDNWKHLNNQYQKN-----	NIHNPAQLLPAEH-----
Deug	MAPLQQLAKVAAESNWCYQDNWKHLNNFYQQSSVNEAQPVPINIYDALELDLNPECKNNN	
Dtak	IAPLQQLAKEAAESNWCYQDNWNHLNNFYQQSRVNEPQPVPINIVDATEELDAPECQNHN	
Dbia	VAPLQQLAKEAAESNWCYQDNWNHLNNIYQRSSVSEPOQPVPINIVDAPEEMNPQC--HN	
Dmel	IVSLHQMANEAAESNWCYQDNWKHLNNFYQQASVNAPKPVPINIYDAPEELAPPGCRTQS	
Dsec	IASLHQMANEAAESNWCYQDNWKHLNNFYQQASVNAPQPVPINIYDAPEELAAPGCRTQS	
Dsim	IASLHQMANEAAESNWCYQDNWKHLNNFYQQASVNAPQPVPINIYDAPEELAAPGCRTQS	
Dyak	SASLHQMANEAAESNWCYQDNWKHLNNFYQQSSVNAPKPVVNIVDAPEELDAPGCRTQN	
Dere	TASLHQMANEAAESNWCYQDNWKHLNNFYQQASVNAPKPVPINIYDAPEEVVDAPGCRTQN	
Drho	ITPLHQQLAKEAAESNWCYQDNWKHLNNFYQQSSANEPOQPVPINIVDPPEEFAVPECRNP	
Dele	AAPLQQLAKEAAESNWCYQDNWKHLNNFYQQSSANEPOQPVPINIVDATEEFAVPEGKNHN	
Dfic	IAPLPQLAKEAAESNWCYQDNWKHLNNYYQQSSANEPOQPVPINIVDATEEFAVPECKNHK	
Dkik	IAALPHMAKAAAESNWCYQDNWNHLKNIYQQSSGNAPHHEPNQIVDPPELLANPECKNHN	
Dana	IATLOKLAKAAGESNWCYQDNWNHLNNFYQQSVNGHQQVPPNIYDTPDRVTDPECQNP	
Dbip	IASLOKLAKAAGESNWCYQDNWNHLNNFYQQNNVNGPQPVPPNIYDTPDQVTDPECQNRN	

Vasa-interacting domain

Dper	IGPLQQLARAAAESWCYQDNWNHLKNCYQRA-----NIFDPPEKLPAQELRNSN
Dpse	IGPLQQLARAAAESWCYQDNWNHLKNCYQRA-----NIFDPPEKLPAQELRNSN
Dimm	-TPVQHLATAAMAPFGVYQDNWKHLNNQYQNS-----NQFDPVQLHPPEQQQORO
Dvir	-SKLQPLATAAAMSNDVYQDNWKHLNNQYQKN-----NIHNPQQLLPAEH----
Deug	MAPLQQLAKVAAESWCYQDNWKHLNNFYQQSSVNEAQPVPINIYDALEDLDNPECKNNN
Dtak	IAPLQQLAKEAAESWCYQDNWNHLNNFYQQASRVNEPQPVPINIYDATEELDAPECQHN
Dbia	VAPLQQLAKEAAESWCYQDNWNHLNNIYQRSSVSEPPQPVPINIYDAPEEMNVQC--HN
Dmel	IVSLHQMANEAAESWCYQDNWKHLNNFYQQASVNAPKPVPINIYDAPEELAPPGCRTQS
Dsec	IASLHQMANEAAESWCYQDNWKHLNNFYQQASVNAPQPVPINIYDAPEELAAPPGCRTQS
Dsim	IASLHQMANEAAESWCYQDNWKHLNNFYQQASVNAPQPVPINIYDAPEELAAPPGCRTQS
Dyak	SASLHQMANEAAESWCYQDNWKHLNNFYQQSSVNAPKPVVINIYDAPEELDAPGCRTQN
Dere	TASLHQMANEAAESWCYQDNWKHLNNFYQQASVNAPKPVPINIYDAPEEVDAPGCRTQN
Drho	ITPLHQOLAKEAAESWCYQDNWKHLNNFYQQSSANEQQPVPINIYDPEEEFADVPECRNPN
Dele	AAPLQQLAKEAAESWCYQDNWKHLNNFYQQSSANEQQPVPINIYDATEEFAVPEGKNHN
Dfic	IAPLQQLAKEAAESWCYQDNWKHLNNYYQQSSANEQQPVPINIYDATEEFAVPECKNHK
Dkik	IAALPHMAKAAAESWCYQDNWNHLKNIYQQSSGNAPHHEPNQIYDPPELLANPECKNHN
Dana	IATLQKLAKAAGESWCYQDNWNHLNNFYQQSVNVNGHQVPPNIYDTPDRVTDPECQNRH
Dbip	IASLQKLAKAAGESWCYQDNWNHLNNFYQQNNVNGPQPVPPNIYDTPDQVTDPECQNRN
Dper	HRDNLKNQRNNEH-HSMQQEPTTISSHDSMFTINSDYDAYLLDFPLLGDFFLYLARMEL
Dpse	HRDNLKNQRNNEH-HSMQQEPTTISSHDSMFTINSDYDAYLLDFPLLGDFFLYLARMEL
Dimm	QQ-----GEYSHKRRHEPTTLSCHDSMFTINSDYDAYLLDFPLLGDFFFLYLARMEL
Dvir	KH-----EYAHKRRHEPTTLSCHDSMFTINSDYDAYLLDFPLLGDFFFLYLARMEL
Deug	HLGNFVNPLNDVNPFKRRQDPTIFTNNSLMTINSDYDAFLDFPLMGDDFMLYLARMEL
Dtak	HLGIVFQPFNDMNLKRRHEPTTLSSNDMSLTINSDYDAFLDFPLMGDDFMLYLARMEL
Dbia	HLGIVFQPFSDMNVLKRRQEPTTLSSNDMSLTINSDYDAYLLDFPLMGDDFMLYLARMEL
Dmel	HLGIVFVHPFNGMNIMKRRHEPTILTSNDSLLTINSDYDAYLLDFPLMGDDFMLYLARMEL
Dsec	HLGIVFVHPFNDMNVLKRRHEPTILTSNDSLLTINSDYDAYLLDFPLMGDDFMLYLARMEL
Dsim	HLGIVFVHPFNDMNVLKRRHEPTILTSNDSLLTINSDYDAYLLDFPLMGDDFMLYLARMEL
Dyak	HLGIVFVHPFNDMNVMKRRHEPTILTSNDSLLTINSDYDAYLLDFPLMGDDFMLYLARMEL
Dere	HLGIVFVHPFSDMNVMKRRHEPTTLSSADSLLTINSDYDAYLLDFPLMGDDFMLYLARMEL
Drho	HLGIVFVNPNDMNVLKRRHDPTTLSSSDSILTISSDYDAYLLDFPLMGDDFMLYLARMEL
Dele	QLGSFVNPFNDMNLKRRHDPTTLSSNDMSLTISSDYDAYLLDFPLMGDDFMLYLARMEL
Dfic	HLGIFMRPYNDMNVLKRRHDPTTLSSADSLLTINSDYDAYLLDFPLMGDDFMLYLARMEL
Dkik	NLDIFRRPRSSCD-LKRRHEPTTMSSNDMSLTISSDYDAYLLDFPLMGDDFLLYLARMEL
Dana	QQDIYNNQNRELFEFKRPRDPTM-TSNDSMRTIISDFDAFLDFPLMGDDFFLYLARMEL
Dbip	HQDLYNNHNRDMLEFKRPRDPTT-TSNDSSMMTIISDFDAFLDFPLMGDDFFLYLARMEL
Dper	RCRFKRTERVLQSGLCVSGQTISGARSRLHHLLVNKGTOIIVNIGSVDIMRGGRPIVQI
Dpse	RCRFKRTERVLQSGLCVSGQTISGARSRLHHLLVNKGTOIIVNIGSVDIMRGGRPIVQI
Dimm	KCRFKKDEKVLSQGLCISGQTINAARQRVQHVELQEMTOIIVNIGSVDIMRGKPLVQIEH
Dvir	KCRFKKFEKVLQSGLCISGQTINAARQRRLVELPEMTQIIVNIGSEDIMRGRSLVQIEH
Deug	KCRFKRHERVLQSGLCASGQTINGARNRLRKVQLPEGTOIIVNIGSVDIMKGKPLVQIEH
Dtak	KCRFRRYERILQSGLCVSGQTINGARNRLKKFHLPEGTOIIVNIGSVDILRGGRPLVQIEH
Dbia	KCRFRRHERVLQSGLCVSGQTINGARMRLKKFHLPEGTOIIVNIGSVDILRGGRPLVQIEH
Dmel	KCRFRRHERVLQSGLCVSGLTINGARNRLKRVQLPEGTOIIVNIGSVDIMRGKPLVQIEH
Dsec	KCRFRRHERVLQSGLCVSGLTINGARNRLKRVQLPEGTOIIVNIGSVDIMRGGRPLVQIEH
Dsim	KCRFRRHERVLQSGLCVSGLTINGARNRLKRVQLPEGTOIIVNIGSVDIMRGGRPLVQIEH
Dyak	KCRFRRHERVLQSGLCVSGLTINGARNRLKRVQLPEGTOIIVNIGSVDIMRGKPLVQIEH
Dere	KCRFRRHERVLQSGLCASGLTINGARNRLRRVQLPEGTOIIVNIGSVDIMRGKPLVQIEH
Drho	KCRFKRHERVLQSGLCVSGQTINGARKRLKKVHLTEGTQIIVNIGSVDILCGRPLVQIEH
Dele	KCRFRRHERILQSGLCVSGQTINGARRRLKKVQLPAGTOIIVNIGSVDILCGRPLVQIEH
Dfic	NCRFKRYERVLQSGLCVSGQTINGARNRLKKANPEGTQIIVNIGSVDIMRGKPLVQIEH
Dkik	KCRFRRYERVLQSGLCVSGQTINAARNLKRVNLPEGTOIIVNIGSVDILRGKPLVQIEH
Dana	NCRFKRYERVLQSGLCVSGQTIIIGARNRFRKVYLPEGTOIIVNIGSVDIMRGKPLVQIEH
Dbip	NCRFKRYERVLQSGLCVSGQTIIIGARNRFRKVYLPENTQIIVNIGSVDIMRGKPLVQIEH

Dper	DFRQLVKDMHNRLVPILTLAPLANYCHDKAMCDKVVKFNFQFIWKEYLKVIDIHSCLVN
Dpse	DFRQLVKDMHNRLVPILTLAPLANYCHDKAMCDKVVKFNFQFIWKEYLKVIDIHSCLVN
Dimm	DFRQLIKEMHNRRFVPVLTTLAPLANYCHDKQTCEKVLRFNKFIRNEHLTVIDIHTCLIN
Dvir	DFRLLVKEMHNRRFVPVLTTLAPLANCRAHDKQTCDKVSRFNKFIRSEHLKVIDIHSCLIN
Deug	DFRLLIKEMHNQRYIPILTNLAPLANYCHDKVLCDKITRFNKFIASEHLKVINIHSCLIN
Dtak	DFRLLIKEMHNQRYVPILTNLAPLANYCHDKDLCDFKVSRFNKFIRAEHLKVVDIHSCLIN
Dbia	DFRLLIKEMHMSQRFVPILTNLAPLANYCHDKELCDKVNRFNKFIRSEHLKVVDIHSCLIN
Dmel	DFRLLIKEMHNMRLVPILTNLAPLGNYCHDKVLCDKIYRFNKFIRSEHLKVIDIHSCLIN
Dsec	DFRLLIKEMHNMRLVPILTNLAPLGNYCHDKVLCDKIYRFNKFIRSEHLKVIDIHSCLVN
Dsim	DFRLLIKEMHNMRLVPILTNLAPLGNYCHDKVLCDKIYRFNKFIRSEHLKVIDIHSCLVN
Dyak	DFRLLIKEMHNMRLVPILTNLAPLANYCHDKVLCDKIHRFNKFIRSEHLKVIDIHSCLIN
Dere	DFRLLIKEMHMSMRYVPILTNLAPLANYCHDKVLCDKVNRFNKFIRSEHLKVIDIHSCLIN
Drho	DFRLLIKEMHMSQRLVPILTNLAPLANYCHDKILCDKVNRFNKFIRSEHLKYIDIHSCLIN
Dele	DFRLLIKEMHMSQRFVPILTNLAPLANYCHDKTMCDFKVNRFNKFIRSEHLKYIDIHSCLIN
Dfic	DFRLLIKEMHNRRFVPILTNLAPLANYCHDTVMCDKMSRFNKFIRSEHLKIIDIHSCLIN
Dkik	DFRLLIKEMHNRRFVPILTNLAPLANYCHDKGLCDKVLRFNNFIRSEHLKVIDIHSCLIN
Dana	DFRLLIKEIHSRCIPILTNLAPLANYCHDKVLCDKITKFNRFVRNEHLKFIDINSCLIN
Dbip	DFRLLIKEIHSRCIPILTNLAPLANYCHDKVLCDKISKFNKFVRNEHLKFIDINSCLIN
Dper	ENGVVRFDCFQSRRNVTGSKESYVFWNKIGRQRLQMIEASLEY
Dpse	ENGVVRFDCFQSRRNVTGSKESYVFWNKIGRQRLQMIEASLEY
dim	ENGVVRFDCFQGPRSVTGSVEPYVFWNKIGRQRLQMIEQNLEY
Dvir	ENGIVRFDCFQGPRSVTGSSEPYVFWNKIGRQRLHMIEENLEY
Deug	ERGIVRFDCFO-----R-----NLIY
Dtak	ESGIVRFDCFO-----REPYLFWNKIGRQRLRIIESLEY
Dbia	ERGVVRFDCFO-----QEPLYFWNKIGRQRLQIIIESLEY
Dmel	ERGVVRFDCFQSPRQVTGSKEPYLFWNKIGRQRLQVIETSLEY
Dsec	ERGVVRFDCFQSPRQVTGSKEPYLFWNKIGRQRLQIIETSLEY
Dsim	ERGVVRFDCFQSPRQVTGSKEPYLFWNKIGRQRLQIIETSLEY
Dyak	ERGVVRFDCFQSPRQVSGSKEPYLFWNKIGRQRLQVIETSLEY
Dere	ERGVVRFDCFQSPRQVSGSKEPYLFWNKIGRQRLQVIETSLEY
Drho	ERGIVRFDCFQSPRQVTGSKEPYLFWNKIGRQRLQVIETSLLEY
Dele	ERGIVRFDCFQSPRQVTGSNQPYLFWNKIGRQRLQTIIESLEY
Dfic	ERGIVRFDCFQSPRQVTGSKEPYLFWNKIGRQRLQTIETNLEY
Dkik	ERGVVRFDCFQAPRSVSGSKEPYLFWNKIGRQRLQMIESSLEY
Dana	EKKNVLFDCFQAPRTVTGSKEPYLFWNKIGRQRLQTIESCLEY
Dbip	EKKNVLFDCFQAPRTVTGSKEPYLFWNKIGRQRLQTIESCLEY

Table S7. Log likelihood and parameter estimates from branch site models for Oskar predicted structural and interaction domains from 18 Drosophilids using PRANK MSA using trimmed alignments.

		ω values under Model: MA				ω values under Model: MA _{fix}			
	Site Class	Proportion	Background	<i>D. virilis</i>	<i>l</i>	Proportion	Background	<i>D. virilis</i>	<i>l</i>
Long Osk	0	0.89	0.19	0.19	-2208.57	0.90	0.19	0.19	-2210.95
	1	0.06	1	1		0.07	1	1	
	2a	0.03	0.19	166.12		0.01	0.19	1	
	2b	0.002	1	166.12		0.001	1	1	
Lasp-binding	0	0.56	0.14	0.14	-1438.50	0.37	0.14	0.14	-1440.61
	1	0.27	1	1		0.17	1	1	
	2a	0.11	0.14	404.24		0.30	0.14	1	
	2b	0.05	1	404.24		0.14	1	1	
Vasa-interacting	0	0.78	0.07	0.079	-6820.33	0.72	0.07	0.07	-6820.79
	1	0.14	1	1		0.13	1	1	
	2a	0.06	0.07	3.14		0.12	0.07	1	
	2b	0.01	1	3.14		0.02	1	1	

Online Resource 5

This file contains amino acid alignments of Oskar from the 5 *melanogaster* subgroup Drosophila species using the MUSCLE MSA, and the results of multiple PAML analyses of this alignment. The following elements comprise this Additional File:

Figure S4: Amino acid alignment.

Table S8: Log likelihood values and parameter estimates under partition models of Oskar.

Table S9. Log likelihood values and parameter estimates under site models of Oskar.

Table S10. Likelihood ratio test statistics for predicted structural and functional Oskar domains of *melanogaster* subgroup *Drosophila* species.

Figure S4: Amino acid alignment of Oskar from *melanogaster* subgroup species generated by MUSCLE MSA used for PAML analyses.

Dmel	MAAVTSEFPSKPISYTSTNTSAKTYYLKSVKKRVTTCFQQLRDKLQSSGSFRK-SSSSCL
Dsec	MATVTSEFPSKPISYTSTNTSAKTYYLKSVKKRLLTCFQQLREKLQPFGSFRK-SSSNRL
Dsim	MATVTSEFPSKPISYTSTNTSAKTYYLKSVKKRLLTCFQQLREKLQPFGSFRKSSSSRL
Dere	MAAVTSEFPSKPISCTSTNTSAKSNYLKSVKKRVTTCFQQLREKLQASGSFRK-SSPSRL
Dyak	MAAVTSEFPSKPISYTSTNTSAKTYYLKSVKKRVTTCFQQLREKLQASGSFRK-SSTSRL
Dmel	NQIFVRSDFSACGERFRKIFKSARKTEPELWKVPLVAHELT-SRQSSQQLQVVARLFSS
Dsec	YKIFVKSDFSACGERFQKIFKSARKTESPELWKVPLVAHELT-SRQSSQQLQIVARLFSS
Dsim	YKIFVKSDFSACGERFQKIFKSARKTESPELWKVPLVAHELT-SRQSSQQLHIVARLFSS
Dere	YKIFVKADFAACGERFQRIFKSARKTOSPELWKVPLAAHDLTNGRQTSQQLRIVARLFSS
Dyak	YKIFVKADFAACGERFQRIFKSARKTESPELWKVPLAHDLTTSRQSSQQLRIVARLFSS
Dmel	TQISTKEITYNSNSN-----TSENNMTIIESNYI SVREEYPDIDSEVRAILL
Dsec	TQISTKEITYSSNNN-----TSETNMTIIESNYI SVREDYPDIDAEVRAILL
Dsim	TQISTKEITYSSNSN-----TSESNMTIIESNYI SVREDYPDIDAEVRAILL
Dere	TQISTKEITYSTSNNNTSSYSFRSRKSECSRNSNMTIIESNYI AVREDYPDIDVEVRAILL
Dyak	TQISTREITYTNNSN--SSYSFRRRKSESHRSNMTIIESNYI GVREDYPDIDTEVRAILL
Dmel	SHAQNGITISSIKSEYRKLTGNPFPPLHDNVTDFLLTIPNVTAECSES GKRIFNLKASLKN
Dsec	SHAQNGITISNIKSEYRKLTGNPFPPLHDNVTDFLLTIPNVTAECSETGKRIFNLKANLKN
Dsim	SHAQNGITISNIKSEYRKLTGNPFPPLHDNVTDFLLTIPNVTAECSETGKRIFNLKANLKN
Dere	SHAQNGITISNIKSEYRKLTGNPFPPLHDNVTDFLLTIPNVTAECSES GKRIFNLKANLKN
Dyak	SHAQNGITISNIKSEYRKLTGNTFPPLHDNVTDFLLTIPNVTAECSES GKRIFNLKANLKN
Dmel	GHLLDMVLNQKERT-----SDYSSGAPSLENIPRAPPRYWKNPFKRRALSQLNTSPRTV
Dsec	GHLLDMVLNQKQRT--TNSTDYSSGAPSLENIPRAPPRYWKNPFKRRALSQLNTSPRTV
Dsim	GHLLDMVLNQKQRT--TNSTDYSSGAPSLENIPRAPPRYWKNAFKRRALSQLNTSPRTV
Dere	HHLLEMVLNQKQRT---GSSDYSSGAPSVESMPRAPPRYWKNPFKRRALSQLN-NLTE
Dyak	RHLLDMVLNQKQRTGTANGTSDYSSGAQSVENMPRAPPRYWKNSFKRRALSQLNDNLNTV
Dmel	PKITDETKDIATR-PVSLHQMANEAAESNWCYQDNWKHLNNFYQQASVNAPKMPVPINI
Dsec	PKITNERTKEIATR-PASLHQMANEAAESNWCYQDNWKHLNNFYQQASVNAPQMPVPINI
Dsim	PKITDEQTKEIPTR-PASLHQMANEAAESNWCYQDNWKHLNNFYQQASVNAPQMPVPINI
Dere	SKITDEQTQETSTR-PASLHQMANEAAESNWCYQDNWKHLNNFYQQASVNAPKMPVPINI
Dyak	PKLTDEQTKESATRQPASLHQMANEAAESNWCYQDNWKHLNNFYQQSSVNAPKMPVPINI

Dmel	YSPDAPEEPINLAPPGHQPS-CRTQSQKTE-PTENRHLGIFVHPFNGMNIMKRRHEM---
Dsec	YS-DAPEEPTHLAAPGQQPSCCRTQSKKTE-PTENRHLGIFVHPFNDMVVLKRRHEM---
Dsim	YS-DAPEEPTHLAAPGQQPSCCRTQSKKPE-PTENRHLGIFVHPFNDMVVLKRRHEM---
Dere	YS-DAPEERIQLDAPGHQPS-CRTQNQKTQHPTENRHLGIFVHPFSDMVVMKRRHEMTPT
Dyak	YS-DAPEERIQLDAPGHQPS-CRTQNQKTQQTENRHLGIFVHPFNDMVVMKRRHEM---
 Dmel	-TPTPTILTSGYNDSLTI NSDYDAYLLDFPLMGDDFMLYLARMELKCRFRRHERVLQS
Dsec	-TPTPTILTSGYNDSLTI NSDYDAYLLDFPLMGDDFMLYLARMELKCRFRRHERVLQS
Dsim	-TPTPTILTSGYNDSLTI NSDYDAYLLDFPLMGDDFMLYLARMELKCRFRRHERVLQS
Dere	PTPTPTTLSSGTYADSLLTINS DYDAYLLDFPLMGDDFMLYLARMELKCRFRRHERVLQS
Dyak	-TPTPTILTSGYSDSLTI NSDYDAYLLDFPLMGDDFMLYLARMELKCRFRRHERVLQS
 Dmel	GLCVSGLTINGARNRLKRVQLPEGTQIIVNIGSVDIMRGKPLVQIEHDFRLLIKEMHNMR
Dsec	GLCVSGLTINGARNRLKRVQLPEGTQIIVNIGSVDIMRGGRPLVQIEHDFRLLIKEMHNMR
Dsim	GLCVSGLTINGARNRLKRVQLPEGTQIIVNIGSVDIMRGGRPLVQIEHDFRLLIKEMHNMR
Dere	GLCASGLTINGARNRLRRVQLPEGTQIIVNIGSVDIMRGKPLVQIEHDFRLLIKEMHSMR
Dyak	GLCVSGLTINGARNRLKRVQLPEGTQIIVNIGSVDIMRGKPLVQIEHDFRLLIKEMHNMR
 Dmel	LVPILTNLAPLGNYCHDKVLCDKIYRFNKFIRSECCHLKVIDIHSSLINERGVVRFD C Q
Dsec	LVPILTNLAPLGNYCHDKVLCDKIYRFNKFIRSECCHLKVIDIHSSLVNERGVVRFD C Q
Dsim	LVPILTNLAPLGNYCHDKVLCDKIYRFNKFIRSECCHLKVIDIHSSLVNERGVVRFD C Q
Dere	YVPILTNLAPLANYCHDKVLCDKVHRFNKFIRSECCHLKVIDIHSSLINERGVVRFD C Q
Dyak	LVPILTNLAPLANYCHDKVLCDKIHRFNKFIRSECCHLKVIDIHSSLINERGVVRFD C Q
 Dmel	ASPRQVTGSKEPYLFWNKIGRQRVLQVIETSLEYX
Dsec	TSPRQVTGSKEPYLFWNKIGRQRVLQIITSLEYX
Dsim	TSPRQVTGSKEPYLFWNKIGRQRVLQIITSLEYX
Dere	SSPRQVSGSKEPYLFWNKIGRQRVLQVIETSLEYX
Dyak	SSPRQVSGSKEPYLFWNKIGRQRVLQVIETSLEYX

Table S8. Log likelihood values and parameter estimates under partition models for Oskar structural and interaction domains using MUSCLE MSA for 5 species analysis.

	Model	np ¹	<i>l</i> ²	<i>rs</i> ₂	<i>rs</i> ₃	<i>rs</i> ₄	<i>rs</i> ₅	<i>rs</i> ₆	K	ω
Mgene=4	E (different <i>rs</i> , κ, ω and π)	28	-6075.47	0.54	0.77	0.83	1.13	0.80	K ₁ = 2.37	ω ₁ = 0.57
									K ₂ = 2.24	ω ₂ = 0.20
									K ₃ = 0.97	ω ₃ = 0.41
									K ₄ = 2.11	ω ₄ =0.07
									K ₅ = 2.71	ω ₅ = 0.16
									K ₆ = 1.77	ω ₆ = 0.13

¹ np: number parameters

¹ *l*: log likelihood value

¹ *rs*: substitution rate. *rs*₁=1 for all models (not shown). Subscript corresponds to six partitions: (1) Long Osk domain; (2) LOTUS domain; (3) Lasp-binding domain; (4) SGNH Hydrolase domain (5) Valois-interacting domain and (6); Vasa-interacting domain

¹ π: equilibrium codon frequencies

¹ K: transition/transversion ratio

¹ ω: dN/dS ratio

Table S9. Log likelihood and parameter estimates from site models for Oskar structural and interaction domains using MUSCLE MSA for 5 *Drosophila* species

	I	K	Parameter estimates
Long Osk			
M1	-1013.77	2.40	$p_0 = 0.43, p_1 = 0.56, \omega_0 = 0, \omega_1 = 1$
M2	-1013.66	2.46	$p_0 = 0.43, p_1 = 0.51, p_2 = 0.04; \omega_0 = 0, \omega_1 = 1, \omega_2 = 2.5$
M7	-1013.83	2.42	$p = 0.007, q = 0.005$
M8	-1013.66	2.46	$p_0 = 0.91, p_1 = 0.08, p = 0.04, q = 0.04, \omega = 2.01$
LOTUS			
M1	-547.81	1.65	$p_0 = 0.83, p_1 = 0.16, \omega_0 = 0, \omega_1 = 1$
M2	-544.67	2.01	$p_0 = 0.95, p_1 = 0.00, p_2 = 0.04; \omega_0 = 0.07, \omega_1 = 1, \omega_2 = 6.12$
M7	-547.95	2.07	$p = 0.005, q = 0.02$
M8	-544.67	2.01	$p_0 = 0.95, p_1 = 0.04, p = 8.18, q = 99.0, \omega = 6.13$
Lasp-binding			
M1	-522.09	0.98	$p_0 = 0.66, p_1 = 0.33, \omega_0 = 0.19, \omega_1 = 1$
M2	-522.09	0.98	$p_0 = 0.66, p_1 = 0.22, p_2 = 0.10; \omega_0 = 0.19, \omega_1 = 1, \omega_2 = 1$
M7	-522.01	0.97	$p = 0.53, q = 0.66$
M8	-522.01	0.97	$p_0 = 0.99, p_1 = 0.00001, p = 0.53, q = 0.66, \omega = 1.00$
SGNH Hydrolase			
M1	-1272.34	2.12	$p_0 = 0.94, p_1 = 0.05, \omega_0 = 0.03, \omega_1 = 1$
M2	-1272.03	2.16	$p_0 = 0.97, p_1 = 0.00, p_2 = 0.02; \omega_0 = 0.04, \omega_1 = 1, \omega_2 = 1.87$
M7	-1272.78	2.14	$p = 0.04, q = 0.48$
M8	-1272.04	2.16	$p_0 = 0.97, p_1 = 0.02; p = 4.42, q = 99.00; \omega = 1.88$
Vasa-interacting			
M1	-8190.91	1.93	$p_0 = 0.92, p_1 = 0.07, \omega_0 = 0.07, \omega_1 = 1$
M2	-8190.91	1.93	$p_0 = 0.92, p_1 = 0.03, p_2 = 0.03; \omega_0 = 0.07, \omega_1 = 1, \omega_2 = 1.01$
M7	-2023.08	1.78	$p = 0.21, q = 1.24$
M8	-2022.84	1.78	$p_0 = 0.93, p_1 = 0.06; p = 8.58, q = 99.0; \omega = 1.02$
Valois-interacting			
M1	-654.94	2.83	$p_0 = 0.86, p_1 = 0.13, \omega_0 = 0.06, \omega_1 = 1$
M2	-654.91	2.87	$p_0 = 0.88, p_1 = 0, p_2 = 0.11; \omega_0 = 0.07, \omega_1 = 1, \omega_2 = 1.21$
M7	-655.11	2.82	$p = 0.13, q = 0.59$
M8	-654.91	2.88	$p_0 = 0.89, p_1 = 0; p = 7.88, q = 99.0; \omega = 1.22$

Table S10. Likelihood Ratio Test statistics for Oskar structural and interaction domains: *melanogaster* subgroup species partitioned analyses, analysis of alignment generated with MUSCLE MSA

	M1 vs M2¹	M7 vs M8
Long Osk	0.23 (0.89)	0.32 (0.85)
LOTUS	6.29*	6.56*
Lasp-binding	0 (1)	0 (1)
SGNH Hydrolase	0.61 (0.74)	1.48 (0.47)
Vasa-interacting	0 (1)	0.48 (0.78)
Valois-interacting	0.05 (.97)	0.40 (0.82)

Values in each column correspond to twice the difference of log likelihood of the two nested models, with P values indicated in parentheses. * significant at $P < 0.05$.

Online Resource 6

This file contains the Oskar amino acid alignment generated with sequences from seven Drosophilids (melanogaster subgroup members plus *D. virilis* and *D. immigrans*) using the PRANK MSA, and the results of multiple PAML analyses of these alignments. The following elements comprise this Additional File:

Figure S5: Amino acid alignments.

Table S11: Log likelihood values and parameter estimates under partition models of Oskar domains.

Table S12. Log likelihood values and parameter estimates under branch site models of Oskar domains.

Figure S5. Amino acid alignment of Oskar from seven Drosophilids (melanogaster subgroup members plus *D. virilis* and *D. immigrans*) aligned using the PRANK MSA.

Dimm/1-553	MATFRSEFNSVPNTHPDR-----LTAFRKKLTTCFHQWRQKL-----
Dvir/1-596	MATFRSEFNSVPNTYNQR-----LTALKKKLTTCFQOWRQQLKLHQQQQQPKQOOO
Dmel/1-607	MAAVTSEFPSKPISYTSTNTSAKTYYLKSVKKRVTTCFQQLRDKL-----
Dsec/1-611	MATVTSEFPSKPISYTSTNTSAKTYYLKSVKKRVTTCFQQLREKL-----
Dsim/1-612	MATVTSEFPSKPISYTSTNTSAKTYYLKSVKKRVTTCFQQLREKL-----
Dyak/1-626	MAAVTSEFPSKPISYTSTNTSAKTYYLKSVKKRVTTCFQQLREKL-----
Dere/1-626	MAAVTSEFPSKPISCTSTNTSAKSNYLKSVKKRVTTCFQQLREKL-----
Dimm/1-553	-----Q-----HF-----KHDYSTLRAKSICWQKFK
Dvir/1-596	QQQHQSQHQHQQQQQQKQKQORGSE-----HFYNIFVRADCSILSEKVTLRKFK
Dmel/1-607	-----QSSGSFRKSSSS-CLNQIFVRSDFS-----ACGERFR
Dsec/1-611	-----QPFGSFRKSSSN-RLYKIFVKSDFS-----ACGERFQ
Dsim/1-612	-----QPFGSFRKSSSSRLYKIFVKSDFS-----ACGERFQ
Dyak/1-626	-----QASGSFRKSSTS-RLYKIFVKADFA-----ACGERFQ
Dere/1-626	-----QASGSFRKSSPS-RLYKIFVKADFA-----ACGERFQ
Dimm/1-553	KIY-TAT-----KLIS-----QEEVEEQOLO
Dvir/1-596	RIF-TARQNLASARTAATAASAKOPVLGRCIG-----Q---EEQOLO
Dmel/1-607	KIFKSAR-----KTEPELWKVPLVAHELT--SRQ---SSQOLO
Dsec/1-611	KIFKSAR-----KTESPELWKVPLVAHELT--SRQ---SSQOLO
Dsim/1-612	KIFKSAR-----KTESPELWKVPLVAHELT--SRQ---SSQQLH
Dyak/1-626	RIFKSAR-----KTESPELWKVPLA-HDLTTTSRQ---SSQQLR
Dere/1-626	RIFKSAR-----KTQSPELWKVPLAAHDLT-NGRQ---TSQQLR
Dimm/1-553	IVATLFSSTLIS----KY-TSSKNSSS-----CNMTIMDDSYIGVREYPDID
Dvir/1-596	IVASLFSSSTLIS---KH-SSSASITS-----SSMTIMDNTYIGVREYPDID
Dmel/1-607	VVARLFSSTQISTKEITY-NSNSNTSE-----NNMTIIIESNYISVREEYPDID
Dsec/1-611	IVARLFSSTQISTKEITY-SSNNNTSE-----TNMTIIIESNYISVREDYPDID
Dsim/1-612	IVARLFSSTQISTKEITY-SSNSNTSE-----SNMTIIIESNYISVREDYPDID
Dyak/1-626	IVARLFSSTQISTREITYTNSSN--SSYSFRRRKSESHRSNMTIIIESNYIGVREDYPDID
Dere/1-626	IVARLFSSTQISTKEITYSTSNNNTSSYSFRSRKSECRSNMTIIIESNYIAVREDYPDID
Dimm/1-553	AEIRAILLANAQKGITISSIKSEYRNMTGNAPLRENITDFLLTIPVTAECCMSGKRIF
Dvir/1-596	TEIRSILLANAQNIGHTISSIKKEYRQLTGTAFPLHDNITDFLLTIPHVTAECCESGKRIF
Dmel/1-607	SEVRAILLSHAQNGITISSIKSEYRKLGNPFPLHDNVTDFLLTIPNVTAECSES GKRIF
Dsec/1-611	AEVRAILLSHAQNGITISNIKSEYRKLGNPFPLHDNVTDFLLTIPNVTAECSETGKRIF
Dsim/1-612	AEVRAILLSHAQNGITISNIKSEYRKLGNPFPLHDNVTDFLLTIPNVTAECSETGKRIF
Dyak/1-626	TEVRAILLSHAQNGITISNIKSEYRKLGNTFPLHDNVTDFLLTIPNVTAECSES GKRIF
Dere/1-626	VEVRAILLSHAQNGITISNIKSEYRKLGNPFPLHDNVTDFLLTIPNVTAECSES GKRIF
Dimm/1-553	NIKPTEDTRHLHDMVLNORORENGNNHNSNNHHVLDQE-----PPRLWRSQYKRRCLON
Dvir/1-596	NIKPTEHTRHLHEMILQQRQRDSVSNP-----IQAQE-----PPRLWRAQYKRRIPQ-
Dmel/1-607	NLKASLKGNGHLLDMVLNQKERT-----SDYSSGAPSLENIPRAPPRYWKNPFKRRALS-
Dsec/1-611	NLKANLKGNGHLLDMVLNQKORT--TNSTDYSSGAPSLENIPRAPPRYWKNPFKRRALS-
Dsim/1-612	NLKANLKNRHLLDMVLNQKORT--TNSTDYSSGAPSLENIPRAPPRYWKNAFKRRALS-
Dyak/1-626	NLKANLKNHLLLEMVLNQKORTG---SSDYSSGAPSVESMPRAPPRYWKNSFKRRALS-
Dere/1-626	NLKANLKNHLLLEMVLNQKORTG---SSDYSSGAPSVESMPRAPPRYWKNPFKRRALS-
Dimm/1-553	ENYNFNLNIC-----EKPPAAK-ITPVQHLATAAMAPFGVYQDNWKHLNNQYQLSQLA
Dvir/1-596	-HFNFnLNTS-----EKPPAVK-ISKLQPLATAAAMSNDVYQDNWKHLNNQYQLPQLN
Dmel/1-607	-QLNTSPRTVPKITDEKTDIATR-PVSLHQMANEAAESNWCYQDNWKHLNNFYQQASVN
Dsec/1-611	-QLNTSPRTVPKITNERTKEIATR-PASLHQMANEAAESNWCYQDNWKHLNNFYQQASVN

Dsim/1-612	-QLNTSPRTVPKITDEQTKEIPTR-PASLHQMANEAAESNWCYQDNWKHLNNFYQQASVN
Dyak/1-626	-QLNDNLNTVPKLTDEQTKEIPTR-PASLHQMANEAAESNWCYQDNWKHLNNFYQQASVN
Dere/1-626	-QLN-NLNTESKITDEQTQETSTR-PASLHQMANEAAESNWCYQDNWKHLNNFYQQASVN
 Dimm/1-553	 TQN----SNOFS-CIRSDPVQLOQIAP-HPPEQ-QHQQQRQQ-Q--QQQIGEYSH-----
Dvir/1-596	AAK----NNIHS--IASNPAQLO-AA-LPAEH-H-----P--SKHVEEYAH-----
Dmel/1-607	APKMPVPINIYSPDAPEEPINLA-PPGHQPSC-RTQSQKTE-PTENRHLGIFVHPFNGMN
Dsec/1-611	APQMPVPINIYS-DAPEEPTHLA-APGQQPSCCRTQSKKTE-PTENRHLGIFVHPFNGMN
Dsim/1-612	APQMPVPINIYS-DAPEEPTHLA-APGQQPSCCRTQSKKPE-PTENRHLGIFVHPFNGMN
Dyak/1-626	APKMPVPVNIYS-DAPEERIHLD-APGHQPSC-RTQNQKTQQTTERHLGIFVHPFNGMN
Dere/1-626	APKMPVPINIYS-DAPEERIQVD-APGHQPSC-RTQNQKTQHPTENRHLGIFVHPFNGMN
 Dimm/1-553	 --KRRHEFTPTPT---TLSCPS-QHDSMFTINSDYDAYLLDFPLLGDDFFLYLARMELK
Dvir/1-596	--KRRHEYTPPTPT---TLSCPSTQHDSMFTINSDYDAYLLDFPLLGDDFFLYLARMELK
Dmel/1-607	IMKRRHEMTPTPT---ILTSGT-YNDSSLTINSDYDAYLLDFPLMGDDFMLYLARMELK
Dsec/1-611	VLKRRHEMTPTPT---ILTSGT-YNDSSLTINSDYDAYLLDFPLMGDDFMLYLARMELK
Dsim/1-612	VLKRRHEMTPTPT---ILTSGT-YNDSSLTINSDYDAYLLDFPLMGDDFMLYLARMELK
Dyak/1-626	VMKRRHEMTPTPT---ILTSGT-YSDSSLTINSDYDAYLLDFPLMGDDFMLYLARMELK
Dere/1-626	VMKRRHEMTPTPTPTLSSGT-YADSSLTINSDYDAYLLDFPLMGDDFMLYLARMELK
 Dimm/1-553	 CRFKKDEKVLSQGLCISGQTINAARQRVQHVELQEMTQIIVNIGSVDIMRGKPLVQIEHD
Dvir/1-596	CRFKKFEKVLSQGLCISGQTINAARQRLRVELPEMTQIIVNIGSEDIMRGRSLVQIEHD
Dmel/1-607	CRFRRHERVLQSGLCVSGLTINGARNRLKRVQLPEGTOIIVNIGSVDIMRGKPLVQIEHD
Dsec/1-611	CRFRRHERVLQSGLCVSGLTINGARNRLKRVQLPEGTOIIVNIGSVDIMGRGRPLVQIEHD
Dsim/1-612	CRFRRHERVLQSGLCVSGLTINGARNRLKRVQLPEGTOIIVNIGSVDIMRGKPLVQIEHD
Dyak/1-626	CRFRRHERVLQSGLCASGLTINGARNRLRRVQLPEGTOIIVNIGSVDIMRGKPLVQIEHD
Dere/1-626	
 Dimm/1-553	 FRQLIKEMHNRRFVPLVTTLAPLANCYCHDKQTCEKVLRFNKFIRNEGRHLTVIDIHTCLI
Dvir/1-596	FRLLVKEMHNRRFVPLVTTLAPLANCRHDQTKCDKVSRFNKFIRSEGRHLKVIDIHSCLI
Dmel/1-607	FRLLIKEHMHNMRVLPILTNLAPLGNYCHDKVLCDFKQYRFNKFIRSECCHLKVIDIHSCLI
Dsec/1-611	FRLLIKEHMHNMRVLPILTNLAPLGNYCHDKVLCDFKQYRFNKFIRSECCHLKVIDIHSCLV
Dsim/1-612	FRLLIKEHMHNMRVLPILTNLAPLGNYCHDKVLCDFKQYRFNKFIRSECCHLKVIDIHSCLV
Dyak/1-626	FRLLIKEHMHNMRVLPILTNLAPLANYCHDKVLCDFKQYRFNKFIRSECCHLKVIDIHSCLI
Dere/1-626	FRLLIKEHMHSRYYVPLTNLAPLANYCHDKVLCDFKQYRFNKFIRSECCHLKVIDIHSCLI
 Dimm/1-553	 NENGVVRFDCFQKGPRSVTGSVEPYVFWNKIGRQRVLQMIEQNLEYHX
Dvir/1-596	NENGIVRFDCFQNGPRSVTGSSEPYVFWNKIGRQRVLHMIEENLEYYX
Dmel/1-607	NERGVVRDFCFQASPRQVTGSKEPYLFWNKIGRQRVLQVIETSLEYX-
Dsec/1-611	NERGVVRDFCFQTSRQVTGSKEPYLFWNKIGRQRVLQVIETSLEYX-
Dsim/1-612	NERGVVRDFDCFQTSRQVTGSKEPYLFWNKIGRQRVLQVIETSLEYX-
Dyak/1-626	NERGVVRDFDCFQSSPRQVSGSKEPYLFWNKIGRQRVLQVIETSLEYX-
Dere/1-626	NERGVVRDFDCFQSSPRQVSGSKEPYLFWNKIGRQRVLQVIETSLEYX-

Table S11. Log likelihood values and parameter estimates under partition models for Oskar predicted structural and interaction domains from seven Drosophilids (melanogaster subgroup members plus *D. virilis* and *D. immigrans*) aligned using PRANK MSA.

	Model	np ¹	<i>l</i> ²	<i>rs</i> ₂	<i>rs</i> ₃	<i>rs</i> ₄	<i>rs</i> ₅	<i>rs</i> ₆	K	ω
Mgene=4	E (different <i>rs</i> , K, ω and π)	28	-10182.10	0.54	1.11	1.06	0.89	0.85	K ₁ = 2.00	ω ₁ = 0.31
									K ₂ = 1.45	ω ₂ = 0.15
									K ₃ = 1.68	ω ₃ = 0.54
									K ₄ = 2.12	ω ₄ =0.09
									K ₅ = 1.6	ω ₅ = 0.13
									K ₆ = 1.68	ω ₆ = 0.07

¹ np: number parameters

¹ *l*: log likelihood value

¹ *rs*: substitution rate. *rs*₁=1 for all models (not shown). Subscript corresponds to six partitions: (1) Long Osk domain; (2) LOTUS domain; (3) Lasp-binding domain; (4) Valois-interacting domain (5) Vasa-interacting domain; and (6) SGNH Hydrolase domain;

¹ π: equilibrium codon frequencies

¹ K: transition/transversion ratio

¹ ω: dN/dS ratio

Table S12. Log likelihood and parameter estimates from branch site models for Oskar predicted structural and interaction domains from seven Drosophilids (melanogaster subgroup members plus *D. virilis* and *D. immigrans*) using PRANK MSA.

		ω values under Model: MA					ω values under Model: MA _{fix}			
	Site Class	Proportion	Background	<i>D. virilis</i>	<i>I</i>	Proportion	Background	<i>D. virilis</i>	<i>I</i>	
Long Osk	0	0.43	0.097	0.19	-1774.30	0.43	0.09	0.19	-1774.30	
	1	0.31	1	1		0.31	1	1		
	2a	0.14	0.097	1		0.14	0.09	1		
	2b	0.10	1	1		0.10	1	1		
LOTUS	0	0.83	0.10	0.10	-918.47	0.80	0.10	0.10	-919.30	
	1	0.10	1.00	1.00		0.09	1.00	1.00		
	2a	0.05	0.10	3.95		0.08	0.10	1.00		
	2b	0.01	1.00	3.95		0.00	1.00	1.00		
Lasp-binding	0	0.18	0	0	-948.59	0.18	0.00	0.00	-948.59	
	1	0.81	1	1		0.81	1	1		
	2a	0	0	1		0.00	0.00	1		
	2b	0	1	404.24		0.00	1	1		
SGNH Hydrolase	0	0.81	0.02	0.02	-2018.07	0.64	0.02	0.02	-2018.76	
	1	0.081	1.00	1.00		0.06	1.00	1.00		
	2a	0.08	0.02	999.0		0.26	0.02	1.00		
	2b	0.008	1.00	999.0		0.02	1.00	1.00		
Vasa-interacting	0	0.65	0.03	0.038	-3354.12	0.49	0.03	0.03	-3354.34	
	1	0.23	1	1		0.18	1	1		
	2a	0.08	0.03	999.0		0.23	0.03	1		
	2b	0.02	1	999.0		0.08	1	1		
Valois-interacting	0	0.67	0.03	0.03	-1022.09	0.54	0.03	0.03	1022.14	
	1	0.16	1.00	1.00		0.12	1.00	1.00		
	2a	0.12	0.03	999.0		0.26	0.03	1.00		

Online Resource 7

Figure S6: Alignment of the amino acid translation of *D. virilis* alleles used in this study as follows:

1. Dvir_GENOMIC: annotated from genomic sequence; NCBI accession XM_002053233.1
2. Dvir_webster: cDNA sequence reported by Webster et al. (1994) ("Macdonald allele"); NCBI accession L22556.1
3. Dvir_sequenced: cDNA sequence obtained from Macdonald lab and verified by Sanger sequencing

Residues that are different between the different alleles and/or under positive selection are indicated as follows:

Outlined: The five sites identified as different between the different alleles:

Blue Outline Genomic allele differs from Macdonald & sequenced alleles; not under positive selection

Red Outline Macdonald allele differs from genomic & sequenced alleles; not under positive selection

Black Outline Genomic allele differs from Macdonald & sequenced alleles; under positive selection based on PRANK analysis of full-length genomic allele (see Online Resource 3)

Blue Shading: The four sites indicated as being under positive selection based on analyses of PRANK alignments of both genomic and Macdonald alleles:

Dark Blue: supported by analyses of

- full length genomic allele (see Online Resource 3)
- conserved sequence block genomic allele (see Online Resource 4)
- full length Macdonald allele (see Online Resource 9)

Medium Blue: supported by analyses of

- conserved sequence block genomic allele (see Online Resource 4)
- full length Macdonald allele (see Online Resource 9)

Light Blue: supported by analyses of

- full length genomic allele (see Online Resource 3)
- full length Macdonald allele (see Online Resource 9)

Pink Shading: The site indicated as being under positive selection based on analyses of MUSCLE alignments:

Pink: supported by analyses of

- full length genomic allele (see Online Resource 2)
- full length Macdonald allele (see Online Resource 8)

Dvir_webster	MATFRSEFNSVPNTYNQRLTALKKLTTCFQQWRQQLLKLHQQQQQPKQQQQQQHQSQH	60
Dvir_sequenced	MATFRSEFNSVPNTYNQRLTALKKLTTCFQQWRQQLLKLHQQQQQPKQQQQQQHQSQH	60
Dvir_GENOMIC	MATFRSEFNSVPNTYNQRLTALKKLTTCFQQWRQQLLKLHQQQQQPKQQQQQQHQSQH	60

Dvir_webster	QHQQQ --K QKORGSEHFYNIFV R ADCSILSEKVTYLRKFKRIFTARQNLSARTAATAAS	118
Dvir_sequenced	QHQQQ --K QKORGSEHFYNIFV R ADCSILSEKVTYLRKFKRIFTARQNLSARTAATAAS	118
Dvir_GENOMIC	QHQQQ QQK QKORGSEHFYNIFV R ADCSILSEKVTYLRKFKRIFTARQNLSARTAATAAS	120

Dvir_webster	AKQPVLGRCIGQEEQQLQIVASLFSSTLISKHSSSASITSSSMIDNTYIGVRDEYPDI	178
Dvir_sequenced	AKQPVLGRCIGQEEQQLQIVASLFSSTLISKHSSSASITSSSMIDNTYIGVRDEYPDI	178
Dvir_GENOMIC	AKQPVLGRCIGQEEQQLQIVASLFSSTLISKHSSSASITSSSMIDNTYIGVRDEYPDI	180

Dvir_webster	DTEIRSILLANAQNGITISSIKKEYRQLTGTAFPLHDNITDFLLTIPHVTAECCESGKRI	238
Dvir_sequenced	DTEIRSILLANAQNGITISSIKKEYRQLTGTAFPLHDNITDFLLTIPHVTAECCESGKRI	238
Dvir_GENOMIC	DTEIRSILLANAQNGITISSIKKEYRQLTGTAFPLHDNITDFLLTIPHVTAECCESGKRI	240

Dvir_webster	FNIKPTEHTRHLHEMILQQRQRDSVSNP IQAQEPPRLWRAQYKRRIPQHFNFLNTSEKP	298
Dvir_sequenced	FNIKPTEHTRHLHEMILQQRQRDSVSNP IQAQEPPRLWRAQYKRRIPQHFNFLNTSEKP	298
Dvir_GENOMIC	FNIKPTEHTRHLHEMILQQRQRDSVSNP IQAQEPPRLWRAQYKRRIPQHFNFLNTSEKP	300

Dvir_webster	PAVKISKLQPLATAAA MSNDVYQDNWKHLNNQYQLPQLNA P KNNI H H IASNPAQLQAL	358
Dvir_sequenced	PAVKISKLQPLATAAA MSNDVYQDNWKHLNNQYQLPQLNA P KNNI H H IASNPAQLQAL	358
Dvir_GENOMIC	PAVKISKLQPLATAAA MSNDVYQDNWKHLNNQYQLPQLNA A KNNI H S -IASNPAQLQAL	359

Dvir_webster	PAEHHPSKHVEEYAHKRRHEYT	R	TPPTTLSCPSTQHDSMFTINSDYDAYLLDFPLLGDDFF	418
Dvir_sequenced	PAEHHPSKHVEEYAHKRRHEYT	P	TPPTTLSCPSTQHDSMFTINSDYDAYLLDFPLLGDDFF	418
Dvir_GENOMIC	PAEHHPSKHVEEYAHKRRHEYT	P	TPPTTLSCPSTQHDSMFTINSDYDAYLLDFPLLGDDFF	419
	*****	*****	*****	*****
Dvir_webster	LYLARMELKCRFKKFEKVLQSGLCISGQTINAARQRLRLVELPEMTQIIVNIGSEDIMRG	478		
Dvir_sequenced	LYLARMELKCRFKKFEKVLQSGLCISGQTINAARQRLRLVELPEMTQIIVNIGSEDIMRG	478		
Dvir_GENOMIC	LYLARMELKCRFKKFEKVLQSGLCISGQTINAARQRLRLVELPEMTQIIVNIGSEDIMRG	479		
	*****	*****	*****	*****
Dvir_webster	RSLVQIEHDFRLLVKEMHNRRFVPVLTTIAPLANC RHD KQTCDKVSRFNK FIRSEG RHLK	538		
Dvir_sequenced	RSLVQIEHDFRLLVKEMHNRRFVPVLTTIAPLANC RHD KQTCDKVSRFNK FIRSEG RHLK	538		
Dvir_GENOMIC	RSLVQIEHDFRLLVKEMHNRRFVPVLTTIAPLANC RHD KQTCDKVSRFNK FIRSEG RHLK	539		
	*****	*****	*****	*****
Dvir_webster	VIDIH SCLINENGIVRFDCFQNGPRSVTGSS SEP YVFWNKIGRQRVLHMIEENLEYY	594		
Dvir_sequenced	VIDIH SCLINENGIVRFDCFQNGPRSVTGSS SEP YVFWNKIGRQRVLHMIEENLEYY	594		
Dvir_GENOMIC	VIDIH SCLINENGIVRFDCFQNGPRSVTGSS SEP YVFWNKIGRQRVLHMIEENLEYY	595		
	*****	*****	*****	*****

Online Resource 8

This file contains amino acid alignments of the Long Oskar domain from 18 Drosophilids including the *D. virilis* allele from Webster et al. (1994) using the MUSCLE MSA, and the results of multiple PAML analyses of these alignments. The following elements comprise this Additional File:

Figure S7: Amino acid alignments.

Table S13. Log likelihood values and parameter estimates under branch site models of Oskar domains.

Figure S7. Amino acid alignment of Long Osk using the *D. virilis* allele from Webster et al (1994) and the MUSCLE MSA. Positively selected sites with BEB probability = 0.99 is indicated in red.

Long Osk domain

Dpse	MATYRSAH-SEPYTSKSNSNSNSSSSRLTALRKRLTTCLQQWCQHL-----
Dimm	MATFRSEFNSVPNTH-----PDRLTAFRKKLTTCFHQWRQKL-----
Dvir	MATFRSEFNSVPNTY-----NQRLTALKKKLTTCFQOWRQQLKLHQQQQOPKOO
Dana	MAAVTSELPREP-----TTTTNTKSDRLRAVKKRVTTCCLQQWCQKL-----
Dbip	MAAVTSELPREP-----TTTTNTKSDRLRAVKKRVTTCCLQQWCQKL-----
Dkik	MAAVRSEFPSE-----SNSNPSTKADSLRSLKHKVATCWHNLCKKL-----
Deug	MATVTSEFPSKPISYNTNTNTSTKTYCLRSVKRRVTTCCLQQLREKL-----
Dmel	MAAVTSEFPSKPISY--TSTNTSAKTYYLKSVKKRVTTCFQQLRDKL-----
Dsec	MATVTSEFPSKPISY--TSTNTSAKTYYLKSVKKRVTTCFQQLREKL-----
Dsim	MATVTSEFPSKPISY--TSTNTSAKTYYLKSVKKRVTTCFQQLREKL-----
Dere	MAAVTSEFPSKPISC--TSTNTSAKSNYLKSVKKRVTTCFQQLREKL-----
Dyak	MAAVTSEFPSKPISY--TSTNTSAKTYYLKSVKKRVTTCFQQLREKL-----
Dfic	MAAVRSEFPSPKN---SNTYTNISKFDLRLRSVKRVTTCCLKLCEKL-----
Dele	MAAVRSEFPSEPNSN--TNTNTSTKTDRLRSVKRVTTCCLKLCEKL-----
Drho	MAAVRSEFPSEPYSN--TNNNTSTKTDCLRSVKRVTTCCLKLCEKL-----
Dbia	MAAVGSEFPSEPISC--RNTNTPAKPDRLRSVKRVTTCCLQQLREKL-----
Dtak	MAAVRSEFPSEPISY--TNTNTSNKSNCLRSVKRVTTCCLQQLREKL-----
 Dper	-----QKPAS F QKCKSGPNHQFYKIFEEVFDSKV-FGDKTKYIROFKKIFKGNKD
Dpse	-----QKPAS F QKCKSGPNHQFYKIFEEVFDSKV-FGDKTKYIROFKKIFKGNKD
Dimm	-----QHFKH-----DYST-LRAKSICWQFKKKIYTATKL
Dvir	QQQQQHQSQHQHQQQ K QKORG--SEHFYNIFVRA-DCSI-LSEKVTYLRKFKRIFTARQN
Dana	-----QQPAS FR K-SS--SSRFYRLFVKA-DYTA-FGDKFKSVRQFHRIFKAGRK
Dbip	-----QQPAS FR K-SS--SSRFYRIFVKA-DYTA-IGDKFKSVRQFHRIFKAGRK
Dkik	-----QQPAS FR K-SS--HSRFYKIFLKA-DCSASFPEKSKFLRQFHRIFKAGRK
Deug	-----PATGS FR K-SS--TSRFHRFFIFIKS-DFSD-FSEK-----LOKIFKSARK
Dmel	-----QSSGS FR K-SS--SSCLNQIFVRS-DFSA-CGER-----FRKIFKSARK
Dsec	-----QPFGS FR K-SS--SNRLYKIFVKS-DFSA-CGER-----FQKIFKSARK
Dsim	-----QPFGS FR KSSS--SSRLYKIFVKS-DFSA-CGER-----FQKIFKSARK
Dere	-----QASGS FR K-SS--PSRLYKIFVKA-DFAA-CGER-----FQRIFKSARK
Dyak	-----QASGS FR K-SS--TSRLYKIFVKA-DFAA-CGER-----FQRIFKSARK
Dfic	-----QAPGS FR K-SS--SSRFHKIFVRA-DYSD-CSK---FVRRFQKIFKSARK
Dele	-----QATGS FR K-SS--SSRFYKIFVKA-DYSG-LREKSKFLRQLQKIYKSTRK
Drho	-----QAPGS FR K-SS--SSRFYKIFVKA-DYSG-FREKSKFVRQFQKIFKSARK
Dbia	-----PAAGS FR K-SS--TGRFYQVFVRA-DFSA-FGER-----FRKIFKSARK
Dtak	-----PASGS FR K-SS--SSRFYKIFVKA-DFSA-FGER-----FQKIFKSARK
 Dper	LAK--KSEPGRRVVLVLVPEKPVDDT-EESQKKQLQLVARLFSSTLISAKEEISCSYN--
Dpse	LAK--KSEPGRRVVLVLVPEKPVDDT-EESQKKQLQLVARLFSSTLISAKEEISCSYN--
Dimm	ISQ-----EEVEEQQLQIVATLFSSTLIS-----KYTS--
Dvir	LAS----ARTAATAASA K QPVLGRCIGQEEQQLQIVASLFSSTLISKHS-----
Dana	GKE-----PAESVDER-----QQLQFVARLFSSTLISTEE---IYGN--
Dbip	KSK-----EPAESVDER-----QQLQFVARLFSSTLISTEE---IYGN--
Dkik	SEGGGEPEPELWKEVPLAD----EQRSQEQLQIVARLFSSTLISTKEEEIITTSRSS
Deug	SES-----PELRR-VPLSTHDQTSQGSSQQLQIVARLFSSTLISTKE---ITFNNHT
Dmel	TEL-----PELWK-VPLVAHELT-SRQSSQQLQIVARLFSSTQISTKE---ITYNS--
Dsec	TES-----PELWK-VPLVAHELT-SRQSSQQLQIVARLFSSTQISTKE---ITYSS--
Dsim	TES-----PELWK-VPLVAHELT-SRQSSQQLHIVARLFSSTQISTKE---ITYSS--
Dere	TQS-----PELWK-VPLAAHDLTNGRQTSQQLRIVARLFSSTQISTKE---ITYSTSN
Dyak	TES-----PELWK-VPLAHDLLTTSRQSSQQLRIVARLFSSTQISTRE---ITYTNSN
Dfic	TES-----PESWK-VPL-AHDVT-RQSSIHQQLQIVARLFSSTQISTEE---IIYNN--
Dele	TES-----PELWK-VPL-AHDVT--GQSSQQLQIVTRLFSSTLISTEE---ITYNTNN

Drho PES-----PELWK-VPL-AHDVT--GQNSQQLQIVARLFSSTLVSTEE---ITYHT--
Dbia TES-----PELWK-VPL-AHDVT--GPSSQHLQIVARLFSSTLISTKE---ISYSN--
Dtak TES-----PELWK-VPL-AHDVT--GQSGQQLQIVARLFSSTLISTKE---IHCNN--

Dper -----SSS
Dpse -----SSS
Dimm -----SKNSS-----SCN
Dvir -----SSASIT-----SSS
Dana -----SSRKS-QSF-NSNS
Dbip ---GKTSYSSRKS-QSY-NSNS
Dkik N--CNTSYTRRKS-QSYCSTST
Deug K--NSFSYKCRKS-EGY--SSN
Dmel -----NSNTS-----ENN
Dsec -----NNNTS-----ETN
Dsim -----NSNTS-----ESN
Dere NNTSSYSFRSRKS-ESC--RSN
Dyak N--SSYSFRRRKS-ESH--RSN
Dfic -----ISFNSRKS-QSH--SSN
Dele K--NNNSCKSRKSHKSY--KSS
Drho ---NINSYKSRKS-QSY--DCD
Dbia ---HRYSCRSGKS-ESY--SSN
Dtak ---NSYSYKSRKS-ESY--SSN

Table S13. Log likelihood and parameter estimates from branch site models for Oskar using MUSCLE MSA alignments of 18 Drosophilids including the *D. virilis* allele sequenced in Webster et al. (1994).

		ω values under Model: MA					ω values under Model: MA _{fix}			
	Site Class	Proportion	Background	<i>D. virilis</i>	<i>I</i>	Proportion	Background	<i>D. virilis</i>	<i>I</i>	
Long Osk (MUSCLE)	0	0.64	0.23	0.23	-4352.28	0.44	0.22	0.22	-4356.5	
	1	0.19	1	1		0.15	1	1		
	2a	0.12	0.23	547.39		0.30	0.22	1		
	2b	0.03	1	547.39		0.1	1	1		

Online Resource 9

This file contains amino acid alignments of the Long Oskar and Lasp-binding domains from 18 Drosophilids including the *D. virilis* allele from Webster et al. (1994) using the PRANK MSA, and the results of multiple PAML analyses of these alignments. The following elements comprise this Additional File:

Figure S8: Amino acid alignments.

Table S14. Log likelihood values and parameter estimates under branch site models of Oskar domains.

Figure S6. Amino acid alignment of Long Osk and Lasp binding region using the *D. virilis* allele from Webster et al (1994) and the PRANK MSA. Positively selected sites with BEB probability = 0.95 are indicated in **red**, BEB = 0.98 in **blue**.

Long Osk domain

Dper	MATYRSAHSEPYTSKSNS-----N-----SNSSSSSRLTALRKRLTTCLQQWCQHL-----
Dpse	MATYRSAHSEPYTSKSNS-----N-----SNSSSSSRLTALRKRLTTCLQQWCQHL-----
Dimm	MATFRSE----FNSVPNT-----H----PD----RLTAFRKKLTTCFHQWRQKL-----
Dvir	MATFRSE----FNSVPNT-----Y----NQ----RLTALKKLTTCFQQWRQQLKLH
Dmel	MAAVTSE----FPSKPISYTST--NTSAKTY-----YLKSVKKRVTTCFQQLRDKL-----
Dsec	MATVTSE----FPSKPISYTST--NTSAKTY-----YLKSVKKRVTTCFQQLREKL-----
Dsim	MATVTSE----FPSKPISYTST--NTSAKTY-----YLKSVKKRVTTCFQQLREKL-----
Dyak	MAAVTSE----FPSKPISYTST--NTSAKTY-----YLKSVKKRVTTCFQQLREKL-----
Dere	MAAVTSE----FPSKPISCTST--NTSAKSN-----YLKSVKKRVTTCFQQLREKL-----
Dtak	MAAVRSE----FPSEPISYTNT--NTSNKSN-----CLRSVKKRVTTCLQQLREKL-----
Dbia	MAAVGSE----FPSEPISCRNT--NTPAKPD-----RLRSVKKRVTTCLQQLREKL-----
Deug	MATVTSE----FPSKPISYTNTNTNTSTKTY-----CLRSVKKRVTTCLQQLREKL-----
Drho	MAAVRSE----FPSEPYNSNTNN--NTSTKTD-----CLRSVKKRVTTCLKQLCEKL-----
Dele	MAAVRSE----FPSEPNNSNTNT--NTSTKTD-----RLRSVKNRVTTCLKQLCEKL-----
Dfic	MAAVRSE----FPSKPNSNTY--NS--KFD-----DLRSVKKRVTTCLKQLCEKL-----
Dkik	MAAVRSE----FPSESNS-----NPSTKAD-----SLRSLKKHVATCWHNLCKKL-----
Dana	MAAVTSE----LPREPTT-----TTNTKSD-----RLRAVKKRVTTCLOQCQKL-----
Dbip	MAAVTSE----LPREPTT-----TTNTKSD-----RLRAVKKRVTTCLOQCQKL-----
Dper	-----QKPASFKCKSG-PNHQFYKIFE E VFD S K
Dpse	-----QKPASFKCKSG-PNHQFYKIFE E VFD S K
Dimm	-----Q-----HF----- K H-DYS
Dvir	QQQQQPKQQQQQQQHQSQHQHQQQKQKQRGSE-----HFYNIFV R A-DCS
Dmel	-----QSSGSFRKSSSS---CLNQIFV R S-DFS
Dsec	-----QPGSFRKSSSN---RLYKIFV K S-DFS
Dsim	-----QPGSFRKSSSS---RLYKIFV K S-DFS
Dyak	-----QASGSFRKSSTS---RLYKIFV K A-DFA
Dere	-----QASGSFRKSSPS---RLYKIFV K A-DFA
Dtak	-----PASGSFRKSSSS---RFYKIFV K A-DFS
Dbia	-----PAAGSFRKSSTG---RFYQVFV R A-DFS
Deug	-----PATGSFRKSSTS---RFHRRFI K S-DFS
Drho	-----QAPGSFRKSSSS---RFYKIFV K A-DYS
Dele	-----QATGSFRKSSSS---RFYKIFV K A-DYS
Dfic	-----QAPGSFRKSSSS---RFHKIFV R A-DYS
Dkik	-----QOPASFRKSSHS---RFYKIFL K A-DCS
Dana	-----QOPASFRKSSSS---RFYRLFV K A-DYT
Dbip	-----QOPASFRKSSSS---RFYRIFV K A-DYT

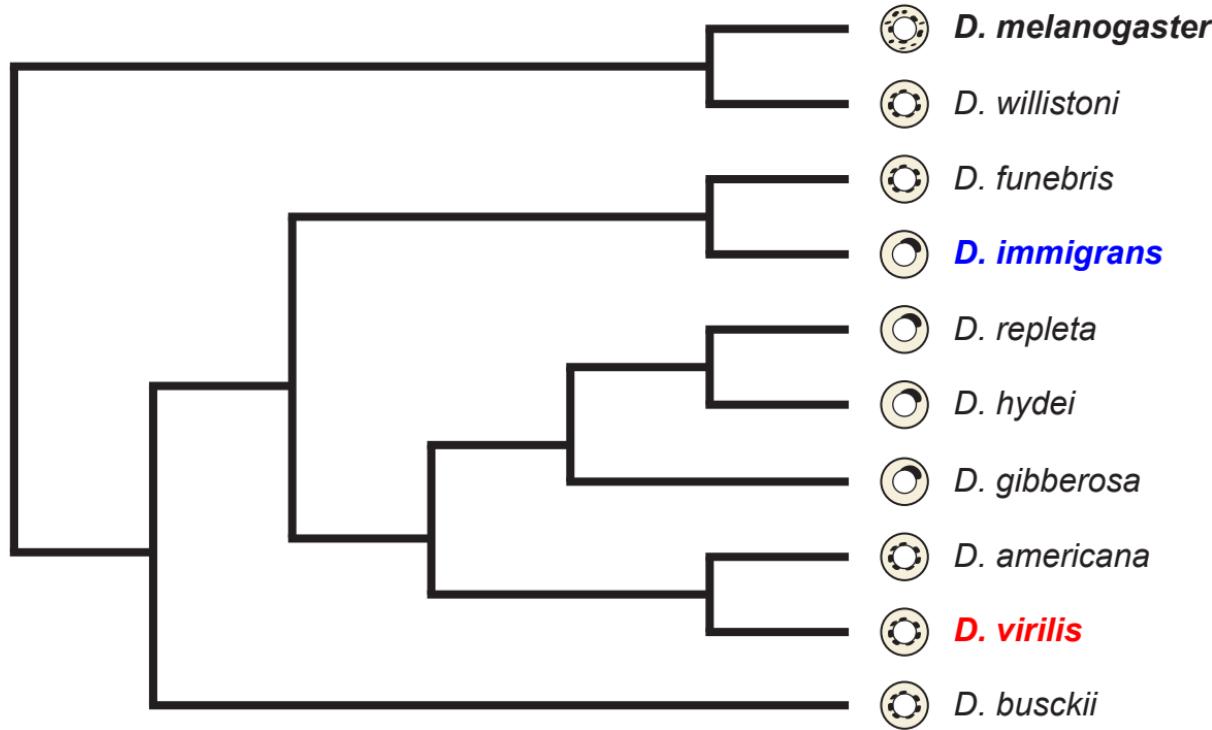
Dper	V-FGDKTKYIRQFKKIFKG----N-----	KD-----L-----
Dpse	V-FGDKTKYIRQFKKIFKG----N-----	KD-----L-----
Dimm	T-LRAKSICWQKFKKIYTA----T-----	KL-----I-----
Dvir	I-LSEKVTYLRFKFRIFTA----RQNLASARTAATAASAKQPVLGRC-----I-----	
Dmel	A-CGE----RFRKIFKSARK-T-----	EL-----PELWK-
Dsec	A-CGE----RFQKIFKSARK-T-----	ES-----PELWK-
Dsim	A-CGE----RFQKIFKSARK-T-----	ES-----PELWK-
Dyak	A-CGE----RFQRIFKSARK-T-----	ES-----PELWK-
Dere	A-CGE----RFQRIFKSARK-T-----	QS-----PELWK-
Dtak	A-FGE----RFQKIFKSARK-T-----	ES-----PELWK-
Dbia	A-FGE----RFRKIFKSARK-T-----	ES-----PELWK-
Deug	D-FSE----KLQKIFKSARK-S-----	ES-----PELRR-
Drho	G-FREKSKFVRQFQKIFKSARK-P-----	ES-----PELWK-
Dele	G-LREKSKFLRQLQKIFYKSTRK-T-----	ES-----PELWK-
Dfic	---DCSKFVRRFQKIFKSARK-T-----	ES-----PESWK-
Dkik	ASFPEKSFKFLRQFHRIFKAGRK-S-----	EGGGEPDPEPELWE
Dana	A-FGDKFKSVRQFHRIFKAGRK-G-----	KE-----P----
Dbip	A-IGDKFKSVRQFHKIFKAGRKK-----	KE-----P----
Dper	---A-KKSEPGRVVLVLVPEKPVDDTEE----SQK-KQLQLVARLFSSSTLISAKEE---	
Dpse	---A-KKSEPGRVVLVLVPEKPVDDTEE----SQK-KQLQLVARLFSSSTLISAKEE---	
Dimm	---S-Q-----EE-----VEE-QQLQIVATLFSSSTLIS-KYT---	
Dvir	---G-Q-----EE-----QQLQIVASLFSSSTLIS-KHS---	
Dmel	VPLVAH-----ELT--SRQSS-QQLQVVARLFSSSTQIS-TKE---	
Dsec	VPLVAH-----ELT--SRQSS-QQLQIVARLFSSSTQIS-TKE---	
Dsim	VPLVAH-----ELT--SRQSS-QQLHIVARLFSSSTQIS-TKE---	
Dyak	VPLA-H-----DLTTTSRQSS-QQLRIVARLFSSSTQIS-TRE---	
Dere	VPLAAH-----DLTN-GRTS-QQLRIVARLFSSSTQIS-TKE---	
Dtak	VPLA-H-----D--V-TGQSG-QQLQIVARLFSSSTLIS-TKE---	
Dbia	VPLA-H-----D--V-TGPSS-QHLQIVARLFSSSTLIS-TKE---	
Deug	VPLSTH-----DQQT-SGQSS-QQLQIVARLFSSSTLIS-TKE---	
Drho	VPLA-H-----D--V-TGQNS-QQLQIVARLFSSSTLVS-TEE---	
Dele	VPLA-H-----D--V-TGQSS-QQLQIVTRLFSSSTLIS-TEE---	
Dfic	VPLA-H-----D--V-TRQSSIHQLOIVARLFSSSTQIS-TEE---	
Dkik	VPLA-D-----EQ----RSQ-EQLQIVARLFSSSTLIS-TKEEI	
Dana	---A-E-----SV----DER-QQLQFVARLFSSSTLIS-TEE---	
Dbip	---A-E-----SV----DER-QQLQFVARLFSSSTLIS-TEE---	
Dper	I-----SC-SY--NSS	
Dpse	I-----SC-SY--NSS	
Dimm	S-----SK-NS--SSC	
Dvir	S-----SA-SI--TSS	
Dmel	ITY-NSNS---NTSE-----NN	
Dsec	ITY-SSNN---NTSE-----TN	
Dsim	ITY-SSNS---NTSE-----SN	
Dyak	ITYTNSNN----SSYSFRRRKSE-SH--RSN	
Dere	ITYSTSNN---NTSSYSFRSRKSE-SC--RSN	
Dtak	IHC-N-NN-----SYSYKSRKSE-SY--SSN	
Dbia	ISY-S-NH-----RYSCRSGKSE-SY--SSN	
Deug	ITF-N-NH---TKNSFSYKCRKSE-GY--SSN	
Drho	ITY-HTNI---N---SYKSRKSQ-SY--DCD	
Dele	ITY-NTNNKNNN---SCKSRKSHKSY--KSS	
Dfic	IIY---NN---I---SFNSRKSQ-SH--SSN	
Dkik	ITT-SRSS---NCNT-SYTRRKSQ-SYCSTST	
Dana	I----YG---N-----SSRKSQ-SF--NSN	
Dbip	I----YG---NGKT-SYSSRKSQ-SY--NSN	

Lasp-binding domain

Dper	IE--DQ-AGPLQQLARAAAESNWCYQDNWNHLKNCYQ-----R-----ANIF----A
Dpse	IE--DQ-AGPLQQLARAAAESNWCYQDNWNHLKNCYQ-----R-----ANIF----A
Dimm	-----K-ITPVQHLATAAMAPFGVYQDNWKHLNNQYQLSQLATQN-----SNQFSCIRS
Dvir	-----K-ISKLQPLATAAAAMSNDVYQDNWKHLNNQYQLPQLNAPK-----NNIHSHIAS
Dmel	IA--TR-PVSLHQMANEAAESNWCYQDNWKHLNNFYQQASVNAPK--MPVPINIV----S
Dsec	IA--TR-PASLHQMANEAAESNWCYQDNWKHLNNFYQQASVNAPQ--MPVPINIV----S
Dsim	IP--TR-PASLHQMANEAAESNWCYQDNWKHLNNFYQQASVNAPQ--MPVPINIV----S
Dyak	SA--TRQPASLHQMANEAAESNWCYQDNWKHLNNFYQQSSVNAPK--MPVPVNIY----S
Dere	TS--TR-PASLHQMANEAAESNWCYQDNWKHLNNFYQQASVNAPK--MPVPINIV----S
Dtak	IN--TK-AAPLQQLAKEAAESNWCYQDNWNHLNNFYQQSRVNEPQ--MPVPINIV----S
Dbia	VD--TR-AAPLQQLAKEAAESNWCYQDNWNHLNNIYQRSSVSEPIQLPVPINIY----S
Deug	MSMHTR-AAPLQQLAKVAAESNWCYQDNWKHLNNFYQQSSVNEAQ--RPVPINIV----T
Drho	IT--AK-TTPLHQQLAKEAAESNWCYQDNWKHLNNFYQQSSANEQ--RPVPINIV----S
Dele	AT--TK-AAPLQQLAKEAAESNWCYQDNWKHLNNFYQQSSANEQ--MPVPINIV----S
Dfic	IA--TK-AAPLPQLAKEAAESNWCYQDNWKHLNNYYQQSSANEQ--IPVPINIV----T
Dkik	IN--TR-SAALPHMAAKAAESNWCYQDNWNHLKNIYQQSSGNAPH--MHEPNQIV----T
Dana	IN--AN-SATLOKLAKAAGEESNWCYQDNWNHLNNFYQQSNVNGHQ--NQVPPNIY----V
Dbip	IK--TN-SASLOKLAKAAGEESNWCYQDNWNHLNNFYQQNNVNGPQ--NPVPPNIY----I
Dper	-DPPEK-P-PPLPA-QE-N-HL-RNSN-----S----QN
Dpse	-DPPEK-P-PPLPA-QE-N-HL-RNSN-----S----QN
Dimm	-DPVQL-QIAPHPP-EQ-Q-H-----QQ
Dvir	-NPAQL-Q-AALPA-E-----
Dmel	PDAPEE-P-INLAPPGH-QPSC-RTQS-----QKTE-PT
Dsec	-DAPEE-P-THLAAPGQ-QPSCCRTQS-----KKTE-PT
Dsim	-DAPEE-P-THLAAPGQ-QPSCCRTQS-----KKPE-PT
Dyak	-DAPEE-R-IHLDAPGH-QPSC-RTQN-----QKTQOTT
Dere	-DAPEE-R-IQVDAPGH-QPSC-RTQN-----QKTQHPT
Dtak	-DATEE-P-IQLDAPEH-HPHC-QNHNEQNQPNQN-QNTK-QN
Dbia	-DAPEE-P-NQMNPQH-HPLC---HNRNQNPSON-QQAK-QD
Deug	-DALEDLP-IQLDNPEH-HPNC-KNNN-----QNIK-QN
Drho	-DPPEE-R-TQFAVPEL-PPHC-RNPN-----QNTK-QN
Dele	-DATEE-R-TQFAVPEP-RPCG-KNHN-----QNNK-PN
Dfic	-DATEE-P-NQFAVPEH-LPPC-KNPK-----QNTK-PN
Dkik	-DPPEL-S-NHLANPELIKPPC-KNHN-----S----PN
Dana	-DTPDR-Q-TQVTDPE--Q-HC-QNPH-----PS----QN
Dbip	-DTPDQ-P-TQVTDPEL-L-HC-QNPN-----PS----QN

Table S13. Log likelihood and parameter estimates from branch site models for Oskar using PRANK MSA alignments of 18 Drosophilids including the *D. virilis* allele sequenced in Webster et al. (1994).

		ω values under Model: MA					ω values under Model: MA _{fix}			
	Site Class	Proportion	Background	<i>D. virilis</i>	<i>I</i>	Proportion	Background	<i>D. virilis</i>	<i>I</i>	
Long Osk (PRANK)	0	0.74	0.22	0.22	-4183.89	0.75	0.22	0.22	-4186.77	
	1	0.16	1	1		0.17	1	1		
	2a	0.07	0.22	89.19		0.05	0.22	1		
	2b	0.01	1	89.19		0.01	1	1		
	0	0.45	0.14	0.14		0.35	0.14	0.14	-2217.65	
Lasp (PRANK)	1	0.42	1	1	-2214.90	0.33	1	1		
	2a	0.06	0.14	40.74		0.15	0.14	1		
	2b	0.06	1	40.74		0.14	1	1		



Germ Plasm Morphology: unipolar perinuclear dispersed

Can assemble germ plasm in *D. melanogaster*: Yes No