Dynamic evolution of the innate immune system in *Drosophila*: Supplementary Materials

Timothy B. Sackton[1], Brian P. Lazzaro[2], Todd A. Schlenke[3], Jay D. Evans[4], Dan Hultmark[5], and Andrew G. Clark[1,6]
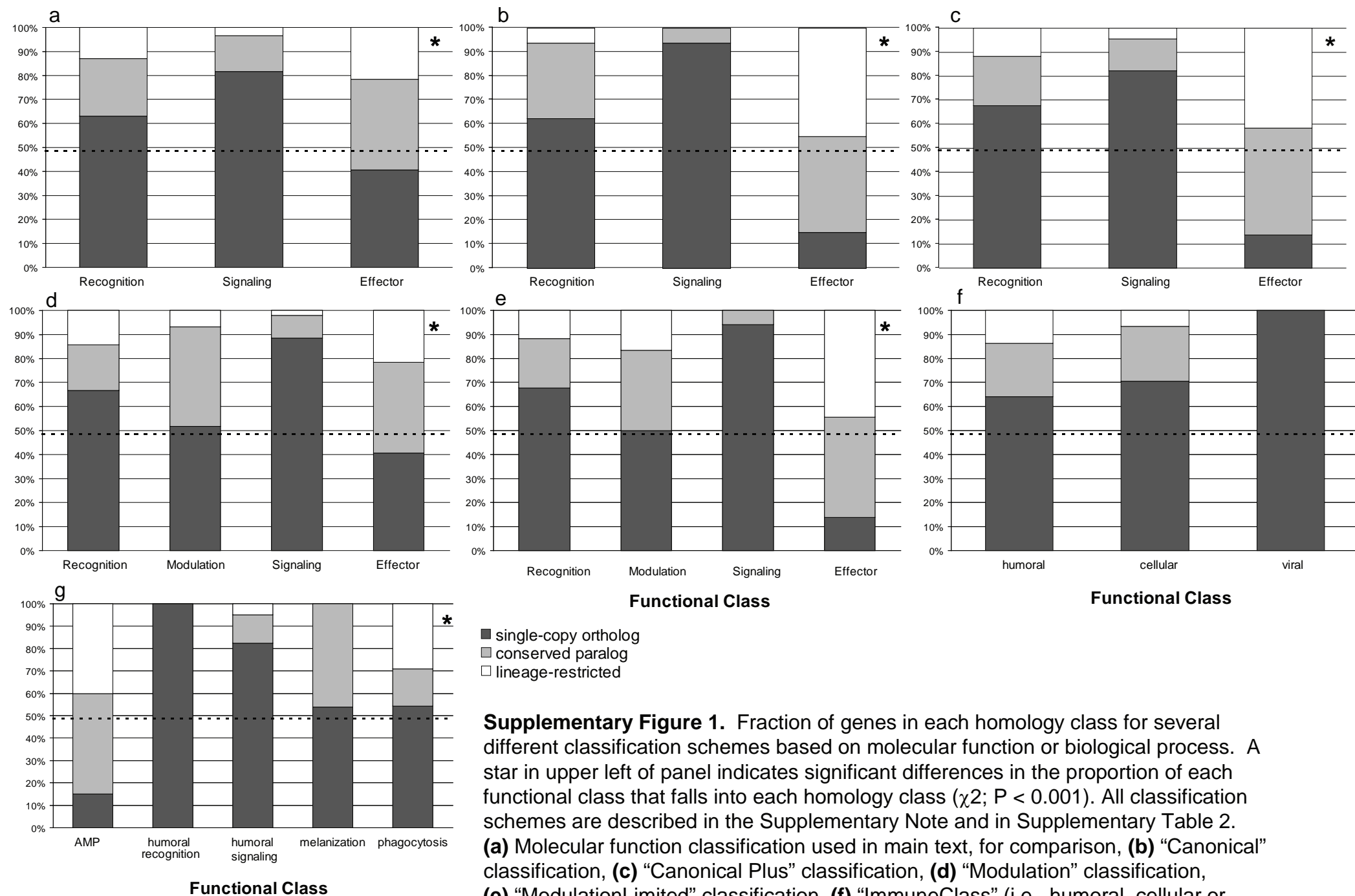
[1]*Field of Ecology and Evolutionary Biology,* [2]*Department of Entomology, and* [6]*Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA;* [3]*Department of Biology, Emory University, Atlanta, Georgia 30322, USA;* [4]*USDA-ARS Bee Research Laboratory, Beltsville, Maryland 20705, USA,* [5]*Umeå Centre for Molecular Pathogenesis, Umeå University, S-901 87 Umeå, Sweden*
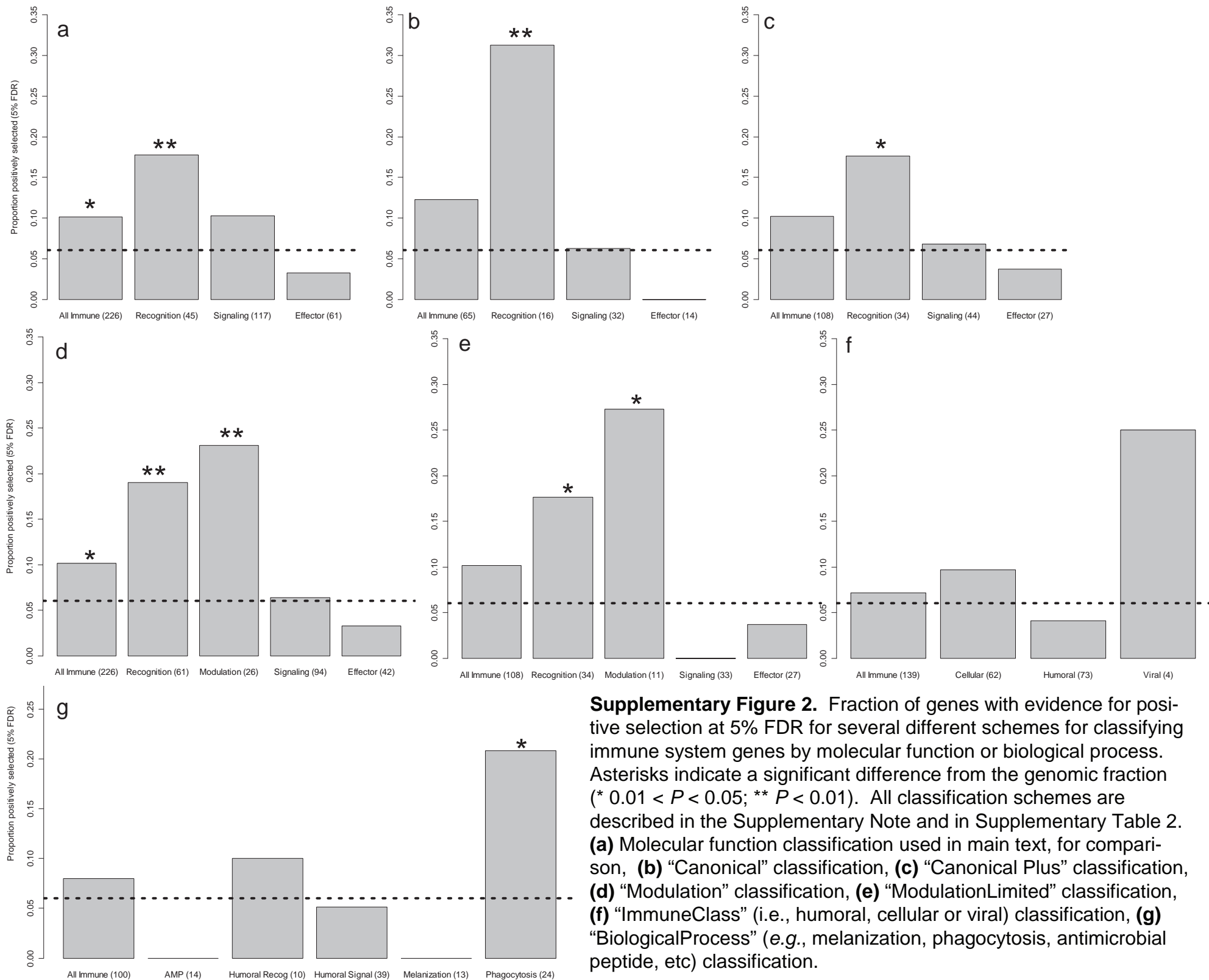
Supplementary Figures 1-7

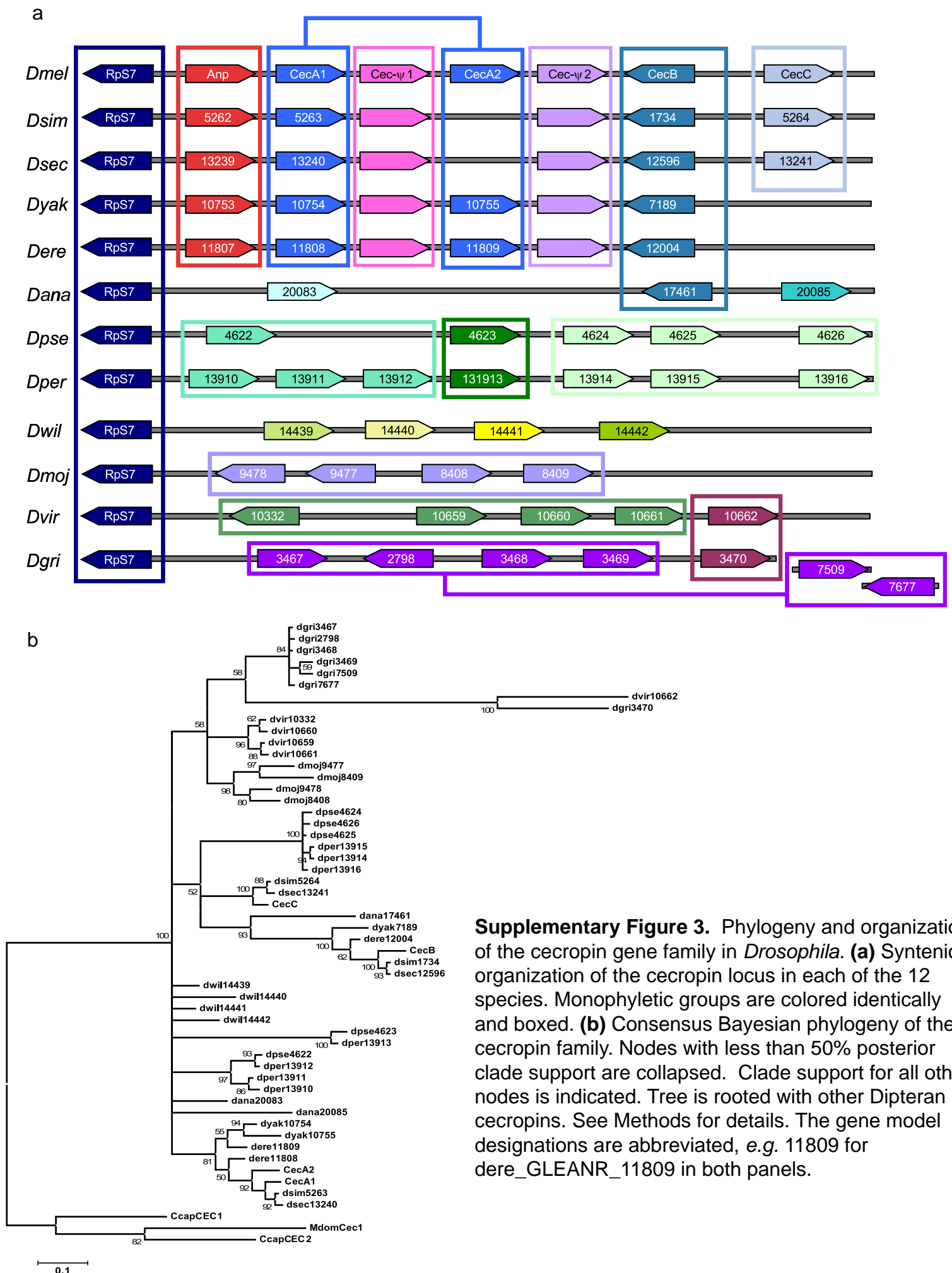**Supplementary Table 5**

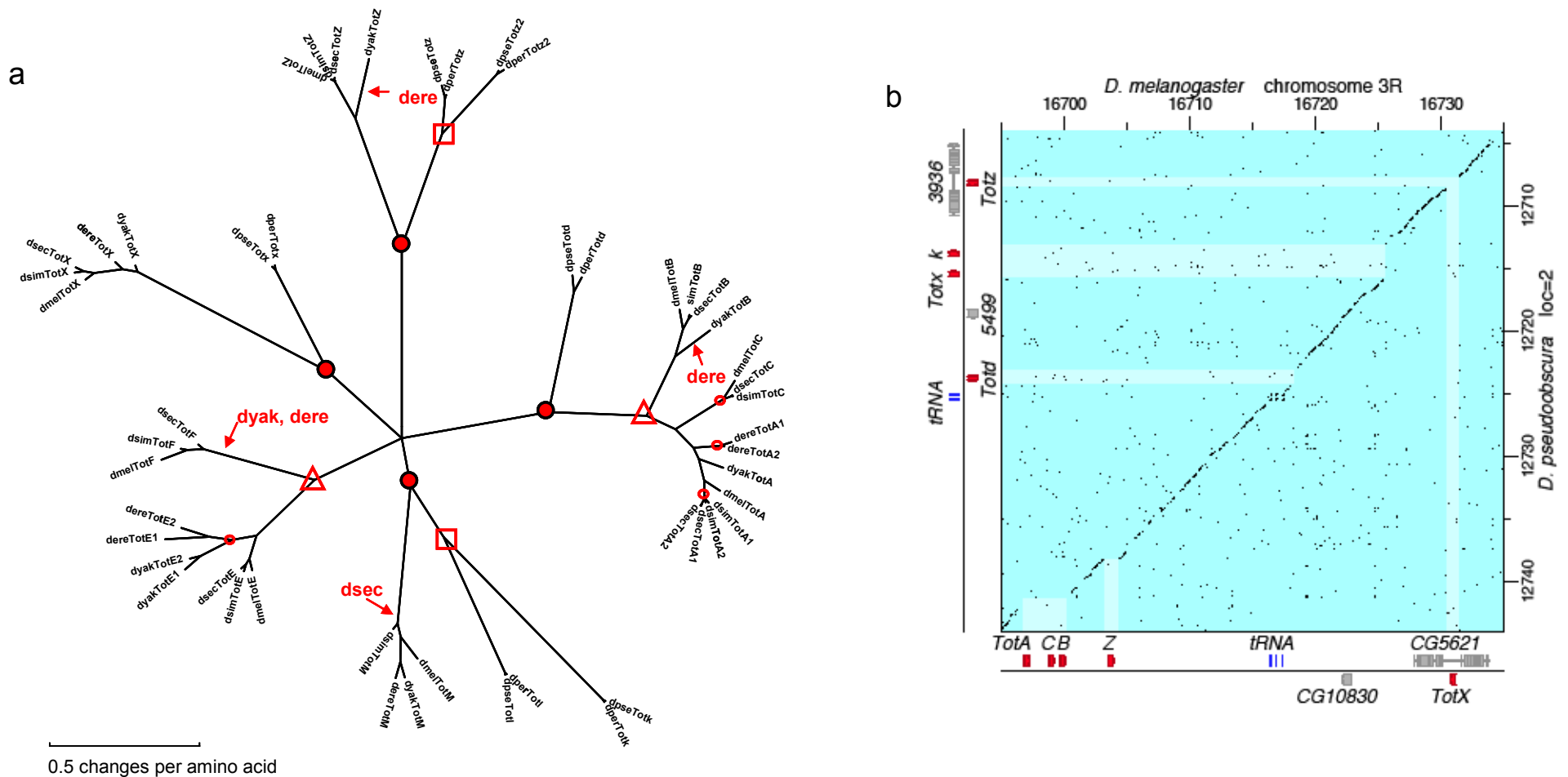**Supplementary Methods**

Supplementary Note

Supplementary Figure 1. Fraction of genes in each homology class for several different classification schemes based on molecular function or biological process. A star in upper left of panel indicates significant differences in the proportion of each functional class that falls into each homology class ($\chi 2$; P < 0.001). All classification schemes are described in the Supplementary Note and in Supplementary Table 2. **(a)** Molecular function classification used in main text, for comparison, **(b)** "Canonical" classification, **(c)** "Canonical Plus" classification, **(d)** "Modulation" classification, **(e)** "ModulationLimited" classification, **(f)** "ImmuneClass" (i.e., humoral, cellular or viral) classification, **(g)** "BiologicalProcess" (i.e., melanization, phagocytosis, antimicrobial peptide, etc) classification.
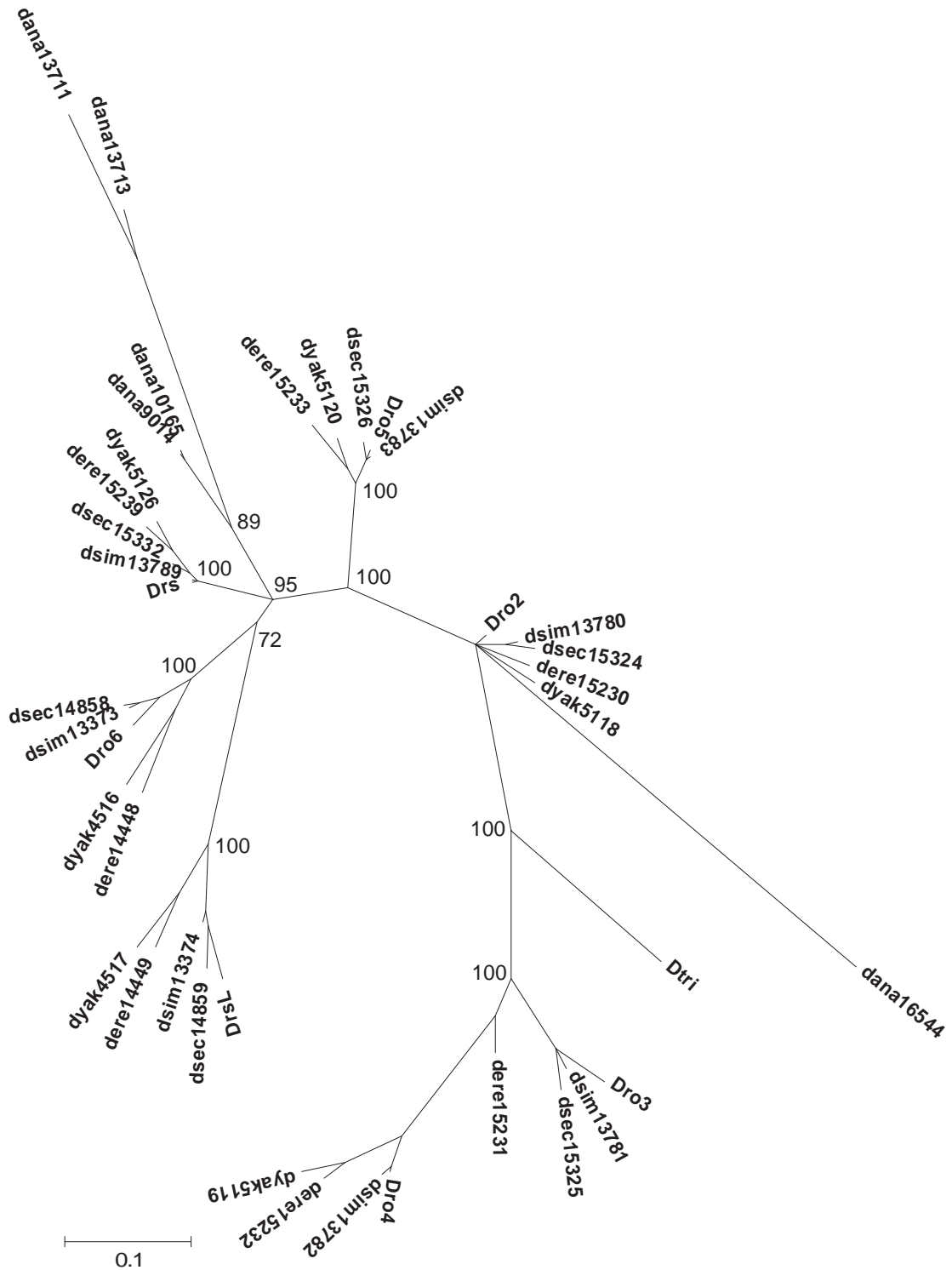
Supplementary Figure 2. Fraction of genes with evidence for positive selection at 5% FDR for several different schemes for classifying immune system genes by molecular function or biological process. Asterisks indicate a significant difference from the genomic fraction (* 0.01 < *P* < 0.05; ** *P* < 0.01). All classification schemes are described in the Supplementary Note and in Supplementary Table 2. **(a)** Molecular function classification used in main text, for comparison, **(b)** "Canonical" classification, **(c)** "Canonical Plus" classification, **(d)** "Modulation" classification, **(e)** "ModulationLimited" classification, **(f)** "ImmuneClass" (i.e., humoral, cellular or viral) classification, **(g)** "BiologicalProcess" (*e.g.*, melanization, phagocytosis, antimicrobial peptide, etc) classification.

Supplementary Figure 3. Phylogeny and organization of the cecropin gene family in *Drosophila*. **(a)** Syntenic organization of the cecropin locus in each of the 12 species. Monophyletic groups are colored identically and boxed. **(b)** Consensus Bayesian phylogeny of the cecropin family. Nodes with less than 50% posterior clade support are collapsed. Clade support for all other nodes is indicated. Tree is rooted with other Dipteran cecropins. See Methods for details. The gene model designations are abbreviated, *e.g.* 11809 for dere_GLEANR_11809 in both panels.
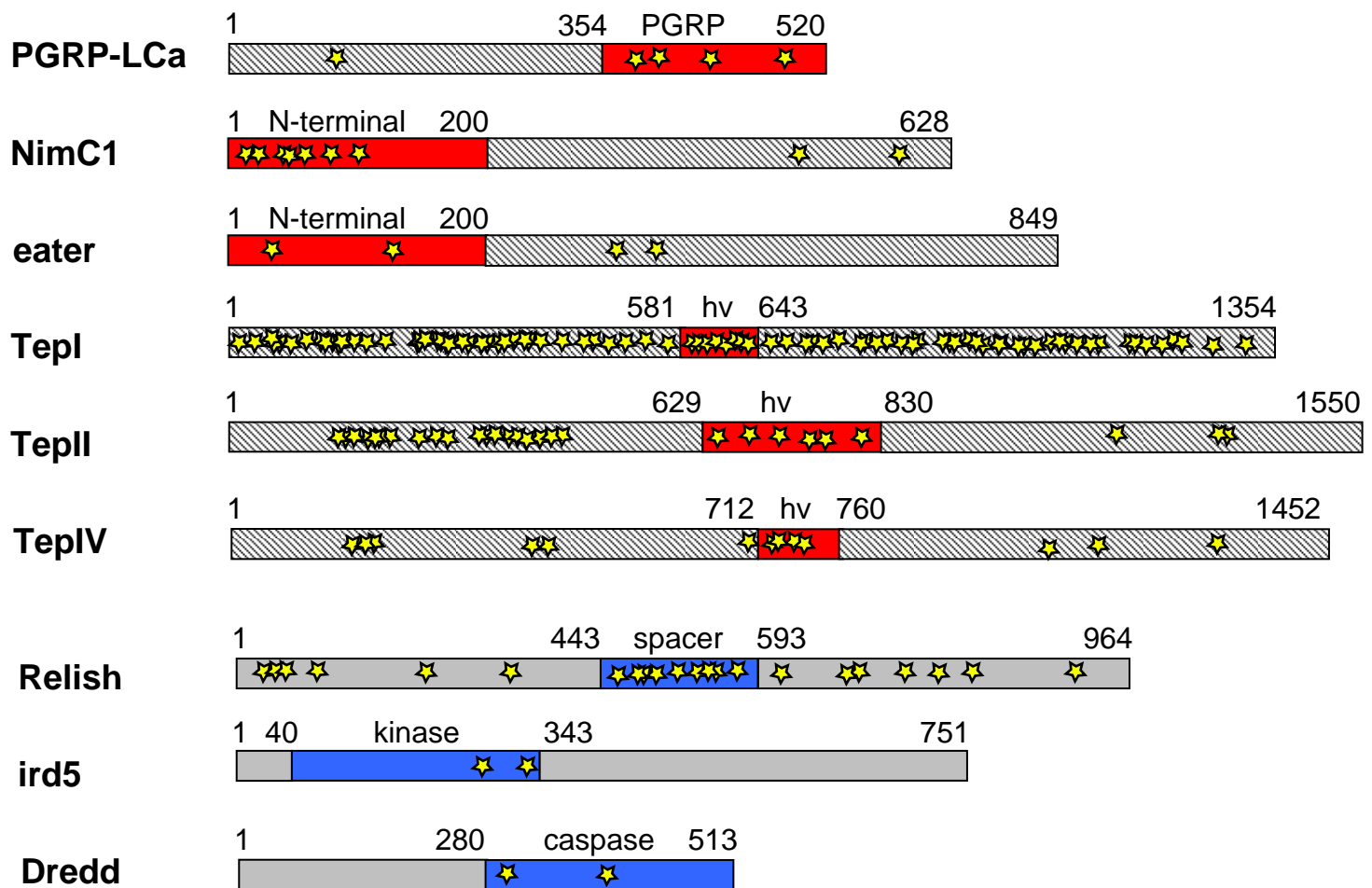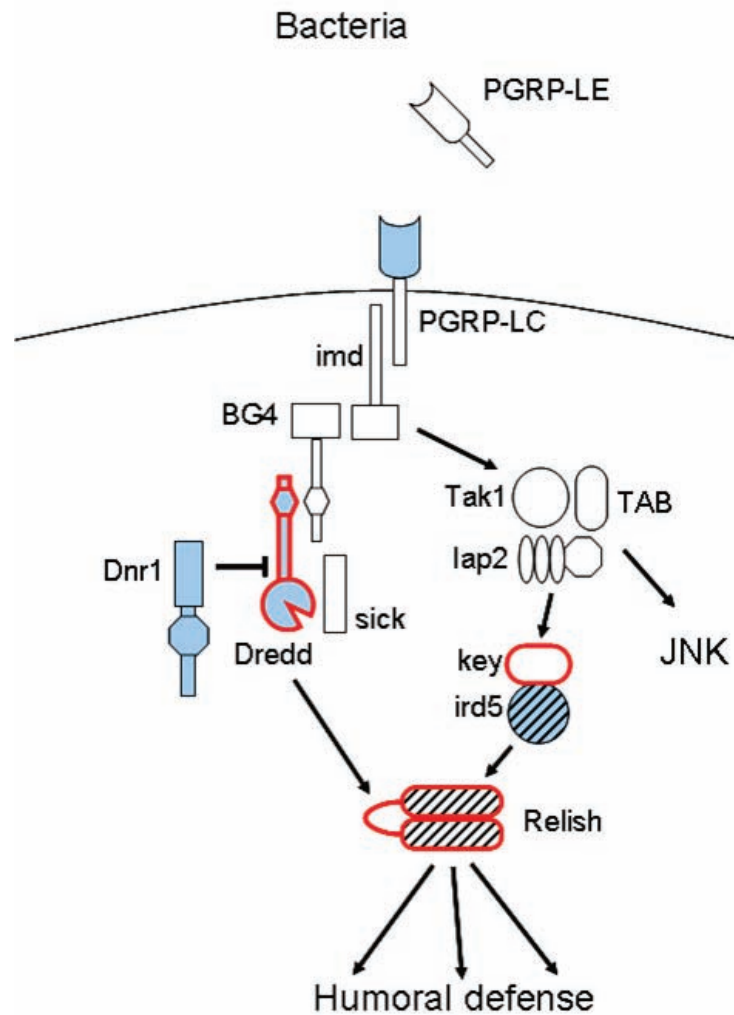
Supplementary Figure 4. Phylogeny and organization of the Turandot genes. **(a)** Consensus Bayesian phylogeny of the Turandot genes in *Drosophila*. Filled red circles indicate ancestral monophyletic clades with 100% posterior support. Empty red squares and triangles indicate specific expansions (>70% posterior support) in the *obscura* and *melanogaster* groups, respectively. Empty red circles indicate specific expansions within the *melanogaster* sub-group. Inferred losses are indicated by red arrows, along with the species in which the loss is inferred to have occurred. **(b)** Dot matrix comparison between the major Turandot loci in *D. melanogaster* and *D. pseudoobscura*. In a sliding 25 bp window, hits with at least 19 identical residues were recorded. Each dot in the graph represents at least one such hit in a 100x100 bp region. Inserted segments that contain Tot genes are highlighted in lighter color. The gene model designations in *D. pseudoobscura* are abbreviated, e.g. 3936 for dpse GLEANR 3936.

Supplementary Figure 5. Bayesian phylogeny of the drosomycin gene family in the genus *Drosophila*. Posterior clade support is indicated for major nodes. GLEANR models are abbreviated with the four-letter species code and the number; the *D. ananassae* models include some non-GLEANR models. The *Dtri* abbreviation represents a drosomycin gene identified in *D. triauraria* (accession number AB059625). No other sequences with similarity to drosomycin are identifiable in the genus *Drosophila*, suggesting that this gene family originated sometime after the divergence of the *melanogaster* and *obscura* groups.

Supplementary Figure 6. Schematic of the genes included in Table 1 and Supplementary Table 5, showing domain structure used in our analysis (not to scale). For recognition proteins, pathogen interaction domains are shown as red bars, and other parts of the protein are represented with cross-hatched bars. For signaling proteins, putative interaction domains are shown as blue bars, and other parts of the protein are represented with grey bars. Coordinates refer to coordinates in the alignments used for PAML. Yellow stars represent approximate location of positively selected sites (posterior probability > 0.75). Selected sites are estimated based on *melangaster* branch-site models for Relish and ird5; for all other genes they are estimated based on the whole phylogeny. hv, hypervariable region; PGRP, peptidoglycan recognition domain

Supplementary Figure 7. Positive selection in the imd pathway. Genes outlined in red have a significantly accelerated ω on the *D. melanogaster* lineage relative to the rest of the tree (P < 0.01). Genes filled with diagonal black lines have significant evidence for positive selection along the *D. melanogaster* branch (P < 0.01; branch-site test). Genes filled in light blue have significant evidence for positive selection on the entire phylogeny (P < 0.02).

Supplementary Table 5. Distribution of positively selected sites among Relish and its interactors

Gene	Domain	Positively Selected Sites[a]	Total Sites (in alignment)	*P*-value[b]
Relish[c]	Spacer	9	149	
	Rest	13	815	0.0033
ird5[c]	Kinase	2	300	
	Rest	0	415	0.176
Dredd[d]	Caspase	2	204	
	Rest	0	255	0.197
Combined	Interaction Domains	13	653	
	Rest	13	1485	0.018

[a]Any site with a Bayesian posterior probability of positive selection greater than 0.75 is considered a "positively selected site."
[b]Calculated by Fisher's Exact Test
[c]In *D. melanogaster* lineage only
[d]In entire phylogeny

Supplementary Methods

Functional classification of immune genes in *D. melanogaster*. Genes were classified into three broad functional categories: "recognition", "signaling", and "effector". The recognition class include thioester-containing proteins[1,2] (TEPs), hemocyte-specific receptors with a role in phagocytosis[3-7], and secreted or membrane-bound recognition proteins that recognize conserved bacterial cell wall components[8,9]. The signaling class includes proteins from the four major signaling pathways with a demonstrated a role in the immune response (Toll, imd, JAK/STAT, and JNK), as well as the p38 stress response pathway[10,11], pathways involved in hemocyte development, differentiation, and proliferation[12-14], and several nuclear pore proteins involved in Nf-κB translocation[15,16]. The effector class primarily includes antimicrobial peptides and enzymes in the phenoloxidase cascade, but we also consider transferrins[17], the Turandot proteins[18,19], lysozymes[20], and proteins involved in coagulation[21] to be effectors. Where molecular function has not been experimentally determined, we have inferred function based on sequence similarity and other indirect means such as expression patterns.

Initial data set, homology assignment and alignments. Our initial gene models and homology assignments derived from computational analysis presented in ref. 22. Briefly, an all-against-all TBLASTN search was run using all annotated *D. melanogaster* proteins and the set of translated consensus gene predictions for each non-*melanogaster* generated by combining several *ab initio* and homology-based gene predictors using GLEAN[23]. The results from this TBLASTN search were combined using a fuzzy reciprocal BLAST algorithm to generate homology clusters, which formed the basis of the primary computational annotation in ref. 22.

In order to study patterns of gene copy number evolution, we started with homology clusters identified by computational algorithms[22], supplemented by manual

TBLASTN searches where necessary. These represent sets of genes that include all orthologs and paralogs of any given *D. melanogaster* gene identifiable in any of the 12 sequenced species. We then assigned these homology clusters to one of three homology classes, based on copy number conservation. Genes that have a single ortholog in all 12 species (*i.e.*, where all pairwise BLAST searches agree on only a single best reciprocal BLAST hit in each species) are assigned to the single-copy orthology class. Genes that have identifiable orthologs and paralogs in both the *Drosophila* and *Sophophora* subgenera (suggesting that these genes were present in the common ancestor), but that are not in the single-copy orthology class, are assigned to the conserved paralogy class. The remaining genes (those that appear to have originated more recently than the common ancestor of the *Drosophila* and *Sophophora* subgenera) are assigned to the lineage-restricted class.

For molecular evolutionary analysis, we generated alignments consisting of the six species in the *melanogaster* group (*D. melanogaster*, *D. simulans*, *D. sechellia*, *D, yakuba*, *D. erecta*, and *D. ananassae*). In some cases gene models could not be found in one or more species, either due to assembly problems or legitimate absence of a gene. For these cases, we used alignments without the missing species. We analyzed gene duplicates that predated the *D. ananassae/D. melanogaster* split separately; for all other paralogs (those that arose after the *D. ananassae / D. melanogaster* split), we included all copies in a single alignment for analysis. In these cases, our estimates of model parameters represent joint estimates of rates in both orthologs and paralogs. When the origin of the gene duplicate was uncertain, we ran the analysis on both versions of the alignment, although in all cases our results are not sensitive to which alignment version we include. For the TEP genes, we could not align the hypervariable region beyond the *D. erecta / D. yakuba* clade, despite the presence of an ortholog in *D. ananassae*, so we present results from the *melanogaster* subgroup alignment only.

Prior to analysis, alignments were masked using the protocol described in ref. 22. Briefly, we generated pairwise sub-alignments for each *Dmel-Dxxx* pair and calculated divergence in a sliding window of 30 nucleotides. Based on the distribution of divergences, and on comparison to expected divergence levels between simulated random alignments, we selected a cutoff level of divergence for masking. Any window with a divergence greater than that cutoff in any given pairwise alignment we then masked out. We manually verified all masked alignments for immune system genes, and any masked codons that could not be unambiguously re-aligned were removed prior to analysis. In addition, all codons with a gap in more than one species were removed prior to analysis.

Phylogenetic analyses. These analyses were carried out on both nucleotide and amino-acid alignments as appropriate, using heuristic parsimony algorithms (PAUP 4.0 or PHYLIP) or Bayesian inference (MrBayes). ModelTest (http://hcv.lanl.gov/content/hcv-db/findmodel/findmodel.html) was used to find best nucleotide model for input into MrBayes, although none of the topologies reported here are sensitive to model assumptions. Amino acid models were run using the 'mixed' prior in MrBayes, which samples over several models of amino acid substitution. We report clade support values from MrBayes unless otherwise noted; topologies reported are generally also recovered by parsimony.

PAML analysis. The primary set of models run are described in the methods section of the main text. In order to estimate p-values, we simulated 10000 alignments under model M7, based on empirical estimates of branch lengths, codon frequencies and kappa in our data set, and analyze the simulated data under both M7 and M8 to generate an empirical null distribution. All observed likelihood ratio test statistics are compared to this empirical null distribution to estimate p-values.

In order to correct for multiple testing, we use two FDR-based approaches that give essentially identical results. Initially, we used the p.adjust function in R to calculate the Benjamini and Hochberg FDR value for the vector of p-values representing the tests of positive selection in immune system genes. These FDR values are interpretable as the fraction of false positives expected based on the number of immune system genes we analyzed. As our data are a subset of the whole genome data, we also used data on the distribution of p-values from ref. 22 to approximate the multiple test correction that would be required if we had included all genomic data in our analysis. For this FDR correction, we used the Q-value method[24]. Results are nearly identical between the two sets, although the latter test is slightly more conservative. Unless otherwise indicated, we report Q-values estimated from the latter procedure, and consider any test with a Q-value of less than 0.05 (or 0.10 in some cases) as significant after multiple test correction. We note that none of our conclusions are qualitatively sensitive to the FDR procedure used or the cutoff used to consider a test significant after multiple test correction (0.05 or 0.10).

Because the topology of the (yak,ere) clade relative to the (mel(sim,sec)) clade is uncertain[25], we ran PAML on all three possible topologies and used the data from the run with the best likelihood. Using only the best supported topology overall – (mel(sim,sec),(yak,ere)) – does not change our results. In order to avoid convergence problems, we ran each analysis three times with different initial values of ω, and used the run with the best likelihood.

To obtain estimates for ω and number of genes with evidence for positive selection, a similar analysis was run on 8510 single-copy orthologs in the *melanogastesr* group[22]. An identical masking procedure was used to screen alignments prior to analysis, although because we could not manually verify all 8510 alignments, we did not attempt to improve gene models or improve the alignment quality in masked regions prior to analysis.

We considered three Bayesian posterior probability of selection cutoffs for determining if a site is positively selected: 0.50, 0.75, and 0.90. We are not primarily interested in determining specific sites evolving by positive selection, but rather assessing the distribution of positively selected sites in a gene, and thus there is a trade-off between having too few positively selected sites to detect significant patterns versus including too many false positives that obliterate any signal of clustering. We use a 0.75 posterior probability cutoff unless otherwise noted, although the patterns we observe are qualitatively similar, though sometimes less significant, when we used either a 0.50 or a 0.90 cutoff.

1.   Blandin, S. & Levashina, E.A. Thioester-containing proteins and insect immunity. *Mol. Immunol.* **40**, 903-8 (2004).
2.   Stroschein-Stevenson, S.L., Foley, E., O'Farrell, P.H. & Johnson, A.D. Identification of Drosophila gene products required for phagocytosis of Candida albicans. *PLoS Biol.* **4**, e4 (2006).
3.   Ramet, M. et al. Drosophila scavenger receptor CI is a pattern recognition receptor for bacteria. *Immunity* **15**, 1027-38 (2001).
4.   Kurucz, E. et al. Hemese, a hemocyte-specific transmembrane protein, affects the cellular immune response in Drosophila. *Proc. Natl. Acad. Sci. USA* **100**, 2622-7 (2003).
5.   Kocks, C. et al. Eater, a transmembrane protein mediating phagocytosis of bacterial pathogens in Drosophila. *Cell* **123**, 335-46 (2005).
6.   Philips, J.A., Rubin, E.J. & Perrimon, N. Drosophila RNAi screen reveals CD36 family member required for mycobacterial infection. *Science* **309**, 1251-3 (2005).
7.   Kurucz, E. et al. Nimrod, a putative phagocytosis receptor with EGF repeats in *Drosophila* plasmatocytes. *Curr. Biol.* **17**, 649-54 (2007)
8.   Kim, Y.S. et al. Gram-negative bacteria-binding protein, a pattern recognition receptor for lipopolysaccharide and beta-1,3-glucan that mediates the signaling for the induction of innate immune genes in Drosophila melanogaster cells. *J. Biol. Chem.* **275**, 32721-7 (2000).
9.   Werner, T. et al. A family of peptidoglycan recognition proteins in the fruit fly Drosophila melanogaster. *Proc. Natl. Acad. Sci. USA* **97**, 13772-7 (2000).
10.   Zhuang, Z.H., Zhou, Y., Yu, M.C., Silverman, N. & Ge, B.X. Regulation of Drosophila p38 activation by specific MAP2 kinase and MAP3 kinase in response to different stimuli. *Cell. Signal.* **18**, 441-8 (2006).
11.   Stronach, B.E. & Perrimon, N. Stress signaling in Drosophila. *Oncogene* **18**, 6172-82 (1999).

12.   Duvic, B., Hoffmann, J.A., Meister, M. & Royet, J. Notch signaling controls lineage specification during Drosophila larval hematopoiesis. *Curr. Biol.* **12**, 1923-7 (2002).

13.   Evans, C.J., Hartenstein, V. & Banerjee, U. Thicker than blood: conserved mechanisms in Drosophila and vertebrate hematopoiesis. *Dev. Cell* **5**, 673-90 (2003).

14.   Muratoglu, S. et al. Regulation of Drosophila friend of GATA gene, u-shaped, during hematopoiesis: a direct role for serpent and lozenge. *Dev. Biol.* **296**, 561-79 (2006).

15.   Uv, A.E. et al. members only encodes a Drosophila nucleoporin required for rel protein import and immune response activation. *Gene. Dev.* **14**, 1945-57 (2000).

16.   Minakhina, S., Yang, J. & Steward, R. Tamo selectively modulates nuclear import in Drosophila. *Genes Cells* **8**, 299-310 (2003).

17.   Dunkov, B. & Georgieva, T. Insect iron binding proteins: insights from the genomes. *Insect Biochem. Molec.* **36**, 300-9 (2006).

18.   Ekengren, S. & Hultmark, D. A family of Turandot-related genes in the humoral stress response of Drosophila. *Biochem. Bioph. Res. Co.* **284**, 998-1003 (2001).

19.   Ekengren, S. et al. A humoral stress response in Drosophila. *Curr. Biol.* **11**, 714-8 (2001).

20.   Hultmark, D. Insect lysozymes. *EXS* **75**, 87-102 (1996).

21.   Scherfer, C. et al. Isolation and characterization of hemolymph clotting factors in Drosophila melanogaster by a pullout method. *Curr. Biol.* **14**, 625-9 (2004).

22.   *Drosophila* 12 Genomes Consortium. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* doi:10.1038/06341 (published online 8 November 2007).

23.   Elsik, C.G. et al. Creating a honey bee consensus gene set. *Genome Biol.* **8**, R13 (2007).

24.   Storey, J.D. A direct approach to false discovery rates. *J. Roy. Stat. Soc. B* **64**, 479-498 (2002).

25.   Pollard, D.A., Iyer, V.N., Moses, A.M. & Eisen, M.B. Widespread discordance of gene trees with species tree in Drosophila: evidence for incomplete lineage sorting. *PLoS Genet.* **2**, e173 (2006).

Supplementary Note

Alternative functional classifications lead to qualitatively similar conclusions

Many of our analyses rely on the assignment of genes to molecular functional classes. Yet some genes have ambiguous or uncertain roles in the immune response, and others may have only limited or weak evidence for a role in immunity. In order to attempt to limit the impact of these issues on our conclusions, we repeated our analyses on four different molecular function classifications (two using the same recognition/signalling/effector framework, two adding a signal modulation class). We also used two biological process schemes: one where we assign immune genes to either humoral, cellular, or viral immunity; and one where we assign immune genes to more precise biological roles within antimicrobial immunity.

Our alternative molecular function classifications are: a "canonical" classification scheme that includes only those proteins with the strongest molecular evidence for a well-defined role in the immune response with a clear molecular function; a "canonical plus" classification that adds paralogs of genes with a clear molecular function, and also some genes with good evidence for at least some immune role, but with a less well-determined molecular function; and two modulation classifications, one that adds a modulation class to the classification used in the main text, and one that adds a modulation class to the "canonical plus" classification just described. These classification schemes, as well as two biological process schemes, are presented as Supplementary Table 2.

For all four molecular function classifications (the two described here plus the one presented in the main text and in Supplementary Table 1) we see a significant deficit of single-copy orthologs among effector proteins and a significant excess of positive

selection among recognition proteins, strongly suggesting that those primary conclusions are not affected by the precise details of the classification schemes used. Furthermore, in most cases we still see significant differences among recognition, signaling, and effector proteins in both the fraction of single-copy orthologs and the fraction of positively selected genes (Supplementary Figs. 1 and 2), suggesting that molecular role in the immune response is an important determinant of both patterns of gene duplication and patterns of positive selection. As discussed in the main text, the 'signal modulation' class appears to have excess positive selection compared to the genomic average, although the addition of this class does not change our main conclusions with respect to the recognition or effector classes. However, our biological process classification of genes into humoral, cellular, and viral responses does not appear to be a significant predictor of either patterns of positive selection or patterns of gene duplication.

The extent to which the immune system as a whole is under excess positive selection compared to the genome as a whole is more uncertain with these additional classifications. Only the molecular function classification scheme presented in the main text has a significant excess of positive selection among the immune system as a whole. However, the fraction of positively selected genes is equal or higher in both of the more restricted molecular classifications, suggesting that it is lack of power, and not absence of an effect, that results in non-significant tests, and that the immune system as a whole, however defined, is more likely to experience positive selection than the average gene in the genomic.

Diversification of class C scavenger receptors (SR-Cs)

The SR-Cs are a small family of scavenger receptor proteins related to SR-CI; these proteins may have a role in phagocytosis, although only SR-CI has been functionally characterized[1]. *D. melanogaster* encodes four SR-Cs, all of which have orthologous

copies in the *melanogaster* subgroup, and two of which (*SR-CIII* and *SR-CIV*) lack transmembrane domains and are probably secreted. *D. ananassae*, and all species outside the *melanogaster* group, have only a single predicted SR-C protein, with a domain structure similar to SR-CI and SR-CII, including a predicted transmembrane domain. *SR-CIII* and *SR-CIV* probably arose from partial or truncated duplications of an ancestral *SR-CI*-like gene.

Diversification the *Turandot* family

The Turandot family, while not functionally well characterized, is strongly induced by infection[2]. Like the drosomycins, it appears to be an evolutionary novelty; within *Drosophila*, the Turandots are limited to the *melanogaster* and *obscura* groups, although they have been secondarily lost in *D. ananassae*. Unlike the drosomycins, there is no trace of any Turandot-like progenitors or relatives outside *Drosophila*, suggesting that the Turandots are an example of a truly novel family within *Drosophila*. The homologous genomic regions in *D. willistoni* and the species of the Drosophila subgenus show no traces of Turandot-like open reading frames, and BLAST searches have failed to identify Turandot homologs in these species or in any other insect. Phylogenetic inference suggests that at least three and perhaps as many as five Turandots were already present before the divergence of the *obscura* and *melanogaster* groups, which subsequently expanded independently in the two groups (Supplementary Fig. 4a). Their dynamic evolutionary history is further illustrated by their different genomic organization in *D. melanogaster* and *D. pseudoobscura*: although the majority of Turandot genes are present in a single physical cluster in both species, the syntenic structure of this cluster with respect to the Turandot genes appears to be completely diverged (Supplementary Fig. 4b).

Differences between immune system genes and other protein-coding genes are not due to different annotation and alignment methodologies

Immune system genes still have a significantly higher ω, albeit only weakly so, when we compared the immune genes in the whole genome dataset (computationally annotated only) to non-immune genes in the same dataset (one-tailed P=0.0487, permutation test). Because the whole genome dataset includes only single-copy orthologs, and includes only the initial computationally predicted set of gene models, only about half of the immune genes in this study are included. ω is not significantly different between the immune genes in the whole genome dataset and the immune genes in this study, suggesting that differences in power, not biases introduced by different alignments and different gene sets, are responsible for the lowered significance.

Pleiotropy does not explain variation in levels of constraint and patterns of positive selection among immune system functional classes

One potential explanation for why immune system functional classes experience different levels of positive selection is the fact that some proteins (such as immune signaling proteins, which are often part of multi-purpose signaling cascades such as the JAK/STAT or Toll pathways) are more likely to have a pleiotropic non-immune function than others (*e.g.*, recognition and effector proteins). These pleiotropic functions may constrain the immune-specific adaptive evolution of signaling genes, if that evolutionary trajectory were to come at a pleiotropic expense[3]. One simple approach to control for pleiotropy is to use information available from FlyBase (http://www.flybase.org/) about mutant phenotypes that have been described in the literature, based on the assumption that genes with a lethal or sterile mutation described in *D. melanogaster* are more likely to have a pleiotropic non-immune function than genes with viable mutations only. While genes that

are mutable to lethal or sterile alleles are marginally more conserved than genes that have only viable mutations described (both across all immune genes and just within the signaling class), the difference is not significant in either case. In addition, there is no significant difference in the fraction of genes that evolve by positive selection between genes with lethal or sterile mutations and genes with only viable mutations. This is consistent with observations from an analysis of the entire protein-coding genome (A. M. Larracuente, T. B. Sackton, A. J. Greenberg, and A. G. Clark, unpublished data) and suggests that variation in levels of pleiotropy is probably not a significant determinant of the patterns of positive selection we observe among recognition, signaling, and effector genes.

Caveats to our study

There are several important limitations to the current study that bear mentioning. Most important, the codon-substitution methodology for detecting positive selection, while very powerful, is limited in the kinds of positive selection in can detect[4]. In particular, recent selective sweeps, and potentially long-term balancing selection, will be missed by our methodology[4]. Second, we cannot rule out the possibility that the signatures of adaptive evolution we see are due to selection on some pleiotropic function. While in many cases we think this is unlikely, either because of the lack of any known non-immune function or because of the localization of positively selected sites to immune-related domains, there are some examples in our data set where this would be plausible (*e.g.,* the nuclear pore proteins[5]). We suggest caution in assuming that positive selection in highly pleiotropic genes is driven by pathogen pressure. Finally, while we have tried to be comprehensive by including the vast majority of proteins with a known role in immunity in our study, the state of knowledge of the *Drosophila* immune system is incomplete, and the observations we make in this study are necessarily based in our current understanding of the system.

1. Ramet, M. et al. Drosophila scavenger receptor CI is a pattern recognition receptor for bacteria. *Immunity* **15**, 1027-38 (2001).
2. Ekengren, S. et al. A humoral stress response in Drosophila. *Curr. Biol.* **11**, 714-8 (2001).
3. Orr, H.A. The genetic theory of adaptation: a brief history. *Nat. Rev. Genet.* **6**, 119-27 (2005).
4. Yang, Z. & Bielawski, J.P. Statistical methods for detecting molecular adaptation. *Trends Ecol. Evol.* **15**, 496-503 (2000).
5. Presgraves, D.C. & Stephan, W. Pervasive adaptive evolution among interactors of the Drosophila hybrid inviability gene, Nup96. *Mol. Biol. Evol.* **24**, 306-14 (2007).