

# Natural Selection Drives *Drosophila* Immune System Evolution

Todd A. Schlenke<sup>1</sup> and David J. Begun

Section of Evolution and Ecology, Division of Biological Sciences, University of California, Davis, California 95616

Manuscript received January 15, 2003

Accepted for publication April 18, 2003

## ABSTRACT

Evidence from disparate sources suggests that natural selection may often play a role in the evolution of host immune system proteins. However, there have been few attempts to make general population genetic inferences on the basis of analysis of several immune-system-related genes from a single species. Here we present DNA polymorphism and divergence data from 34 genes thought to function in the innate immune system of *Drosophila simulans* and compare these data to those from 28 nonimmunity genes sequenced from the same lines. Several statistics, including average  $K_A/K_S$  ratio, average silent heterozygosity, and average haplotype diversity, significantly differ between the immunity and nonimmunity genes, suggesting an important role for directional selection in immune system protein evolution. In contrast to data from mammalian immunoglobulins and other proteins, we find no strong evidence for the selective maintenance of protein diversity in *Drosophila* immune system proteins. This may be a consequence of *Drosophila*'s generalized innate immune response.

**I**NVESTIGATING the relationship between functional properties of genes and patterns of evolution is a major goal of evolutionary genetics. For example, genes functioning in host-pathogen interactions may be targets of directional or balancing selection more often than genes from other functional categories. Evidence of these unusual evolutionary forces may be manifest in the distribution of DNA sequence variation in host immunity genes within and between species (*e.g.*, TANAKA and NEI 1989; HUGHES *et al.* 1990; BISHOP *et al.* 2000; STAHL and BISHOP 2000). Population genetic data from immunity proteins may also clarify the mechanisms of host response to pathogens over evolutionary time and the underlying population dynamics of pathogen virulence.

*Drosophila* has an innate immune system, which mediates a rapid, generalized response to invading pathogens. Innate immunity is conserved from insects to vertebrates (for reviews see KIMBRELL and BEUTLER 2001; SILVERMAN and MANIATIS 2001), although vertebrates have added the acquired immune response, which allows more specificity and “memory” of previous infections. *Drosophila* immunity has two main components: the cellular response and the humoral (systemic) response. The cellular response relies on hemocytes, which possess cell-surface pattern recognition receptors

and which engulf or encapsulate foreign cells. The humoral response is mounted in the fat body, the immunity organ of *Drosophila*. The presence of pathogen antigens in the hemolymph triggers a signaling cascade across fat body cell membranes. Transduction of the signal culminates in the translocation of transcription factors into the nucleus, where they elevate transcription of immunity peptides. These peptides are released from the fat body cells into the hemolymph where they lyse invading microbial cells.

Advances in the genetic and biochemical description of *Drosophila* immunity have allowed us to pursue a large-scale molecular population genetic investigation of immune system genes. Here we present polymorphism and divergence data from 34 genes thought to be involved in recognition and signaling in the cellular and humoral immune responses of *Drosophila simulans*. Previously published data from 28 nonimmunity genes sequenced from the same *D. simulans* lines serve as a basis for comparison with the immunity data, such that the genomic effects of demographic history may be distinguished from the gene-specific effects of positive selection. Our goal is to investigate the relative importance of directional selection, balancing selection, and genetic drift in determining the evolution of host immune system proteins.

## MATERIALS AND METHODS

All *D. simulans* sequence data are from a set of highly inbred lines made from field-caught inseminated females collected in the Wolfskill Orchard, Winters, California, in summer 1995 (BEGUN and WHITLEY 2000a). Sampling from a single population avoids complications in population genetic inference that can arise when genes are sampled from disparate populations

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AF544231–AF544239, AY349649–AY349675, AY349684–AY349736, AY349745–AY349752, AY349761–AY349932, AY352227–AY352265, AY354407–AY354454.

<sup>1</sup>Corresponding author: Section of Evolution and Ecology, Storer Hall, University of California, Davis, CA 95616.  
E-mail: taschlenke@ucdavis.edu

**TABLE 1**  
Summary of the data collected

| Gene                | Location <sup>c</sup> | N <sup>d</sup> | No. of sites <sup>e</sup> | YAK <sup>f</sup> |
|---------------------|-----------------------|----------------|---------------------------|------------------|
| Immunity            |                       |                |                           |                  |
| <i>I8w</i>          | 56F                   | 8              | 1769                      | *                |
| <i>cact</i>         | 35F                   | 8              | 1696                      | *                |
| <i>CG3212</i>       | 23F                   | 8              | 818                       | *                |
| <i>CG11833</i>      | 98F                   | 8              | 897                       | *                |
| <i>crq</i>          | 21C                   | 8              | 1140                      | *                |
| <i>Dif</i>          | 36C                   | 8              | 1610                      | *                |
| <i>dorsal</i>       | 36C                   | 8              | 1478                      | *                |
| <i>Dredd</i>        | 1B                    | 8              | 1237                      | *                |
| <i>GNBP1</i>        | 75D                   | 8              | 863                       | *                |
| <i>Hmu</i>          | 97F                   | 8              | 1901                      | *                |
| <i>IKKβ</i>         | 88B                   | 8              | 1171                      | *                |
| <i>IKKγ</i>         | 60E                   | 8              | 1046                      | *                |
| <i>imd</i>          | 55C                   | 8              | 1183                      | *                |
| <i>κB-Ras</i>       | 43C                   | 7              | 489                       | *                |
| <i>Lectin-galC1</i> | 37D                   | 8              | 518                       | *                |
| <i>Mvl</i>          | 85D                   | 8              | 2044                      | *                |
| <i>pelle</i>        | 97E                   | 8              | 1237                      | *                |
| <i>PGRP-LA</i>      | 67A                   | 7              | 1317                      | *                |
| <i>PGRP-SA</i>      | 10C                   | 8              | 1048                      | *                |
| <i>PGRP-SD</i>      | 66A                   | 8              | 419                       | *                |
| <i>ref(2)P</i>      | 37F                   | 8              | 1983                      | *                |
| <i>Relish</i>       | 93D                   | 7              | 2730                      | *                |
| <i>Ser7</i>         | 9A                    | 8              | 797                       | *                |
| <i>Spn43Ac</i>      | 43A                   | 8              | 1072                      | *                |
| <i>spz</i>          | 97E                   | 7              | 896                       | *                |
| <i>Sr-CI</i>        | 24D                   | 8              | 1903                      | *                |
| <i>Sr-CII</i>       | 48E                   | 8              | 1499                      | *                |
| <i>Sr-CIII</i>      | 24D                   | 8              | 779                       | *                |
| <i>Tak1</i>         | 19D                   | 8              | 2590                      | *                |
| <i>Tehao</i>        | 34C                   | 8              | 2048                      | *                |
| <i>Tep1</i>         | 35D                   | 8              | 1859                      | *                |
| <i>Thor</i>         | 23F                   | 7              | 909                       | *                |
| <i>Toll</i>         | 97D                   | 8              | 3075                      | *                |
| <i>tube</i>         | 82B                   | 8              | 1027                      | *                |

(continued)

that may have different demographic and selective histories. *D. melanogaster* sequence data were taken from GenBank. For a subset of the genes, a *D. yakuba* sequence was used as the outgroup to the *D. melanogaster/D. simulans* species pair. PCR products were directly sequenced on an Applied Biosystems (Foster City, CA) 377 automated sequencer. Information on primer sequences used for both PCR and sequencing reactions for all genes are available upon request from T.A.S. The analysis section in the DnaSP program (ROZAS and ROZAS 1999) was used for most sequence analyses.

The 34 immunity genes surveyed are located throughout the *D. simulans* genome; an average of 7.85 alleles with an average of 1384 bp were sequenced per gene. Theoretical results suggest that sample sizes in this range allow for reasonable estimation of population heterozygosity, at least under the neutral model (TAJIMA 1983; KLIMAN and HEY 1993). A subset of the genes surveyed here are linked to the immune system not by direct experimentation but by virtue of homology to known immunity proteins. The known or inferred function of each immunity gene is provided in supplementary material A at <http://www.genetics.org/supplemental>. Data from 28 nonimmunity genes located on both chromosome

**TABLE 1**  
(Continued)

| Gene                          | Location <sup>c</sup> | N <sup>d</sup> | No. of sites <sup>e</sup> | YAK <sup>f</sup> |
|-------------------------------|-----------------------|----------------|---------------------------|------------------|
| Non-immX <sup>a</sup>         |                       |                |                           |                  |
| <i>bnb</i>                    | 17D                   | 8              | 1015                      |                  |
| <i>ct</i>                     | 7B                    | 6              | 1082                      |                  |
| <i>dec-1</i>                  | 7C                    | 7              | 1415                      | *                |
| <i>garnet</i>                 | 12B                   | 7              | 1071                      | *                |
| <i>mei-218</i>                | 15E                   | 8              | 1177                      | *                |
| <i>mei-9</i>                  | 4B                    | 6              | 1284                      |                  |
| <i>otu</i>                    | 7F                    | 6              | 1146                      |                  |
| <i>ovo</i>                    | 4E                    | 8              | 1356                      |                  |
| <i>Pgd</i>                    | 2D                    | 7              | 912                       | *                |
| <i>rudimentary</i>            | 14F                   | 6              | 1197                      |                  |
| <i>singed</i>                 | 7D                    | 8              | 1431                      | *                |
| <i>sog</i>                    | 13E                   | 8              | 1233                      | *                |
| <i>sqh</i>                    | 5E                    | 7              | 765                       |                  |
| <i>X</i>                      | 5E                    | 7              | 1425                      | *                |
| <i>Yp3</i>                    | 12C                   | 8              | 1219                      |                  |
| Non-imm3R <sup>b</sup>        |                       |                |                           |                  |
| <i>Aats-glu<sup>pro</sup></i> | 95C                   | 7              | 1337                      |                  |
| <i>AP50</i>                   | 94A                   | 8              | 1398                      | *                |
| <i>CP190</i>                  | 88E                   | 7              | 1222                      | *                |
| <i>eld</i>                    | 87B                   | 7              | 971                       | *                |
| <i>fzo</i>                    | 94E                   | 8              | 1360                      |                  |
| <i>Hsc70</i>                  | 88E                   | 7              | 1292                      | *                |
| <i>hyd</i>                    | 92E                   | 8              | 1480                      | *                |
| <i>mir</i>                    | 86A                   | 6              | 1200                      | *                |
| <i>nos</i>                    | 86D                   | 7              | 1050                      |                  |
| <i>Osbp</i>                   | 96B                   | 8              | 1060                      | *                |
| <i>pit</i>                    | 93F                   | 7              | 1123                      |                  |
| <i>T-cpl</i>                  | 94B                   | 8              | 1138                      | *                |
| <i>tld</i>                    | 96A                   | 7              | 953                       |                  |

<sup>a</sup> Nonimmunity X-linked genes.<sup>b</sup> Nonimmunity autosomal genes.

<sup>c</sup> Cytological location. The immunity genes *IKKβ*, *Mvl*, and *Relish* and the nonimmunity genes *CP190*, *eld*, *Hsc70*, *hyd*, *mir*, and *nos* are located within a fixed inversion difference between *D. melanogaster* and *D. simulans* (85A; 93F6). Their cytological locations in *D. simulans* are estimated (BEGUN and WHITLEY 2000b).

<sup>d</sup> Number of alleles sampled.<sup>e</sup> Number of base pairs sequenced per allele per gene.

<sup>f</sup> An asterisk indicates that a *D. yakuba* allele was collected for the gene.

3R and the X chromosome were also included in this study (BEGUN and WHITLEY 2000b; BEGUN 2001). Thirteen of the nonimmunity genes are located on chromosome 3R; the average sample size is 7.31 *D. simulans* alleles with an average of 1199 bp sequenced per allele per gene. For the X chromosome, 15 nonimmunity genes are analyzed with an average of 7.13 alleles and 1182 bp sequenced per gene. These nonimmunity genes were originally chosen on the basis of their cytological locations, not because of *a priori* hypotheses regarding the action of selection (BEGUN and WHITLEY 2000b). Thus, patterns of variation and divergence in a set of genes belonging to a coherent functional category, the immune system, can be compared to this functionally diverse, "random" set of genes. The cytological location, the sample size, the number of sites sequenced (coding and noncoding), and the presence/

TABLE 2

Comparison of replacement divergence and silent divergence

|                             | $K_A/K_S^a$ | Replacement divergence <sup>b</sup> | Silent divergence <sup>c</sup> |
|-----------------------------|-------------|-------------------------------------|--------------------------------|
| Immunity                    | 0.193       | 0.022                               | 0.117                          |
| Nonimmunity                 | 0.133       | 0.015                               | 0.109                          |
| Mann-Whitney <i>U</i> -test | $P = 0.049$ | $P = 0.02$                          | $P = 0.34$                     |

<sup>a</sup> Average per gene  $K_A/K_S$  ratio between *D. simulans* and *D. melanogaster*.

<sup>b</sup> Average per gene replacement divergence between *D. simulans* and *D. melanogaster*.

<sup>c</sup> Average per gene silent divergence between *D. simulans* and *D. melanogaster*.

absence of a *D. yakuba* allele for all genes analyzed are listed in Table 1. Immunity gene sequences (excluding *Relish*; BEGUN and WHITLEY 2000a) have been deposited in GenBank under accession nos. AF544231–AF544239, AY349684–AY349699, AY349705–AY349736, AY349745–AY349752, AY349761–AY349932, AY352227–AY352265, AY354407–AY354454.

For some analyses, fixed differences between *D. simulans* and *D. melanogaster* were polarized using parsimony. For example, if *D. melanogaster* and *D. yakuba* share a base at a particular site but all *D. simulans* alleles have a different base, then a mutation is inferred to have arisen and fixed along the *D. simulans* lineage. Mutations were assigned to the *D. simulans* lineage only if the *D. melanogaster* and *D. yakuba* alleles were identical. Polymorphisms within *D. simulans* were polarized in a similar manner (e.g., AKASHI 1996). Silent site mutations were classified as preferred or unpreferred through the use of the outgroup method as described by AKASHI (1996).

Silent and replacement divergence levels do not significantly differ between autosomal and sex-linked genes in *D. simulans* and *D. melanogaster* (BAUER and AQUADRO 1997; BEGUN and WHITLEY 2000b; BEGUN 2002). BEGUN and WHITLEY (2000b) and BEGUN (2001), however, found that silent heterozygosity levels and the polymorphism frequency distribution do significantly differ across chromosome arms. Thus, for comparisons of heterozygosity and the polymorphism frequency distribution between immunity and nonimmunity genes, nonimmunity chromosome 3R and X-linked data will be treated separately. In these cases, the most relevant comparison is immunity genes *vs.* autosomal nonimmunity genes because 30 of the 34 sampled immunity genes are autosomal. For all other analyses, the nonimmunity chromosome 3R and the X-linked data are combined. Polymorphism and divergence data for individual genes are provided in supplementary material B at <http://www.genetics.org/supplemental>.

## RESULTS

### Tests for recurrent directional selection

**$K_A/K_S$  ratio:** The  $K_A/K_S$  ratio compares the number of replacement (amino acid altering) substitutions per site and silent (synonymous) substitutions per site among different DNA sequences (LI *et al.* 1985). This ratio will tend to be higher for genes experiencing recurrent fixation of beneficial replacement mutations. The average  $K_A/K_S$  value between *D. simulans* and *D. melanogaster* is significantly greater for immunity genes than for

TABLE 3

 $K_A/K_S$  values along the *D. simulans* and *D. melanogaster* lineages

|                        | $K_A/K_S^a$        | Mann-Whitney <i>U</i> -test |
|------------------------|--------------------|-----------------------------|
| <i>D. simulans</i>     |                    |                             |
| Immunity               | 0.268 <sup>b</sup> | $P = 0.022$                 |
| Nonimmunity            | 0.082 <sup>c</sup> |                             |
| <i>D. melanogaster</i> |                    |                             |
| Immunity               | 0.207              | $P = 0.053$                 |
| Nonimmunity            | 0.172              |                             |

<sup>a</sup> The divergence data are derived from fixations that have been polarized to the *D. simulans* or *D. melanogaster* lineages using the outgroup *D. yakuba*.

<sup>b</sup> The genes *PGRP-SA*, *Ser7*, and *Thor* were excluded because  $K_S$  along the *D. simulans* lineage was zero.

<sup>c</sup> The gene *eld* was excluded because  $K_S$  along the *D. simulans* lineage was zero.

nonimmunity genes (Mann-Whitney *U*-test,  $P = 0.049$ , Table 2). Given that silent divergence is not significantly different between these groups (Mann-Whitney *U*-test,  $P = 0.34$ , Table 2), the higher ratio is attributable primarily to elevated replacement divergence (Mann-Whitney *U*-test,  $P = 0.02$ , Table 2). This finding is consistent with that of a similar study comparing rates of mammalian immunity-gene evolution to rates of evolution of other classes of genes (MURPHY 1993).  $K_A/K_S$  estimates along the *D. simulans* and *D. melanogaster* lineages may be made independently using *D. yakuba* as an outgroup. Average  $K_A/K_S$  along the *D. simulans* lineage is significantly greater for immunity than for nonimmunity genes (Mann-Whitney *U*-test,  $P = 0.022$ , Table 3). For *D. melanogaster*, the difference is nearly significant (Mann-Whitney *U*-test,  $P = 0.053$ , Table 3).

**McDonald-Kreitman tests:** Although the elevated  $K_A/K_S$  ratio in immunity genes is consistent with directional selection on immunity proteins, it is also consistent with a higher neutral mutation rate at replacement sites in immunity genes. Joint consideration of polymorphic and fixed mutations can provide a means of distinguishing these alternatives. Under the neutral model of molecular evolution, the ratio of replacement to silent fixations between species should equal the ratio of replacement to silent polymorphisms within a species, regardless of the level of functional constraint on a gene (KIMURA 1983; McDONALD and KREITMAN 1991). However, because of the short sojourn time of replacement mutations under directional selection, such mutations make a relatively smaller contribution to standing protein polymorphism than to protein divergence, thereby elevating the replacement fixation to silent fixation ratio (relative to the replacement-polymorphism-to-silent-polymorphism ratio). The McDonald-Kreitman test (McDONALD and KREITMAN 1991) uses a  $2 \times 2$  contingency table to test for differences in these ratios.

**TABLE 4**  
**McDonald-Kreitman test results with and without outlier genes**

|                                  | Polymorphism             |                     | Divergence <sup>d</sup> |        | $\chi^2$ test          |
|----------------------------------|--------------------------|---------------------|-------------------------|--------|------------------------|
|                                  | Replacement <sup>b</sup> | Silent <sup>c</sup> | Replacement             | Silent |                        |
| Immunity                         | 128                      | 480                 | 160                     | 194    | 62.19, $P \ll 10^{-5}$ |
| Without seven genes <sup>a</sup> | 104                      | 279                 | 80                      | 137    | 6.15, $P = 0.01$       |
| Nonimmunity                      | 35                       | 238                 | 50                      | 122    | 18.03, $P < 10^{-4}$   |
| Without <i>mei218</i>            | 30                       | 232                 | 17                      | 103    | 0.56, $P = 0.45$       |

<sup>a</sup> The seven excluded immunity genes are *dorsal*, *Dredd*, *imd*, *Relish*, *Spn43Ac*, *Tehao*, and *Toll*.

<sup>b</sup> Number of replacement mutations in *D. simulans*.

<sup>c</sup> Number of silent mutations in *D. simulans*.

<sup>d</sup> The divergence data are fixations that have been polarized to the *D. simulans* lineage using the outgroups *D. melanogaster* and *D. yakuba*.

Table 4 shows the total numbers of replacement and silent polymorphisms and replacement and silent fixed differences for immunity and nonimmunity genes. Fixed differences between *D. simulans* and *D. melanogaster* were included only if they could be polarized to the *D. simulans* lineage (*i.e.*, only genes in which a *D. yakuba* allele has been sequenced were analyzed: 33 immunity genes and 15 nonimmunity genes). Both immunity and nonimmunity genes as groups show highly significant deviations from the neutral expectation in these  $2 \times 2$  contingency tables ( $\chi^2 = 62.19$ ,  $P \ll 10^{-5}$ ;  $\chi^2 = 18.03$ ,  $P < 10^{-4}$ , respectively, Table 4). In principle, such deviations from neutrality may be caused by any of the cells in the  $2 \times 2$  table. Results of the type observed in our data, however, are usually interpreted as evidence for adaptive protein evolution, *i.e.*, an excess of replacement fixations. We address the possibility of selection on silent sites in a later section.

Because individual genes are not equally weighted in

our McDonald-Kreitman tests, ratios of polymorphic-to-fixed mutations pooled across genes may be skewed by data from a small number of genes. Such a pattern could obscure the fact that most genes are evolving neutrally. This appears to be the case for the nonimmunity genes, as the departure from neutrality in the pooled McDonald-Kreitman test is attributable to one gene, *mei-218*, which has 33 of the 50 nonimmunity replacement fixations (although *mei-218* data alone are not significantly different from neutral expectation). However, the significant McDonald-Kreitman test for immunity genes cannot be explained by data from a small number of outliers. Seven immunity genes (*dorsal*, *Dredd*, *imd*, *Relish*, *Spn43Ac*, *Tehao*, and *Toll*) individually deviate from neutrality in the McDonald-Kreitman test ( $P < 0.05$ ). Data from the remaining 26 immunity genes still significantly deviate from the neutral expectation in a direction consistent with excess replacement fixations (Table 4).

Selection on silent sites is thought to have led to the high amounts of codon bias typically observed in *Drosophila* genes. Although immunity genes have slightly lower levels of codon bias than nonimmunity genes, average values for three different measures of codon bias do not significantly differ between immunity and nonimmunity genes (Table 5), suggesting that the intensity of selection for codon bias is similar in these gene groups. AKASHI (1996, 1999) hypothesized that selection on silent sites could contribute to the pattern typically attributed to adaptive protein evolution in McDonald-Kreitman tests. This model posits that many silent mutations in *Drosophila* are borderline deleterious (unpreferred) and thus may make a larger contribution to polymorphism than to divergence (SHARP and LI 1986; BULMER 1988; AKASHI 1995). Table 6 compares the ratios of preferred to unpreferred silent polymorphisms with the ratios of preferred to unpreferred silent fixations. As expected (AKASHI 1996, 1999; TAKANO 1998; BEGUN 2001), nonimmunity genes show a significant deviation from the neutral expectation in the direc-

**TABLE 5**

**Comparison of average codon bias between gene groups**

|                        | ENC <sup>a</sup>                             | CBI <sup>b</sup> | % GC 3rd <sup>c</sup> |
|------------------------|--|------------------|-----------------------|
| Immunity               | 49.60  | 0.39             | 0.65                  |
| Non-immX <sup>d</sup>  | 44.54  | 0.48             | 0.70                  |
| Non-imm3R <sup>e</sup> | 45.85  | 0.44             | 0.67                  |
|                        | Mann-Whitney <i>U</i> -test <i>P</i> -values |                  |                       |
| Imm vs. 3R             | 0.18   | 0.45             | 0.59                  |
| X vs. 3R               | 0.50   | 0.45             | 0.56                  |
| Imm vs. X              | 0.08   | 0.07             | 0.11                  |

<sup>a</sup> Effective number of codons (WRIGHT 1990). Low values indicate high codon bias.

<sup>b</sup> Codon bias index (MORTON 1993). High values indicate high codon bias.

<sup>c</sup> Percentage of GC content at third positions. High values are usually associated with high codon bias in *Drosophila*.

<sup>d</sup> Nonimmunity X-linked genes.

<sup>e</sup> Nonimmunity autosomal genes.

**TABLE 6**  
**Silent site evolution and modified McDonald-Kreitman test**

|             | Polymorphism   |                |                | Divergence <sup>d</sup> |    |     | 2 × 2 $\chi^2$ test   |                                   |
|-------------|----------------|----------------|----------------|-------------------------|----|-----|-----------------------|-----------------------------------|
|             | P <sup>a</sup> | U <sup>b</sup> | R <sup>c</sup> | P                       | U  | R   | P/U                   | P/R                               |
| Immunity    | 78             | 167            | 128            | 53                      | 74 | 160 | 3.59, <i>P</i> = 0.06 | 8.21, <i>P</i> < 10 <sup>-2</sup> |
| Nonimmunity | 28             | 108            | 35             | 31                      | 56 | 50  | 6.17, <i>P</i> = 0.01 | 0.56, <i>P</i> = 0.45             |

<sup>a</sup> Number of preferred silent mutations.

<sup>b</sup> Number of unpreferred silent mutations.

<sup>c</sup> Number of replacement mutations.

<sup>d</sup> The divergence data are fixations that have been polarized to the *D. simulans* lineage using the outgroups *D. melanogaster* and *D. yakuba*.

tion of “too many” unpreferred polymorphisms ( $\chi^2 = 6.17$ , *P* = 0.01). The immunity genes do not show a significant deviation ( $\chi^2 = 3.59$ , *P* = 0.06), although these data also show a trend toward an excess of unpreferred polymorphisms.

One strategy for minimizing the potential complications of weakly deleterious silent mutations on interpretation of the McDonald-Kreitman test is to use only preferred silent mutations. Under the mutation-selection-drift model, preferred codons are beneficial. Thus, if the ratio of replacement to preferred silent fixations is significantly greater than the ratio of replacement to preferred silent polymorphisms, one might have more confidence in the inference that replacement sites are under directional selection. Comparison of replacement and preferred silent mutations in immunity genes reveals a highly significant deviation in the direction of an excess of replacement fixations ( $\chi^2 = 8.21$ , *P* < 10<sup>-2</sup>). However, replacement and preferred silent mutations from nonimmunity genes show no deviation ( $\chi^2 = 0.56$ , *P* = 0.45). This result provides additional support for the notion that directional selection plays a greater role in immunity protein evolution than in nonimmunity protein evolution. This result also suggests that selection on silent sites may have inflated McDonald-Kreitman estimates of the effects of positive selection observed in previous studies (FAY *et al.* 2002; SMITH and EYRE-WALKER 2002) utilizing pooled nonimmunity *Drosophila* data.

### Tests for recent selection

**Polymorphism levels:** Several types of selection may lead to reductions of linked heterozygosity (MAYNARD-SMITH and HAIGH 1974; KAPLAN *et al.* 1989; GILLESPIE 1997, 1999). Thus, if immunity genes experience directional selection more often than nonimmunity genes, immunity genes should have lower nucleotide diversity (relative to divergence) than nonimmunity genes. The ratio of silent polymorphism to divergence is significantly lower for immunity genes (0.19) than for autosomal nonimmunity genes (0.32, Mann-Whitney *U*-test,

*P* = 0.03, Table 7). Although BEGUN and WHITLEY (2000b) previously reported that X-linked genes have reduced heterozygosity compared to autosomal genes in *D. simulans*, heterozygosity in the subset of their X-linked and autosomal data analyzed here are not significantly different (Mann-Whitney *U*-test, *P* = 0.08). Silent heterozygosity-to-divergence ratios for immunity and X-linked genes are not different (Mann-Whitney *U*-test, *P* = 0.79). Reduced silent heterozygosity in immunity genes supports the hypothesis that linked selection has a greater effect on these genes than on a random sample of autosomal genes in *D. simulans*. Although greater effects of linked purifying selection (*i.e.*, background selection; CHARLESWORTH *et al.* 1993) on immunity proteins could also contribute to this pattern, this seems unlikely given the higher rate of protein divergence in immunity genes compared to nonimmunity genes.

**Polymorphic-site allele frequencies:** In addition to reducing linked heterozygosity, directional selection may skew the frequency of derived alleles at linked polymorphic sites toward an excess of high- and low-frequency alleles (BRAVERMAN *et al.* 1995; FAY and WU 2000; PRZEWORSKI 2002). This skew may occur if a selected allele does not fix or if recombination dissociates alleles at linked polymorphic sites from the selected allele before fixation. Table 8 compares the number of

**TABLE 7**

### Average silent heterozygosity-to-divergence ratios

|           | Het/Div <sup>a</sup> |                   | Mann-Whitney <i>U</i> -test |
|-----------|----------------------|-------------------|-----------------------------|
| Immunity  | 0.19                 | Imm <i>vs.</i> 3R | <i>P</i> = 0.03             |
| Non-immX  | 0.18                 | X <i>vs.</i> 3R   | <i>P</i> = 0.08             |
| Non-imm3R | 0.32                 | Imm <i>vs.</i> X  | <i>P</i> = 0.79             |

<sup>a</sup> Silent heterozygosity in *D. simulans* standardized by silent divergence between *D. simulans* and *D. melanogaster*. X-linked values have been multiplied by 4/3 to account for chromosomal population size differences (assuming equal numbers of males and females).

**TABLE 8**  
**Numbers of derived silent polymorphisms at low, intermediate, and high frequency**

|            | Low <sup>a</sup><br>0.0 < P ≤ 0.2 | Intermediate<br>0.2 < P < 0.8 | High<br>0.8 ≤ P < 1.0 | 2 × 3 homogeneity<br>test   |
|------------|-----------------------------------|-------------------------------|-----------------------|-----------------------------|
| Immunity   | 150 (0.86) <sup>b</sup>           | 174 (1.00)                    | 26 (0.15)             |                             |
| Non-immX   | 15 (0.28)                         | 53 (1.00)                     | 2 (0.04)              |                             |
| Non-imm3R  | 42 (0.64)                         | 66 (1.00)                     | 0 (0.00)              |                             |
| Imm vs. 3R |                                   |                               |                       | 10.38, P < 10 <sup>-2</sup> |
| X vs. 3R   |                                   |                               |                       | 8.48, P = 0.01              |
| Imm vs. X  |                                   |                               |                       | 15.94, P < 10 <sup>-3</sup> |

<sup>a</sup> Derived alleles are identified using the outgroups *D. melanogaster* and *D. yakuba*.

<sup>b</sup> The proportion of polymorphisms of each frequency class relative to the number of intermediate frequency polymorphisms is shown in parentheses.

derived polymorphisms in low-frequency ( $\leq 0.2$ ), high-frequency ( $\geq 0.8$ ), and intermediate-frequency classes among immunity and nonimmunity X-linked and autosomal genes. Given that between six and eight *D. simulans* alleles were sequenced for each gene, the low-frequency class corresponds to singletons, while the high-frequency class includes only the highest possible frequency mutations in the sample. The ratio of low-to-intermediate-to-high-frequency silent polymorphisms is significantly heterogeneous for the three pairwise comparisons among the three categories of genes (Table 8), with the immunity genes showing the highest proportion of both high- and low-frequency polymorphisms. This pattern is even more extreme when the proportions of low- and high-frequency polymorphisms are calculated on a gene-by-gene basis and then averaged, giving equal weight to every gene (data not shown). Nonimmunity X-linked genes have by far the smallest proportion of low-frequency mutants (see also Table 3 of BEGUN 2001).

**Polymorphic-site allele distributions:** Several statistics are available for determining whether the distributions of alleles at polymorphic sites among sampled chromosomes conform to the neutral equilibrium expectation. We focus on one such statistic, haplotype diversity ( $H_d$ ), given by equation 8.4 of NEI (1987), but we replace  $2n$  by  $n$ :  $H_d = n(1 - \sum x^2)/(n - 1)$ , where  $x$  is the frequency of each haplotype and  $n$  is the number of alleles sampled. Recent directional selection events that rapidly elevate the frequency of a favored haplotype tend to

decrease overall haplotype diversity at a locus. Immunity genes have the lowest average haplotype diversity, and both immunity genes and X-linked nonimmunity genes have significantly lower haplotype diversities than the autosomal nonimmunity genes (Mann-Whitney *U*-test,  $P < 10^{-2}$ ,  $P < 10^{-2}$ , respectively; Table 9). Low haplotype diversity in immunity genes cannot be explained by a lack of segregating sites, since average  $S$  (including non-coding mutations) for immunity genes (31.1) is greater than that for nonimmunity autosomal genes (29.3). There is no evidence for different average haplotype diversities between immunity and nonimmunity X-linked genes, although nonimmunity X-linked genes have significantly lower average  $S$  (14.4, Mann-Whitney *U*-test,  $P < 10^{-2}$ ).

### Balancing selection

If balancing selection (*e.g.*, overdominance, negative frequency-dependent selection) were common in immunity proteins, one might predict that compared to nonimmunity genes, immunity genes would show elevated replacement heterozygosity. Table 10 shows that the average replacement heterozygosity (relative to divergence) for immunity genes is not significantly greater than that observed for autosomal nonimmunity genes (Mann-Whitney *U*-test,  $P = 0.21$ ). In fact, the ratio for immunity genes (0.12) is substantially less than that for autosomal nonimmunity genes (0.31). However, this difference is due primarily to the “outlier” replacement heterozygosity value for the nonimmunity gene *Hsc70* (1.89). If data from *Hsc70* are omitted, replacement heterozygosity for immunity and autosomal nonimmunity genes is very similar (0.12 and 0.16, respectively). This analysis would seem to provide no evidence for excess protein polymorphism in immunity genes. A complication in interpreting this result, however, is the evidence for reduced silent heterozygosity in immunity genes (Table 7). If reduced silent heterozygosity in immunity genes results from linked selection, then we

**TABLE 9**  
**Average per gene haplotype diversity ( $H_d$ )**

|           | $H_d$ |            | Mann-Whitney<br><i>U</i> -test |
|-----------|-------|------------|--------------------------------|
| Immunity  | 0.77  | Imm vs. 3R | $P < 10^{-2}$                  |
| Non-immX  | 0.81  | X vs. 3R   | $P < 10^{-2}$                  |
| Non-imm3R | 0.94  | Imm vs. X  | $P = 0.57$                     |

TABLE 10

## Average replacement heterozygosity-to-divergence ratios

|           | Het/Div <sup>a</sup> |            | Mann-Whitney<br><i>U</i> -test |
|-----------|----------------------|------------|--------------------------------|
| Immunity  | 0.12 <sup>b</sup>    | Imm vs. 3R | <i>P</i> = 0.21                |
| Non-immX  | 0.19 <sup>c</sup>    | X vs. 3R   | <i>P</i> = 0.49                |
| Non-imm3R | 0.31 <sup>d</sup>    | Imm vs. X  | <i>P</i> = 0.83                |

<sup>a</sup> Replacement heterozygosity in *D. simulans* standardized by replacement divergence between *D. simulans* and *D. melanogaster*. X-linked values have been multiplied by 4/3 to account for chromosomal population size differences.

<sup>b</sup> The gene *Mvl* was excluded because replacement divergence was zero.

<sup>c</sup> The gene *sqh* was excluded because replacement divergence was zero.

<sup>d</sup> The gene *AP50* was excluded because replacement divergence was zero.

would also expect to observe reduced replacement heterozygosity in immunity genes. Furthermore, our analyses of polymorphism and divergence provide evidence that directional selection elevates rates of amino acid divergence in immunity genes. Overall, then, we might expect replacement heterozygosity-to-divergence ratios for immunity genes to be significantly lower than those for nonimmunity genes. Thus, one interpretation of the fact that replacement heterozygosity is not clearly reduced in immunity genes is that some level of diversity-enhancing selection acts on amino acid variants in immune system genes.

## DISCUSSION

Molecular population genetic analyses of large numbers of genes allow investigation of the relative importance of various evolutionary forces acting on proteins in different functional classes. Comparing genes from different functional classes also allows one to move beyond the assumptions and testing of the neutral equilibrium model. This is important because deviations from neutral model expectations can often be explained by either selection or demographic phenomena (*e.g.*, population bottlenecks, population expansion, population subdivision). Comparisons of different classes of genes within a single population sample can help distinguish between these alternatives because effects of selection tend to be gene specific while effects of demography tend to be genome wide.

The *D. simulans* immune system DNA sequence data presented here are mainly from proteins thought to be involved in recognition of pathogens (in the cellular and/or humoral immune responses) and in upstream humoral response-signaling pathways. The distribution of DNA sequence variation within and between species suggests that directional selection plays a more important role in immunity gene evolution than in nonimmu-

TABLE 11

## Potential effects of physical location on silent heterozygosity and haplotype diversity

|                    | Het/Div <sup>a</sup>  | Mann-Whitney<br><i>U</i> -test |
|--------------------|-----------------------|--------------------------------|
| ImmLo <sup>b</sup> | 0.18                  |                                |
| ImmHi <sup>c</sup> | 0.19                  |                                |
| Non-imm3R          | 0.32                  |                                |
| ImmHi vs. ImmLo    |                       | <i>P</i> = 0.59                |
| ImmHi vs. 3R       |                       | <i>P</i> = 0.06                |
| ImmLo vs. 3R       |                       | <i>P</i> = 0.05                |
|                    | <i>H</i> <sub>d</sub> | Mann-Whitney<br><i>U</i> -test |
| ImmLo              | 0.82                  |                                |
| ImmHi              | 0.74                  |                                |
| Non-imm3R          | 0.94                  |                                |
| ImmHi vs. ImmLo    |                       | <i>P</i> = 0.23                |
| ImmHi vs. 3R       |                       | <i>P</i> < 10 <sup>-2</sup>    |
| ImmLo vs. 3R       |                       | <i>P</i> = 0.03                |

<sup>a</sup> X-linked values have been multiplied by 4/3 to account for chromosomal population size differences.

<sup>b</sup> Data from 12 immunity genes in potentially low recombination regions (see text).

<sup>c</sup> Data from the remaining 22 immunity genes.

nity genes. Immunity genes have a significantly higher average  $K_A/K_S$  ratio (Table 2), a greater proportion of genes contributing to significant McDonald-Kreitman test results (Table 4), a significantly lower average standardized silent heterozygosity (Table 7), a greater proportion of both low- and high-frequency silent polymorphisms (Table 8), and significantly lower average haplotype diversity (Table 9) than nonimmunity genes sampled from the same set of inbred lines from a single California population.

Some population genetic differences between immunity and nonimmunity genes might be explained by differences in the levels of recombination experienced by these gene groups, as opposed to higher levels of positive selection in the immunity genes. For example, reduced silent heterozygosity in immunity genes could potentially be explained by lower levels of recombination in immunity genes (BEGUN and AQUADRO 1992; AQUADRO *et al.* 1994). Genetic data from *D. simulans*, although limited, suggest that there is little heterogeneity in recombination rates across the genome (STURTEVANT 1929; OHNISHI and VOELKER 1979; TRUE *et al.* 1996). Although reduced recombination tends to be associated with proximity to centromeres and telomeres, the centromere effect of reduced crossing over encompasses a smaller physical region in *D. simulans* than in *D. melanogaster*. These genetic data are consistent with population genetic data showing that genes located within only five cytological divisions from centromeres, or within two

TABLE 12

Comparison of replacement and silent variation in homologous immunity genes in *D. simulans* and *D. melanogaster*

|                        | Polymorphism <sup>a</sup> |        | Divergence  |        | $\chi^2$ test          |
|------------------------|---------------------------|--------|-------------|--------|------------------------|
|                        | Replacement               | Silent | Replacement | Silent |                        |
| <i>D. simulans</i>     | 36                        | 163    | 63          | 54     | 43.79, $P \ll 10^{-5}$ |
| <i>D. melanogaster</i> | 41                        | 69     | 88          | 113    | 1.24, $P = 0.27$       |

<sup>a</sup> Data from the immunity genes *Relish*, *Spn43Ac*, *spz*, *Sr-CI*, and *Toll* (see text).

cytological divisions from telomeres, have “normal” levels of polymorphism (HASSON *et al.* 1998; BEGUN and WHITLEY 2000b; BEGUN 2002).

Of the 34 immunity genes considered here, 12 (*crq*, *Dif*, *dorsal*, *Dredd*, *IKK $\gamma$* ,  $\kappa$ *B-Ras*, *Lectin-galCI*, *Mvl*, *ref(2)P*, *Spn43Ac*, *Tak1*, and *tube*) are located within five cytological divisions of a centromere or within two cytological divisions of a telomere. The same is true of only 2 of the 28 nonimmunity genes (*bnb*, *Pgd*). However, neither the average standardized silent heterozygosity nor the average haplotype diversity of these 12 immunity genes is significantly different from that of the remaining 22 immunity genes (Mann-Whitney *U*-test,  $P = 0.59$ ,  $P = 0.23$  respectively; Table 11). Although the 22 immunity genes from regions of normal recombination no longer have levels of silent heterozygosity significantly lower than those of the autosomal nonimmunity genes (Mann-Whitney *U*-test,  $P = 0.06$ , Table 11), the similarity in average standardized silent heterozygosity levels between the immunity gene groups (0.19 *vs.* 0.18, respectively) suggests that this loss of significance is more reflective of a loss of power due to the removal of the 12 immunity genes. The 22 immunity genes in regions of normal recombination still have significantly lower average haplotype diversity than the autosomal nonimmunity genes (Mann-Whitney *U*-test,  $P < 10^{-2}$ , Table 11). Overall, there is little evidence that a difference in recombination rates between sampled immunity and nonimmunity genes contributes significantly to their different population genetic patterns.

The fact that a subset of humoral response immunity genes surveyed here also functions in early development (BELVIN and ANDERSON 1996) could compromise our inference that the unusual population genetic data from “immunity” genes are attributable to their role in immunity. However, patterns of variation in immunity genes that also have developmental roles (*18w*, *cactus*, *dorsal*, *pelle*, *spz*, *Toll*, and *tube*) are indistinguishable from patterns observed in other immunity genes (supplementary material B at <http://www.genetics.org/supplemental>). Removing these genes does not affect any of our inferences regarding the population genetics of immunity genes *vs.* nonimmunity genes (results not shown).

The evidence for positive selection in immunity protein evolution spans genes from both the cellular and

the humoral immune responses. A growing body of literature reveals that pathogen genomes encode a wide array of immunomodulatory molecules specifically designed to interfere with host proteins involved in recognition and attack of pathogens and in immunity-signaling pathways (SPRIGGS 1996; HUECK 1998). For example, pathogenic bacteria utilize type III secretion systems to inject proteins into the cytoplasm of host cells that specifically disrupt immunity signal transduction (*e.g.*, SCHESSER *et al.* 1998). The importance of directional selection for *D. simulans* immune system evolution supports the notion that pathogen proteins interact with and exert strong selection pressures on *D. simulans* immune system proteins from multiple pathways and functional groups. Our data are consistent with models of coevolution such as an arms race (DAWKINS and KREBS 1979), in which a specialist pathogen constantly evolves to escape its host’s own evolving defenses, or with a turnover of pathogen species to which the host must repeatedly evolve resistance. Distinguishing between these models will require much additional data regarding the number, frequencies, and infection characteristics of the microbial pathogens that infect *D. simulans* in nature.

Given the effects of directional selection on *D. simulans* immune system genes, it seems reasonable to propose that immunity genes of other host species are also strongly influenced by pathogen-mediated directional selection. Previous studies of antimicrobial peptides (the most downstream steps in the humoral response pathways) from the sister species, *D. melanogaster*, provided no evidence for adaptive protein divergence (CLARK and WANG 1997; DATE *et al.* 1998; RAMOS-ONSINS and AGUADÉ 1998). However, without comparable data from *D. simulans* antimicrobial peptides we cannot be confident that this difference in evolutionary histories results from a lineage effect. Instead, the simplest pathogen virulence strategy may be to interfere with upstream signaling and recognition proteins rather than with the smaller and more numerous downstream antimicrobial peptides.

To better compare and contrast *D. simulans* and *D. melanogaster* immunity protein evolution, we collected *D. melanogaster* polymorphism data for five of the immunity genes that were sampled in *D. simulans* (*Relish*, *Spn43Ac*,

*spz*, *Sr-CI*, *Toll*—GenBank accession nos. AY349649–AY349675, AY349700–AY349704 with the exception of *Relish*; BEGUN and WHITLEY 2000a). The McDonald-Kreitman test from these five genes along the *D. simulans* lineage is highly significant ( $\chi^2 = 43.79$ ,  $P < 10^{-4}$ , Table 12), while the same test on *D. melanogaster* data is not significant ( $\chi^2 = 1.24$ ,  $P = 0.27$ ). Thus, neither the signaling and recognition proteins nor the antimicrobial peptides of *D. melanogaster* show evidence of adaptive protein divergence. Although additional data from *D. melanogaster* immunity genes will be required to make strong statements regarding the comparative population genetics of the immune system in these two species, one possibility is that *D. simulans* immunity genes have recently experienced an unusually intense bout of selection that *D. melanogaster* immunity genes have not. Alternatively, population genetic characteristics unique to *D. melanogaster*, such as a recent elevation of the silent-site substitution rate (AKASHI 1995, 1996) or an excess of replacement polymorphism (ANDOLFATTO 2001), could obscure the signal of adaptive protein evolution that one might otherwise observe along the *D. melanogaster* lineage using the McDonald-Kreitman test.

In contrast to genes involved in the mammalian acquired immune response (HUGHES and NEI 1988, 1989; TANAKA and NEI 1989; HUGHES *et al.* 1990), we found little evidence for an important role of balancing selection in the population genetics of host-pathogen interactions in *D. simulans*. However, the *Drosophila* immune system relies on generalized pathogen-associated molecular patterns (such as lipopolysaccharide of gram-negative bacterial cell walls) to recognize microbes and apparently has little specificity for distinguishing members of large classes of pathogen species (LEMAITRE *et al.* 1997). Therefore, lack of evidence for selective maintenance of protein variation in *D. simulans* immunity genes may not be surprising. It will be interesting to investigate whether mammalian innate immunity proteins that interact directly with pathogens evolve in a manner more akin to the specialized mammalian acquired immune system genes or in a manner more akin to the generalized *Drosophila* innate immune system genes.

Two anonymous reviewers provided useful comments. This work was supported by National Institutes of Health grant GM-55298, the National Science Foundation, and the Sloan Foundation. T.A.S. is cofounder of the Institute of *Drosophila* Immunomics, Ithaca, NY.

#### LITERATURE CITED

- AKASHI, H., 1995 Inferring weak selection from patterns of polymorphism and divergence at “silent” sites in *Drosophila* DNA. *Genetics* **139**: 1067–1076.
- AKASHI, H., 1996 Molecular evolution between *Drosophila melanogaster* and *D. simulans*, reduced codon bias, faster rates of amino acid substitution, and larger proteins in *D. melanogaster*. *Genetics* **144**: 1297–1307.
- AKASHI, H., 1999 Inferring the fitness effects of DNA mutations from polymorphism and divergence data: statistical power to detect directional selection under stationarity and free recombination. *Genetics* **151**: 221–238.
- ANDOLFATTO, P., 2001 Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. *Mol. Biol. Evol.* **18**: 279–290.
- AQUADRO, C. F., D. J. BEGUN and E. C. KINDAHL, 1994 Selection, recombination, and DNA polymorphism in *Drosophila*, pp. 46–56 in *Non-Neutral Evolution: Theories and Molecular Data*, edited by B. GOLDING. Chapman & Hall, New York.
- BAUER, V. L., and C. F. AQUADRO, 1997 Rates of DNA sequence evolution are not sex-biased in *Drosophila melanogaster* and *D. simulans*. *Mol. Biol. Evol.* **14**: 1252–1257.
- BEGUN, D. J., 2001 The frequency distribution of nucleotide variation in *Drosophila simulans*. *Mol. Biol. Evol.* **18**: 1343–1352.
- BEGUN, D. J., 2002 Protein variation in *Drosophila simulans*, and comparison of genes from centromeric versus noncentromeric regions of chromosome 3. *Mol. Biol. Evol.* **19**: 201–203.
- BEGUN, D. J., and C. F. AQUADRO, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356**: 519–520.
- BEGUN, D. J., and P. WHITLEY, 2000a Adaptive evolution of *Relish*, a *Drosophila* NF- $\kappa$ B/I $\kappa$ B protein. *Genetics* **154**: 1231–1238.
- BEGUN, D. J., and P. WHITLEY, 2000b Reduced X-linked nucleotide polymorphism in *Drosophila simulans*. *Proc. Natl. Acad. Sci. USA* **97**: 5960–5965.
- BELVIN, M. P., and K. V. ANDERSON, 1996 A conserved signaling pathway: the *Drosophila* toll-dorsal pathway. *Annu. Rev. Cell Dev. Biol.* **12**: 393–416.
- BISHOP, J. G., A. M. DEAN and T. MITCHELL-OLDS, 2000 Rapid evolution in plant chitinases: molecular targets of selection in plant-pathogen coevolution. *Proc. Natl. Acad. Sci. USA* **97**: 5322–5327.
- BRAVERMAN, J. M., R. R. HUDSON, N. L. KAPLAN, C. H. LANGLEY and W. STEPHAN, 1995 The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* **140**: 783–796.
- BULMER, M., 1988 Are codon usage patterns in unicellular organisms determined by selection-mutation balance? *J. Evol. Biol.* **1**: 15–26.
- CHARLESWORTH, B., M. T. MORGAN and D. CHARLESWORTH, 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**: 1289–1303.
- CLARK, A. G., and L. WANG, 1997 Molecular population genetics of *Drosophila* immune system genes. *Genetics* **147**: 713–724.
- DATE, A., Y. SATTA, N. TAKAHATA and S. I. CHIGUSA, 1998 Evolutionary history and mechanism of the *Drosophila* cecropin gene family. *Immunogenetics* **47**: 417–429.
- DAWKINS, R., and J. R. KREBS, 1979 Arms races between and within species. *Proc. R. Soc. Lond. B Biol. Sci.* **205**: 489–511.
- FAY, J. C., and C.-I. WU, 2000 Hitchhiking under positive Darwinian selection. *Genetics* **155**: 1405–1413.
- FAY, J. C., G. J. WYCKOFF and C.-I. WU, 2002 Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* **415**: 1024–1026.
- GILLESPIE, J. H., 1997 Junk ain’t what junk does: neutral alleles in a selected context. *Gene* **205**: 291–299.
- GILLESPIE, J. H., 1999 The role of population size in molecular evolution. *Theor. Popul. Biol.* **55**: 145–156.
- HASSON, E., I. N. WANG, L. W. ZENG, M. KREITMAN and W. F. EANES, 1998 Nucleotide variation in the triosephosphate isomerase (*Tpi*) locus of *Drosophila melanogaster* and *Drosophila simulans*. *Mol. Biol. Evol.* **15**: 756–769.
- HUECK, C. J., 1998 Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol. Mol. Biol. Rev.* **62**: 379–433.
- HUGHES, A. L., and M. NEI, 1988 Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**: 167–170.
- HUGHES, A. L., and M. NEI, 1989 Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc. Natl. Acad. Sci. USA* **86**: 958–962.
- HUGHES, A. L., T. OTA and M. NEI, 1990 Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. *Mol. Biol. Evol.* **7**: 515–524.
- KAPLAN, N. L., R. R. HUDSON and C. H. LANGLEY, 1989 The “hitchhiking” effect revisited. *Genetics* **123**: 887–899.
- KIMBRELL, D. A., and B. BEUTLER, 2001 The evolution and genetics of innate immunity. *Nat. Rev. Genet.* **2**: 256–267.

- KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK.
- KLIMAN, R. M., and J. HEY, 1993 DNA sequence variation at the *period* locus within and among species of the *Drosophila melanogaster* complex. *Genetics* **133**: 375–387.
- LEMAITRE, B., J.-M. REICHHART and J. A. HOFFMANN, 1997 *Drosophila* host defense: differential induction of antimicrobial peptide genes after infection by various classes of microorganisms. *Proc. Natl. Acad. Sci. USA* **94**: 14614–14619.
- LI, W. H., C. I. WU and C. C. LUO, 1985 A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* **2**: 150–174.
- MAYNARD-SMITH, J., and J. HAIGH, 1974 The hitch-hiking effect of a favorable gene. *Genet. Res.* **23**: 23–35.
- MCDONALD, J. M., and M. KREITMAN, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**: 652–654.
- MORTON, B. R., 1993 Chloroplast DNA codon use: evidence for selection at the *psbA* locus based on tRNA availability. *J. Mol. Evol.* **37**: 273–280.
- MURPHY, P. M., 1993 Molecular mimicry and the generation of host defense protein diversity. *Cell* **72**: 823–826.
- NEI, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.
- OHNISHI, S., and R. A. VOELKER, 1979 Comparative studies of allozyme loci in *Drosophila simulans* and *D. melanogaster*. II. Gene arrangement on the 3rd chromosome. *Jpn. J. Genet.* **54**: 203–209.
- PRZEWORSKI, M., 2002 The signature of positive selection at randomly chosen loci. *Genetics* **160**: 1179–1189.
- RAMOS-ONSINS, S., and M. AGUADÉ, 1998 Molecular evolution of the *cecropin* multigene family in *Drosophila*: functional genes *vs.* pseudogenes. *Genetics* **150**: 157–171.
- ROZAS, J., and R. ROZAS, 1999 DnaSP 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**: 174–175.
- SCHESSER, K., A. K. SPIIK, J. M. DUKUZUMUREMYI, M. F. NEURATH, S. PETTERSSON *et al.*, 1998 The *yopJ* locus is required for *Yersinia*-mediated inhibition of NF-kappaB activation and cytokine expression: YopJ contains a eukaryotic SH2-like domain that is essential for its repressive activity. *Mol. Microbiol.* **28**: 1067–1079.
- SHARP, P. M., and W.-H. LI, 1986 An evolutionary perspective on synonymous codon usage in unicellular organisms. *J. Mol. Evol.* **24**: 28–38.
- SILVERMAN, N., and T. MANIATIS, 2001 NF- $\kappa$ B signaling pathways in mammalian and insect innate immunity. *Genes Dev.* **15**: 2321–2342.
- SMITH, N. G. C., and A. EYRE-WALKER, 2002 Adaptive protein evolution in *Drosophila*. *Nature* **415**: 1022–1024.
- SPRIGGS, M. K., 1996 One step ahead of the game: viral immunomodulatory molecules. *Annu. Rev. Immunol.* **14**: 101–130.
- STAHL, E. A., and J. G. BISHOP, 2000 Plant-pathogen arms races at the molecular level. *Curr. Opin. Plant Biol.* **3**: 299–304.
- STURTEVANT, A. H., 1929 *Contributions to the Genetics of Drosophila simulans and Drosophila melanogaster*, No. 399. Carnegie Institute, Washington, DC.
- TAJIMA, F., 1983 Evolutionary relationship of DNA sequences in finite populations. *Genetics* **105**: 437–460.
- TAKANO, T. S., 1998 Rate variation of DNA sequence evolution in the *Drosophila* lineages. *Genetics* **149**: 959–970.
- TANAKA, T., and M. NEI, 1989 Positive Darwinian selection observed at the variable-region genes of immunoglobulins. *Mol. Biol. Evol.* **6**: 447–459.
- TRUE, J. R., J. M. MERCER and C. C. LAURIE, 1996 Differences in crossover frequency and distribution among three sibling species of *Drosophila*. *Genetics* **142**: 507–523.
- WRIGHT, F., 1990 The “effective number of codons” used in a gene. *Gene* **87**: 23–29.

Communicating editor: W. STEPHAN