

# The Evolution of Antifungal Peptides in *Drosophila*

Francis M. Jiggins<sup>1</sup> and Kang-Wook Kim

*Institute of Evolutionary Biology, University of Edinburgh, Edinburgh EH9 3JT, Scotland*

Manuscript received May 10, 2005

Accepted for publication August 31, 2005

## ABSTRACT

An essential component of the immune system of animals is the production of antimicrobial peptides (AMPs). In vertebrates and termites the protein sequence of some AMPs evolves rapidly under positive selection, suggesting that they may be coevolving with pathogens. However, antibacterial peptides in *Drosophila* tend to be highly conserved. We have inferred the selection pressures acting on *Drosophila* antifungal peptides (drosomycins) from both the divergence of drosomycin genes within and between five species of *Drosophila* and polymorphism data from *Drosophila simulans* and *D. melanogaster*. In common with *Drosophila* antibacterial peptides, there is no evidence of adaptive protein evolution in any of the drosomycin genes, suggesting that they do not coevolve with pathogens. It is possible that this reflects a lack of specific fungal and bacterial parasites in *Drosophila* populations. The polymorphism data from both species differed from neutrality at one locus, but this was not associated with changes in the protein sequence. The synonymous site diversity was greater in *D. simulans* than in *D. melanogaster*, but the diversity both upstream of the genes and at nonsynonymous sites was similar. This can be explained if both upstream and nonsynonymous mutations are slightly deleterious and are removed more effectively from *D. simulans* due to its larger effective population size.

GENES involved in host-parasite interactions are often subject to strong balancing or directional selection. For example, parasite antigens commonly evolve rapidly, and natural selection can maintain many different alleles in a population (ESCALANTE *et al.* 1998). Similarly, vertebrate MHC genes are highly polymorphic and have elevated levels of nonsynonymous substitutions (HUGHES and NEI 1988). However, most of our understanding of the molecular evolution of immune systems comes from studies of the acquired immune response of vertebrates. Acquired immune responses detect and eliminate many different parasites by generating a huge repertoire of receptor molecules with different specificities by somatic rearrangement. This type of immune system is a relatively recent evolutionary innovation of vertebrates; invertebrates instead rely on an innate immune response for defense against pathogens. The innate immune response also remains an essential component of the vertebrate immune response.

The innate immune response relies on a limited number of germline-encoded receptor and effector molecules. Despite this, it is still highly effective in defending against a diverse array of pathogens. This is thought

to be because parasites are recognized using highly conserved molecular patterns and eliminated using responses that are effective across a broad range of parasite taxa (MEDZHITOV and JANEWAY 1997). Presumably these parasite molecules are highly conserved because they are under strong functional constraints, unlike protein antigens that commonly evolve very quickly. Under this scenario, there may be less opportunity for rapid host-parasite coevolution between pathogens and the innate immune system than between pathogens and the acquired immune system.

There is, however, indirect evidence of coevolution between parasites and invertebrate hosts (which possess only innate immune systems). For example, parasites are often adapted to their local host population, as is predicted by most theoretical models of host-parasite coevolution (EBERT 1994; MORAND *et al.* 1996; LIVELY and DYBDAHL 2000). However, it is currently unclear whether these patterns result from coevolution between parasites and components of the innate immune system. Alternatively, coevolution could occur between parasites and other host molecules involved in host-parasite interactions (*e.g.*, cell surface molecules exploited by pathogens to enter cells).

If the innate immune system does coevolve with pathogens, then we would expect immune system genes to show patterns of rapid adaptive evolution or elevated polymorphism. This is the case in *Drosophila simulans*, where a survey of immune system genes showed evidence of stronger directional selection than was the case for nonimmunity genes (SCHLENKE and BEGUN 2003).

Sequence data from this article have been deposited with the EMBL/GenBank data libraries under the accession nos. AJ885048–AJ885086 and AJ965498.

<sup>1</sup>Corresponding author: Institute of Evolutionary Biology, School of Biological Sciences, University of Edinburgh, Ashworth Lab, King's Bldgs., West Mains Rd., Edinburgh EH9 3JT, Scotland.  
E-mail: francis.jiggins@ed.ac.uk

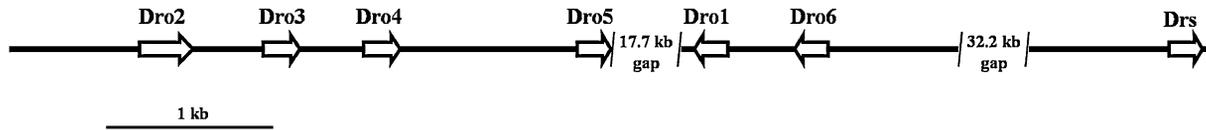


FIGURE 1.—The arrangement of drosomycin genes in *D. melanogaster*, showing the region sequenced.

This could be the result of coevolution, but there is currently no evidence of reciprocal changes in parasite genes. Alternatively, directional selection may result from ecological factors that change the type of opportunistic infections acquired by flies or alter the costs of mounting an immune response.

Clues as to how innate immune systems adapt to novel parasites or parasite genotypes can be gained by comparing the lists of genes under directional and purifying selection. For example, peptidoglycan recognition proteins, which include receptors that bind to bacteria leading to the expression of antimicrobial peptides (AMPs), tend to be highly conserved (JIGGINS and HURST 2003). In contrast, thioester-containing proteins and scavenger receptors, some members of which bind to pathogens and are involved in their subsequent phagocytosis or encapsulation, can be under strong directional selection (LAZZARO 2005; LITTLE and COBBE 2005).

In this study we have focused on the evolution of AMPs, which are an important component of the innate immune response. Most AMPs are thought to exploit the fact that the outer surface of bacterial membranes contains negatively charged phospholipid headgroups that are absent from animal and plant cells (ZASLOFF 2002). Because AMPs have an amphipathic structure (separate hydrophobic and hydrophilic domains) they can first carpet and then integrate into the outer layer of the membrane. This then allows them either to disrupt the integrity of the cell membrane (*e.g.*, by lysing the membrane or forming pores) or to enter the cell to disrupt some intracellular target (ZASLOFF 2002).

In vertebrates, AMPs appear to be a hotspot of rapid adaptive evolution. The most dramatic case is found in frogs, each species of which produces 10–20 AMPs in their dermal gland secretions. These AMPs differ among closely related species in their size, sequence, and antimicrobial specificity, and this rapid diversification has been driven by diversifying selection (DUDA *et al.* 2002). In various groups of mammals, both  $\alpha$ - and  $\beta$ -defensins have diversified under positive selection, and this has led to changes in their antimicrobial specificity (HUGHES and YEAGER 1997; MORRISON *et al.* 2003; SEMPLE *et al.* 2003; ANTICHEVA *et al.* 2004; LYNN *et al.* 2004). These patterns strongly suggest that these AMPs are subject to continually changing selection pressures, presumably due a changing pathogen environment. The changing selection pressures may be the result of parasites coevolving with the AMP molecules.

Studies of insect AMPs have produced more varied results. In termites, the antifungal peptide termicin

evolves rapidly under positive selection (BULMER and CROZIER 2004). However, several studies of six different families of antibacterial peptides in *Drosophila* have consistently failed to produce evidence for rapid adaptive evolution of the amino acid sequence (CLARK and WANG 1997; RAMOS-ONSINS and AGUADE 1998; LAZZARO and CLARK 2003). Although these studies have detected some evidence of positive selection, it is clear that the rate of adaptive protein evolution is dramatically less than that in vertebrates and termites [an exception, andropin, is discussed later (DATE-ITO *et al.* 2002)].

We wanted to test whether this pattern of evolution was general across *Drosophila* AMPs. Therefore, we have investigated the evolution of the drosomycins, a family of AMPs that are active against fungi rather than against bacteria and that show no homology to previously studied antibacterial peptides. Drosomycin (*Drs*) strongly inhibits the growth of filamentous fungi, but has no effect on the growth of a range of bacteria (FEHLBAUM *et al.* 1994). This makes it the only purely antifungal peptide characterized in *Drosophila*. At low concentrations, drosomycin causes the cell cytoplasm to be extruded along the hyphae, suggesting that it lyses cell membranes (FEHLBAUM *et al.* 1994). However, drosomycin's exact mechanism of action is unknown. There are also six drosomycin-like genes in the *D. melanogaster* genome, all found within a 56-kb region of the left arm of chromosome 3 (Figure 1). Unlike drosomycin itself, the antifungal activity of these genes has not been tested experimentally, although it is known that one of them (*Dro5*) is upregulated following fungal infection (DE GREGORIO *et al.* 2001).

## MATERIALS AND METHODS

**Fly lines:** We sequenced the drosomycin genes from 20 *D. melanogaster* chromosomes and 20 *D. simulans* chromosomes. The flies were all originally collected as isofemale lines from natural populations. We used *D. melanogaster* lines from Gabon in west central Africa and from The Netherlands. The *D. simulans* lines were collected in Kenya and France.

The third chromosome of the *D. melanogaster* stocks was made isogenic by standard crosses to the balancer stock TM6/*Sb*. Population genetics analyses in *D. melanogaster* can be confounded by linkage to chromosomal inversions. Therefore, we crossed the isogenic chromosomes to an inversion-free stock and inspected the salivary gland chromosomes of the F<sub>1</sub> progeny for inversion loops. Inversions were uncommon in the region containing the drosomycin genes (1 of 21 Gabon chromosomes and 0 of 15 Netherlands chromosomes). Chromosomes containing inversions were discarded. The *D. simulans* lines were inbred by sib mating for six to nine generations.

**DNA sequence data:** PCR primers were designed from the genome sequences of *D. melanogaster* and *D. simulans* using the program Primer 3. From each chromosome we sequenced the three regions shown in Figure 1. These were amplified in multiple overlapping amplicons. In some cases mutations in the primer binding site prevented the PCR primers from working. In all such cases new primers were designed to ensure that the complete sequence was obtained from all 40 chromosomes. The PCR products were purified either with QIAGEN (Chatsworth, CA) PCR cleanup columns or by using exonuclease I and shrimp alkaline phosphatase to digest unused PCR primers and dNTPs. The PCR products were then sequenced directly using big dye reagents on an ABI capillary sequencer. The sequence chromatograms were inspected by eye to confirm the validity of all differences within and between species. The sequences were then aligned using Clustal W and the alignment was corrected by eye.

Sequences also were obtained from three other species in the *melanogaster* group, *D. yakuba*, *D. ananassae*, and *D. erecta*, whose genomes have been sequenced but are currently not annotated (SMITH 2004; WILSON 2004). The genome assemblies used were *D. yakuba* Langley Group assembly 22/5/2004, *D. erecta* Agencourt assembly 28/10/2004, and *D. ananassae* Agencourt assembly 6/12/2004. The contigs of these species were searched by Blast using the protein sequences of all the *D. melanogaster* drosomycin genes. Unique coding sequences were then downloaded and aligned with the *D. simulans* and *D. melanogaster* sequences. These sequences may contain some errors. However, these are not expected to occur preferentially at either synonymous or nonsynonymous sites. Therefore, any errors will tend to bias estimates of  $d_N/d_S$  toward 1 (neutrality).

**Tree reconstruction:** The phylogenetic relationships of the drosomycin sequences from different species were reconstructed by maximum likelihood using PAUP\* v.4.0b10 (SWOFFORD 1998). We used the HKY85 model of sequence evolution and allowed gamma distributed rate heterogeneity between sites (HASEGAWA *et al.* 1985). The transition-to-transversion ratio and the shape of the gamma distribution were first estimated from a maximum parsimony tree. The phylogeny was then reconstructed using a heuristic search with nearest-neighbor interchanges. The robustness of the topology was assessed by 1000 nonparametric bootstrap replicates. The tree was rooted using the sequences from *D. ananassae*.

**The  $d_N/d_S$  ratio:** We tested the hypothesis that functionally distinct regions of the protein have different ratios of non-synonymous-to-synonymous substitutions ( $d_N/d_S$  ratios). This analysis used the sequences, alignment, and tree topology shown in Figures 2 and 3 (with the exclusion of *D. ananassae* sequence D). Therefore, we are looking at the divergence of both paralogs and orthologs. The analysis was performed using the Codeml program in the PAML v. 3.14 package (YANG 1997), which fits a maximum-likelihood model of codon substitution along the phylogenetic tree. First, we fitted a model under the null hypothesis that all the codons have the same  $d_N/d_S$  ratio. Then, using a likelihood-ratio test, we compared this to a model in which the signal peptide and mature peptide have different  $d_N/d_S$  ratios. Finally, this was compared to a three-ratio model in which the active site has a  $d_N/d_S$  ratio different from that of the rest of the mature peptide.

It is possible that positive selection could act on sites that cannot be predicted *a priori*. To detect such sites, we used a model that allows the  $d_N/d_S$  ratio to vary between codons (YANG *et al.* 2000). Our null model, M8A, forces the  $d_N/d_S$  ratio of all the codons to be  $<1$ . In this model, sites either have 1 of 10 different  $d_N/d_S$  ratios calculated from a beta distribution bounded by 0 and 1 or belong to an eleventh class in which the  $d_N/d_S$  ratio is fixed at 1 (SWANSON *et al.* 2003). This is then compared to model M8, in which the eleventh  $d_N/d_S$  ratio is

free to vary above 1 (*i.e.*, positive selection is allowed). In all the PAML analyses, statistical significance was assessed by comparing the likelihood-ratio statistic ( $2\Delta l$ ) to the  $\chi^2$  distribution, with the degrees of freedom being the number of additional parameters in the more complex model.

**Tests of neutrality:** The majority of population genetics analyses were performed using the program DnaSP (ROZAS and ROZAS 1999). The neutral distribution of Fay and Wu's  $H$ , Tajima's  $D$ , and Fu and Li's  $F$  and  $D$  statistics was obtained by 2000 coalescent simulations. The simulations were performed on the basis of the number of segregating sites ( $S$ ), the length of the sequence, and the per-site recombination parameter ( $C$ ). In *D. simulans* and *D. melanogaster*, there is no recombination in males, and therefore for autosomal genes  $C = 2Nc$  (where  $N$  is the effective population size and  $c$  is the crossing-over rate/base pair/generation in females). We assumed that in *D. simulans*  $N = 2 \times 10^6$  and that in *D. melanogaster*  $N = 10^6$  (ANDOLFATTO and PRZEORSKI 2000). We used the estimate of  $c$  ( $3.2 \times 10^{-8}$ ) for this region in *D. melanogaster* obtained by MARAIS *et al.* (2003) using the polynomial method of HEY and KLIMAN (2002). We conservatively assumed the same value of  $c$  for *D. simulans*. Therefore, in *D. simulans*  $C = 0.128$  and in *D. melanogaster*  $C = 0.064$ .

The HKA test was performed using the program HKA written by Jody Hey (<http://lifesci.rutgers.edu/~hey/lab>). The test statistics were compared to a neutral distribution generated from 10,000 coalescent simulations (HUDSON *et al.* 1987; WANG and HEY 1996).

## RESULTS

**Origins of the gene family:** We began by reconstructing the patterns of gene duplication and loss leading to today's drosomycin gene family. We found six drosomycin genes in *D. yakuba*, seven in *D. simulans*, *D. melanogaster*, and *D. erecta*, and four in *D. ananassae*. The fourth *D. ananassae* gene (D in Figure 2) could not be aligned at the 3'-end and lacked the disulphide bridges that are essential for maintaining the tertiary structure of the peptide. Therefore, this gene was omitted from further analyses. The phylogenetic tree of these sequences shows that the *D. ananassae* sequences form a single monophyletic group, suggesting that they diverged after the split from the *erecta/yakuba/melanogaster/simulans* lineage (Figure 3). Similarly, assuming that the tree is rooted with the *D. ananassae* sequences, the seven *D. melanogaster* drosomycins duplicated after the split from *D. ananassae* (18–30 million years ago; POWELL 1997) but before *D. erecta* split from the *simulans/melanogaster* lineage (6–15 million years ago; POWELL 1997). The *Dro3* gene either has been lost from *D. yakuba* or is simply missing from the currently available sequence. The relationship of the seven drosomycin genes roughly mirrors their arrangement in the genome, with related genes being nearby in the genome (Figures 1 and 3).

The phylogenetic relationships of the *melanogaster* subgroup have been the subject of much debate (KO *et al.* 2003). Some studies have grouped *D. yakuba* into a clade containing *D. melanogaster* and *D. simulans* but not *D. erecta*. Other work groups *D. yakuba* with *D. erecta*. Our data clearly support the latter tree topology, as the



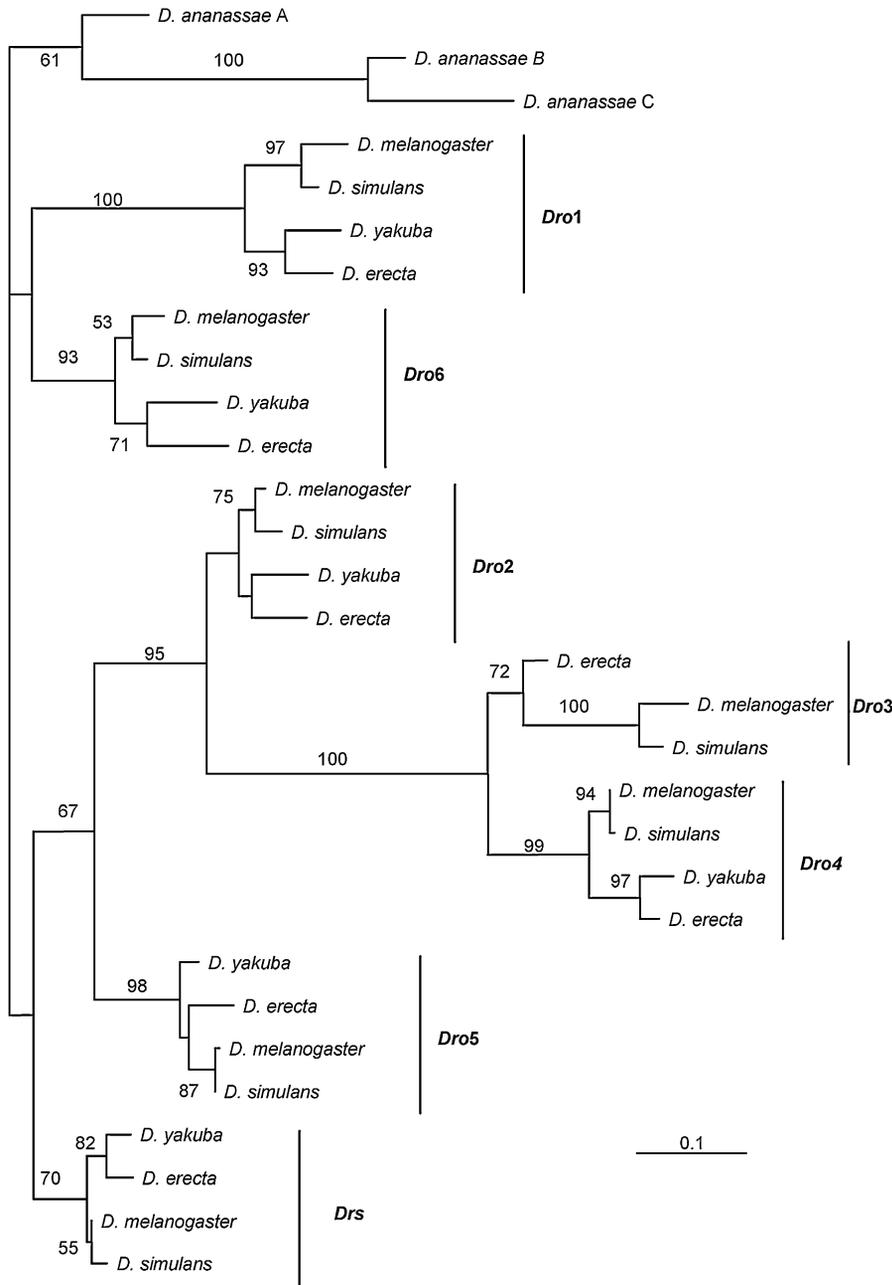


FIGURE 3.—Phylogeny of the drosomycin genes. Percentage of bootstrap support (if >50) is given next to nodes.

that have been copied from upstream of *Dro2* to upstream of *Dro5*. This rearrangement consists of a 182- to 258-bp deletion ~644 bp before the *Dro5* gene in *D. simulans* relative to *D. melanogaster*. The original length of this sequence has been largely restored by an ~226-bp insertion just 42 bp before the deletion. This inserted sequence has been copied from the region ending 6 bp before the *Dro2* coding sequence. Unfortunately, it is impossible to align this region with an outgroup, making the direction (insertion *vs.* deletion) uncertain. This same region also has many indels segregating within populations, including a 604- to 680-bp deletion in one of the *D. melanogaster* chromosomes.

We found three polymorphisms in natural populations that probably result in nonfunctional proteins.

First, 4 of 20 *D. simulans* *Dro2* alleles contain a 2-bp deletion that introduces a frameshift into the mature peptide. Second, 1 of 20 *D. melanogaster* *Dro1* alleles contains an internal stop codon. Finally, 2 of 20 *D. melanogaster* *Dro1* alleles have a mutation that results in the loss of a disulphide bridge. The presence of several null alleles suggests that there is not strong selection maintaining the function of these loci. It is notable that null alleles have been recorded in attacins and class C scavenger receptors in *D. melanogaster* (HEDENGREN *et al.* 2000; LAZZARO and CLARK 2001; LAZZARO 2005).

**Amino acid divergence of drosomycin genes:** If hosts and parasites are coevolving, this can drive the rapid divergence of amino acid sequences. Conversely, stabilizing selection may conserve amino acids that are

**TABLE 1**  
**Estimated  $d_N/d_S$  ratios for different regions of the drosomycin protein**

Model	Class sites	No. of codons	$d_N/d_S$	Likelihood
One ratio	All	68	0.1460	-2308.71
Two ratios	Signal	24	0.2303	-2235.62
	Mature	44	0.1033	
Three ratios	Signal	24	0.2302	-2205.28
	Mature excluding active site	36	0.1191	
	Active site	8	0.4858	

The one-ratio model fits a single ratio for all sites. The two-ratio model fits separate ratios for the signal *vs.* mature peptide. The three-ratio model estimates separate ratios for the signal peptide, the active site, and the remaining sites in the mature peptide.

essential for the function or structural integrity of the molecule. In this section we have compared the patterns of divergence between different regions of the drosomycin molecule and inferred the selection pressures acting on them.

In species where natural selection is driving rapid divergence of AMPs, the mature peptide is much less conserved than the signal peptide (DUDA *et al.* 2002). However, in the amino acid alignment of drosomycin, it is clear that the signal peptide is slightly more variable than the mature peptide (Figure 2). We can test whether different selection pressures act on the two protein domains by comparing different models of codon substitution with a likelihood-ratio test (Table 1). The model in which the signal and mature peptides have different  $d_N/d_S$  ratios is a significantly better fit to the data than the model in which they have the same ratio (Table 1; one ratio *vs.* two ratio:  $2\Delta l = 73.1$ , d.f. = 1,  $P < 0.001$ ). The mature peptide has the lower  $d_N/d_S$  ratio, indicating that it is under stronger stabilizing selection than the signal peptide.

The structure of drosomycin, in common with most other AMPs, includes a cluster of hydrophobic amino acids that probably interact directly with the fungal cell membrane (Figure 2) (LANDON *et al.* 2000). This structure, combined with comparisons to better-characterized AMPs, has been used to identify the probable active site of the molecule (Figure 2) (LANDON *et al.* 2000). If drosomycin is coevolving with pathogens, then the active site might be a target of selection. Inspection of Figure 2 does show several mutations in the active site, including mutations from hydrophobic to hydrophilic amino acids. Furthermore, the  $d_N/d_S$  ratio in the active site is significantly higher than in the rest of the mature peptide (Table 1; two ratio *vs.* three ratio:  $2\Delta l = 30.3$ , d.f. = 1,  $P < 0.001$ ). However, this ratio is still  $< 1$  ( $d_N/d_S =$

0.49), and therefore this difference can be explained by either relaxed selective constraints or positive selection on the active site.

Next we tested whether positive selection acts on amino acids scattered elsewhere in the drosomycin molecule. Including a fraction of positively selected sites in the model of codon substitution did not alter the likelihood score (M8A *vs.* M8:  $2\Delta l = 0.00$ , d.f. = 1, NS). Therefore, the variation in the rate of amino acid substitution in different regions of the protein appears to be due to variation in selective constraints rather than to positive selection.

Finally, we looked at the evolution of cysteine residues. The drosomycin molecule contains four disulphide bridges between eight cysteine residues that stabilize the structure of the molecule (MICHAUT *et al.* 1996; LANDON *et al.* 1997). These cysteines have been conserved in the same location in all the drosomycin genes across all five species (Figure 2), confirming that they are essential to the function of drosomycin. The only exception is "sequence D" in *D. ananassae*, which has lost three of the four disulphide bridges. It is therefore unlikely that this gene produces a functional antifungal peptide.

**Nucleotide polymorphism:** In both *D. melanogaster* and *D. simulans*, the African population tends to be more diverse than the European one, but the differences are small and not consistent (Table 2). This is typical of the pattern observed at other autosomal loci in *D. melanogaster*, but the difference between continents is normally larger in *D. simulans* (BEGUN and AQUADRO 1993; HAMBLIN and VEUILLE 1999; ANDOLFATTO 2001).

The diversity of coding sequence is substantially higher in *D. simulans* than in *D. melanogaster*, which is similar to the pattern reported for other genes (MORIYAMA and POWELL 1996; ANDOLFATTO 2001). The higher diversity in *D. simulans* is entirely due to a larger number of synonymous polymorphisms, as the two species have similar numbers of nonsynonymous polymorphisms (Tables 3 and 4). Therefore the two species have different ratios of synonymous-to-nonsynonymous segregating sites (Table 4; Europe:  $G = 5.41$ ,  $P = 0.02$ ; Africa:  $G = 0.95$ , NS).

The diversity of noncoding intergenic sequence is very similar in *D. simulans* and *D. melanogaster* (Tables 2-4). The ratio of intergenic-to-replacement polymorphisms is similar in the two species (Table 4; Europe:  $G = 0.06$ , NS; Africa:  $G = 0.94$ , NS). However, the ratio of synonymous (coding region) to intergenic polymorphisms is very different (Table 4; Europe:  $G = 11.01$ ,  $P = 0.001$ ; Africa:  $G = 6.95$ ,  $P = 0.01$ ). In summary, the two species have similar diversities at nonsynonymous sites and in intergenic regions, but *D. simulans* has a much higher diversity at synonymous sites.

The divergence between the two species also differs among the three classes of sites. As expected from the  $d_N/d_S$  analysis above, synonymous divergence is far greater

**TABLE 2**  
**The nucleotide diversity of different species, populations, and genes**

Gene/region	No. of sites						No. of segregating sites (S)						$\theta \times 10^{-2}$						$\pi \times 10^{-2}$						$\pi \times 10^{-2}$					
	<i>melanogaster</i>			<i>simulans</i>			<i>melanogaster</i>			<i>simulans</i>			<i>melanogaster</i>			<i>simulans</i>			<i>melanogaster</i>			<i>simulans</i>			<i>melanogaster</i>			<i>simulans</i>		
	Afr	Eur	Afr	Eur	Afr	Eur	Afr	Eur	Afr	Eur	Afr	Eur	Afr	Eur	Afr	Eur	Afr	Eur	Afr	Eur	Afr	Eur	Afr	Eur	Afr	Eur	Afr	Eur		
<i>Dro2</i> upstream	701	699	754	774	22	16	21	17	1.21	0.76	0.99	0.78	1.21	0.69	0.87	0.78														
<i>Dro3</i> upstream	446	446	298	347	18	22	15	16	1.56	1.63	1.78	1.63	1.22	1.74	1.32	2.05														
<i>Dro4</i> upstream	384	384	369	369	14	16	20	18	1.41	1.38	1.92	1.72	1.77	1.54	2.13	1.87														
<i>Dro5</i> upstream	851	898	1046	1041	44	57	40	30	2.11	2.10	1.35	1.02	2.33	2.39	1.08	1.03														
<i>Dro1</i> downstream	48	48	73	73	3	3	0	0	2.41	2.07	0.00	0.00	2.23	1.33	0.00	0.00														
<i>Dro1</i> upstream	355	353	352	354	11	10	9	8	1.20	0.94	0.91	0.80	1.20	1.10	1.03	0.74														
<i>Dro6</i> upstream	772	770	756	760	13	11	23	8	0.65	0.47	1.08	0.37	0.44	0.48	0.85	0.49														
<i>Drs</i> upstream	873	874	802	846	38	39	59	55	1.55	1.36	2.60	2.30	1.63	1.15	2.69	2.33														
<i>Drs</i> downstream	52	56	56	56	1	4	5	2	0.74	2.37	3.16	1.26	0.82	1.19	2.54	0.71														
All intergenic	3969	4536	4517	4631	139	175	192	154	1.35	1.28	1.50	1.18	1.35	1.30	1.39	1.24														
<i>Dro2</i>	294	305	313	313	5	4	7	5	0.66	0.43	1.13	0.68	0.53	0.31	1.14	0.87														
<i>Dro3</i>	213	213	213	213	3	6	8	6	0.54	0.93	1.33	1.00	0.50	1.02	1.39	1.34														
<i>Dro4</i>	213	213	213	213	3	1	2	2	0.54	0.16	0.33	0.33	0.35	0.08	0.19	0.34														
<i>Dro5</i>	189	189	189	189	3	3	3	2	0.61	0.53	0.56	0.37	0.57	0.63	0.32	0.38														
<i>Dro1</i>	207	207	210	210	4	2	6	3	0.75	0.32	1.01	0.51	0.62	0.34	1.05	0.84														
<i>Dro6</i>	215	215	215	215	2	3	3	4	0.36	0.46	0.49	0.66	0.45	0.68	0.35	0.84														
<i>Drs</i>	210	210	210	210	0	0	11	8	0.00	0.00	1.85	1.35	0.00	0.00	1.47	0.98														
All coding	1551	1562	1570	1570	21	20	42	31	0.52	0.42	0.95	0.70	0.47	0.44	0.85	0.81														

Data split into Europe (Eur) and Africa (Afr).  $\theta$  per site was estimated from the number of segregating sites (S). Excludes sites with alignment gaps only if in sequences under consideration.

**TABLE 3**  
**Nucleotide diversity and divergence in coding and intergenic sequence**

Gene/region	$\pi \times 10^{-2}$				Divergence $\times 10^{-2}$	
	<i>melanogaster</i>		<i>simulans</i>		Africa	Europe
	Africa	Europe	Africa	Europe		
Intergenic	1.42	1.17	1.22	1.09	7.09	7.20
Synonymous	0.94	0.80	2.58	2.72	9.21	8.99
Nonsynonymous	0.27	0.30	0.28	0.23	1.25	1.30

Sites with gaps anywhere in the alignment were excluded, resulting in slight differences from Table 2. The mean number pairwise substitutions between the two species were calculated using the method of NEI and GOJOBORI (1986).

than nonsynonymous divergence (Table 3). Interestingly, the intergenic divergence is also lower than the synonymous divergence, suggesting that intergenic regions have either lower mutation rates or higher functional constraints (Table 3).

**Frequency of polymorphic sites:** A selective sweep or population expansion will result in an increase in the number of low-frequency polymorphisms relative to those at intermediate frequency. This can be detected using Tajima's test statistic  $D$ , which is negative if there is an excess of rare mutations and positive if there is an excess of intermediate-frequency mutations (TAJIMA 1989). In our data set there are 10 significantly negative estimates of  $D$  and 1 significantly positive estimate (Table 5). Although it is clear from this that our data do not fit the neutral equilibrium model, the multiple tests performed make it difficult to identify specific loci that might be under selection. We also performed Fu and Li's  $D$  and  $F$  tests with an outgroup (FU and LI 1993) by comparing the number of mutations on terminal and internal branches of the genealogy. These tests produced qualitatively similar results to Tajima's  $D$  (data not shown).

Genetic hitchhiking during a selective sweep results in an excess of derived variants at high frequency (FAY and WU 2000). This effect occurs at some distance away from the selected site, as recombination must have partially broken down the association with the actual site under selection. Furthermore, the effect is relatively short

lived. We found an excess of high-frequency-derived variants in the intergenic sequence just downstream of *Drs* in the European population of *D. melanogaster* (Table 5; Fay and Wu's  $H = -6.67$ ;  $P < 0.00001$ ). This result is still highly significant after correction for multiple tests and when different rates of recombination are assumed (data not shown). It stems from the fact that all four ancestral-state sites segregating in this region are singletons, all of which are found in a single haplotype. Therefore, this could reflect a single recombination event during a selective sweep. It is notable that only one other estimate of  $H$  is marginally significantly different from the neutral distribution. Furthermore, this statistic is thought to be less sensitive to demographic processes than Tajima's  $D$ . Therefore, it is possible that there has been a recent selective sweep in this region. Interestingly, there are no fixed differences between *Drs* in *D. simulans* and *D. melanogaster*. Therefore, the data cannot be explained by positive selection on the *Drs* locus fixing nonsynonymous mutations.

**Polymorphism and divergence:** Under the neutral model, the ratio of synonymous-to-nonsynonymous polymorphisms will be the same as the ratio of synonymous-to-nonsynonymous differences between species. However, positive selection can increase the number nonsynonymous differences between species due to the fixation of favorable mutations. This can be detected using a McDonald-Kreitman test, which simply compares the two ratios in a  $2 \times 2$  contingency table using Fisher's exact test (MCDONALD and KREITMAN 1991). The McDonald-Kreitman test was not significant for any of the genes individually or for the combined data set of all genes (data in Table 6, all comparisons not significant). The same was true when the sites in intergenic regions were included as synonymous sites. If there are slightly deleterious amino acid polymorphisms that contribute to polymorphism but not to divergence, this can reduce the power of this test. Therefore, we removed singletons from the data set and repeated the test, but the result was still nonsignificant (data not shown).

Selective sweeps will reduce the genetic diversity of a gene, and balancing selection may increase it. It is possible to detect these effects as neutral theory predicts

**TABLE 4**

**Intergenic, synonymous, and replacement segregating mutations**

Population	Species	Coding		
		Synonymous	Replacement	Intergenic
Europe	<i>simulans</i>	24	7	105
Europe	<i>melanogaster</i>	8	10	132
Africa	<i>simulans</i>	27	14	136
Africa	<i>melanogaster</i>	10	9	134

Sites with gaps anywhere in the data set were excluded.

**TABLE 5**  
**Tests of neutrality based on the frequency of mutations**

Gene/region	Tajima's <i>D</i>				Fay and Wu's <i>H</i>			
	<i>melanogaster</i>		<i>simulans</i>		<i>melanogaster</i>		<i>simulans</i>	
	Africa	Europe	Africa	Europe	Africa	Europe	Africa	Europe
<i>Dro2</i> upstream	-0.24	-0.64	-0.55	0.04	2.29	-2.00	2.13	-0.53
<i>Dro2</i>	-0.84	-1.02	0.03	1.15	-0.86	-1.03	-1.96	-0.18
<i>Dro3</i> upstream	-1.13*	0.29	-1.19*	1.19*	-1.29	-1.39	-0.71	-0.09
<i>Dro3</i>	-0.30	0.35	0.19	1.41	-3.00	-1.67	1.24	-0.89
<i>Dro4</i> upstream	0.88	-0.05	0.52	0.40	-0.71	-1.73	0.36	1.16
<i>Dro4</i>	-1.45	-1.14*	-1.40*	0.12	0.64	0.15	0.36	0.36
<i>Dro5</i> upstream	0.58	0.46	-0.96**	0.03	0.00	1.12	-1.07	-1.24
<i>Dro5</i>	-0.30	0.67	-1.56*	0.02	0.43	0.76	0.53	0.53
<i>Dro6</i> upstream	-1.60**	0.03	-1.01*	1.38	-2.50	-0.67	1.24	-1.42
<i>Dro6</i>	0.93	1.57	-1.03	1.11	0.29	0.42	0.62	0.53
<i>Dro1</i> upstream	-0.41	0.69	0.61	-0.35	0.36	-0.45	-0.53	0.98
<i>Dro1</i>	-0.73	0.22	0.15	0.93	-0.07	-1.52	0.18	0.36
<i>Dro1</i> downstream	-0.30	-1.18			0.43	0.55		
<i>Drs</i> upstream	0.28	-0.71	-0.16	-0.11	3.64	-3.79	1.16	-2.13
<i>Drs</i>			-1.23	-1.16			2.49*	1.69
<i>Drs</i> downstream	0.33	-1.75*	-0.78	-1.40*	-0.86	-6.67***	0.18	-1.42

Sites with alignment gaps in the same species and population are excluded. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.00001$ .

that both interspecific divergence and intraspecific polymorphism are proportional to the neutral mutation rate and therefore positively correlated across loci. Deviations from this prediction can be detected in a contingency table (where the table rows are loci and the table columns are polymorphism and divergence) using an HKA test (HUDSON *et al.* 1987). In our case, this test is conservative as it makes the assumption that there is free recombination between loci and no recombination within loci, which clearly does not apply to our data.

We have used the HKA test to compare polymorphism and divergence across all our coding and intergenic sequences (Table 7). The sum of deviations was signif-

icantly greater than expected in *D. simulans* but not in *D. melanogaster* (*D. simulans*: sum deviations = 24.3,  $P = 0.01$ ; *D. melanogaster*: sum deviations = 9.1, NS). If only a single locus has been under selection, it is more powerful to test whether the maximum standardized discrepancy (MSD) for any one of the loci is greater than expected (WANG and HEY 1996). This approach also has the advantage of identifying specific loci that differ from the rest of the data set. This test was significant only in *D. simulans*, due to the *Drs* locus (MSD = 5.43, *Drs* locus,  $P = 0.003$ ). The unusual pattern of polymorphism in *Drs* can be clearly seen in Tables 2 and 7. This gene has both the lowest interspecific divergence of all the loci and the

**TABLE 6**  
**Polymorphism and divergence at synonymous and nonsynonymous sites**

Gene	Polymorphism				Fixed differences	
	<i>D. melanogaster</i>		<i>D. simulans</i>		Synonymous	Nonsynonymous
	Synonymous	Nonsynonymous	Synonymous	Nonsynonymous		
<i>Dro2</i>	2	2	5	3	8	0
<i>Dro3</i>	1	5	5	3	3	1
<i>Dro4</i>	3	0	1	2	1	0
<i>Dro5</i>	1	2	4	0	1	0
<i>Dro6</i>	3	0	2	2	4	3
<i>Dro1</i>	1	3	5	2	4	5
<i>Drs</i>	0	0	9	4	0	0
All coding sequence	11	12	31	16	21	9
All coding and intergenic	192	12	251	16	183	9

TABLE 7

Polymorphic sites and mean interspecific divergence of genes and intergenic sequence

Gene	Length	Segregating sites		Mean interspecific divergence
		<i>melanogaster</i>	<i>simulans</i>	
<i>Dro2</i> upstream	677	24	18	31.19
<i>Dro2</i>	285	4	8	11
<i>Dro3</i> upstream	287	18	17	14.48
<i>Dro3</i>	213	6	8	8.9
<i>Dro4</i> upstream	369	19	22	34.25
<i>Dro4</i>	213	3	3	1.55
<i>Dro5</i> upstream	324	26	9	57.81
<i>Dro5</i>	189	3	4	2.09
<i>Dro6</i> upstream	748	20	24	37.35
<i>Dro6</i>	216	3	4	8.4
<i>Dro1</i> upstream	346	16	10	30.1
<i>Dro1</i>	210	5	6	11.45
<i>Dro1</i> downstream	47	3	0	3.5
<i>Drs</i> upstream	743	41	53	62.09
<i>Drs</i>	210	0	12	1.4
<i>Drs</i> downstream	52	1	5	6.95

Sites with alignment gaps in either species were excluded. Data set included all 20 alleles of both species.

lowest intraspecific diversity ( $\theta$ ) in *D. melanogaster*, and yet in *D. simulans*  $\theta$  is the highest of all the drosomycin genes.

The polymorphic sites in the *D. simulans Drs* sequences are shown in Figure 4. They include four nonsynonymous mutations and eight synonymous mutations. The nonsynonymous mutations are all in the signal peptide. Interestingly, the most common haplotype is identical to that found in *D. melanogaster*. Therefore, the derived mutations tend to be at a lower frequency than expected under neutrality, resulting in a significantly positive value of Fay and Wu's *H* in the African population (Table 5). Overall, the segregating mutations tend to be found at a low frequency, as reflected by negative estimates of Tajima's *D* (Table 5, NS).

DISCUSSION

**Adaptive protein evolution:** The protein sequences of antimicrobial peptides in vertebrates and termites have extremely high rates of evolution, driven by long-term positive selection (see Introduction). This is in stark contrast to the drosomycin genes, whose amino acid sequence is principally under purifying selection. Although not all the drosomycins appear to evolve neutrally, it is clear that their rate of adaptive evolution is far lower than that of AMPs in other species. This pattern is similar to that reported for antibacterial peptides in *Drosophila*, which also tend to have conserved amino acid sequences (see Introduction).

Why do AMPs show such differing patterns of evolution? One hypothesis is that there are differences

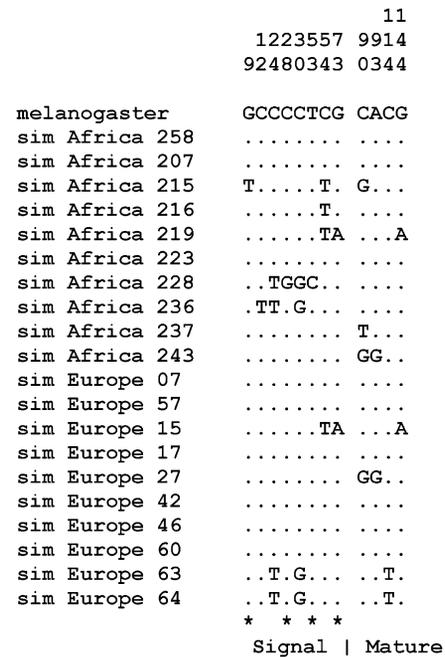


FIGURE 4.—Polymorphic sites in *Drs* from *D. simulans*. Nonsynonymous sites are marked as (\*).

intrinsic to the peptides themselves that alter their mode of evolution. For example, certain AMPs could have strong structural constraints, or they may target pathogen molecules that are highly conserved. However, there are now data from several structurally unrelated peptides in vertebrates and *Drosophila*, and consistent differences have emerged between AMP evolution in the two groups of hosts. Although more comparisons are needed, it currently seems likely that it is the ecology or physiology of the species that explains the different modes of evolution.

One explanation is that vertebrates and termites co-evolve with specialist bacterial and fungal pathogens, but *Drosophila* does not. It is striking that no specialist fungal or bacterial pathogens of *D. melanogaster* have been described, despite this being one of the most studied of all species. If such pathogens do not exist, it may be related to the ecology of *D. melanogaster* and related species. They have short life spans and live on ephemeral food patches. Within each food patch there are unlikely to be many overlapping generations before dispersal to a new food patch, making it difficult for a pathogen to persist by transmission solely between individuals of *D. melanogaster*. Therefore, it is possible that most *D. melanogaster* pathogens are generalists infecting many species, which in turn may restrict the opportunity for coevolution. This contrasts with vertebrates and termite colonies, which are long lived with overlapping generations in the same habitat patch, facilitating the spread of specialist pathogens. This hypothesis could be tested by a comparative analysis of AMP evolution across taxa with differing ecologies.

This hypothesis is not inconsistent with the observation of rapid evolution in other components of the

*Drosophila* immune system (BEGUN and WHITLEY 2000; SCHLENKE and BEGUN 2003; LAZZARO 2005). First, many of these genes may also be involved in defenses against specialist parasites such as parasitoids or viruses. For example, the Toll pathway is important in antiviral as well as antibacterial defenses (CONTAMINE *et al.* 1989; ZAMBON *et al.* 2005). Also, many of these genes are involved in signaling pathways, and selection may be acting to alter patterns and the magnitude of immune responses rather than adapting to novel parasite genotypes.

The one notable exception to the pattern above is a peptide called andropin in *D. melanogaster*, which evolves rapidly under positive selection (DATE-ITO *et al.* 2002). This peptide, which is related to the AMP cecropin and has antibacterial activity, is unusual in being expressed in the male ejaculatory duct and transferred to females during mating (SAMAKOVLIS *et al.* 1991; LUNG *et al.* 2001). Proteins transferred from males to females often evolve very rapidly, and it is therefore unclear whether the rapid evolution is associated with its antibacterial properties or with some other postmating function (SWANSON *et al.* 2001). If andropin does coevolve with microbes, infections of the reproductive tract might be *D. melanogaster* specialists if they are sexually transmitted.

**Selection on *Drs*:** Although we did not find rapid adaptive protein evolution, there was some indication of recent selection at or near *Drs* in both species. In *D. melanogaster*, there was an excess of high-frequency-derived mutations just downstream of *Drs*. This could result from a selective sweep at linked sites, such as the *Drs* coding sequence or regulatory elements. A selective sweep could also explain the low diversity of the *Drs* coding sequence. However, there are no fixed differences between the *Drs* coding sequence of the two species, suggesting that selection favoring beneficial changes to the amino acid sequence is unlikely. This result was also significant only in Europe, and we are unable to exclude the possibility that it is an artifact of some demographic process.

In *D. simulans*, an HKA test showed that *Drs* had an anomalous polymorphism-to-divergence ratio. This was the result of *Drs* having both the lowest interspecific divergence (the most common allele is identical to *D. melanogaster*) and the highest intraspecific polymorphism of any gene. The first explanation of this data is that balancing selection has increased the diversity at this locus. This is consistent with the flanking intergenic sequence also having fairly high levels of polymorphism (Table 2). However, the high polymorphism is due to low-frequency-derived mutations that have occurred on multiple different haplotypes, and there are no non-synonymous polymorphisms in the mature peptide. It seems unlikely that such a pattern would arise from balancing selection. This suggests that some recent change has led to the accumulation of variation. This could be a relaxation of selective constraints, a local increase in the mutation rate, gene conversion, or recent

positive selection increasing the frequency of certain mutations.

An alternative hypothesis is that the *Drs* locus has introgressed between the two species, resulting in reduced interspecific divergence. A similar process has been postulated in the *Cecropin* gene region of *D. melanogaster* (DATE *et al.* 1998). In our case, this would have to be coupled to a selective sweep reducing the diversity of *Drs* in *D. melanogaster*, as in this species the polymorphism-to-divergence ratio is normal. This is possible, given the evidence for selection near *Drs* (see above) and might be expected if selection favored the introgressing allele.

The observation that the *D. melanogaster* *Drs* sequence is identical to the most common *D. simulans* allele is certainly unusual and is compatible with the introgression hypothesis. Introgression could potentially introduce derived *D. simulans* polymorphisms into *D. melanogaster*. However, using *D. yakuba* as an outgroup, we found that the shared *melanogaster/simulans* allele is ancestral to all the other *simulans* alleles. Introgression might also be expected to involve some of the flanking intergenic sequence. However, the interspecific divergence of the flanking sequence was typical elsewhere in the drosomycin region (sliding-window analysis; data not shown). Instead, the short length of the *Drs* gene means that the absence of fixed differences between species may simply have occurred by chance. Assuming that the occurrence of fixed differences between the species follows a Poisson distribution, two of the other drosomycin loci have divergences that fall within the 95% confidence interval of divergence at *Drs*. Similarly, the upper 95% confidence limit on *K*'s between the most common *D. simulans* allele and *D. melanogaster* is 0.079, which is only slightly lower than the divergence estimated at the other loci. In conclusion, although the *Drs* locus does not fit the neutral model, the causes of the unusual patterns of polymorphism and divergence remain unclear.

#### **Polymorphism in coding and intergenic sequences:**

The data that we collected are ideal for comparing sequence evolution in *D. simulans* and *D. melanogaster*, as we sequenced similar numbers of alleles from a single genomic region that encompassed both coding and intergenic sequence. The two species have similar non-synonymous diversities, but *D. simulans* has much greater synonymous site diversity. Similar patterns have been observed previously and been attributed to *D. simulans* having a larger population size than *D. melanogaster* (MORIYAMA and POWELL 1996; ANDOLFATTO 2001). This is expected to increase the diversity of neutrally evolving synonymous sites. However, if nonsynonymous mutations in drosomycin tend to be slightly deleterious, they will be removed more effectively from *D. simulans* than from *D. melanogaster*.

The intergenic sequence has diverged less between species than between synonymous sites. This has been

observed before and has been attributed to either higher mutation rates in transcribed regions or to selective constraints on intergenic regions (HALLIGAN *et al.* 2004; KERN and BEGUN 2005). The intergenic sequence has similar diversities in the two species (*i.e.*, it behaves most like the nonsynonymous sites). This is the result of the nucleotide diversity of intergenic regions being markedly lower than the diversity of synonymous sites in *D. simulans*, while in *D. melanogaster* synonymous and intergenic diversities are more similar. This difference in diversity cannot be easily explained by higher mutation rates in transcribed regions, as this would increase the synonymous diversity but not alter the intergenic-to-synonymous polymorphism ratio in the two different species. The pattern can be explained, however, if polymorphisms in the intergenic sequence tend to be slightly deleterious and are therefore removed by selection more efficiently in *D. simulans* (KERN and BEGUN 2005). Similar observations were reported by KERN and BEGUN (2005), who found similar intergenic but different synonymous diversities in the two species. However, when they compared African *D. melanogaster* with a sample of *D. simulans* of mixed origin they found similar synonymous diversities in the two species. In our data we have compared both ancestral (African) and derived (European) populations of both species and found that the original pattern holds in both cases.

The isofemale lines were kindly supplied by Penny Haddrill. This work was funded by a Wellcome Trust Research Career Development Fellowship.

#### LITERATURE CITED

- ANDOLFATTO, P., 2001 Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. *Mol. Biol. Evol.* **18**: 279–290.
- ANDOLFATTO, P., and M. PRZEWORSKI, 2000 A genome-wide departure from the standard neutral model in natural populations of *Drosophila*. *Genetics* **156**: 257–268.
- ANTCHEVA, N., M. BONIOTTO, I. ZELEZETSKY, S. PACOR, M. V. V. FALZACAPPA *et al.*, 2004 Effects of positively selected sequence variations in human and *Macaca fascicularis* beta-defensins 2 on antimicrobial activity. *Antimicrob. Agents Chemother.* **48**: 685–688.
- BEGUN, D. J., and C. F. AQUADRO, 1993 African and North-American populations of *Drosophila melanogaster* are very different at the DNA level. *Nature* **365**: 548–550.
- BEGUN, D. J., and P. WHITLEY, 2000 Adaptive evolution of relish, a *Drosophila* NF-kappaB/IkappaB protein. *Genetics* **154**: 1231–1238.
- BULMER, M. S., and R. H. CROZIER, 2004 Duplication and diversifying selection among termite antifungal peptides. *Mol. Biol. Evol.* **21**: 2256–2264.
- CLARK, A. G., and L. WANG, 1997 Molecular population genetics of *Drosophila* immune system genes. *Genetics* **147**: 713–724.
- CONTAMINE, D., A. M. PETITJEAN and M. ASHBURNER, 1989 Genetic-resistance to viral-infection: the molecular cloning of a *Drosophila* gene that restricts infection by the rhabdovirus sigma. *Genetics* **123**: 525–533.
- 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.
- DATE-ITO, A., K. KASAHARA, H. SAWAI and S. I. CHIGUSA, 2002 Rapid evolution of the male-specific antibacterial protein andropin gene in *Drosophila*. *J. Mol. Evol.* **54**: 665–670.
- DE GREGORIO, E., P. T. SPELLMAN, G. M. RUBIN and B. LEMAITRE, 2001 Genome-wide analysis of the *Drosophila* immune response by using oligonucleotide microarrays. *Proc. Natl. Acad. Sci. USA* **98**: 12590–12595.
- DUDA, T. F., JR., D. VANHOYE and P. NICOLAS, 2002 Roles of diversifying selection and coordinated evolution in the evolution of amphibian antimicrobial peptides. *Mol. Biol. Evol.* **19**: 858–864.
- EBERT, D., 1994 Virulence and local adaptation of a horizontally transmitted parasite. *Science* **265**: 1084–1086.
- ESCALANTE, A. A., A. A. LAL and F. J. AYALA, 1998 Genetic polymorphism and natural selection in the malaria parasite *Plasmodium falciparum*. *Genetics* **149**: 189–202.
- FAY, J. C., and C.-I. WU, 2000 Hitchhiking under positive Darwinian selection. *Genetics* **155**: 1405–1413.
- FEHLBAUM, P., P. BULET, L. MICHAUT, M. LAGUEUX, W. BROEKAERT *et al.*, 1994 Insect immunity. Septic injury of *Drosophila* induces the synthesis of a potent antifungal peptide with sequence homology to plant antifungal peptides. *J. Biol. Chem.* **269**: 33159–33163.
- FU, Y.-X., and W.-H. LI, 1993 Statistical tests of neutrality of mutations. *Genetics* **133**: 693–709.
- HALLIGAN, D. L., A. EYRE-WALKER, P. ANDOLFATTO and P. D. KEIGHTLEY, 2004 Patterns of evolutionary constraints in intronic and intergenic DNA of *Drosophila*. *Genome Res.* **14**: 273–279.
- HAMBLIN, M. T., and M. VEUILLE, 1999 Population structure among African and derived populations of *Drosophila simulans*: evidence for ancient subdivision and recent admixture. *Genetics* **153**: 305–317.
- HASEGAWA, M., H. KISHINO and T. YANO, 1985 Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **21**: 160–174.
- HEDENGRÉN, M., K. BORGE and D. HULTMARK, 2000 Expression and evolution of the *Drosophila* attacin/diptericin gene family. **279**: 574.
- HEY, J., and R. M. KLIMAN, 2002 Interactions between natural selection, recombination and gene density in the genes of *Drosophila*. *Genetics* **160**: 595–608.
- HUDSON, R. R., M. KREITMAN and M. AQUADE, 1987 A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**: 153–159.
- 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. YEAGER, 1997 Coordinated amino acid changes in the evolution of mammalian defensins. *J. Mol. Evol.* **44**: 675–682.
- JIGGINS, F. M., and G. D. D. HURST, 2003 The evolution of parasite recognition genes in the innate immune system: purifying selection on *Drosophila melanogaster* peptidoglycan recognition proteins. *J. Mol. Evol.* **57**: 598–605.
- KERN, A. D., and D. J. BEGUN, 2005 Patterns of polymorphism and divergence from noncoding sequences of *Drosophila melanogaster* and *D. simulans*: evidence for nonequilibrium processes. *Mol. Biol. Evol.* **22**: 51–62.
- KO, W. Y., R. M. DAVID and H. AKASHI, 2003 Molecular phylogeny of the *Drosophila melanogaster* species subgroup. *J. Mol. Evol.* **57**: 562–573.
- LANDON, C., P. SODANO, C. HETRU, J. HOFFMANN and M. PTAK, 1997 Solution structure of drosomycin, the first inducible antifungal protein from insects. *Protein Sci.* **6**: 1878–1884.
- LANDON, C., A. PAJON, F. VOVELLE and P. SODANO, 2000 The active site of drosomycin, a small insect antifungal protein, delineated by comparison with the modeled structure of Rs-AFP2, a plant antifungal protein. *J. Pept. Res.* **56**: 231–238.
- LAZZARO, B. P., 2005 Elevated polymorphism and divergence in the class C scavenger receptors of *Drosophila melanogaster* and *D. simulans*. *Genetics* **169**: 2023–2034.
- LAZZARO, B. P., and A. G. CLARK, 2001 Evidence for recurrent paralogous gene conversion and exceptional allelic divergence in the Attacin genes of *Drosophila melanogaster*. *Genetics* **159**: 659–671.
- LAZZARO, B. P., and A. G. CLARK, 2003 Molecular population genetics of inducible antibacterial peptide genes in *Drosophila melanogaster*. *Mol. Biol. Evol.* **20**: 914–923.
- LITTLE, T. J., and N. COBBE, 2005 The evolution of immune related genes from disease carrying mosquitoes: diversity in a peptidoglycan- and thioester-recognising protein. *Insect Mol. Biol.* (in press).

- LIVELY, C. M., and M. F. DYBDAHL, 2000 Parasite adaptation to locally common host genotypes. *Nature* **405**: 679–681.
- LUNG, O., L. KUO and M. F. WOLFNER, 2001 *Drosophila* males transfer antibacterial proteins from their accessory gland and ejaculatory duct to their mates. *Dev. Biol.* **47**: 617.
- LYNN, D. J., A. T. LLOYD, M. A. FARES and C. O'FARRELLY, 2004 Evidence of positively selected sites in mammalian alpha-defensins. *Mol. Biol. Evol.* **21**: 819–827.
- MARAIS, G., D. MOUCHIROUD and L. DURET, 2003 Neutral effect of recombination on base composition in *Drosophila*. *Genetical Res.* **81**: 79–87.
- MCDONALD, J. H., and M. KREITMAN, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**: 652–654.
- MEDZHITOV, R., and C. A. JANEWAY, 1997 Innate immunity: the virtues of a nonclonal system of recognition. *Cell* **91**: 295–298.
- MICHAUT, L., P. FEHLBAUM, M. MONIATTE, A. VANDORSSELAER, J. M. REICHHART *et al.*, 1996 Determination of the disulfide array of the first inducible antifungal peptide from insects: drosomycin from *Drosophila melanogaster*. *FEBS Lett.* **395**: 6–10.
- MORAND, S., S. D. MANNING and M. E. J. WOOLHOUSE, 1996 Parasite-host coevolution and geographic patterns of parasite infectivity and host susceptibility. *Proc. R. Soc. Lond. Ser. B Biol. Sci.* **263**: 119–128.
- MORIYAMA, E. N., and J. R. POWELL, 1996 Intraspecific nuclear DNA variation in *Drosophila*. *Mol. Biol. Evol.* **13**: 261–277.
- MORRISON, G. M., C. A. M. SEMPLE, F. M. KILANOWSKI, R. E. HILL and J. R. DORIN, 2003 Signal sequence conservation and mature peptide divergence within subgroups of the murine beta-defensin gene family. *Mol. Biol. Evol.* **20**: 460–470.
- NEL, M., and T. GOJOBORI, 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**: 418–426.
- POWELL, J. R., 1997 *Progress and Prospects in Evolutionary Biology: The Drosophila Model*. Oxford University Press, Oxford.
- RAMOS-ONSINS, S., and M. AGUADE, 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 version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**: 174–175.
- SAMAKOVLIS, C., P. KYLSTEN, D. A. KIMBRELL, A. ENGSTROM and D. HULTMARK, 1991 The andropin gene and its product, a male-specific antibacterial peptide in *Drosophila melanogaster*. *EMBO J.* **10**: 163–169.
- SAWYER, S., 1989 Statistical tests for detecting gene conversion. *Mol. Biol. Evol.* **6**: 526–538.
- SCHLENKE, T. A., and D. J. BEGUN, 2003 Natural selection drives *Drosophila* immune system evolution. *Genetics* **164**: 1471–1480.
- SEMPLE, C. A. M., M. ROLFE and J. R. DORIN, 2003 Duplication and selection in the evolution of primate beta-defensin genes. *Genome Biol.* **4**: R31.
- SHIMODAIRA, H., and M. HASEGAWA, 1999 Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol. Biol. Evol.* **16**: 1114–1116.
- SMITH, D. R., 2004 *Drosophila ananassae* and *D. erecta* Whole-Genome Shotgun Reads. Agencourt Bioscience, Beverly, MA.
- SWANSON, W. J., A. G. CLARK, H. M. WALDRIP-DAIL, M. F. WOLFNER and C. F. AQUADRO, 2001 Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **98**: 7375–7379.
- SWANSON, W. J., R. NIELSEN and Q. YANG, 2003 Pervasive adaptive evolution in mammalian fertilization proteins. *Mol. Biol. Evol.* **20**: 18–20.
- TAJIMA, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**: 585–595.
- WANG, R. L., and J. HEY, 1996 The speciation history of *Drosophila pseudoobscura* and close relatives: inferences from DNA sequence variation at the period locus. *Genetics* **144**: 1113–1126.
- WILSON, R. K., 2004 *D. yakuba* Whole-Genome Shotgun Sequence. Washington University Genome Sequencing Center, St. Louis.
- YANG, Z., 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput. Appl. Biosci.* **13**: 555–556.
- YANG, Z. H., R. NIELSEN, N. GOLDMAN and A. M. K. PEDERSEN, 2000 Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**: 431–449.
- ZAMBON, R. A., M. NANDAKUMAR, V. N. VAKHARIA and L. P. WU, 2005 The Toll pathway is important for an antiviral response in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **102**: 7257–7262.
- ZASLOFF, M., 2002 Antimicrobial peptides of multicellular organisms. *Nature* **415**: 389–395.

Communicating editor: D. BEGUN