Adaptive Evolution of Relish, a Drosophila NF-κB/IκB Protein

David J. Begun and Penn Whitley

Section of Integrative Biology and Institute for Cellular and Molecular Biology, University of Texas, Austin, Texas 78712

Manuscript received September 9, 1999
Accepted for publication November 12, 1999

ABSTRACT

NF-κB and IκB proteins have central roles in regulation of inflammation and innate immunity in mammals. Homologues of these proteins also play an important role in regulation of the Drosophila immune response. Here we present a molecular population genetic analysis of Relish, a Drosophila NF-κB/IκB protein, in *Drosophila simulans* and *D. melanogaster*. We find strong evidence for adaptive protein evolution in *D. simulans*, but not in *D. melanogaster*. The adaptive evolution appears to be restricted to the IκB domain. A possible explanation for these results is that Relish is a site of evolutionary conflict between flies and their microbial pathogens.

A possible consequence of host-pathogen interactions is an "arms race" resulting in rapid evolution; pathogens evolve to evade host defenses while host defenses evolve to circumvent such evasion (Levin and Lenski 1983). Proteins having a role in such an arms race might be expected to evolve quickly under the influence of natural selection. Insect cecropins are small antibacterial proteins that insert into bacterial cell walls, causing leakage and cell death (Kylsten *et al.* 1990; Durell *et al.* 1992). Therefore, one plausible arena for an arms race is the interaction between Drosophila cecropins and Drosophila pathogens. However, molecular evolutionary analysis of cecropins in *Drosophila melanogaster* and its close relatives provided no evidence for adaptive protein evolution between species (Clark and Wang 1997; Date *et al.* 1998; Ramos-Onsins and Aguadé 1998). Therefore, proteins in Drosophila that might be evolving as a direct or indirect result of selection pressures from microbial pathogens remain unknown.

Rel/NF-κB proteins and IκB proteins play an important role in vertebrate innate immunity and inflammation, and in regulation of the Drosophila immune response (Hoffmann and Reichhart 1997; Dushay and Eldon 1998; Ghosh *et al.* 1998). Rel/NF-κB domains function in dimerization and DNA binding. IκB domains are composed primarily of ankyrin repeats, which function in protein-protein interactions. These domains interact to control the subcellular localization of NF-κB (Ghosh *et al.* 1998). IκB proteins form a complex with NF-κB proteins, maintaining the latter in an inactive cytoplasmic form, probably through interaction with a nuclear localization signal (Baeuerle 1998; Ghosh *et al.* 1998; Huxford *et al.* 1998; Jacobs and Harrison 1998). Signal-dependent degradation of IκB results in unmasking of the nuclear localization signal and subsequent translocation of NF-κB to the nucleus, where it transcriptionally upregulates several genes. Thus, IκB proteins usually function as inhibitors of NF-κB activity. Rel proteins are found complexed with their IκB inhibitors in the cytoplasm of uninfected animals, thereby allowing initiation of signal-induced immune response in the absence of additional production of Rel proteins. Such a mechanism allows rapid induction of the immune response.

Drosophila Relish is an unusual member of the Rel family of proteins (Dushay *et al.* 1996), as it possesses both Rel/NF-κB domains and an inhibitory IκB domain (Figure 1). The mammalian p100 and p105 genes have a similar structure; however, in most cases these domains are found in different genes (Ghosh *et al.* 1998). Though IκB and NF-κB proteins are known to interact, there is no experimental evidence bearing on the question of whether the two functional domains of Relish participate in direct interactions with one another. *Relish* is transcriptionally upregulated in response to microbial infection (Dushay *et al.* 1996). Experiments done in Drosophila cell culture suggest that Relish transcriptionally upregulates the antibacterial gene, cecropinA1 (Dushay *et al.* 1996; it is not known if Relish can transcriptionally upregulate other antibacterial or antifungal genes). The IκB domain of Relish is hypothesized to belong to the γ subfamily of IκB proteins (Dushay *et al.* 1996), the specific functional properties of which are poorly known (Inoue *et al.* 1992; Ghosh *et al.* 1998). We report here the results of our molecular population genetic analysis of *Relish* in *D. simulans*, *D. melanogaster*, and *D. yakuba*.

MATERIAL AND METHODS

D. melanogaster alleles ($n = 6$) were sampled at random from homozygous chromosome III stocks made from isofemale lines from Zimbabwe. *D. simulans* alleles ($n = 7$) were sampled from a set of highly inbred lines made from individual females

Corresponding author: David Begun, Section of Integrative Biology, Patterson Labs C0930, University of Texas, Austin, TX 78712.
E-mail: djbegun@mail.utexas.edu

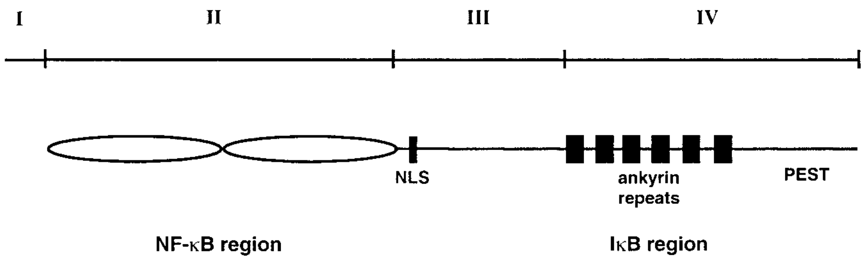

Figure 1.—Region I is nucleotides 1–253; region II is nucleotides 254–1449; region III is nucleotides 1450–1905; region IV is nucleotides 1906–2756. All coordinates refer to our GenBank entries. The NF-κB domains include bases 254–1449; the ankyrin repeats include bases 1906–2556. These regions correspond to bases 807–1700 (NF-κB) and 2151–2801 (ankyrin repeats) of the original *D. melanogaster* GenBank entry. NLS, nuclear localization signal; PEST, PEST domain.

caught at the Wolfskill Orchard, Winters, California in summer of 1995. A *D. yakuba* allele was isolated from an isofemale line obtained from the Drosophila Species Center at Bowling Green State University. The *Relish* region was amplified in two fragments. The first fragment was amplified using PCR primers cccggcggcaattcaccacac (forward560) and cccggcggcaattcaccacac (reverse1560); the second fragment was amplified using PCR primers gtgtgggaggcatacgcaaagttccg (forward1543) and gttgggttaaccagtagggcgtaagc (reverse3246). Numbering of primers refers to the most 3′ nucleotide of the primer using the coordinates of GenBank entry U62005. PCR products from *Relish* were directly sequenced using an ABI 377 automated sequencer. We analyzed 803 codons of *Relish* from all three species (the entire protein is 971 amino acids long in *D. melanogaster*). The region surveyed corresponds to bases 561–2999 of GenBank entry U62005. A 303-bp intron is located between bases 1271 and 1272 of the GenBank entry (which was derived from a cDNA). Sequences reported here can be found under GenBank accession nos. AF204277–AF204290. All sequences were easily aligned, with the exception of alignment of the *D. yakuba* intron with the intron of both *D. simulans* and *D. melanogaster*; none of the analyses presented here depend on proper alignment of this region. Analyses were carried out using the SITES (Hey and Wakeley 1997), DnaSP (Rozas and Rozas 1999), PAML (Yang 1999), and Molecular Evolutionary Analysis (E. Moriyama, unpublished results) programs. Codons harboring ambiguous bases were excluded from all analyses. Variable sites in exons were classified as replacement (nonsynonymous) or silent (synonymous). For some analyses, fixed differences in exons between the *D. simulans* and *D. melanogaster* samples were assigned to one lineage or the other under the parsimony criterion. *D. yakuba* was used as the outgroup in such analyses; only fixed differences at sites for which the *D. yakuba* sequence was identical to either *D. melanogaster* or *D. simulans* were used in these analyses (*e.g.*, sites at which each of the three species had a different base were excluded). Silent mutations were classified as preferred or unpreferred (Sharp and Lloyd 1993) through use of the outgroup method as described by Akashi (1996). We classified silent mutations that were from preferred to preferred codons or from unpreferred to unpreferred codons as "no change" mutations. Analyses of evolution in the three species lineages were also carried out by inferring the ancestral *Relish* sequence for the *D. simulans*/*D. melanogaster* pair using the baseml program of the PAML package. This hypothetical sequence was then compared to extant sequences from each of the three species.

RESULTS

Figure 2 and Tables 1–5 show summaries of variation at the *Relish* gene within and between *D. simulans* and *D. melanogaster*. Silent site heterozygosity at *Relish* in *D. simulans* is close to the average value for genes located in regions of normal recombination in this species (Moriyama and Powell 1996). Silent heterozygosity in *D. simulans* is about five times greater than silent heterozygosity in *D. melanogaster* (Table 1). The "average" gene is ~2.5 times more variable at silent sites in *D. simulans* than in *D. melanogaster* (Moriyama and Powell 1996). Given that *Relish* is not particularly polymorphic in *D. simulans* compared to other genes, there is some evidence that *D. melanogaster Relish* is less polymorphic than one would expect. Silent divergence between species is unremarkable, well within the range of values previously documented in this species pair (*e.g.*,

D. simulans

```
                                  1111111111122222222222222222222222
                 111333445777788899022333567890133444445666666666777
                1118046890347704713005599996656348145678012478889124
                5233923354467896294501958354397938854960102332584514

Site type       SRRSSSSSSSIIIIIIIIIISSSSSSSSSSSSSSSRSSSSSSSSRSSRSRSR

Sim1            GGGTTCCGCCGACGTCTACGGTTTGAGGCGGTATTAATCCGATCCGGGCGGC
Sim2            ..................................CGCCT.T..GG...A.A.
Sim3            A...CA...T.GA.C..GG...G.CGT.TAA.T.CG..T..G..GCA.A.A.
Sim4            ..................................CGCCT.T..GG...A.A.
Sim6            ..............................NC..CG..TA........ATAT
Sim7            ...A...ATTAG..CT.G.A.CG.C.........CGCCT.T..GG...A.A.
Sim8            .AC.C.T..T.G.T..AG..C.GCCGTA..A..CCGCCT...A....CATAT
```

D. melanogaster

```
                        112222222
                 1356789042344666
                43098979021211499
                43344770920508889

Site type       RRSSSIIIISRSSSR?R

Zim1            ATCAAATGCGCGCTGAG
Zim2            G.....C.....T.A..
Zim3            .CTGGTC.A..T..A..
Zim5            ..T...CT.....CANN
Zim7            ......C......CA..
Zim8            ......C..TT..CA.A
```

Figure 2.—S, R, and I refer to silent, replacement, and intron polymorphisms, respectively. Site 2067 in *D. simulans* and sites 2698 and 2699 in *D. melanogaster* (these two sites are the first and second positions of a single codon) were not included in the analyses because of ambiguous bases. Dots represent identity to the Sim1 allele and Zim1 allele in *D. simulans* and *D. melanogaster*, respectively. Coordinates are those of our GenBank entries. See Figure 1 legend for assignment of polymorphisms to different domains of the protein according to coordinates.

TABLE 1

Polymorphism and divergence per site, and Tajima's *D* statistics at the *Relish* gene of *D. simulans*, *D. melanogaster*, and *D. yakuba*

	θ		π		Tajima's *D*	
Polymorphism	Silent	Replacement	Silent	Replacement	Silent	Replacement
D. simulans	0.0255	0.0015	0.0221	0.0016	−0.775	0.173
D. melanogaster	0.0056	0.0009	0.0052	0.0007	−0.504	−1.295
Divergence			Silent	Replacement		
Lineage						
D. simulans			0.0364	0.0221		
D. melanogaster			0.0618	0.0291		
D. yakuba			0.2386	0.0427		
Pairwise						
sim. vs. mel.			0.0987	0.0519		
yakuba vs. sim.			0.2821	0.0618		
yakuba vs. mel.			0.3052	0.0676		

Lineage divergence estimates are pairwise differences per site between the population sample and the hypothetical ancestral sequence of the *D. simulans*/*D. melanogaster* species pair (reconstructed through maximum likelihood). The numbers of silent and replacement sites surveyed varied very slightly among species or analyses; the numbers of silent and replacement sites surveyed in *D. simulans* and *D. melanogaster* were 544 and 1865, respectively. Pairwise divergence is the average number of differences between species for all pairs of alleles. All divergence estimates were corrected for multiple hits using the Jukes-Cantor formula. θ and π were estimated according to Watterson (1975) and Nei (1987), respectively.

Bauer and Aquadro 1997). Replacement heterozygosity is low in both species; however, replacement divergence per site of ~5% is fairly high. Frequency distributions of replacement and silent polymorphisms in each species as measured by Tajima's *D* (Tajima 1989) are compatible with the distribution expected under a neutral equilibrium model (Table 1).

Table 2 shows the number of silent and replacement polymorphisms within *D. melanogaster* and *D. simulans*, and the number of silent and replacement fixations between species. Under the null hypothesis that polymorphisms and fixations are neutral, the ratio of silent to replacement polymorphism should be roughly equal to the ratio of silent to replacement fixations (Kimura 1983). A homogeneity test (McDonald and Kreitman 1991) of the polymorphic and fixed sites at *Relish* in *D. melanogaster* and *D. simulans* rejects the null hypothesis of neutral evolution ($P < 10^{-5}$). Many of the fixed differences between *D. melanogaster* and *D. simulans* can be assigned to one lineage or the other under the parsimony criterion, with *D. yakuba* serving as the outgroup (Table 2). Homogeneity tests (Table 2) of polymorphisms and fixations in the two resulting 2 × 2 contingency tables (one for each species) result in a highly

TABLE 2

Polymorphic and fixed, silent and replacement variants in the *Relish* gene of *D. simulans* and *D. melanogaster*, and associated test statistics

	Silent	Replacement	*G*-test
D. simulans and *D. melanogaster*			
Polymorphic	41	10	
Fixed	40	89	37.50, $P < 10^{-5}$
D. simulans			
Polymorphic	34	6	
Fixed	8	29	33.66, $P < 10^{-5}$
D. melanogaster			
Polymorphic	7	4	
Fixed	26	32	1.32, $P = 0.25$

Fixed differences in each lineage were determined under the parsimony criterion using *D. yakuba* as the outgroup. The sum of the fixed differences along the two lineages does not equal the number of fixed differences in the pooled data because only fixations that could be unambiguously assigned to one lineage or the other under the parsimony criterion were used.

TABLE 3

Polymorphisms and fixations in different regions of the Relish protein

	D. simulans				*D. melanogaster*				Pooled			
	Sil.		Repl.		Sil.		Repl.		Sil.		Repl.	
Region	Poly.	Fix.	Poly.	Fix.	Poly.	Fix.	Poly.	Fix.	Poly.	Fix.	Poly.	Fix.
I	2	2	1	6	0	2	2	5	2	4	4	11
II	12	3	0	0	4	12	0	3	15	17	0	4
III	4	1	0	15	0	5	0	18	4	10	0	49
IV	16	2	5	8	3	7	2	7	19	9	7	25

Coordinates of regions I–IV are in the Figure 1 legend. Polymorphisms and fixations of silent and replacement variation in regions III and IV are significantly heterogeneous in the *D. simulans* lineage and in the pooled data. No other regions are significantly heterogeneous in either lineage or in the pooled data. Sil., silent; Repl., replacement; Poly., polymorphism; Fix., fixation.

significant rejection of the null hypothesis in *D. simulans* ($P < 10^{-5}$), but not in *D. melanogaster* ($P = 0.25$).

Polymorphisms and fixations in different regions of *Relish* can be analyzed separately to determine whether significant heterogeneity of polymorphisms and fixations in *D. simulans* is attributable to unusual evolution throughout the gene or rather to evolution in particular regions. We divide up the sequenced region of Relish into four domains (Figure 1). Two easily recognizable functional domains of Relish are the NF-κB region and the region from the first ankyrin repeat to the termination codon (Dushay *et al.* 1996). We also define the "spacer" region between these two domains as a separate domain, though based on sequence similarity it is not obviously homologous to known domains from other proteins (Dushay *et al.* 1996). Finally, we analyze the region 5′ of the first codon of the NF-κB domain as our fourth domain, though again there are no data suggesting a particular function. We refer to regions I and II as the NF-κB region (composed primarily of two Rel-homology domains), and regions III and IV as the IκB region (composed primarily of ankyrin repeats).

Table 3 shows the numbers of silent and replacement polymorphisms and fixed differences, as well as numbers of fixed differences in each of the two lineages for each of the four regions of *Relish.* The main conclusion from analyses of these data is that polymorphism and divergence are significantly heterogeneous for both region III (the "spacer") and region IV (the ankyrin repeats) in *D. simulans* and in the pooled data. Homogeneity tests of polymorphism and divergence for other subsets of the data are not significant. In terms of functional domains, the IκB region of *Relish* (corresponding roughly to region III and IV; Dushay *et al.* 1996) is not evolving neutrally in *D. simulans*, while the NF-κB domain shows no evidence of deviations from neutrality in either species or in the pooled data.

Akashi (1996, 1999) has proposed that rejection of the null hypothesis of homogeneity for contingency tables of silent and replacement polymorphisms and fixations could be attributable to selection on silent sites rather than selection on replacement sites. Categorization of silent mutations into putative fitness classes can help address this possibility. For a given amino acid, preferred codons are those that are significantly more abundant in genes with a high degree of codon bias compared to their abundance in genes exhibiting less codon bias (*e.g.*, Sharp and Lloyd 1993). Preferred codons are hypothesized to have slightly higher average fitness than unpreferred codons. Thus, unpreferred codons are hypothesized to be maintained by a balance between mutation (which introduces them into populations), weak purifying selection (which tends to remove them), and genetic drift (which can fix them). Preferred mutations (polymorphisms or fixations) are those for which analysis of outgroups suggests that the ancestral state is unpreferred, while unpreferred mutations result from mutations from preferred to unpreferred codons (Akashi 1996).

Table 4 shows the numbers of preferred and unpreferred mutations at *Relish.* Homogeneity tests of the 2×3 contingency tables of polymorphic and fixed, preferred, unpreferred, and no change mutations are not significant in either *D. simulans* ($P = 0.78$) or *D. melanogaster* ($P = 0.71$). Addition of amino acid variation to the analyses results in significant rejection of homogeneity for the 2×4 contingency table from *D. simulans*

TABLE 4

Polarized silent and replacement variation at *Relish* in *D. simulans* and *D. melanogaster*[a]

	Polymorphic				Fixed			
	P	U	NC	R	P	U	NC	R
D. simulans	9	11	8	5	3	2	3	29
D. melanogaster	0	3	0	2	1	15	3	32

P, U, NC, and R are preferred, unpreferred, no change, and replacement variants, respectively.

[a] Replacement polymorphisms were not polarized.

TABLE 5

Numbers of silent and replacement fixations at the *Relish* gene and eight other genes (Takano 1998) between each of three species and the hypothetical ancestor of the *D. simulans*/*D. melanogaster* species pair, and estimates of codon bias and base composition in *Relish*

	Silent	Replacement	ENC[a]	%GC[b]
D. simulans	19 (103)	40 (29)	49.4	67.5
D. melanogaster	32 (158)	54 (39)	50.3	64.8
D. yakuba	114 (326)	79 (98)	48.0	66.8

Alleles Sim1, Zim1, and yakuba were used to reconstruct the ancestral sequence of the *D. simulans*/*D. melanogaster* species pair (the baseml program in the PAML package was used; Yang 1999). The number of differences between this hypothetical ancestral allele and each of the three extant alleles was determined, without correction for multiple hits. Numbers in parentheses are parsimony-based estimates from Takano (1998; Table 2) for eight genes, rounded to the nearest integer.

[a] ENC, estimated as described by Wright (1990) for the Sim1, Zim1, and yakuba alleles.

[b] GC content at fourfold degenerate sites for the Sim1, Zim1, and yakuba alleles.

($P < 10^{-5}$) but not for the comparable contingency table from *D. melanogaster* ($P = 0.71$). This suggests that rejection of homogeneity in *D. simulans* is attributable to selection on replacement sites.

Analysis of the three Drosophila lineages provides a moderate degree of additional insight into the evolutionary history of *Relish*. In the absence of sequence data from outgroups we are unable to distinguish between fixations from the common ancestor of the three species to *D. yakuba*, and the fixations from the common ancestor of the three species to the common ancestor of *D. simulans*/*D. melanogaster*. For convenience we refer to the lineage connecting the common ancestor of *D. simulans*/*D. melanogaster* with *D. yakuba* as the *D. yakuba* lineage. Silent divergence along the *D. simulans* lineage is about twice as great as the silent divergence along the *D. melanogaster* lineage, as was previously observed for other genes located in regions of normal rates of crossing-over in these two species (Akashi 1996; Takano 1998). Replacement divergence at *Relish* is also higher in the *D. simulans* lineage than in the *D. melanogaster* lineage. The relative rates of silent to replacement site evolution at *Relish* are roughly 2 to 1 in the *D. simulans* and *D. melanogaster* lineages (Table 1). The relative rate of silent to replacement evolution along the *D. yakuba* lineage, ~6 to 1 (Table 1), appears to be different from the relative rates in the other two lineages.

Table 5 shows the numbers of silent and replacement differences between the hypothetical ancestor of *D. simulans*/*D. melanogaster* and each of the three species in our analysis. The ratio of silent to replacement fixations in the *D. yakuba* lineage is significantly higher than the ratio in the other lineages. It is difficult to decide whether the silent fixations, the replacement fixations, or both kinds of fixations contribute to the lineage differences. Our estimates of silent substitutions per site along the *D. simulans* and *D. melanogaster* lineages are similar to the average rate estimated for eight genes (Takano 1998); our estimate of silent substitution rate at *Relish* in the *D. yakuba* lineage (Table 1) is about twice as large as the the average estimate for eight genes (Takano 1998). This suggests some elevation of the silent substitution rate in the *D. yakuba* lineage, though there is no evidence for relaxed selection on codon bias in *D. yakuba* compared to the other species (*cf.* Akashi 1996); in fact, the degree of codon bias in *D. yakuba Relish* is slightly higher than the degree of bias in *D. simulans* and *D. melanogaster* (Table 5). However, there also appear to be proportionally fewer amino acid fixations in the *D. yakuba* lineage in *Relish* compared to pooled data from eight other genes (Takano 1998). Sequence data from additional species will be required to help us disentangle evolution on distinct lineages and to help us determine whether adaptive protein evolution of *Relish* is a general phenomenon.

DISCUSSION

The configuration of polymorphisms and fixations at silent and replacement sites in *Relish* provides extremely strong evidence that a neutral model of molecular evolution cannot explain evolution of this gene; departures from predictions of the neutral model are primarily attributable to evolution in the *D. simulans* lineage. Furthermore, separate analysis of distinct structural/functional domains reveals that nonneutral evolution is apparent only in the IκB region of Relish.

Rejections of the null hypothesis of homogeneity in analyses of contingency tables of polymorphism and divergence can be difficult to interpret because, in principle, any observation or combination of observations can contribute to rejection of the null hypothesis (*e.g.*, Hudson *et al.* 1987; Akashi 1996). The reasoning most often used in interpreting such tests is that silent variation is likely to be under much weaker selection than replacement variation. Therefore, deviations from expectations under neutral evolution are usually interpreted in terms of selection on replacement polymorphisms or fixations. Following this reasoning, a configuration of silent and replacement variation such as that seen in Table 2 has been interpreted as a consequence of adaptive protein evolution, or too many amino acid fixations (McDonald and Kreitman 1991; Eanes *et al.* 1996). In our case, we would further conclude that adaptive protein evolution at *Relish* has been more important in the *D. simulans* lineage than in the *D. melanogaster* lineage.

However, Akashi (1996, 1999) has proposed that weak selection on silent variation cannot be dismissed

as a possible cause of significant heterogeneity tests of polymorphism and divergence. Specifically, Akashi proposes that dynamics of unpreferred silent polymorphisms in *D. simulans* fit a slightly deleterious model of evolution (*e.g.*, Ohta 1992). Slightly deleterious mutations are (by definition) sufficiently weakly selected such that they can drift to appreciable frequencies in populations, yet are sufficiently strongly selected such that they are unlikely to fix. Such mutations are expected to make a disproportionate contribution to polymorphism, compared to their contribution to divergence (*e.g.*, Kimura 1983). Akashi's analyses suggest that unpreferred silent polymorphisms belong to this slightly deleterious class. If this is true, then we might just as easily explain our significant heterogeneity tests of *Relish* in *D. simulans* as a consequence of too many silent polymorphisms (of the unpreferred type) rather than as a consequence of too many replacement fixations.

Multiple lines of evidence, however, render this explanation unlikely. Among the eight genes analyzed in Akashi (1999) the number of unpreferred and preferred polymorphisms were 87 and 24, respectively; the number of polymorphisms in each category for *Relish* are 11 and 9 (Table 4). Thus, if anything, *Relish* shows proportionally fewer unpreferred silent polymorphisms than other genes in the species. Furthermore, unlike the pattern seen in most other *D. simulans* genes (Akashi 1996), *Relish* has roughly equal numbers of preferred and unpreferred polymorphisms. Pooled data from several *D. simulans* genes provide evidence that there are roughly equal numbers of unpreferred and preferred fixations along this lineage (Akashi 1996). Thus, the ratio of unpreferred to preferred polymorphisms at *Relish* is not significantly different from the overall ratio of unpreferred to preferred fixations in *D. simulans*. Furthermore, as noted in results, the 2×3 contingency table of unpreferred, preferred, and no change silent mutations is not significantly heterogeneous in *D. simulans*, yet addition of the replacement variation to the analysis yields a 2×4 contingency table that is significantly heterogeneous. Finally, *Relish*, along with *Zw* (Eanes *et al.* 1996), stands out among all other *D. simulans* genes from regions of normal recombination in having a relatively low level of amino acid polymorphism in spite of a relatively high rate of amino acid divergence. All of these observations favor the interpretation that evolution at replacement sites is the cause of rejections of the null hypothesis in *D. simulans*.

Akashi (1999) provides evidence that a particular model of evolution of weak selection on silent sites and neutral evolution of replacement sites can cause some fraction of heterogeneity tests to be statistically significant in the direction seen in our data. However, there is no evidence that such a model can cause deviations from homogeneity of the magnitude seen in the *D. simulans Relish* data. Finally, relaxed functional constraints on Relish early in the *D. simulans* lineage could contribute to increased rates of protein evolution. However, the special timing required for changes in purifying selection to account for the data in these recently separated species renders such an explanation unlikely (*e.g.*, McDonald and Kreitman 1991). Adaptive protein evolution at *Relish* in *D. simulans* is the best explanation for our data.

As we noted earlier, silent heterozygosity at *Relish* in *D. melanogaster* is low relative to silent heterozygosity at *Relish* in *D. simulans*. Comparison of the ratios of polymorphism to divergence (Hudson *et al.* 1987) of silent and nonprotein-coding sites in Zimbabwe *D. melanogaster* samples of *Relish vs. vermilion* (Begun and Aquadro 1995) suggests that the ratio is lower at *Relish* than at *vermilion*, though not quite significantly so ($\chi^2 = 3.18$, $P = 0.07$). Genes experiencing lower recombination rates are expected to be less variable (Begun and Aquadro 1992; Aquadro *et al.* 1994) as a result of selection at linked sites (Maynard Smith and Haigh 1974; Kaplan *et al.* 1989; Charlesworth *et al.* 1993). Therefore, a possible explanation for reduced variation at *Relish* in *D. melanogaster* is that the recombination rate at *Relish* in *D. melanogaster* is lower than the rate in *D. simulans*. A mechanistic explanation for such a phenomenon is the fixed inversion difference between species on chromosome arm 3R. *Relish* has been localized to polytene band position 85C in *D. melanogaster* (Dushay *et al.* 1996), and thus is probably sufficiently close to the centromere of chromosome 3 to experience reduced recombination compared to genes located more distally (*e.g.*, Kliman and Hey 1993; Kindahl 1994). As a result of the fixed inversion difference between species, which has breakpoints at approximately 84F and 93F (*e.g.*, Ashburner 1989), *Relish* is considerably further from the centromere in *D. simulans* than in *D. melanogaster*; a rough approximation is that *Relish* in *D. simulans* is located at a physical position equivalent to 93D of *D. melanogaster*. The *D. simulans* karyotype is probably ancestral (*e.g.*, Ashburner 1989), suggesting that there has been a recent drop in recombination rates in the *Relish* region in the *D. melanogaster* lineage. Despite this drop in recombination rates in *D. melanogaster*, the relative numbers of unpreferred and preferred fixations in *D. melanogaster Relish* (Table 4) are not significantly different from the ratio for an independent sample of eight genes from regions of normal recombination in this species (Takano 1998; Fisher's exact test, $P = 0.11$, one-tailed), though the difference in the ratio for *Relish vs.* the other eight genes (*Relish* has proportionally more unpreferred fixations) is in the direction predicted by Akashi (1996).

If the high rate of protein evolution at *Relish* in *D. simulans* is a consequence of directional selection, then how are we to explain the finding that a similar rate of protein evolution in *D. melanogaster* leaves us with no statistical support for adaptive evolution in this lineage? Comparison of silent and replacement divergence along

each lineage (Tables 1 and 2) shows that the main difference between lineages is the much higher rate of silent site evolution in *D. melanogaster*. As noted above, this might be attributable to differences in the recombinational environment of *Relish* in *D. simulans* and *D. melanogaster* resulting from the fixed inversion difference between species, as well as from the reduction of recombination that presumably occurred in ancestral *D. melanogaster* populations as this inversion increased in frequency on its way toward fixation. Therefore, one interpretation is that protein evolution has proceeded rapidly in both lineages as a consequence of directional selection but that statistical support for adaptation in *D. melanogaster* has been obscured by increased rates of silent site evolution in this lineage compared to the rate in *D. simulans* (Akashi 1996; Takano 1998).

The high silent heterozygosity at *Relish* in *D. simulans* is interesting in light of our inference of recurrent, directional selection at this gene in this lineage. This is something of a paradox, as directional selection can have large effects on reducing silent heterozygosity at tightly linked sites (Maynard Smith and Haigh 1974; Kaplan *et al.* 1989). How might our historical inference of directional selection be reconciled with our observations of polymorphism? At least four factors determine the extent of reduction of neutral variation by directional selection at linked sites (Maynard Smith and Haigh 1974; Kaplan *et al.* 1989). These factors are the selection coefficient, the recombination rate, the initial frequency of the mutant when selection begins, and the time since the most recent selective fixation.

Directional selection can result in rapid fixation times relative to the neutral expectation, yet can still have a fairly restricted impact on reducing linked sequence variation (Kaplan *et al.* 1989; Eanes *et al.* 1996; Tsaur *et al.* 1998). We can use the results from Kaplan *et al.* (1989) as an example. Though values of parameters from Kaplan *et al.* (1989) may not be very close to the true values in *D. simulans*, the model is still illustrative. For example, if $2N = 10^8$, the selection coefficient of a new beneficial mutant is 10^{-4}, and the recombination rate per base per generation is 10^{-8}, then the expected window of reduced polymorphism caused by selective fixation of a beneficial mutant is only 200 bases. Nevertheless, the expected fixation time of such a mutant (conditional on its fixation) is $\sim 2/s \ln 2N$ (Nei 1987) or $\sim 3.7 \times 10^5$ generations. In terms of N generations the expected fixation time is ~ 0.007 compared to 4.0 for neutral mutants. The point is simply that there may be a broad range of selection coefficients of new beneficial mutants that accommodate rapid fixation yet result in hitchhiking effects over only small physical distances. Another potential explanation for the absence of severely reduced, linked silent variation is that amino acids fixing under directional selection do not start out at as unique or extremely low frequency mutants but rather are sampled from a set of previously neutral or balanced amino acid polymorphisms. Finally, one could explain the data if most selected amino acid fixations in *Relish* occurred in a rapid burst of evolution early in the history of the *D. simulans* lineage, with few or no adaptive fixations in the more recent past. In this case, silent polymorphism might have been severely reduced, but recovered since to a level near the expected equilibrium value. It is worth noting that we now have examples of two genes, *Relish* and *G6pd* (Eanes *et al.* 1996), for which there is evidence for large numbers of "excess" amino acid fixations in the *D. simulans* lineage, yet no evidence for dramatically reduced silent polymorphism at tightly linked sites. This pattern may prove to be a common one in Drosophila, thus increasing the importance of determining which models of evolution might best explain such data.

We have convincingly established a history of directional selection on amino acid variation in *D. simulans*. How might our analysis impinge on broader issues of the evolution of fly immunity and the biological role of *Relish*? One potentially relevant finding is that there is strong evidence for adaptive evolution in the IκB domain, yet no evidence for adaptive protein evolution in the NF-κB domain. Models of IκB function posit that such proteins are modulated primarily through kinase-dependent phosphorylation and subsequent ubiquitin-dependent targeting to proteolytic degradation pathways (Ghosh *et al.* 1998). An interesting issue is whether adaptive amino acid evolution at large numbers of residues throughout the IκB region of Relish is likely to be caused strictly by selection resulting from interactions of this protein with internal signaling components. If this is thought to be unlikely, an alternative possibility is that selection pressures acting on the IκB domain of Relish arise from direct interactions with other molecules; those deriving directly from pathogens are obvious candidates.

One can speculate that microbial pathogens could benefit by interfering with activation of the Drosophila immune response. Pathogenic bacteria possessing type III secretion systems are able to carry out contact-mediated transport of proteins directly into the cytoplasm of host cells. These bacterial proteins can specifically interfere with host-cell signal transduction or other processes (Hueck 1998). Thus, there is a well-established mechanistic basis for specific manipulation of animal cytoplasmic proteins by microbial pathogens, though there has been no exploration of the phenomenon in Drosophila. Manipulation of IκB proteins such that nuclear translocation of NF-κB proteins (which regulate transcription of other immune system proteins) is inhibited would be a potential mechanism whereby microbial pathogens could suppress the Drosophila immune response. Drosophila populations would experience strong natural selection to evade such strategies. In this scenario, a putative arms race is manifested in an evolutionary conflict (mediated through interactions with

IκB proteins) between fly and pathogen over control of subcellular localization of NF-κB proteins. These hypotheses must be considered to be very speculative. Our ability to formulate evolutionary hypotheses about Relish is limited by our poor understanding of the biology of this protein and its precise role in the Drosophila immune response.

Nevertheless, the data presented here provide at least one potential experimental foothold into the evolutionary or ecological genetics of Drosophila-microbe interactions. For example, analysis of phenotypic consequences of standing variation at *Relish* could prove interesting from both a mechanistic and evolutionary/ecological perspective. Experiments to elucidate functional consequences of interspecific differences in *Relish* in the context of natural pathogens might also be interesting. The recent discovery of numbers of Drosophila mutants affecting nuclear localization of Rel proteins (Wu and Anderson 1998) suggests that there could be numerous arenas for conflict between flies and their microbial pathogens.

Anonymous reviewers provided useful comments. This work was supported by National Institutes of Health grant GM55298.

LITERATURE CITED

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.

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 and Hall, New York.

Ashburner, M., 1989 *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

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., 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 C. F. Aquadro, 1995 Molecular variation at the *vermilion* locus in geographically diverse populations of *Drosophila melanogaster* and *D. simulans*. Genetics **140:** 1019–1032.

Baeuerle, P. A., 1998 IκB-NF-κB structures: at the interface of inflammation control. Cell **95:** 729–731.

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.

Durell, S. R., G. Raghunathan and H. R. Guy, 1992 Modeling the ion channel structure of cecropin. Biophys. J. **63:** 1623–1631.

Dushay, M. S., and E. Eldon, 1998 Drosophila immune responses as models of human immunity. Am. J. Hum. Genet. **62:** 10–14.

Dushay, M. S., B. Asling and D. Hultmark, 1996 Origins of immunity: Relish, a compound Rel-like gene in the antibacterial defense of Drosophila. Proc. Natl. Acad. Sci. USA **93:** 10343–10347.

Eanes, W. F., M. Kirchner, J. Yoon, C. H. Biermann, I. N. Wang *et al.*, 1996 Historical selection, amino acid polymorphism and lineage-specific divergence at the *G6pd* locus in *Drosophila melanogaster* and *D. simulans*. Genetics **144:** 1027–1041.

Ghosh, S., M. J. May and E. B. Kopp, 1998 NF-κB and Rel proteins: evolutionarily conserved mediators of immune responses. Annu. Rev. Immunol. **16:** 225–260.

Hey, J., and J. Wakeley, 1997 A coalescent estimator of the population recombination rate. Genetics **145:** 833–846.

Hoffmann, J. A., and J.-M. Reichhart, 1997 Drosophila immunity. Trends Cell Biol. **7:** 309–316.

Hudson, R. R., M. Kreitman and M. Aguade, 1987 A test of neutral molecular evolution based on nucleotide data. Genetics **116:** 153–159.

Hueck, C. J., 1998 Type III protein secretion systems in bacterial pathogens of plants and animals. Microbiol. Mol. Biol. Rev. **62(2):** 379–433.

Huxford, T., D.-B. Huang, S. Malek and G. Ghosh, 1998 The crystal structure of the IκBα/NF-κB complex reveals mechanisms of NF-kB inactivation. Cell **95:** 759–770.

Inoue, J., L. D. Kerr, A. Kakizuka and I. M. Verma, 1992 IκBγ, a 70kd protein identical to the C-terminal half of p110 NF-κB: a new member of the IkB family. Cell **68:** 1109–1120.

Jacobs, M. D., and S. C. Harrison, 1998 Structure of an IκBα/NF-κB complex. Cell **95:** 749–758.

Kaplan, N. L., R. R. Hudson and C. H. Langley, 1989 The "hitchhiking" effect revisited. Genetics **123:** 887–899.

Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, United Kingdom.

Kindahl, E. C., 1994 Recombination and DNA polymorphism on the third chromosome of *Drosophila melanogaster*. Ph.D. Thesis, Cornell University, Ithaca, NY.

Kliman, R. M., and J. Hey, 1993 Reduced natural selection associated with low recombination in *Drosophila melanogaster*. Mol. Biol. Evol. **10:** 1239–1258.

Kylsten, P., C. Samakovlis and D. Hultmark, 1990 The cecropin locus in Drosophila: a compact gene cluster involved in the response to infection. EMBO J. **9:** 217–224.

Levin, B. R., and R. E. Lenski, 1983 Coevolution in bacteria and their viruses and plasmids, pp. 99–127 in *Coevolution*, edited by D. Futuyma and M. Slatkin. Sinauer, Sunderland, MA.

Maynard Smith, J., and J. Haigh, 1974 The hitchhiking effect of a favourable 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.

Moriyama, E., and J. R. Powell, 1996 Intraspecific nuclear DNA variation in Drosophila. Mol. Biol. Evol. **13:** 261–277.

Nei, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

Ohta, T., 1992 The nearly neutral theory of molecular evolution. Annu. Rev. Ecol. Syst. **23:** 263–286.

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.

Sharp, P. M., and A. T. Lloyd, 1993 Codon usage, pp. 378–397 in *An Atlas of Drosophila Genes: Sequences and Molecular Features*, edited by G. Maroni. Oxford University Press, Oxford.

Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123:** 585–595.

Takano, T. S., 1998 Rate variation of DNA sequence evolution in the Drosophila lineages. Genetics **149:** 959–970.

Tsaur, S.-C., C.-T. Ting and C.-I. Wu, 1998 Positive selection driving the evolution of a gene of male reproduction, *Acp26Aa*, of Drosophila: II. Divergence vs. polymorphism. Mol. Biol. Evol. **15:** 1040–1046.

Watterson, G. A., 1975 On the number of segregating sites in genetic models without recombination. Theor. Popul. Biol. **7:** 256–276.

Wright, F., 1990 The "effective number of codons" used in a gene. Gene **87:** 23–29.

Wu, L. P., and K. V. Anderson, 1998 Regulated nuclear import of Rel proteins in the Drosophila immune response. Nature **392:** 93–96.

Yang, Z., 1999 Phylogenetic analysis by maximum likelihood (PAML), Version 2.0. University College, London.