J Mol Evol (2003) 57:598–605
DOI: 10.1007/s00239-003-2506-6

JOURNAL OF MOLECULAR EVOLUTION

The Evolution of Parasite Recognition Genes in the Innate Immune System: Purifying Selection on *Drosophila melanogaster* Peptidoglycan Recognition Proteins

Francis M. Jiggins,[1] Gregory D.D. Hurst[2]

[1] Department of Genetics, University of Cambridge, Downing Street, Cambridge CB2 3EH, UK
[2] Department of Biology, University College London, 4 Stephenson Way, London NW1 2HE, UK

Received: 17 January 2003 / Accepted: 27 May 2003

Abstract. Genes involved in the recognition of parasites by the acquired immune system are often subject to intense selection pressures. In some cases, selection to recognize a diverse range of parasites has resulted in high levels of polymorphism, while elsewhere the protein sequence has changed rapidly under directional selection. We tested whether parasite recognition genes in the innate immune system show similar patterns of evolution. We sequenced seven peptidoglycan recognition protein genes (*PGRP*s) from 12 lines of *Drosophila melanogaster* and one line of *D. simulans* and used a variety of tests to determine whether the observed mutations were selectively neutral. We were unable to detect either balancing or directional selection. This suggests that the molecular cues used by insects to detect parasites are highly conserved and probably under strong functional constraints which prevent their evolving to evade the host immune response. Therefore, interactions between these genes are unlikely to be the focus of host–parasite coevolution, at least in *Drosophila*. We also found evidence of gene conversion occurring between two genes, *PGRP-SC1A* and *PGRP-SC1B*.

Key words: Coevolution — Peptidoglycan recognition protein — PGRP — Innate immunity — Positive selection — Balancing selection

Correspondence to: Francis M. Jiggins; *email:* fmj1001@mole.bio.cam.ac.uk

Introduction

Host–parasite interactions are commonly associated with rapid evolution. This is thought to occur because hosts and parasites enter into an evolutionary arms race, in which parasite adaptations, are met by host counter-adaptations, and vice versa. The molecular basis of coevolution is best understood in vertebrates, where genes involved in parasite recognition are often subject to strong selection pressures. For example, balancing selection often maintains multiple alleles of major histocompatibility complex (MHC) genes over millions of years (Hughes and Nei 1988). The MHC molecules present antigens to the acquired immune system, and different alleles at the same locus are able to bind different antigens (Doherty and Zinkernagel 1975). Therefore, overdominant selection probably maintains high levels of polymorphism because heterozygotes can recognize a greater diversity of antigens than homozygotes. The situation is somewhat different in chimpanzees, where several MHC loci have been subject to a recent selective sweep, probably in response to a past pandemic (Groot et al. 2002).

In vertebrates, the acquired immune system can recognize a diverse range of parasites by producing a huge variety of different recognition receptors. However, this is a recent evolutionary innovation, and the vast majority of animals instead rely upon an innate immune system. Parasite detection in invertebrates is far less well understood, but it is generally thought that they are recognized by molecular

structures that are conserved across a broad range of pathogens (Medzhitov and Janeway 1997). Therefore, a lower diversity of parasite recognition molecules need be produced. The fact that recognition by the immune system is presumably disadvantageous to the parasite implies that strong functional constraints must prevent parasites evolving to evade detection. Therefore, there appears to be little scope for host–parasite coevolution between parasite recognition molecules and their targets.

This caricature of the innate immune system must, however, be treated with great caution, as the molecular mechanisms of parasite recognition in invertebrates are still poorly understood. Indeed, studies of the success of parasites and pathogens in propagating within *Daphnia* have found host genotype × parasite genotype interactions, which suggests the presence of some specific interactions within arthropods at either the recognition or the immune reaction level (Carius et al. 2001). Studies of *Anopheles gambiae* have similarly indicated genetic variation for susceptibility to *Plasmodium* within mosquito populations (Niare et al. 2002). This led the authors to suggest, by analogy with plants, that this variation may be associated with differences between pattern recognition receptors. Further, comparison between the genome sequence of *Anopheles gambiae* and that of *Drosophila* indicates that while core signaling processes are conserved within immune responses, there is a diversification in receptor and effector molecules between the taxa, with a deficit of orthologues and an excess of gene family expansions (Christophides et al. 2002). This further suggests that coevolutionary interactions may be important in the evolution of the innate immune system.

Our aim is to test whether innate immune system parasite recognition molecules coevolve with parasites. The null hypothesis of coevolutionary interaction predicts that, unlike their counterparts in the acquired immune system, these molecules will be subject to neither directional nor balancing selection. To this end, we have investigated the peptidoglycan recognition proteins (PGRPs) of *Drosophila melanogaster*. The first PGRP was isolated from the moth *Bombyx mori* due to its ability to bind bacterial peptidoglycan and activate an immune response (Yoshida et al. 1996). Subsequent analyses found that the *D. melanogaster* genome contains at least 12 *PGRP* genes (Werner et al. 2000), at least 5 of which are up regulated following immune challenge (PGRP-SA, -SB1, -SC2, -SD, and -LB [De Gregorio et al. 2001; Irving et al. 2001; Werner et al. 2000]). Seven of the *Drosophila PGRP* genes produce short transcripts (*PGRP-S*) that are predicted to encode extracellular proteins, while the remaining five produce longer transcripts and encode membrane-spanning proteins (*PGRP-L*). It is now known that *PGRP-SA* is required to activate the immune response to gram-positive bacteria, and *PGRP-LC* elicits a response to gram-negative bacteria (Choe et al. 2002; Gottar et al. 2002; Michel et al. 2001; Ramet et al. 2002). Although it is clear that several of the PGRP genes play a role in pattern recognition, some may have other functions. For example, PGRP-SC1B, which is expressed in the gut, has enzymatic activity and can degrade peptidoglycan (Mellroth et al. 2003). It is possible that this protein is acting as a scavenger, digesting immunogenic molecules from the gut to prevent them inducing an unnecessary immune response. The five amino acid residues necessary for enzymatic activity are conserved in PGRP-SC1A, -SB1, -SB2, -SC2, and -LB, suggesting that these genes may play a similar role (Mellroth et al. 2003).

We have sequenced seven *PGRP* genes from 12 *D. melanogaster* lines and a single *D. simulans* line. We then inferred the selection pressures acting on the genes from patterns of intraspecific polymorphism and interspecific divergence. We found that there was strong purifying selection acting to conserve the protein sequence, and no evidence of either directional selection or balancing selection.

Table 1. *Drosophila melanogaster* lines

Line	Origin
Loua	Democratic Republic Congo
Texas	Texas, USA
KY024	Kenya
Draveil	France
Monty5	Montpellier, France
Tahiti	Tahiti
ZW141	Zimbabwe
DI7	Australia
P. Bourg	Petit Bourg, Guadeloupe
KY038	Kenya
S30	Seattle WA, USA
genome sequence	

Methods

The genes were sequenced from a worldwide collection of *D. melanogaster* isofemale lines (Table 1) and *D. simulans* line C167.4. In two cases we were unable to amplify a particular gene from certain lines, and another line was substituted (SC1B was sequenced from the Seattle S14 rather than Monty5, and SC2 was sequenced from S14 rather than S30). The X chromosome genes were sequenced from a DNA extracted from a single male individual. The genes on chromosomes 2 and 3 were sequenced from fly lines that were first made isogenic for these chromosomes via balancer chromosome-mediated chromosome substitution.

Seven *PGRP* genes were sequenced, all of which encode extracellular proteins (Table 2). They were chosen as they have the shortest genomic sequences (both the fewest introns and the shortest exons), which allowed us to maximize the number of genes sequenced. The PCR primers were designed in the noncoding flanking sequence from the *D. melanogaster* genome sequence (Table 2) (Adams et al. 2000). In two cases where these primers did not amplify the gene in *D. simulans*, alternative primers were designed and used. *PGRP-SA* in *D. simulans* was amplified using the forward

Table 2. *PGRP* genes, their chromosomal location, function, and PCR primers (Adams et al. 2000; Michel et al. 2001; Werner et al. 2000)

Gene	Chromosome	Map location	Function	Primers
PGRP-SA	X	10C	Gram + bacteria recognition Binds peptidoglycan	F 5′-GATAAATCCGGCAGATAGCCC R 5′-CTGATATGCTCATTCTGGCAGA
PGRP-SB1	3	73C	Function unknown	F 5′-ATAAACACAGCCGTAGCGGA R 5′-TTGTACATTACCAAATAAACACATT
PGRP-SB2	3	73C	Function unknown	F 5′-CTGGCCAGATATITCGGTTG R 5′-GCTACAGAGAATGGGTTTCAAGA
PGRP-SC1A	2	44E	Function unknown	F 5′-CATCGACGAGGGCTGTAAAT R 5′-CCCGACTGCCTAATAAACCA
PGRP-SC1B	2	44E	Digests peptidoglycan	F 5′-TACCCGATAAGCAGCGATTT R 5′-ACTTTGCAATTGGATGTCCC
PGRP-SC2	2	44E	Function Unknown	F 5′-GTAGCCCGAACCTGACACAT R 5′-ATTCCCCGGAGAAAAATACC
PGRP-SD	3	66A	Function Unknown	F 5′-ACTATCTCCTTGGGGGCACT R 5′-AGGAGTGGTCTTTACGTGTCG

Table 3. Intraspecific polymorphism in *PGRP* genes

		Exons								Neutrality tests		
		Polymorphisms					Introns			Tajima	Fu and Li	
Gene	No. alleles	Length	Silent	Replace	π	θ_w	Length	π	θ_w	*D*	*D*	*F*
PGRP-SA	12	609	4	0	0.0014	0.0021	119	0.0014	0.0027	−1.29	−1.55	−1.70
PGRP-SB1	12	571	13	3	0.0102	0.0092	64	0.0239	0.0207	0.50	0.99	1.03
PGRP-SB2	12	546	7	4	0.0196	0.0066	60	0.0108	0.0110	0.07	−0.66	−0.55
PGRP-SC1A	12	555	2	1	0.0011	0.0017	0	—	—	−1.18	−1.12	−1.34
PGRP-SC1B	12	555	5	2	0.0039	0.0041	0	—	—	−0.26	−0.43	−0.46
PGRP-SC2	12	552	2	4	0.0043	0.0036	0	—	—	0.84	0.69	0.87
PGRP-SD	12	558	3	4	0.0051	0.0041	0	—	—	0.92	0.20	0.47

primer in Table 2 and the reverse primer 5′-GCCATTACTT GTAGTTGATTTAGGGAT. *PGRP-SB1* in *D. simulans* was amplified using the forward primer 5′-GGTATTTAACGC-TTCGTTTCCCGTA and the reverse primer 5′-TTGCTTTCTAG TTTTGCCTCCAGT. Both of these reverse primers overlap with the coding sequence, so several codons were missing from the final sequences. All the other sequences included the entire coding sequence and any introns. They have been submitted to the EMBL database under the accession numbers AJ556551–AJ556634.

DNA was extracted by digesting whole flies with Chelex 100 ion exchange resin (Bio-Rad), dithiothreitol, and proteinase-K, and the boiled supernatant used directly for PCR. The genes were amplified by PCR using *Taq* DNA polymerase and 35 thermal cycles (95°C for 30 s, 56°C for 1 min, and 72°C for 1 min). The PCR products were then run on an agarose gel and extracted using a gel extraction kit (Sigma). The fragments were then sequenced directly using the PCR primers and the ABI PRISM Dye Terminator cycle sequencing ready reaction kit (Perkin Elmer) on an ABI 384 automated sequencer (Perkin Elmer). These sequences were then aligned by eye together with the sequence from the Berkeley genome project (Adams et al. 2000).

Results

Polymorphism

The nucleotide variation observed in the *PGRP* genes is summarized in Table 3. The nucleotide diversity is estimated both as the average pairwise nucleotide difference per site (π [Nei 1987]), and from the proportion of segregating sites (θ_w [Watterson 1975]). The level of variation observed in most of the *PGRP* genes is typical of that observed for other loci in *Drosophila* (Moriyama and Powell 1996). However, *PGRP-SB1* and *PGRP-SB2* show unusually high levels of variation at synonymous sites, especially when measured as π.

The intraspecific and interspecific synonymous variation is fairly evenly distributed along the coding sequence (Fig. 1). However, the nonsynonymous variation is primarily found in the N-terminal and, to a lesser extent, the C-terminal regions of the protein. The central region is also the most conserved region when protein sequences are compared (Figure 1). It is possible that this reflects a region of functional importance.

Concerted Evolution

Two of the genes showed evidence of gene conversion. The loci *PGRP-SC1A* and *PGRP-SC1B*, which are tandemly arranged and separated by 3,124 bp, are more similar to each other within the same species than to the homologous locus in the other species. There is only one fixed difference between the *SC1A*

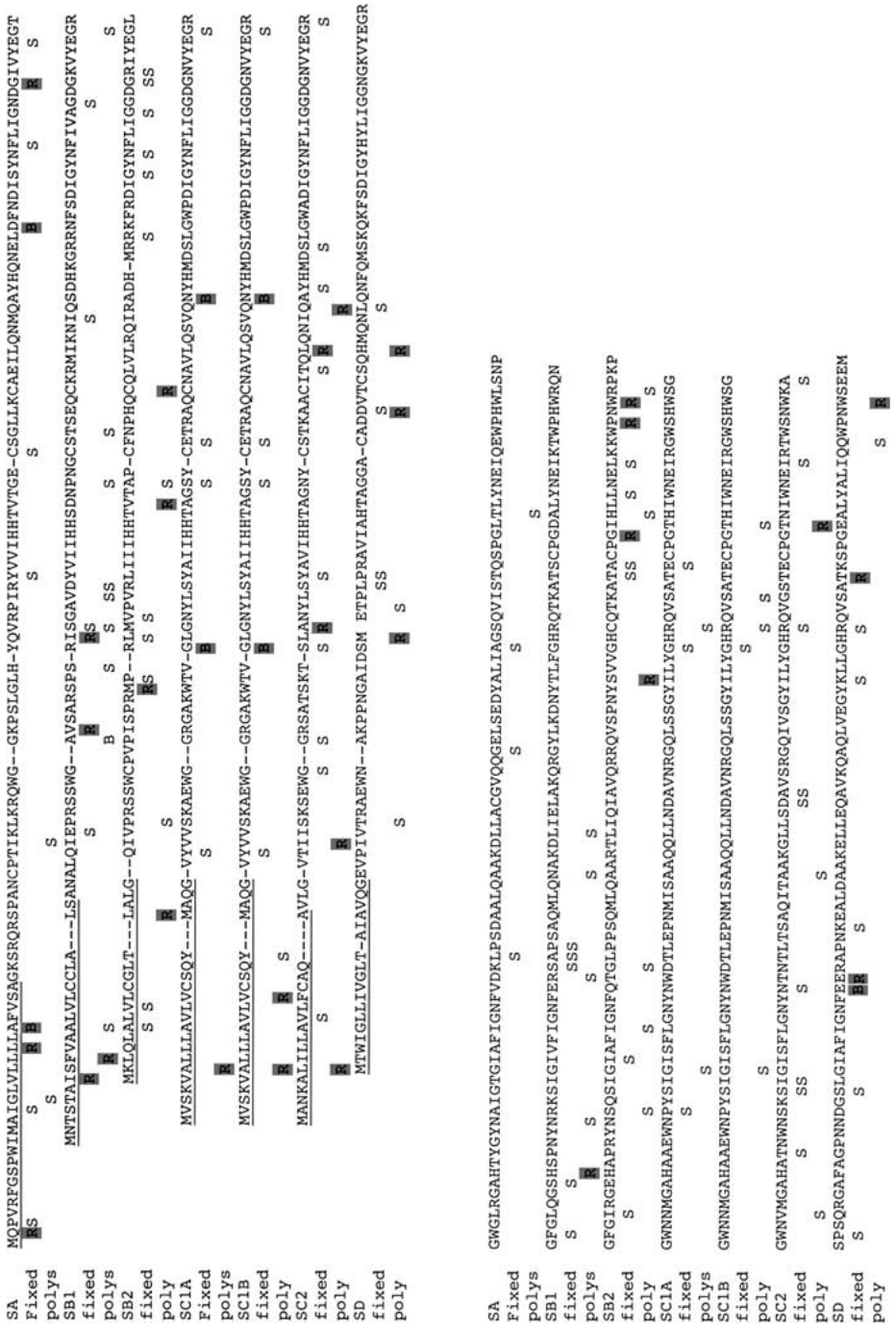

Fig. 1. Alignment of the PGRP proteins. The signal peptide is *underlined* (Werner et al. 2000). Codons which contained either a replacement substitution (R), a silent substitution (S), or both (B) are shown underneath the protein sequence. These differences are either fixed between *D. simulans* and *D. melanogaster* or polymorphic (Poly) within *D. melanogaster*.

and *SC1B* sequences in *D. melanogaster*, and three sites differ between each locus in *D. simulans* (Fig. 2). However, there is a higher level of divergence between the species (Fig. 2). In total there are nine substitutions between species that are fixed at both loci. The interpretation that this results from gene conversion is supported by the phylogeny of the sequences, which provides strong bootstrap support for the two loci being more similar to each other within a species than they are to their homologue in the other species (Fig. 3). There is also some suggestion of gene conversion in the polymorphism data from *D. melanogaster*. There are three low frequency derived mutations in *D. melanogaster PGRP-SC1A*, and all of these are also present at a low frequency in *D. melanogaster PGRP-SC1B*.

Neutrality Tests

The first approach we took to testing whether these genes are evolving neutrally was to examine whether the pattern of variation observed within *D. melanogaster* differed from that predicted by models of neutral evolution. First, we applied Tajima's (1989) *D*

```
SC1A melanogaster   a-genome GGCTAGCGAAGCCTCCACC
                    a-S30    --------------------
                    a-Tahiti --------------------
                    a-Drav   --------------------
                    a-Texas  --------------------
                    a-Loua   --------------------
                    a-zw141  --------------------
                    a-ky024  ----------A----
                    a-ky038  A----------T-A----
                    a-Pbourg --------------------
                    a-DI7    --------------------
                    a-Monty5 --------------------

SC1B melanogaster   genome   --T--------------TT
                    S30      --T--------------TT
                    Tahiti   ---------------T--T
                    Drav     ----------------TT
                    Texas    --T--------------TT
                    Loua     --T--------------TT
                    zw141    ------------T--T--T
                    ky024    AA-------------T--T
                    ky038    A-------------A--TT
                    Pbourg   --T--------------TT
                    DI7      -----------------TT
                    S14      --T--------------TT

SC1A simulans       ---CTCTAGCAA-C--G--
SC1B simulans       ---CTCTAGCA--C----T
```

Fig. 2. Variable sites in the *PGRP-SC1A* and *PGRP-SC1B* genes showing evidence of gene conversion.

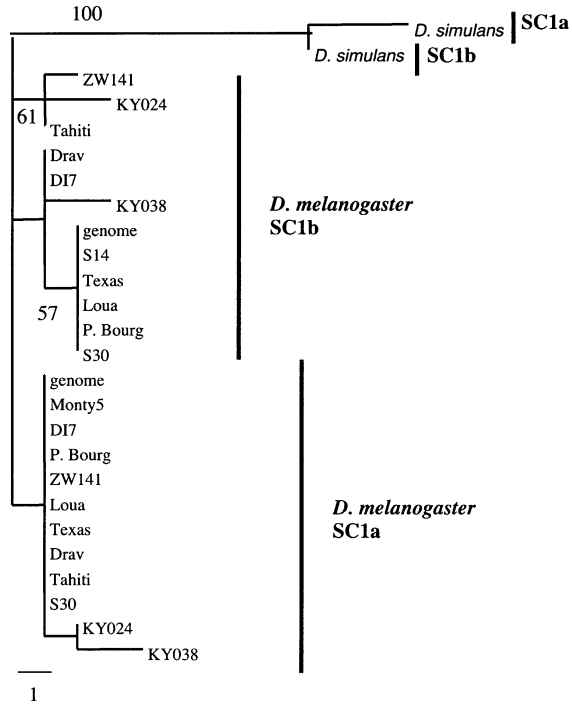

Fig. 3. Maximum parsimony tree of *PGRP-SC1A* and *PGRP-SC1B* illustrating that the two loci from the same species cluster together.

test, which examines the frequency distribution of polymorphisms. We then used Fu and Li's (1993) *D** and *F** tests, using *D. simulans* as an outgroup to determine the ancestral state. None of these tests showed a significant deviation from neutrality at any of the loci (Table 3).

An alternative approach to detect selection is to compare the levels of polymorphism at different loci. Neutral theory predicts that, when different genes are compared, there will be a positive correlation between the level of intraspecific polymorphism at a locus and the amount that the locus has diverged from an outgroup (Hudson et al. 1987). However, directional selection will reduce the amount of within species variation relative to an outgroup, while balancing selection will increase it. In order to test whether any of the loci were under selection, we performed a multilocus HKA test (Hudson et al. 1987). The interspecific divergence and intraspecific polymorphism of the different genes are shown in Table 4. If all the genes are included, there was no significant heterogeneity in the levels of polymorphism and divergence at the different loci ($\chi^2 = 4.32$, df $= 6$, $p = 0.63$). Concerted evolution can influence levels of polymorphism, so this test was repeated without the *PGRP-SC1A* and *PGRP-SC1B* genes, but once again, the result was not significant ($\chi^2 = 3.56$, df $= 4$, $p = 0.47$).

A final test for selection is to compare patterns of variation at synonymous and non-synonymous sites within a sequence. It is clear from the data that there are far more synonymous substitutions than nonsynonymous ones between *D. simulans* and *D. melanogaster*. However, a more powerful approach is to consider patterns of polymorphism at the two classes of sites. If polymorphism and divergence at both synonymous and nonsynonymous sites are due to neutral mutations, then the ratio of synonymous to nonsynonymous polymorphisms within a species will be the same as the ratio of synonymous to nonsynonymous differences between species (McDonald and Kreitman 1991). These statistics are shown for both each gene separately and the combined dataset in Table 5. Only *PGRP-SC2* differs significantly from neutrality, and this is due to an excess of either replacement polymorphisms or synonymous fixed differences. This pattern could be caused by gene conversion between the SC1 and the SC2 loci. However, the SC2 sequences are no more similar to the SC1 sequences within a species than between species, as would be expected if gene conversion were occurring. However, if the McDonald–Kreitman tests are Bonferroni corrected for multiple comparisons, then none of the data is significant. Therefore, we are unable to conclude that selection is acting on the mutations in *PGRP-SC2*.

Conclusions

We have been unable to detect any evidence of either directional or balancing selection acting on the

Table 4. HKA test data showing the number of sites that are polymorphic within *D. melanogaster* and the mean pairwise divergence from *D. simulans*

Gene	Inheritance	Length excl. gaps	No. alleles	Intraspecific polymorphism	Interspecific divergence
PGRP-SA	X	697	12	4	26
PGRP-SB1	Autosome	627	12	20	24.5
PGRP-SB2	Autosome	607	12	13	27.92
PGRP-SC1A	Autosome	558	12	3	11.33
PGRP-SC1B	Autosome	558	12	7	10.92
PGRP-SC2	Autosome	555	12	6	22.83
PGRP-SD	Autosome	561	12	7	15.92

Table 5. McDonald–Kreitman Test[a]

		Intraspecific polymorphism			Interspecific fixed differences			
Gene	No. alleles	Synonymus coding	Synonymous intron	Replacement	Synonymus coding	Synonymous intron	Replacement	*p*
PGRP-SA	12	3	1	0	11	9	5	0.57
PGRP-SB1	12	13	4	3	9	3	4	0.68
PGRP-SB2	12	7	2	4	17	2	5	0.69
PGRP-SC1A	12	2	—	1	8	—	3	1
PGRP-SC1B	12	5	—	2	6	—	3	1
PGRP-SC2	12	2	—	4	18	—	2	0.02
PGRP-SD	12	3	—	4	10	—	3	0.17
All genes	12	35	7	18	79	14	25	0.27

[a] The probability (*p*) of observing the data if all the observed mutations are neutral is calculated using a Fisher's exact test, and includes synonymous sites both from within the coding sequence and from the introns.

extracellular PGRP proteins in *D. melanogaster*. This result is in striking contrast to the acquired immune system, where genes involved in parasite recognition, such as MHC molecules and immunoglobulins, are subject to both directional and balancing selection (Groot et al. 2002; Hughes and Nei 1988; Su and Nei 1999; Tanaka and Nei 1989).

Instead, our study suggests that parasite recognition genes in the *Drosophila* innate immune system evolve in the way envisaged by Medzhitov and Janeway (1997). A simple explanation would be that the innate immune system does recognize conserved motifs in invading pathogens through molecular patterns that are conserved across a broad taxonomic range of parasites and represent a general surveillance mechanism. But how then do we explain observations of genetic diversity for pathogen resistance within populations of *Daphnia* and *Anopheles* and variation in the mechanism of detection and resistance between *Drosophila* and *Anopheles*?

The first possible explanation is that *Drosophila* is not a general model for the evolutionary genetics of immunity. It could be argued that *Anopheles*, as a blood feeder, is exposed to a range of pathogens that specifically target entry and proliferation into mosquitoes, as the insect can act as a vector. These pathogens may be specialized upon a few related host species, and they may be common within particular mosquito populations, producing host–pathogen interactions that are likely to coevolve. In contrast, being a nonbiting species, *Drosophila* may not face strong coevolutionary interactions, rather being challenged by a wide variety of pathogens, each of which interacts rarely with *Drosophila*, and are therefore not particularly specialized upon it. In common parlance, *Anopheles* is subject to tight coevolution with its pathogens, of an arms race type, and *Drosophila* is subject to diffuse coevolution with an array of nonspecialist pathogens (Janzen 1980).

The second possibility is that the specificity that is observed in host–parasite interactions does not arise from pathogen detection via PGRP in any system but that the PGRP detection system is only responsible for dealing with rare opportunistic infections as a general mechanism. In addition, some of the PRGP proteins may function as scavengers rather than receptors, as we discussed in the Introduction. In this case, the pattern exhibited for PGRP will be general across taxa, but other detection/reaction mechanisms will be responsible for creating the specificity observed in natural populations.

There have been several studies similar to this one that have investigated patterns of variation in genes encoding the effector molecules rather than receptors. In *Drosophila* cecropins, attacins, and diptericin, all of which are antimicrobial peptides, are mostly conserved and do not have the high levels of variation observed in some vertebrate immune genes (Clark

and Wang 1997; Date et al. 1998; Lazzaro and Clark 2001; Ramos-Onsins and Aguade 1998; Zhou et al. 1997). There are, however, some indications that these genes may sometimes be subject to positive selection, as the pattern of variation at the *diptericin* locus differs from neutrality, and one *attacin* allele has recently increased in frequency (Clark and Wang 1997; Lazzaro and Clark 2001). A further case occurs in mammals, where the defensin family of antimicrobial peptides has been relatively recently duplicated, and then these copies diversified under positive selection (Hughes and Yeager 1997). This may indicate occasional coevolution between parasites and the effector molecules of the innate immune system.

Differentiation between these hypotheses requires comparative analysis of PGRP evolution in different taxa. If PGRP genes are under purifying selection in mosquitoes as well as *Drosophila*, then it will be fair to accept that PGRP genes represent a surveillance mechanism for detecting opportunistic infections and are not involved in creating the specificity observed in natural populations. If, on the other hand, PGRP genes are under balancing or directional selection in *Anopheles*, then we would suggest that the lack of directional selection on PGRP in *Drosophila* is likely to be a result of lack of strongly coevolved interactions in this system. We would then predict that species that can act as vectors, such as blood and phloem feeding insects, will in general show elevated rates of evolution in PGRP genes, whilst PGRP genes in others are likely to be under purifying selection.

The two tandemly arranged PGRP-SC1 loci are subject to high rates of concerted evolution, which has homogenized the copies within *D. simulans* and *D. melanogaster*. Similar patterns have been observed in two other multicopy *Drosophila* immune gene families, *Attacins A* and *B*, and the *Cecropin* genes (Date et al. 1998; Lazzaro and Clark 2001; Ramos-Onsins and Aguade 1998). We do not know the cause of gene conversion in PGRP-SC1, but it could have important consequences for the evolution of these genes if, in other insects, they are subject to strong positive selection exerted by parasites. Beneficial mutations arising at one locus can spread to other loci by gene conversion, accelerating the rate of adaptive evolution (Arnheim 1983). Furthermore, multiple-copy genes will result in higher levels of genetic variation in the population when summed across the loci, which may also increase the ability of the population to respond to selection (Li 1997). However, gene conversion can also oppose the divergence of duplicated genes, which may prevent adaptation to a diverse range of parasites (Li 1997).

Acknowledgments. We wish to thank Dr. John Brookfield for supplying *Drosophila* lines and two anonymous reviewers for helpful comments. F.M.J. is funded by a Research Fellowship from Emmanuel College, Cambridge, and a Natural Environment Research Council small grant. G.D.D.H. is supported by a Biotechnology and Biological Sciences Research Council D. Phillips Fellowship.

References

Adams MD, Celniker SE, Holt RA, et al. (2000) The genome sequence of *Drosophila melanogaster*. Science 287:2185–2195

Arnheim N (1983) Concerted evolution of multigene families. In: Nei M, Koehn RK (eds) Evolution of genes and proteins. Sinauer Associates, Sunderland, MA, pp 33–61

Carius HJ, Little TJ, Ebert D (2001) Genetic variation in a host-parasite association: Potential for coevolution and frequency-dependent selection. Evolution 55:1136–1145

Choe KM, Werner T, Stoven S, Hultmark D, Anderson KV (2002) Requirement for a peptidoglycan recognition protein (PGRP) in relish activation and antibacterial immune responses in *Drosophila*. Science 296:359–362

Christophides GK, Zdobnov E, Barillas-Mury C, et al. (2002) Immunity-related genes and gene families in *Anopheles gambiae*. Science 298:159–165

Clark AG, Wang L (1997) Molecular population genetics of *Drosophila* immune system genes. Genetics 147:713–724

Date A, Satta Y, Takahata N, Chigusa SI (1998) Evolutionary history and mechanism of the *Drosophila* cecropin gene family. Immunogenetics 47:417–429

De Gregorio E, Spellman PT, Rubin GM, Lemaitre B (2001) Genome-wide analysis of the *Drosophila* immune response by using oligonucleotide microarrays. Proc Natl Acad Sci USA 98:12590–12595

Doherty PC, Zinkernagel RM (1975) Enhanced immunologic surveillance in mice heterozygous at the *H-2* gene complex. Nature 256:50–52

Fu Y-X, Li W-H (1993) Statistical tests of neutrality of mutations. Genetics 133:693–709

Gottar M, Gobert V, Michel T, Belvin M, Duyk G, Hoffmann JA, Ferrandon D, Royet J (2002) The *Drosophila* immune response against Gram-negative bacteria is mediated by a peptidoglycan recognition protein. Nature 416:640–644

Groot NGd, Otting N, Doxiadis GGM, Balla-Jhagjhoorsingh SS, Heeney JL, Rood JJv, Gagneux P, Bontrop RE (2002) Evidence for an ancient selective sweep in the MHC class I gene repertoire of chimpanzees. Proc Natl Acad Sci USA 99:11748–11753

Hudson RR, Kreitman M, Aquade M (1987) A test of neutral molecular evolution based on nucleotide data. Genetics 116:153–159

Hughes AL, Nei M (1988) Pattern of neucleotide substitution at major histocompatibility complex Class-I loci reveals overdominant selection. Nature 335:167–170

Hughes AL, Yeager M (1997) Coordinated amino acid changes in the evolution of mammalian defensins. J Mol Evol 44:675–682

Irving P, Troxler L, Heuer TS, Belvin M, Kopczynski C, Reichhart JM, Hoffmann JA, Hetru C (2001) A genome-wide analysis of immune responses in *Drosophila*. Proc Natl Acad Sci USA 98:15119–15124

Janzen DH (1980) When is it coevolution? Evolution 34:611–612

Lazzaro BP, Clark AC (2001) Evidence for recurrent paralogous gene conversion and exceptional allelic divergence in the *Attacin* genes of *Drosophila melanogaster*. Genetics 159:659–671

Li W-H (1997) Molecular evolution. Sinauer Associates, Sunderland MA

McDonald JH, Kreitman M (1991) Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature 351:652–654

Medzhitov R, Janeway CA (1997) Innate immunity: The virtues of a nonclonal system of recognition. Cell 91:295–298

Mellroth P, Karlsson J, Steiner H (2003) A scavenger function for *Drosophila* peptidoglycan recognition protein. J Biol Chem 278:7059–7064

Michel T, Reichhart JM, Hoffmann JA, Royet J (2001) *Drosophila* Toll is activated by Gram-positive bacteria through a circulating peptidoglycan recognition protein. Nature 414:756–759

Moriyama EN, Powell JR (1996) Intraspecific nuclear DNA variation in *Drosophila*. Mol Biol Evol 13:261–277

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Niare O, Markianos K, Volz J, Oduol F, Toure A, Bagayoko M, Sangare D, Traore SF, Wang R, Blass C, Dolo G, Bouare M, Kafatos FC, Kruglyak L, Toure YT, Vernick KD (2002) Genetic loci affecting resistance to human malaria parasites in a West African mosquito vector population. Science 298:213–216

Ramet M, Manfruelli P, Pearson A, Mathey-Prevot B, Ezekowitz RAB (2002) Functional genomic analysis of phagocytosis and identification of a *Drosophila* receptor for *E. coli*. Nature 416:644–648

Ramos-Onsins S, Aguade M (1998) Molecular evolution of the *Cecropin* multigene family in *Drosophila:* Functional genes vs. pseudogenes. Genetics 150:157–171

Su C, Nei M (1999) Fifty-million-year-old polymorphism at an immunoglobulin variable region gene locus in the rabbit evolutionary lineage. Proc Natl Acad Sci USA 96:9710–9715

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

Tanaka T, Nei M (1989) Positive Darwinian selection observed at the variable-region genes of immunoglobulins. Mol Biol Evol 6:447–459

Watterson GA (1975) On the number of segregating sites in models without recombination. Theor Pop Biol 7:256–276

Werner T, Liu G, Kang D, Ekengren S, Steiner H, Hultmark D (2000) A family of peptidoglycan recognition proteins in the fruit fly *Drosophila melanogaster*. Proc Natl Acad Sci USA 97:13772–13777

Yoshida H, Kinoshita K, Ashida M (1996) Purification of a peptidoglycan recognition protein from hemolymph of the silkworm, *Bombyx mori*. J Biol Chem 271:13854–13860

Zhou XJ, Nguyen T, Kimbrell DA (1997) Identification and characterization of the *Cecropin* antibacterial protein gene locus in *Drosophila virilis*. J Mol Evol 44:272–281