Insect Molecular Biology (2005) **14**(6), 599–605

The evolution of immune-related genes from disease carrying mosquitoes: diversity in a peptidoglycan- and a thioester-recognizing protein

T. J. Little* and N. Cobbe†

**Institute of Evolutionary Biology, School of Biology, University of Edinburgh, Edinburgh, UK;* †*Wellcome Trust Centre for Cell Biology, University of Edinburgh, University of Edinburgh, Edinburgh, UK*

Abstract

Adaptive polymorphism may be common in immune system genes as co-evolutionary interactions foster diversity; either through ongoing positive selection (arms races), or balancing selection. DNA sequence diversity in two putative immune system genes was examined in species of the genus *Anopheles* and from *Aedes aegypti*. For one gene, encoding the peptidoglycan recognizing protein PGRPLB, there was evidence of purifying selection, suggesting that selection acts to eliminate sequence variation. For another gene, encoding the thioester-containing protein TEP3, higher levels of amino acid replacement were found than would be expected under neutral models of evolution – an indication that this gene has been subject to repeated bouts of positive selection.

Keywords: selection, *Anopheles*, malaria, arms race, innate immunity, pathogen, parasite.

Introduction

Genome projects and expression studies of invertebrates have greatly expanded knowledge of the innate immune system. For example, genes that the insects *Drososphila* and *Anopheles* up- or down-regulate in response to pathogen invasion have been extensively studied (Oduol *et al*., 2000; Christophides *et al*., 2002, 2004; Holt *et al*., 2002; Osta *et al*., 2004). However, the importance of genetic polymorphism in the immune-related genome is not yet clear. Polymorphism of immune-related genes may reflect phenotypic variation in disease burden or vector competence. Compared to other parts of the genome, adaptive polymorphism may be common in immune system genes, as coevolutionary interactions with pathogens or parasites promote rapid, ongoing adaptation (Schlenke & Begun, 2003).

Analyses of DNA polymorphism typically focus on the detection of positive or balancing selection, both of which foster diversity, albeit at different levels (Yang & Bielawski, 2000; Ford, 2002; Olson, 2002). Positive selection, for example, is evident as an elevated rated of amino acid divergence among species, accompanied by a lack of heterozygosity within species (Hurst & Smith, 1999; Bishop *et al*., 2000; Ford, 2002; Olson, 2002). This arises when coevolution drives repeated bouts of selective sweeps, a process often termed an 'arms race'. Alternatively, antagonistic coevolution can maintain polymorphism through balancing selection, which is evident as high heterozygosity and/or the deep divergence of alleles at single loci within populations (Hill *et al*., 1991, 1992; Hughes & Nei, 1992; Stahl *et al*., 1999; Stahl & Bishop, 2000). Arms races or balancing selection may act simultaneously (Bergelson *et al*., 2001), but both are easily distinguished from purifying selection as this generates gene sequences with low levels of amino acid polymorphism, because new mutations are pruned from the population.

Among the arthropoda, immune genes from *Drosophila* and the crustacean *Daphnia* have been subject to molecular population genetic analyses of polymorphism. A genome-wide study by Schlenke & Begun (2003) comparing *Drosophila melanogaster* to *Drosophila simulans* indicated that immune system genes are subject to positive selection to a greater extent than are other parts of the genome. When particular immune genes have been the target of study in *Drosophila*, results have been mixed: genes for antimicrobial peptides and peptidoglycan recognizing proteins largely showed evidence of purifying selection (Jiggins & Hurst, 2003; Lazzaro & Clark, 2003), but the transcription factor relish gave evidence of positive selection (Begun & Whitley, 2000). Studies of *Daphnia* concerned two genes, one of which, a gram negative binding protein gene showed evidence of purifying selection, while another

doi: 10.1111/j.1365-2583.2005.00588.x

Received 12 April 2005; accepted after revision 23 May 2005.

Correspondence: Tom Little, Institute of Evolutionary Biology, School of Biology, University of Edinburgh, Kings Buildings, Edinburgh, EH9 3JT, UK. Tel.: +44 131 6507781; fax: +44 131 6506465; e-mail: Tom.Little@ed.ac.uk

```
                 **-**--**--**--****---------*----****------*--------**-----
An. arabiensis   GLQCIAAMQSMQKMHQDVRQWNDIGYSFAVGGDGRVYQGRGFNVIGAHAPRYNNRSVGIC  60
An. atroparvus   GLKCIEAMQQMQKMHQEQRLWNDIGYSFAVGGDGRVYQGRGFNVIGAHAPRYNNRSVGIC  60
An. gambiae      GLQCIAAMQSMQKMHQDERQWNDIGYSFAVGGDGHVYQGRGFNVIGAHAPRYNNRSVGIC  60
An. stephensi    GLQCIAAMQQMQKMHQEVRQWNDIGYSFAVGGDGRICQGRGFNVVGAHAPRYNNRSVGIC  60
An. funestus     GLQCIAAMQQMQKMHQEVRQWNDIGYSFAIGGDGRVYQGRGFNVVGAHAPRYNDRSVGIC  60
An. albimanus    GTHCIEAMQQMQTMHQDVRSWNDIGYSFAVGGDGRIYQGRGFNVVGAHAPRYNNKSVGIC  60
Aedes aegypti    GSECCEAMLSRQKFHQQDRGWNDIGYSFAVGGDGRVYEGRGFNVVGAHAPRYNDKSVGIC  60

                 *-----***---**-**-***---------***-**--*********-***---*------
An. arabiensis   LIGDW AADLPPKNMLTAAQNLIEYGVRNGLIAQNYTLLGHRQVRTTECPGDRLFEEIKTWP 120
An. atroparvus   LIGDW AADLPPKNMRAATQQLIEYGVRHGIIAQNYTLLGHRQVRSTECPGDRLFEEIKTWP 120
An. gambiae      LIGDW VADLPPKNMLTAAQNLIEYGVRNGLIAQNYTLLGHRQVRTTECPGDRLFEEIKTWP 120
An. stephensi    LIGDW PADLPPKNMLAAAKNLIEYGVRNGLIAQNYTRIAHRQVRTTECPGDRLFKEIKTWP 120
An. funestus     LIGDW TANLPPKNMLTAAQNLIEYGVRNSLIAQNYTLLGHRQVRSTECPGDRLFEEIKTWP 120
An. albimanus    LIGDW IAELPPKKMLAATKTLIEYGVRNGIIASNYTLLGHRQVRIDGCPGRRLFEEIKTWP 120
Aedes aegypti    MIGDW LVDLPPENMIAAAQSLIEYGVRNNIIASNYTLIGHRQVRATECPGERLFKEIKTWP 120
```

Figure 1. The deduced amino acid sequences of DNA encoding a peptidoglycan recognizing protein (PGRPLB) for six species of *Anopheles* mosquitoes and *Ae. aegypti*. Asterisks indicate sites with non-synonymous nucleotide substitutions. Introns are indicated by gaps.

gene, an *alpha-2-macroglobulin*, showed evidence of positive selection (Little *et al*., 2004).

Mosquito–pathogen interactions, as with most hosts and pathogens (Little, 2002), show genetic variation for parasite or pathogen resistance (Hogg & Hurd, 1995; Yan *et al*., 1997; Bosio *et al*., 2000; Ferguson & Read, 2002). The genetic basis of resistance-variation is yet to be fully described, although progress has been made regarding the genes that determine the *Plasmodium* sp. invasion success (Levashina *et al*., 2001; Osta *et al*., 2004). This study presents the first molecular population genetic analysis of immune system genes from disease-carrying mosquitoes. Levels of polymorphism and rates of amino acid substitution were determined in two immune system genes, with the aim of establishing if these genes have been under purifying, directional or balancing selection. By analogy with plant and vertebrate systems, host proteins, which recognize and/ or directly interact with pathogens, are prime candidates for the detection of adaptive polymorphism (e.g. Hill *et al*., 1991; Stahl & Bishop, 2000) due to balancing selection or directional selective sweeps. Therefore, genes that coded for proteins known to be involved in the recognition of invading pathogens were isolated for study. Specifically, DNA polymorphism at a peptidoglycan recognizing protein (PGRP) encoding gene and a thioester containing protein (TEP) encoding gene were analyzed.

Results

Polymorphism and divergence at PGRPLB

An ~500 base pair (bp) fragment of *PGRPLB* was amplified in *Ae. aegypti* and each *Anopheles* species. To this was added sequence information from the published genome of *Anopheles gambiae*. Thus, our data set for comparisons among species at *PGRPLB* contained seven taxa (*An. gambiae*, *Anopheles arabiensis*, *Anopheles albimanus*, *Anopheles atroparvus*, *Anopheles funestus*, *Anopheles stephensi*, and *Ae. aegypti* (Fig. 1)).

Comparing among these species, K_a:K_s ratios at *PGRPLB* averaged 0.068, with pair-wise comparisons ranging from 0.029 (*An. arabiensis* vs. *An. funestus*) to 0.23 (*An. arabiensis* vs. *An. gambiae*). The CODEML program of PAML and the gene tree of these seven taxa were used to test if K_a:K_s ratios were above 1.0 at particular codons. It was determined that the model incorporating among site variation was significantly more likely than one that assumed all sites to be evolving at the same rate (e.g., comparing NSsites M0 [log likelihood = –3431.52] to M2 [log likelihood = –3409.06], $\chi^2 = 45$, d.f. = 2, $P < 0.0001$). To test for positive selection a likelihood-ratio test was used to compare M7 – a model for which ω is described by a beta distribution and does not allow for positively selected sites ($\omega < 1.0$) – to M8, which permits a proportion of sites to fall outside this beta distribution (Yang *et al*., 2000), i.e. some sites may have $\omega > 1.0$. The likelihood ratio test comparing these two models was not significant (M7 [log likelihood = –3409.31] to M8 [log likelihood = –3408.92], $\chi^2 = 0.78$, d.f. = 2, n.s.).

Sample sizes and the number of polymorphic sites in the African populations are given in Table 1. Basic polymorphism and divergence data are shown in Table 2. Within population tests for selection (Tajima's D, Fu and Li's D) in *An. gambiae* (from the Cameroon or Kenyan populations) and *An. funestus* (from the Sudanese population) were not significant. Comparing among populations of *An. gambiae* from Kenya and Cameroon revealed ample polymorphism within each population (Tables 2 and 3), but surprisingly for

Table 1. Polymorphism parameters for DNA sequences encoding a peptidoglycan recognizing protein (PGRPLB) and a thioester containing protein (TEP3) from three populations of malaria-carrying mosquitoes in the genus *Anopheles*

Gene	Species (population)	Total length (bp)	Exon length (bp)	Number of individuals	Number of haplotypes	Polymorphic sites	
						Silent	Replacement
PGRPLB	*An. gambiae* (Cameroon)	492	408	7	13	17	4
	An. gambiae (Kenya)	492	408	5	10	8	4
	An. funestus (Sudan)	492	408	9	15	8	0
TEP3	*An. gambiae* (Cameroon)	1443	1287	9	18	37	0
	An. gambiae (Kenya)	1465	1287	6	12	34	0
	An. funestus (Sudan)	1024	948	5	10	18	5

Table 2. Nucleotide polymorphism (π) and divergence parameters comparing three populations of malaria-carrying mosquitoes in the genus *Anopheles* at two immune-related genes (PGRPLB and TEP3)

Gene	Population	Total	Silent	Replacement
Polymorphism				
PGRPLB	*An. gambiae* (Cameroon)	0.017	0.053	0.004
	An. gambiae (Kenya)	0.010	0.027	0.005
	An. funestus (Sudan)	0.009	0.037	0.000
TEP-3	*An. gambiae* (Cameroon)	0.011	0.045	0.000
	An. gambiae (Kenya)	0.011	0.045	0.000
	An. funestus (Sudan)	0.007	0.024	0.002
Divergence				
PGRPLB	*An. gambiae* (Cameroon vs. Kenya)	0.016	0.042	0.007
	An. gambiae (Cameroon) vs. *An. funestus*	0.176	0.600	0.044
	An. gambiae (Kenya) vs. *An. funestus*	0.175	0.592	0.044
TEP-3	*An. gambiae* (Cameroon vs. Kenya)	0.010	0.042	0.000
	An. gambiae (Cameroon) vs. *An. funestus*	0.173	0.562	0.056
	An. gambiae (Kenya) vs. *An. funestus*	0.171	0.552	0.056

Table 3. Polymorphism and divergence data used to compare populations of malaria-carrying mosquitoes for two immune-related genes (PGRPLB and TEP3). The MacDonald–Kreitman Test was used to test for positive selection by comparing *An. funestus* to each of two populations of *An. gambiae.* The probability of observing the data if all the observed mutations are neutral is calculated using a contingency table analysis and a G-test, and only includes sites from the coding sequence. NA indicates not applicable because populations are of the same species

	Intraspecific polymorphism		Interspecific fixed differences			
	Silent	Replacement	Silent	Replacement	G	*P*
PGRP-LB						
An. gambiae (Cameroon vs. Kenya)	20	7	0	0	NA	
An. gambiae (Cameroon) vs. *An. funestus*	22	4	46	11	0.189	0.66
An. gambiae (Kenya) vs. *An. funestus*	14	4	50	11	0.154	0.69
TEP-3						
An. gambiae (Cameroon vs. Kenya)	45	0	0	0	NA	
An. gambiae (Cameroon) vs. *An. funestus*	44	4	109	34	6.19	0.013
An. gambiae (Kenya) vs. *An. funestus*	42	5	108	34	4.26	0.039

populations separated by < 2000 km, not a single fixed difference between populations (Table 3). A MacDonald–Kreitman test comparing the *An. funestus* population to the two populations of *An. gambiae* indicated that divergence between species was not due to positive selection (Table 3), i.e. there was no apparent excess of non-synonymous substitutions. In summary, the patterns of polymorphism and divergence at PGRPLB suggest the action of purifying selection.

Polymorphism and divergence at TEP3

1645 bp of sequence information from the *TEP3* were obtained for each of *An. gambiae*, *An. arabiensis*, *An. atroparvus*, and *An. stephensi* (Fig. 2). For *An. funestus,* sequence information from the first 1024 bp only was obtained. It was not possible to acquire PCR product from other species. K_a:K_s ratios at this gene were similar to PGRPLB averaging 0.061, with pair-wise comparison

```
                -------*-------*---*-*---*-*-*-------**--------------------*-*---*--*------
An. arabiensis VRFRVLVLDPNTKPLQKADNISVHINDAKANRIKQWKEGKLVKGVFESELTLSTAPVLGAWTINVEVLGS KHNKV   75
An. atroparvus VRFRVLVLDPNTKPLLKADRINVHITDGKGNRIKQWKAAKLVKGVFESELTLSTAPVLGTWTINVEALGS KHNKV   75
An. gambiae    VRFRVLVLDPNTKPLQKADNISVHINDAKANRIKQWKEGKLVKGVFESELTLSTAPVLGAWTINVEVLGS KHNKV   75
An. stephensi  VRFRVLVPDPNTKPLPKADSINVHINDAKANRIKQWKDGKLVKGVFESELTLSTAPVLGAWTINVEVLGT KHNKV   75
An. funestus   VRFRVLVLDPNTKPLQKADNISVHINDAKANRIKQWKEGKLVKGVFESELTLSTAPVLGAWSINVNVLGT KHNKV   73

                ------------------------------*-*-----------------*----*-------*---*----*---
An. arabiensis FEVDEYVLPKFEVTVESPGITTFKDGKVKAIIRSKYTYGKPVKGEATVSVSPEFQFHYVQPFAKDVITRKVIPID   150
An. atroparvus FEVDEYVLPKFEVTVESPGITTFKDGKVKVIVRSKYTYGKPVKGEATVSISPEFRFHYVQPFEKDVTTRKVIPID   150
An. gambiae    FEVDEYVLPKFEVTVESPGITTFKDGKVKAIIRSKYTYGKPVKGEATVSVSPEFQFHYVQPFAKDVITRKVIPID   150
An. stephensi  FEVDEYVLPKFEVTVESPGITTFKDGKVKAIIRSKYTYGKPVKGEATVSASPEFQFHYVQPFAKDVITRKVVPID   150
An. funestus   FEVDEYVLPKFEVTVESPGITTFKDGKVKAIIRSKYTYGKPVKGEATVSASPEFGFHYVQPFAKDVITRKVVPID   148

                ----------***-*---*-**----------------***--*-*-----*--*-------------*----*
An. arabiensis GKGSVEFDLREDIHLEGDYSRNIVIEAVVEEELTGRKQNASAKVMIYDRRYKMELVKSDDNFKPGLPYTAWLKVS   225
An. atroparvus GKGSVEFDLREELRLDGDYTRTVVIEAVVEEELTGRKQNTTGKVTIHDRRYKIELVKSDDNFKPGLPYTAWLKVT   225
An. gambiae    GKGSVEFDLREDIHLEGDYSRNIVIEAVVEEELTGRKQNASAKVMIYDRRYKMELVKSDDNFKPGLPYTAWLKVS   225
An. stephensi  GKGSVEFDLREELRLEGDYTRNIVIEAVVEEELTGRKQNASAKVMIYDRRYKMELIKSDDNFKPGLPYTAWLKVT   225
An. funestus   GKGSVEFDLREELRLEGDYTRNIVIEAVVEEELTGRKQNASAKVMIYDRRVKMELVKSDDNFKPGLPYTTWLKVT   223

                *---*-----*--------*********-----------*-*-**--*---***-----*---*-----*----
An. arabiensis YQDGAPVQDQTNPVEVKQSSFESTTSVQNYTLDQNGMAKLEINTEVNSSYINVVGVYLGQEFYLHGISKAESDVD   300
An. atroparvus YQDGTPVQDQTNPVEVKQSFYGSDEISQNYTLDQNGMAKLVIDTAGNSSYINMNAVYLGQNFYLNGISKADSDVD   300
An. gambiae    YQDGAPVQDQTNPVEVKQSSFESTTSVQNYTLDQNGMAKLEINTEVNSSYINVVGVYLGQEFYLHGISKAESDVD   300
An. stephensi  YQDGAPVQDQTNPVEVKQSTYESTQFLRNYTLDQNGMAKLEIHTDVNSTYINVVGVYLGQEFYLSGISKADSDVD   300
An. funestus   HQDGTPVQDQVNPVEVKQSTYESSQFVRNYTLDQNGMAKLDINTDVNSTYINVVGVYLGQEFYLSGISKADSDVD   298

                -----*-*-------**--*---**-*---*-------------*--*----*-***---*-------------*-
An. arabiensis SYIRAQVLTEM PLVGKDVLVEVTSTSPMKYFTYQLLGRGDVLLSNTIAVPESKTQSFKFPATFAMVPRAKLVVYY   375
An. atroparvus SYIRARVLTEL PLVNKDVVVEVAATSPMKYFTYQLLGRGDVLLGNTVAVPETKIHTFKFLATFAMVPRAKLVVYY   375
An. gambiae    SYIRAQVLTEM PLVGKDVLVEVTSTSPMKYFTYQLLGRGDVLLSNTIAVPESKTQSFKFPATFAMVPRAKLVVYY   375
An. stephensi  SYIRARVLTEM PLVGSDVVVEVSATAPMKFFTYQLLGRGDVLLSNTVAVPESKTHKFKFPATFAMVPRAKLVVFY   375
Sudan          SYIRARVVTEM ------------------------------------------------------------   309

                ------------*-----------------*-*-----*----------------------
An. arabiensis IAPNGDMVSDSKVITFDSELQNFMKVSLSKEQSKPGQDG----------------------   414
An. atroparvus IAPNGDMVSDSKEITFDSELQNFMKVSLSKDQTKPGQDVEISISTNPDSYV---------   426
An. gambiae    IAPNGDMVSDSKVITFDSELQNFMKVSLSKEQSKPGQDVEISISTNPDSYVGLLGVDQSV   435
An. stephensi  IAPNGDMVSDSKVITFDSELQNFMKVSLSKDQTKPGQDVEISISTNPDSYVGLLGVDQSV   434
Sudan          ------------------------------------------------------------   309
```

Figure 2. The deduced amino acid sequences of DNA encoding a thioester containing protein (tep3) for five species of *Anopheles* mosquitoes. Asterisks indicate sites with non-synonymous nucleotide substitutions. Introns are indicated by gaps.

ranging from 0.028 (*An. arabiensis* vs. *An. gambiae*) to 0.11 (*An. arabiensis* vs. *An. atroparvus*). Given the small number of species for which data were available, further tests for elevated K_a:K_s ratios in PAML were not performed.

Sample sizes and the number of polymorphic sites in the three populations samples are given in Table 1. Basic polymorphism and divergence data are in Table 2. Within population tests for selection (Tajima's D, Fu and Li's D) were not significant. As with PGRPLB, comparing among populations of *An. gambiae* from Kenya and Cameroon at TEP3 revealed ample polymorphism within each population (Tables 2 and 3), but not a single fixed difference between populations (Table 3). Comparisons between either of these populations and the population of *An. funestus* from Sudan indicated strong divergence. By contrast with the analysis of PGRPLB, however, a MacDonald–Kreitman test comparing the *An. funestus* population to *An. gambiae* from either Kenya or Cameroon indicated the presence of positive selection (Table 3). In particular, there appeared to be an excess of amino acid replacements contributing to divergence (Table 3), although our estimate of K_s may be biased by the sizable evolutionary distance separating *An. funestus* from *An. gambiae*.

Discussion

This study documents the first DNA polymorphism data from the immune system genes of *Anopheles* mosquitoes. A number of intriguing patterns of polymorphism and divergence were particularly notable. First, populations of *An. gambiae* may show almost no divergence across a distance of over 2000 km. This confirms previous studies on mitochondrial DNA (Besansky *et al.*, 1997) and suggests that *Anopheles* populations can exist as very large panmictic

entities. Secondly, in tests for the imprint of natural selection, a gene encoding a peptidoglycan recognizing protein, showed evidence of purifying selection, which indicates that selection tends to eliminate amino acid variation at this gene. However, the other studied gene, encoding a thioester containing protein, appeared to be under positive selection, which is evidence that this gene is part of a host-pathogen arms race. Specifically, for TEP3, the ratio of replacement to silent substitutions between species was significantly greater than expected from levels of within-species polymorphism (The MacDonald–Kreitman Test) when comparing the *An. funestus* population to either of two widely separated populations of *An. gambiae*. Neither the PGRPLB or TEP3 encoding genes showed evidence of balancing selection as some immune system genes of plants and vertebrates do.

This evidence for positive selection in one gene but not the other may be linked to function. Peptidoglycans are an essential component of bacterial cell membranes, and as such cannot be changed to be unrecognizable by the host (Janeway & Medzhitov, 2002). If peptidoglycans cannot be changed or removed as to make the bacteria undetectable, then their will be no evolutionary pressure on hosts to modify their PGRPs (Medzhitov & Janeway, 1997; Janeway & Medzhitov, 2002). This hypothesis is supported by the present data, as well as a study on *Drosophila* (Jiggins & Hurst, 2003), which found PGRPs to be subject only to purifying selection. Moreover, the study of a *Daphnia* gram negative binding protein (GNBP) encoding gene, which also recognizes essential carbohydrate-based pathogen-associated molecular patterns, also showed evidence of purifying selection.

TEPs are serine protease inhibitors and thus recognize and bind proteins. In general, gene products involved in protein–protein interactions (as opposed to the protein–carbohydrate interactions typified by PGRPs or GNBPs) seem more likely candidates for arms races. The precise function of *Anopheles* TEP3 is not well established, although an important role for TEP-1 in *Anopheles* immunity has been shown (Levashina *et al*., 2001; Blandin *et al*., 2004), and the role of TEPs in immunity of other invertebrate taxa is also well studied (Armstrong & Quigley, 1999; Kanost, 1999). For example TEP family members called alpha-2-macroglobulins are involved in the recognition of pathogen serine proteases, which are important virulence factors (Han *et al*., 2000). A study of a *Daphnia* alpha-2-macroglobulin gene showed evidence of positive selection indicating that it was subject to a host parasite arms race centred on host evolution to produce TEPs that inhibit parasite serine proteases and parasite evolution to produce serine proteases that go unrecognized by hosts. Diversity in alpha-2-macroglobulins is highest in the bait region (which directly interacts with, and is cleaved by, serine proteases), but *Anopheles* TEP3, despite showing evidence of positive selection, did not show such a concentration of amino acid substitutions similar to the bait region in *Daphnia* (Fig. 2). Nevertheless, given the combined results of the *Daphnia* study and the present one, it may be tempting to speculate that TEP's are common foci of coevolutionary arms races.

Presently, there are few available data on polymorphism in arthropod immune-genes to draw robust conclusions about which parts of the immune-related genome are under positive or other forms of diversifying selection. It is hoped that the genomic era will further include the study of polymorphism at genes known to be turned on in response to pathogen invasion so that we may gain a general impression of tempo and mode of selection on particular immune genes, functional classes of genes, and the entire immune-related genome. The thorough identification of coevolutionary hotspots in the invertebrate immune system looms as an exciting and feasible prospect, at least for those taxa with sequenced genomes. Ultimately, it may even be possible to associate phenotypic patterns of disease burden with immune system polymorphism as has been accomplished in plant-pathogen systems (Stahl *et al*., 1999; Stahl & Bishop, 2000; Bergelson *et al*., 2001)

Experimental procedures

Candidate genes and mosquito sampling

Both TEPs and PGRPs are part of moderately sized multigene families, with a total of 15 TEPs and 6 PGRP genes in the genome of *Anopheles gambiae* (Christophides *et al*., 2002). From GenBank, DNA sequences from all members of each gene family were acquired and alignments made using CLUSTALW. Gene trees were constructed using the neighbour-joining algorithm in PAUP, and from these trees, suitable gene copies identified for study. In particular, stretches of sequence were identified from which PCR primers could be designed that would (1) amplify parts of one gene copy (i.e. one locus) to the exclusion of others – it was important that comparisons be among orthologues – and (2) be sufficiently conserved such that sequence from a variety of species could be studied. For this second aim, inclusion of *Drosophila* genes in this alignment facilitated the identification of conserved sites for the design of degenerate primers. With this approach, analysis of the gene copies identified by Christophides *et al*. (2002) as PGRPLB and TEP3 were carried out.

Primers for the TEP3 locus, which produced a ~1650 bp product, were forward primer 5′-GCAAACGGACAAATCCATCTAC-3′ and reverse primer 5′-ACACCTGCTGCTTCGTAATATC-3′. This fragment was difficult to amplify in some cases, and so two internal primers to amplify smaller fragments were developed. These were forward primer 5′-ATCACCACGTTCAAGGATGG-3′ and reverse primer 5′-GTAGGTGAAGTACTTCATCGG-3′. Primers for the *PGRPLB* locus were forward primer 5′-CCGTACGTCATCATACATCATTCG-3′ and reverse primer 5′-GAAATGGGGCCAGGTTTTGATCTC-3′ which produced a ~500 bp PCR product.

Laboratory strains of *An. arabiensis*, *An. stephensi*, *An. atroparvus*, *An. albimanus*, and *Aedes aegypti*, with *An. gambiae* sequence from GenBank, were used for among-species analyses of TEP3 and PGRPLB. In addition, population samples from three natural populations were aquired. These were *An. gambiae* from Cameroon (the mount Cameroon region), *An. gambiae* from the Kenyan shore

of Lake Victoria (Mbita), and *An. funestus* from Gedaref agricultural plain in eastern Sudan. Between 10 and 20 haplotypes from each population sample were isolated and used for analysis. In any cases where species identification was uncertain, the ITS-2 region was amplified using the primers of Porter & Collins (1991) and sequences compared this PCR fragment with those in GenBank.

Using a modification of the CTAB method (Colbourne *et al*., 1998), DNA was extracted from decapitated adult mosquitoes. PCR amplification conditions for all primers used a 58 C annealing temperature for 30 s with an extension time of 45 s. Magnesium chloride concentration was 2.5 mM in all reactions. PCR reactions produced a single discrete band in each case. Each PCR product was excised from an agarose gel, cleaned using the Qiagen Gel extraction kit, cloned using the TOPO TA 2.1 vector (Invitrogen Life Technologies, Carlsbad, CA, USA). Six clones containing plasmids with mosquito DNA inserts were picked and sequenced in both forward and reverse directions, yeilding 12 sequences from each mosquito, identifying both haplotypes per individual when they were heterozygous. Sequencing was performed with the ABI Prism BigDye Terminator Cycle Sequencing Kit, Version 3.0. Sequences were compared to the non-redundant protein sequence database using the BLASTX alignment program (Altschul *et al*., 1997) to confirm that sequences were indeed TEP3 or PGRPLB.

Analyses

Basic analyses of polymorphism and divergence were performed using the computer program *DNAsp* (Rozas & Rozas, 1999). For each of the three population samples, departures from neutrality were tested using Tajima's D, and Fu and Li's D as implemented in DNAsp. Among species were compared by examining the number of replacement nucleotide substitutions (substitutions that result in amino acid substitutions, K_a) and the number of synonymous (or silent) substitutions (K_s) separating pairs of species. The rate of silent substitution is an (approximately) neutral expectation, and thus K_a/K_s ratios much less than one could indicate purifying selection, but higher ratios may be evidence of positive selection. However, simply examining K_a/K_s ratios is a conservative estimate of positive selection. More powerful procedures to test for elevated K_a/K_s ratios use maximum likelihood approaches, which account for phylogenetic structure, and identify particular codons that have been evolving rapidly (as opposed to averaging across all codons in a pair of sequences) (Yang & Bielawski, 2000). Sequences from each study species were used to generate gene trees using PAUP (Swofford, 2002) and then implemented maximum likelihood methods using the PAML package of programs (Yang, 1997).

Having population polymorpism data for both *An. gambiae* and *An. funestus* enabled the use of the MacDonald–Kreitman test (McDonald & Kreitman, 1991) which detects selection by finding statistical discrepancy in the relative level of synonymous and non-synonymous substitutions within and between species. The basic assumption of MacDonald–Kreitman analysis is that the ratio of replacement to synonymous substitution between species will be the same as the ratio of replacement to synonymous polymorphism within species if divergence and polymorphism are due solely to random drift acting on neutral mutations.

Acknowledgements

This work was financially supported by The Wellcome Trust UK and by the School of Biology, University of Edinburgh. Thanks to Hamza Babiker, Jacamo, Chris Curtis, Barabra Saywer, Samuel Wanji, Heather Ferguson and Judi Allen for help with mosquito collections. Deborah Callister provided laboratory assistance.

References

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Armstrong, P.B. and Quigley, J.P. (1999) Alpha (2) -macroglobulin: an evolutionarily conserved arm of the innate immune system. *Dev Comparative Immunol* **23**: 375–390.

Begun, D.J. and Whitley, P. (2000) Adaptive evolution of relish, a *Drosophila* NF-kappa B/I kappa B protein. *Genetics* **154**: 1231–1238.

Bergelson, J., Kreitman, M., Stahl, E.A. and Tian, D.C. (2001) Evolutionary dynamics of plant R-genes. *Science* **292**: 2281–2285.

Besansky, N.J., Lehmann, T., Fahey, T., Fontenille, D., Braack, L.E.O., Hawley, W.A. and Collins, F.H. (1997) Patterns of mitochondrial variation within and between african malaria vectors, *Anopheles gambiae and An. arabiansis,* suggest extensive gene flow. *Genetics* **147**: 1817–1828.

Bishop, J.G., Dean, A.M. and Mitchell-Olds, T. (2000) Rapid evolution in plant chitinases: molecular targets of selection in plant-pathogen coevolution. *Proc Natl Acad Sci USA* **97**: 5322–5377.

Blandin, S., Shiao, S.H., Moita, L.F., Janse, C.J., Waters, A.P., Kafatos, F.C. and Levashina, E.A. (2004) Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector *Anopheles gambiae*. *Cell* **116**: 661–670.

Bosio, C.F., Fulton, R.E., Salasek, M.L., Beaty, B.J. and Black, W.C. (2000) Quantitative trait loci that control vector competence for dengue-2 virus in the mosquito *Aedes aegypti*. *Genetics* **156**: 687–698.

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

Christophides, G.K., Vlachou, D. and Kafatos, F.C. (2004) Comparative and functional genomics of the innate immune system in the malaria vector *Anopheles gambiae*. *Immunol Rev* **198**: 127–148.

Colbourne, J.K., Crease, T.J., Weider, L.J., Hebert, P.D.N., Dufresne, F. and Hobaek, A. (1998) Phylogenetics and evolution of a circumarctic species complex (Cladocera: *Daphnia pulex*). *Biol J Linn Soc* **65**: 347–365.

Ferguson, H.M. and Read, A.F. (2002) Genetic and environmental determinants of malaria parasite virulence in mosquitoes. *Proc R Soc Lond Series B-Biol Sci* **269**: 1217–1224.

Ford, M.J. (2002) Applications of selective neutrality tests to molecular ecology. *Mol Ecol* **11**: 1245–1262.

Han, Y.S., Thompson, J., Kafatos, F.C. and Barillas-Mury, C. (2000) Molecular interactions between *Anopheles stephensi* midgut cells and *Plasmodium berghei*: the time bomb theory of ookinete invasion of mosquitoes. *EMBO J* **19**: 6030–6040.

Hill, A.V.S., Allsopp, C.E.M., Kwiatkowski, D., Anstey, N.M., Twumasi, P., Rowe, P.A., Bennett, S., Brewster, D., McMichael, A.J. and Greenwood, B.M. (1991) Common West African HLA antigens are associated with protection from severe malaria. *Nature* **352**: 595–600.

Hill, A.V.S., Kwiatkowski, D., McMichael, A.J., Greenwood, B.M. and Bennett, S. (1992) Maintenance of MHC polymorphism – reply. *Nature* **355**: 403–403.

Hogg, J.C. and Hurd, H. (1995) Malaria-Induced Reduction of Fecundity During the First Gonotrophic Cycle of *Anopheles stephensi* Mosquitos. *Med Vet Entomol* **9**: 176–180.

Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R., Nusskern, D.R. *et al.* (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* **298**: 129–149.

Hughes, A.L. and Nei, M. (1992) Maintenance of MHC polymorphism. *Nature* **355**: 402–403.

Hurst, L.D. and Smith, N.G.C. (1999) Do essential genes evolve slowly? *Curr Biol* **9**: 747–750.

Janeway, C.A. and Medzhitov, R. (2002) Innate immune recognition. *Annu Rev Immunol* **20**: 197–216.

Jiggins, F.M. and Hurst, G.D.D. (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.

Kanost, M.R. (1999) Serine proteinase inhibitors in arthropod immunity. *Dev Comparative Immunol* **23**: 291–301.

Lazzaro, B.P. and Clark, A.G. (2003) Molecular population genetics of antibacterial peptide genes in *Drosophila melanogaster*. *Mol Biol Evol* **20**: 914–923.

Levashina, E.A., Moita, L.F., Blandin, S., Vriend, G., Lagueux, M. and Kafatos, F.C. (2001) Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae*. *Cell* **104**: 709–718.

Little, T.J. (2002) The evolutionary significance of parasitism: do parasite-driven genetic dynamics occur *ex silico*? *J Evol Biol* **15**: 1–9.

Little, T.J., Colbourne, J.K. and Crease, T.J. (2004) Molecular evolution of *Daphnia* immunity genes: polymorphism in a Gram negative binding protein and an alpha-2-macroglobulin. *J Mol Evol* **59**: 498–506.

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

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

Oduol, F., Xu, J.N., Niare, O., Natarajan, R. and Vernick, K.D. (2000) Genes identified by an expression screen of the vector mosquito *Anopheles gambiae* display differential molecular immune response to malaria parasites and bacteria. *Proc Natl Acad Sci USA* **97**: 11397–11402.

Olson, S. (2002) Seeking the signs of selection. *Science* **298**: 1324–1325.

Osta, M.A., Christophides, G.K. and Kafatos, F.C. (2004) Effects of mosquito genes on *Plasmosium* development. *Science* **303**: 2030–2032.

Porter, C.H. and Collins, F.H. (1991) Species-diagnostic differences in a ribosomal DNA internal transcribed spacer from the sibling species *Anopheles freeborni* and *Anopheles hermsi* (Diptera: Culicidae). *Am J Trop Med Hyg* **45**: 271–279.

Rozas, J. and Rozas, R. (1999) Dnasp, Version 3: and. integrated program for population genetics and molecular evolution analysis. *Bioinformatics*, **15**: 174–175.

Schlenke, T.A. and Begun, D.J. (2003) Natural selection drives *Drosophila* immune system evolution. *Genetics* **164**: 1471–1480.

Stahl, E.A. and Bishop, J.G. (2000) Plant-pathogen arms races at the molecular level. *Curr Opin Plant Biol* **3**: 299–304.

Stahl, E.A., Dwyer, G., Mauricio, R., Kreitman, M. and Bergelson, J. (1999) Dynamics of disease resistance polymorphism at the Rpm1 locus of *Arabidopsis*. *Nature* **400**: 667–671.

Swofford, D.L. (2002) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4. Sinauer Associates, Sunderland, MA.

Yan, G., Christensen, B.M. and Severson, D.W. (1997) Comparisons of genetic variability and genome structure among mosquito strains selected for refractoriness to a malaria parasite. *J Hered* **88**: 187–194.

Yang, Z. (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Cabios* **13**: 555–556.

Yang, Z. and Bielawski, J.P. (2000) Statistical methods for detecting molecular adaptation. *Trends Ecol Evol* **15**: 496–502.

Yang, Z., Nielsen, R., Goldman, N. and Pedersen, A.-M.K. (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**: 431–449.