Population genetics of *Plasmodium* resistance genes in *Anopheles gambiae*: no evidence for strong selection

D. J. OBBARD,* Y.-M. LINTON,† F. M. JIGGINS,* G. YAN‡ and T. J. LITTLE*

*Institute of Evolutionary Biology, University of Edinburgh, Kings Buildings, West Mains Road, Edinburgh, UK, †Department of Entomology, Natural History Museum, Cromwell Road, London, UK, ‡College of Health Sciences, University of California at Irvine, Irvine, CA 92697, USA

Abstract

*Anopheles* mosquitoes are the primary vectors for malaria in Africa, transmitting the disease to more than 100 million people annually. Recent functional studies have revealed mosquito genes that are crucial for *Plasmodium* development, but there is presently little understanding of which genes mediate vector competence in the wild, or evolve in response to parasite-mediated selection. Here, we use population genetic approaches to study the strength and mode of natural selection on a suite of mosquito immune system genes, CTL4, CTLMA2, LRIM1, and APL2 (LRRD7), which have been shown to affect *Plasmodium* development in functional studies. We sampled these genes from two African populations of *An. gambiae* s.s., along with several closely related species, and conclude that there is no evidence for either strong directional or balancing selection on these genes. We highlight a number of challenges that need to be met in order to apply population genetic tests for selection in *Anopheles* mosquitoes; in particular the dearth of suitable outgroup species and the potential difficulties that arise when working within a closely-related species complex.

Keywords: *Anopheles gambiae*, APL2, CTL4, CTLMA2, innate immunity, LRIM1, natural selection, *Plasmodium falciparum*, sequence evolution

Received 4 January 2007; revision received 2 April 2007; accepted 20 April 2007

Introduction

Mosquitoes belonging to the *Anopheles gambiae* Giles species complex include the primary vectors of human-pathogenic *Plasmodium* species in sub-Saharan Africa, and as such are indirectly responsible for the deaths of more than one million people annually (World Malaria Report, WHO 2005). This has driven research into the molecular basis of mosquito–*Plasmodium* interaction, most notably the sequencing of the *An. gambiae* genome (Christophides et al. 2002; Holt et al. 2002) and the subsequent identification of genes that mediate mosquito susceptibility to *Plasmodium* infections (e.g. Blandin et al. 2004; Osta et al. 2004; Abraham et al. 2005; Michel & Kafatos 2005; Michel et al. 2005; Vlachou & Kafatos 2005; Dong et al. 2006a, b). These genes have largely been identified through expression or gene knockout studies, and while it is clear that such genes are critically involved in *Plasmodium* development and/or transmission, these studies do not indicate which genes are key players in the ecology and evolution of mosquito–*Plasmodium* interactions. In particular, it is generally not known which immune system genes are targets of parasite adaptation, or which mosquito genes mediate variation in vector competence in the field, and thus human exposure to *Plasmodium* (but see Riehle et al. 2006).

Analysis of DNA polymorphism and divergence can identify which resistance genes are targets of strong selection (reviewed in Nielsen 2005). On the one hand, strong directional selection (as in an arms-race) is expected to increase the rate of amino-acid substitution between species (reviewed in Yang & Bielawski 2000), and the concomitant spread of new advantageous alleles will reduce the level of within-species genetic diversity around the selected locus (a ‘selective sweep’, Maynard Smith & Haigh 1974; Kaplan et al. 1989; Braverman et al. 1995; Kim & Stephan 2000). On the other hand, if heterozygous genotypes are advantageous, if rare alleles experience a selective advantage, or if there are strong costs to resistance alleles in the absence of the parasite, then genetic diversity can be maintained for extended periods of time and
divergence between alternate haplotypes may be extreme (e.g. Seger 1988; Stahl et al. 1999; Charlesworth 2006). Importantly, the evolutionary history and population genetics of resistance genes may have implications for the control of malaria through the use of genetically modified mosquitoes (Little 2006). For example, if host–parasite interactions maintain adaptive diversity at a particular host locus, then artificially driving a ‘resistance’ allele from this locus (e.g. as identified by a small experimental study) through the mosquito population could facilitate the future spread of a new, or currently rare, strain of Plasmodium (Little 2006; see also Slate 2005). The degree of threat posed by this process could be elucidated through population genetic studies that shed light on the rate of adaptation or the nature of coevolutionary interactions that resistance genes are subject to. Generally, there is little understanding of how transient resistance polymorphisms are likely to be in natural populations.

However, population-genetic approaches to understanding patterns of molecular evolution demand a number of prerequisites that may be difficult to fulfil for some taxa. Firstly, analyses involving between-species comparisons, such as those based on the rates of synonymous substitution (substitutions that do not alter an amino acid: \(K_s\)) and nonsynonymous substitution (those that cause an amino acid change: \(K_a\)) require the availability of outgroup sequences with an appropriate level of divergence (e.g. Yang & Bielawski 2000). Secondly, analyses based on within-lineage variation (including genetic diversity, haplotype structure, and the distribution of allele frequencies) assume that loci share the same history of population size and growth, which may not be the case for loci which reside on chromosomal inversions that fluctuate in frequency (Andolfatto et al. 2001) or which have introgressed from other lineages. We discuss below how both may be problematic for An. gambiae.

Anopheles gambiae is a member of the Pyretophorus Series of Anopheles (subgenus Cellia), which comprises both African (e.g. the gambiae complex) and Oriental taxa (such as An. vagus Dönitz), including some of the most notorious malaria vectors in both regions. The Afrotropical species include An. daudi Coluzzi, An. christyi Newstead & Carter and the eight members of the gambiae complex, namely: An. gambiae, An. arabiensis Patton, An. brasiliensis White, An. comorensis Brunhes, le Goff & Geoffroy, An. melas Theobald, An. merus Dönitz and An. quadriannulatus ‘A’ Theobald, and An. quadriannulatus ‘B’ (Hunt et al. 1998). Unfortunately, despite their biomedical importance, surprisingly little is known of the internal systematics of the Pyretophorus Series (Foley et al. 1998; Anthony et al. 1999; Harbach 2004), making a priori selection of outgroups difficult. Sequence similarity between members of the gambiae complex is known to be high (e.g. Besansky et al. 1994, e.g. Besansky et al. 2003a), while outside of the Pyretophorus Series other well-studied Anopheles mosquitoes appear to be highly divergent, such that \(K_s\) for nuclear genes is approaching saturation (e.g. Little & Cobbe 2005). In addition, members of the gambiae complex may share considerable ancestral polymorphism, and/or may not be fully reproductively isolated from each other (e.g. Besansky et al. 1994; Besansky et al. 2003a; Donnelly et al. 2004). For example, even rare matings and partial hybrid sterility, as seen between complex members (White 1971; White 1974), may allow homogenization of genetic diversity for neutral or globally favourable alleles. Furthermore, An. gambiae s.s. itself comprises differentiated lineages that may constitute a case of incipient speciation (i.e. M and S molecular forms: Slotman et al. 2007; della Torre et al. 2002; Turner et al. 2005; but see also Yawson et al. 2007). Shared ancestral polymorphism, incipient speciation, and recent introgression or intermittent gene flow between the complex members make it especially difficult to unambiguously ascribe otherwise unusual patterns of diversity to selection. Therefore, the identification of suitable (and unsuitable) outgroups is a priority for studies of molecular evolution in An. gambiae.

Although it is probable that many genes across the Anopheles genome mediate interactions with Plasmodium, early reports suggested that the polymorphic chromosomal inversion 2La may be strongly associated with a plasmodium susceptibility phenotype in some mosquito lineages (e.g. Vernick & Collins 1989; Petrarca & Beier 1992). Efforts were made to map susceptibility genes in this region (Zheng et al. 1997) and a number of candidate genes have since been identified. Following a functional screen of candidate genes, a leucine rich repeat gene (LRIM1) located in the 2La inversion, and two C-type lectins (CTL4 and CTLMA2) also located on 2L, but outside the inversion, were identified as giving strong Plasmodium-related phenotypes in the non-natural host-parasite combination of An. gambiae and P. berghei (Osta et al. 2004; but see also Cohuet et al. 2006). In this species combination, CTL4 and CTLMA2 appear to be essential for successful Plasmodium invasion, as RNAi knockdowns resulted in an increased Plasmodium melanization from almost zero to 97% (CTL4) and 53% (CTLMA2) (Osta et al. 2004). Conversely, LRIM1 may be involved in resistance to Plasmodium infection, as knockdowns of this gene resulted in a ~3.6-fold increase in oocyte numbers (Osta et al. 2004). Recently, Riehle et al. (2006) measured Plasmodium-susceptibility in the progenies of wild-caught female mosquitoes, and found a quantitative effect associated within a region overlapping the 2La inversion. Following a database search for candidate genes within this quantitative trait locus, they identified two candidate genes (of 976 loci within the 15Mbp interval) which were designated APL1 and APL2. APL2 is synonymous with LRRD7 (Dong et al. 2006a), and in that study was found to be induced in response to Plasmodium infection,
and to display increased levels of Plasmodium infection upon RNAi knock-down (Dong et al. 2006a).

Segregating chromosomal inversions, such 2La, affect selection in a number of ways. Genetic recombination is reduced within inversions, allowing long-term and longer range association between alleles at different loci (reviewed in Schaeffer & Anderson 2005; Kirkpatrick & Barton 2006). These associations may be adaptive, and potentially contribute to the speciation process (e.g. Ayala & Coluzzi 2005; Butlin 2005; Kirkpatrick & Barton 2006). For example the frequency of the 2La inversion An. gambiae is known to correlate with the abiotic conditions (Fig. 2 in Coluzzi 1992), and is thought to be advantageous in arid environments. However, the increase in haplotype structure associated with inversions can also complicate, or even mislead, population-genetic inference of selection. Selection at linked loci will have a greater impact on local diversity and the allele-frequency spectrum that it otherwise would, and in the case of 2La, association with very strongly selected traits such as resistance to the pesticide dieldrin (Brooke et al. 2000; but see also Brooke et al. 2006) might dominate patterns of diversity. Indeed, it has been suggested that 2La inversion polymorphism is adaptively maintained in An. gambiae, even within laboratory populations (Brooke et al. 2000), and such balancing selection acting on the whole inversion might maintain highly divergent alleles at genes such as LRIM1, even in the absence of any direct selection acting on them.

Here we take a population-genetic approach to understanding the evolution of genes implicated in Plasmodium resistance or susceptibility in An. gambiae, as identified by Osta et al. (2004; CTL4, CTLMA2, LRIM1), and Riehle et al. (2006) and Dong et al. (2006a; APL2/LRRD7). First, by considering the degree of substitution saturation at synonymous sites, we screened potential outgroups for their suitability for use in studies of sequence evolution in Anopheles gambiae s.s. Second, we surveyed An. gambiae populations from East and West Africa for the level and distribution of genetic diversity in these genes. Third, we used population-genetic tests based on the observed levels genetic differentiation between population and patterns of diversity to test whether these genes show evidence of strong or recent adaptive evolution.

**Materials and methods**

**Origin and identification of samples**

The identification of all An. gambiae complex members was verified by diagnostic PCR (Scott et al. 1993). An. gambiae s.s. individuals fall into two groups that are reported to be partially reproductively isolated, known as the M and S molecular forms, and we assayed for the presence of M and S from IGS sequences using PCR-RFLP (e.g. Favia et al. 1997). An. gambiae individuals were collected from three sites in Africa: Mount Cameroon region (Cameroon, provided by S. Wanjir, University of Buea), Mbita (Suba district, Western Kenya, provided by H. M. Ferguson, University of Glasgow, UK) and Paziani (Malindi district, coastal Kenya). As expected from their known geographical distributions (della Torre et al. 2002), all surveyed Mbita (n = 26) and Paziani individuals were S form, and all but one of the Cameroon individuals were M form (n = 38; S-form individual not used in this study).

Other species were sampled from field collections and/or laboratory strains. Specifically, An. arabiensis individuals were sourced from a laboratory strain maintained at London School of Hygiene and Tropical Medicine (Strain ‘Dondotha’, provided by C. Curtis) and from field collections in Ahero (Kenya); An. merus individuals were sourced from a laboratory strain maintained by the Medical Research Council of South Africa (provided by R. Maharaj; MRC, Durban, South Africa) and from field collections in Kilifi (Kenya). An. quadriannulatus A individuals were sourced from a laboratory strain maintained at the University of Wageningen (Strain ‘Sangaqua’, Zimbabwe, provided by W. Takken). An. harsus (Bundibugyo, Uganda), An. christyi (Runkugiri District, Uganda) and An. vagus (Vietnam) were collected for ongoing taxonomic work by Y. M. Linton. An. albimanus and An. stephensi individuals were sourced from a laboratory strain maintained at the LSHTM (provided by C. Curtis, strains ‘PANAMA’ and ‘DUB 234’, respectively). Aedes aegypti sequences were derived from the publicly available draft genome sequence (strain LVP, Broad Institute and The Institute for Genomic Research).

**Loci analysed**

We amplified three An. gambiae loci that have been experimentally associated with resistance or susceptibility to Plasmodium berghei infection (LRIM1, CTL4, CTLMA2), and one candidate locus that experimentally associated with both P. bersei and P. falciparum phenotypes (APL2/LRRD7). We also attempted to amplify APL1, but were unable to reliably attribute amplified sequences to a single locus; therefore APL1 was not included for further study. In An. gambiae the four loci are spread over approximately 27 Mb on chromosome arm 2L, between cytological bands 24B and 21F (Fig. 1). In addition to these infection-associated loci, we also sequenced two putative ‘housekeeping’ genes from the same region; ENSANGG00000019219 that has homology to Drosophila ubiquitin C-terminal hydrolase CG3431 (EC 3.4.19.12), and ENSANGG00000008286 that has homology to kynurenine 3-monoxygenases (EC 1.14.13.9). These loci are likely to be engaged in routine cell function and not associated with resistance to parasite infection.
PCR assay based on the cloned inversion breakpoints
Mount Cameroon populations using a recently published karyotypes of individuals sampled from the Mbita and inversion forms 2L+a and 2La, respectively. These are designed to paired with competing primers outside of the inversion multiplex PCR with a single primer inside the inversion, 2La/2L+a amplification fragments were in Hardy–locus, as within the polymorphic Mbita population to be PCR artefacts or to derive from a second unlinked c. found the primers also amplified fragments of lengths 687 bp, 672 bp, 760 bp and 1020 bp. These do not appear to be expected fragment lengths, we amplified as a second fragment. Following PCR, unincorporated PCR primers and dNTPs were removed using ‘QIAquick’ spin-column kits (QIAGen), and products were cloned using TOPO Cloning Kits for sequencing (Invitrogen). The PCR products were then sequenced in both directions using BigDye™ Terminator Cycle Sequencing Kit (v3.1, Applied BioSystems) reagents and an ABI capillary sequencer. The sequence chromatograms were inspected by eye to confirm the validity of all differences within and between species, and assembled using SeqManII (DNASTAR Inc., Madison USA). All sequences have been submitted to GenBank as aligned sets using ambiguity codes to indicate heterozygous sites in direct sequences. Sequence accession numbers span the range EF519345–EF519529.

2La inversion status
Two of the sequenced loci reside in the region of the 2La chromosomal inversion (ENSANGG0000008286 and LRIM1; Fig. 1), which is polymorphic in An. gambiae, and fixed in the other members of the gambiae complex: 2La in arabiensis, 2L+a in the other complex members (reviewed by Coluzzi et al. 2002). Reduced recombination within the inversion and fluctuations in inversion frequency are expected to affect patterns of genetic diversity, particularly in populations where the inversion is polymorphic. We therefore attempted to identify the 2La/2L+a inversion karyotypes of individuals sampled from the Mbita and Mount Cameroon populations using a recently published PCR assay based on the cloned inversion breakpoints (Sharakhov et al. 2006; White et al. 2007). This assay uses a multiplex PCR with a single primer inside the inversion, paired with competing primers outside of the inversion breakpoints, one at either end. These are designed to amplify fragments of 207 bp and 492 bp, indicative of inversion forms 2L+a and 2La, respectively.

In addition to these expected fragment lengths, we found the primers also amplified fragments of lengths c. 687 bp, 672 bp, 760 bp and 1020 bp. These do not appear to be PCR artefacts or to derive from a second unlinked locus, as within the polymorphic Mbita population the 2La/2L+a amplification fragments were in Hardy–Weinberg equilibrium (51 individuals, \( \chi^2 = 0.44, 1\text{df.}, P = 0.51 \)). Furthermore, direct sequencing of these fragments from a small subset of individuals allowed us to identify them as insertion/deletion derivatives of expected PCR products of the published assay. Therefore we were able to tentatively assign 2La/2L+a inversion status based on fragment length, contingent on the assay primers being truly diagnostic in these populations. The Mount Cameroon population appeared to be fixed for the 2L+a variant (20 individuals assayed), whilst the inversion was polymorphic in the Mbita population (54% 2La, 46% 2L+a, 51 individuals assayed). The inversion-assay sequences have been submitted to GenBank under accession numbers EF519331–EF519344. Unfortunately, because we lacked sufficient DNA from some sampled individuals, we were unable to obtain molecular karyotypes for all the Mbita individuals genotyped at other loci. However, sufficient individuals were assayed to confirm that that the inversion karyotype did not have a large impact upon our conclusions (see Results and Discussion).

PCR and sequencing
Genomic DNA was extracted from single mosquitoes using DNeasy kits (QIAGen). PCR primers for each amplified region were designed from the published genome sequence of An. gambiae (Holt et al. 2002), and primer sequences are given in Table S1 (Supplementary material). Where possible, primers were positioned in protein-coding DNA sequence to facilitate amplification in species other than An. gambiae s.s. Consequently, sequences do not represent entire genes (see Table 2 for amplified lengths). Where amplification repeatedly failed, additional primers were investigated, but amplification success was generally sporadic outside of the gambiae complex.

Approximately 40% of sequences from CTL4, CTLMA2 and LRIM1 were cloned and each allele sequenced separately to allow determination of phase between heterozygous sites. All other sequences were obtained by directly sequencing PCR products derived from heterozygous individuals, and are therefore unphased. For cloned PCR products, LRIM1 was amplified as a single fragment, and CTL4: CTLMA2 (which are close neighbours) were amplified as a second fragment. Following PCR, unincorporated PCR primers and dNTPs were removed using ‘QIAquick’ spin-column kits (QIAGen), and products were cloned using TOPO Cloning Kits for sequencing (Invitrogen). From the remainder of the PCR fragments, that were not cloned and also for colony PCR products that derived from cloned fragments, unincorporated PCR primers and dNTPs were removed using exonuclease I (New England BioLabs) and shrimp alkaline phosphatase (Amersham). The PCR products were then sequenced in both directions using BigDye™ Terminator Cycle Sequencing Kit (v3.1, Applied BioSystems) reagents and an ABI capillary sequencer. The sequence chromatograms were inspected by eye to confirm the validity of all differences within and between species, and assembled using SeqManII (DNASTAR Inc., Madison USA). All sequences have been submitted to GenBank as aligned sets using ambiguity codes to indicate heterozygous sites in direct sequences. Sequence accession numbers span the range EF519345–EF519529.

---

Fig. 1 Genomic locations of sequenced loci. Vertical lines indicate the positions of amplified loci. The grey box indicates the location of the 2La inversion (note that the inversion extends beyond the region shown in this diagram).
Genetic divergence between lineages

Many population-genetic tests for selection require the number of substitutions between species to be estimated (e.g. that of Hudson et al. 1987), and/or that homologous synonymous and nonsynonymous sites can be correctly identified between species (e.g. the McDonald-Kreitman test, 1991). If divergence is too high, substitutions may be saturated (particularly at synonymous sites and any nonsynonymous sites under positive selection) and homology between codons difficult to establish. Conversely, if divergence it low, then (weak) selection may not have had time to drive a detectable number of substitutions. Therefore, in order to select outgroups we estimated divergence between An. gambiae and the other lineages using two methods.

First, for all comparisons we calculated the Jukes-Cantor corrected number of substitutions using dnasp Version 4.10.3 (Rozas et al. 2003). Second, divergence between more distantly related species outside of the An. gambiae complex was estimated using paml Version 3.15 (Yang 1997), using both pairwise estimates (runmode = –2) and, for housekeeping locus ENSANGG0000008286 only, a tree-based analysis (runmode = 0, Model 1) where tree topology was determined using the neighbour-joining method, and was based on nonsynonymous sites only (as implemented in MEGA Version 3.1; Kumar et al. 2004). Jukes-Cantor correction assumes all changes are equally likely, and will tend to underestimate the number of substitutions when the sequence is approaching saturation; whereas the maximum-likelihood (ML) method should be less susceptible to this.

McDonald-Kreitman (MK) tests for selection

MK tests (McDonald & Kreitman 1991) are intended to identify selection through an excess of amino acid substitution between species. Briefly, if it is assumed that synonymous sites are selectively neutral (or close to neutrality) and that polymorphic nonsynonymous sites are close to neutrality (a reasonable assumption as selected sites will be only transiently polymorphic) then departures from independence in a 2 × 2 contingency table for synonymous and nonsynonymous fixed differences and polymorphisms can be ascribed to selection acting on amino acid substitutions. However, if there have been multiple substitutions at each synonymous site (i.e. divergence is approaching mutation saturation) there is a danger that the number of synonymous substitutions will be underestimated, leading to an erroneous inference of positive selection. We therefore only applied MK tests to comparisons within the An. gambiae complex, where \( K_S \ll 1 \). Tests were performed using dnasp Version 4.10.3 (Rozas et al. 2003).

Genetic diversity and differentiation

As a new allele spreads to fixation it carries with it neighbouring variants, reducing genetic diversity in the surrounding region (a ‘selective sweep’, e.g. Maynard Smith & Haigh 1974; Kaplan et al. 1989; Braverman et al. 1995). Thus a local reduction in diversity can provide evidence for recent selection. We calculated pairwise diversity (\( \pi \)) at synonymous and nonsynonymous sites in An. gambiae for all sequenced loci using dnasp. This expected reduction in diversity following a selective sweep forms the basis of the Hudson-Kreitman-Aguadé (HKA; Hudson et al. 1987) test for selection, in which diversity is compared between loci hypothesized to be under selection, and loci believed to be evolving according to the neutral model. Since variation in diversity between loci may also be due to variation in mutation rates, or to differing proportions of very highly constrained sites, HKA tests also take into account silent-site divergence between loci (Hudson et al. 1987).

Because of the potential for multiple substitutions at synonymous sites, we chose not to apply HKA tests to species comparisons outside the An. gambiae species complex; therefore we used An. merus as the outgroup for all HKA tests, as this taxon showed the largest average synonymous site divergence from An. gambiae seen within the complex (\( K_S = 3.9\% – 11.1\% \), depending on locus). We assumed that the ‘housekeeping’ loci ENSANGG0000019219 and ENSANGG0000008286 would not be under strong positive or balancing selection, and we tested these ‘control’ loci against each putative Plasmodium resistance/susceptibility gene (CTL4, CTLMA2, LRIM1, and APL2) using synonymous sites only. First, we pooled An. gambiae data from both Mbital and Cameroon populations and applied HKA tests as implemented in dnasp, which assesses significance based on the Chi-squared distribution. Second, we applied tests separately to Mbital and Cameroon populations using the program hka (Hey 2004), which assesses significance by coalescent simulation (10,000 replicates for each test), conservatively assuming samples come from a single panmictic population, and that loci are unlinked from each other but otherwise nonrecombining.

The distribution of genetic diversity within and between populations may also be informative regarding selection (reviewed in Beaumont 2005), as global selective sweeps or strong balancing selection will tend to reduce differentiation between populations at selected loci, while local adaptation will tend to increase differentiation. Therefore, to identify any large differences in differentiation, we calculated \( F_{ST} \) between Mount Cameroon and Mbital, Kenya (West and East Africa, respectively) for each of the analysed loci. Rare introgression and/or ancestral polymorphism mean that the gambiae complex members can often share variation (Besansky et al. 1994; Besansky et al. 2003a;
Slotman et al. 2005), such that the calculation of $F_{ST}$ between complex members becomes meaningful (e.g. Besansky et al. 2003a). Therefore, where sample sizes allowed, we additionally calculated $F_{ST}$ between An. gambiae and both An. merus and An. arabiensis.

**Deviations from the neutral allele-frequency distribution**

Diversity will be regained by mutation after a selective sweep, leading to an excess of new rare variants. Tajima’s $D$ statistic (1989b), which measures the difference between average pairwise diversity (which in turn is dependent on allele frequencies) and Watterson’s estimate of $\theta$ (which depends only the number of segregating sites; Watterson 1975), is sensitive to this excess of rare variants and can be used to infer recent selection. Fu & Li (1993) have proposed a number of related test statistics that similarly identify deviations from the expected site frequency spectrum that are likely to be associated with recent selection. Fu and Li’s $D$ statistic (1993) is sensitive to the ratio of new to old mutations, as distinguished by polarizing alleles using an outgroup, and their $D^*$ statistic is intended to do the same in the absence of an outgroup. Fay & Wu (2000) have proposed a statistic, $H$, which is sensitive to a selective sweep through its effect on the number of high-frequency derived variants.

We calculated all four of these test statistics for each of the loci using DNASP. Because these statistics are also sensitive to population structure, we applied tests separately to the Mount Cameroon and Mbite populations, as well as to all *An. gambiae* individuals combined. Significance was assessed on the basis of coalescent simulations (as implemented in DNASP) using 1000 replicates for each test. Simulations were conditional on the number of segregating sites and conservatively assumed no recombination within loci.

**Results**

**Cross-species amplification and levels conservation**

We were able to amplify all loci from five species of the *An. gambiae* complex (*An. gambiae* s.s., *An. arabiensis*, *An. merus*, *An. bwambae* and *An. quadriannulatus*), with the exception of ‘housekeeping’ locus ENSANGG00000019219 from *An. bwambae*. Across species outside of the *An. gambiae* complex the housekeeping locus ENSANGG0000008286 proved the easiest to amplify (Table 1, Fig. 2), but amplification of the other loci was (Table 1). In particular, we were unable to amplify a product for LRIM1 for any species outside of the complex, and we were able to amplify only three loci from *An. vagus* and two from *An. christyi*.

The putative ‘housekeeping’ genes ENSANGG00000019219 and ENSANGG0000008286 were the most highly conserved between species; for example, homologues of the *An. gambiae* sequences could be identified in the draft *Aedes aegypti* genome (c. 80% amino acid identity; Table 1). In contrast, APL2, CTLMA2 and CTL4 were less conserved: we were unable to identify clear orthologues of these genes in the *Aedes* genome and there was relatively low amino acid identity between *An. gambiae* and other *Anopheles* mosquitoes (e.g. 71% in CTL4 between *An. gambiae* and *An. vagus*). Furthermore, our failure to amplify these loci from more than one species outside of the *An. gambiae* complex, and our failure to amplify LRIM1 from any species outside of the complex is, was suggestive of reduced sequence conservation.

As expected, all methods indicated that synonymous substitution between *An. gambiae* and *Ae. aegypti* was approaching saturation, i.e. $K_S >> 1$ (Table 1). Comparisons between *An. gambiae* and other *Anopheles* species gave $K_S$ estimates in the range c. 0.4–1.5, and in general ML estimates using codeml (PAML) were much higher than those based on the Jukes-Cantor correction. This suggests that multiple substitutions have occurred at many synonymous sites, such that raw counts, or simple corrections thereof, are a poor estimator of the number of synonymous substitutions between *An. gambiae* and the nearest relatives of the *An. gambiae* complex. In contrast, within the complex, synonymous site divergence was low, in the range $K_S \sim 0.03–0.11$, depending on locus and species comparison (Table 2). For most loci (data not shown, but see Fig. 1
Table 1  Divergence between *An. gambiae* s.s. and species from outside the complex

| Locus                  | *An. gambiae* vs. | AA identity* | JC†  | PAML‡  | K_c | S.E. | K_s | SE  |
|------------------------|------------------|--------------|------|--------|-----|------|-----|-----|
| ENSANGG0000008286      | *An. albimanus*   | 0.872        | 0.082| 0.883  | 0.066| 0.015| 1.537| 0.384|
| *An. stephensi*        | 0.976            | 0.037        | 0.626| 0.031  | 0.010| 1.023| 0.237|
| *An. vagus*            | 0.939            | 0.027        | 0.396| 0.024  | 0.009| 0.839| 0.219|
| *An. christyi*         | 0.939            | 0.027        | 0.681| 0.025  | 0.009| 1.398| 0.469|
| *Ac. aegypti*§         | 0.798            | 0.134        | 1.744| 0.108  | 0.019| 6.299| 4.263|
| ENSANGG0000019219      | *An. albimanus*   | 0.848        | 0.091| 0.885  | 0.080| 0.016| 1.736| 0.438|
| *Ac. aegypti¶         | 0.859            | 0.104        | ..** | 0.089  | 0.017| 8.459| 8.036|
| APL2                   | *An. christyi*    | 0.868        | 0.075| 0.454  | 0.070| 0.012| 0.480| 0.072|
| CTL4                   | *An. vagus*       | 0.707        | 0.206| 1.170  | 0.209| 0.047| 1.086| 0.331|
| CTLMA2                 | *An. vagus*       | 0.808        | 0.109| 0.861  | 0.125| 0.022| 0.948| 0.192|

*Proportion of amino acids (AA) that are identical; †Jukes-Cantor corrected synonymous and non-synonymous substitutions per site (dnasp); ‡Maximum-likelihood estimate of synonymous and non-synonymous substitutions per site (paml runmode = –2); §*Ac. aegypti* gene identifier AAEL008879; ¶*Ac. aegypti* gene identifier AAEL010966; **Divergence not calculated by dnap.

Table 2  Divergence and McDonald-Kreitman tests within the *An. gambiae* complex

| Locus                  | Speciation         | n*  | bp†  | K_a (%) | K_s (%) | K_a/K_s | Fix  | Poly | Fix  | Poly | Nf‡  | P§  |
|------------------------|--------------------|-----|------|---------|---------|---------|------|------|------|------|------|-----|
| CTL4 (ENSANGG0000018677) | *An. arabiensis*   | 49  | 10   | 483     | 3.9     | 0.073   | 0    | 21   | 0    | 16   | ..   | ..  |
|                        | *An. bwambae*      | 49  | 2    | 481     | 1.1     | 0.202   | 4    | 15   | 3    | 13   | 1.156| 1.00|
|                        | *An. merus*        | 49  | 9    | 468     | 1.6     | 0.145   | 8    | 15   | 5    | 13   | 1.387| 0.74|
|                        | *An. quadriann.*   | 49  | 1    | 483     | 0.9     | 0.118   | 7    | 15   | 3    | 13   | 2.022| 0.47|
| CTLMA2 (ENSANGG0000018421) | *An. arabiensis*   | 42  | 10   | 445     | 0.2     | 0.058   | 0    | 16   | 0    | 18   | ..   | ..  |
|                        | *An. bwambae*      | 42  | 2    | 443     | 0.1     | 0.028   | 1    | 12   | 0    | 15   | ..   | 0.46|
|                        | *An. merus*        | 42  | 3    | 445     | 1.3     | 0.117   | 7    | 15   | 5    | 13   | 2.8  | 0.28|
|                        | *An. quadriann.*   | 42  | 1    | 445     | 1.2     | 0.163   | 3    | 12   | 4    | 15   | 1.56 | 0.71|
| ENSANGG000000019219    | *An. arabiensis*   | 38  | 2    | 519     | 0       | 0.000   | 4    | 18   | 0    | 0    | ..   | ..  |
|                        | *An. merus*        | 38  | 2    | 519     | 0.1     | 0.032   | 1    | 18   | 0    | 1    | ..   | 1.00|
|                        | *An. quadriann.*   | 38  | 2    | 519     | 0.1     | 0.033   | 1    | 19   | 0    | 1    | ..   | 1.00|
| APL2 (ENSANGG00000019333) | *An. arabiensis*   | 40  | 4    | 741     | 0.7     | 0.149   | 1    | 32   | 0    | 19   | ..   | 1.00|
|                        | *An. bwambae*      | 40  | 2    | 750     | 0.9     | 0.320   | 1    | 26   | 3    | 14   | 0.179| 0.28|
|                        | *An. merus*        | 40  | 2    | 750     | 0.6     | 0.075   | 9    | 26   | 1    | 14   | 4.846| 0.15|
|                        | *An. quadriann.*   | 40  | 2    | 750     | 0.6     | 0.112   | 11   | 26   | 1    | 14   | 2.692| 0.65|
| ENSANGG000000008286    | *An. arabiensis*   | 44  | 2    | 411     | 0.3     | 8.0     | 0.043 | 1    | 18   | 0    | 4    | ..   | 1.00|
|                        | *An. bwambae*      | 44  | 2    | 411     | 0.3     | 3.6     | 0.097 | 0    | 18   | 0    | 4    | ..   | ..  |
|                        | *An. merus*        | 44  | 8    | 411     | 0.2     | 6.6     | 0.035 | 0    | 21   | 0    | 4    | ..   | ..  |
|                        | *An. quadriann.*   | 44  | 2    | 411     | 0.3     | 8.0     | 0.043 | 1    | 18   | 0    | 4    | ..   | 1.00|
| LRIM1 (ENSANGG00000010552) | *An. arabiensis*   | 44  | 2    | 1212    | 1.4     | 5.5     | 0.250 | 3    | 44   | 2    | 21   | 0.716| 1.00|
|                        | *An. bwambae*      | 44  | 2    | 1221    | 0.5     | 4.4     | 0.109 | 3    | 44   | 1    | 21   | 1.432| 1.00|
|                        | *An. merus*        | 44  | 6    | 1373    | 1.2     | 5.4     | 0.231 | 3    | 59   | 4    | 31   | 0.394| 0.42|
|                        | *An. quadriann.*   | 44  | 2    | 1215    | 0.7     | 4.2     | 0.157 | 2    | 47   | 1    | 24   | 1.021| 1.00|

*Number of sequences analysed from *An. gambiae* and the listed outgroup, respectively; †Length of the sequences analysed in base pairs; ‡The Neutrality Index, where NI < 1 is indicative of positive selection (Rand & Kann 1996); §P-value calculated using Fisher’s exact test.
for ENSANGG0000008286) sequences derived from some or all of the other complex members nest within the *An. gambiae* s.s. gene tree, indicative either of recent introgression, or shared ancestral polymorphism.

**McDonald-Kreitman tests for selection**

Synonymous site divergence was too high for MK tests to be usefully applied between *An. gambiae* and taxa outside of the *gambiae* complex (including both *An. vagus* and *An. christyi*). Consequently, MK tests were only applied between *An. gambiae* s.s and other members of the *gambiae* complex. No amino acid substitutions were observed within the *An. gambiae* complex for either of the two 'housekeeping' loci (Table 2). Of 16 (nonindependent) MK tests performed on the four genes putatively associated with potential resistance or susceptibility, CTL4, CTLMA2, LRIM1, and APL2, only three were in the direction of positive selection (APL2 in comparison with *An. bwambae*, and LRIM1 in comparison with *An. arabiensis* and *An. merus*; Table 2), and none were statistically significant. However, it should be noted that the extremely small numbers of fixed differences between *An. gambiae* s.s and the other complex members mean that there is very little power to identify an excess of amino acid substitutions.

**Patterns of genetic diversity**

Synonymous site diversity in *An. gambiae*, as measured by \( \pi \) (the average pairwise differences between sequences), ranged between approximately 2 and 4% at synonymous sites, and 0.2 and 0.5% at nonsynonymous sites (Table 3). Diversity was slightly, but nonsignificantly, higher in the Kenyan population (Mbita) than the Mount Cameroon population (\( P = 0.17 \), Mann–Whitney test across loci; Table 3). HKA tests measure heterogeneity in diversity between loci, relative to differentiation between species (e.g. as summarized by \( S/K \) in Table 4). All HKA tests for differences between the housekeeping loci and each of the candidate loci, tested independently, were nonsignificant. We found low, but significantly greater than zero, genetic differentiation between the Kenyan population

---

**Table 3 Genetic diversity within *Anopheles gambiae* s.s.**

| Gene Accession       | All Coding (plus noncoding) length | N† | S‡ | \( \pi_A \)§ | \( \pi_S \)§ | \( \pi_{non-cod} \) § |
|----------------------|------------------------------------|----|----|-------------|-------------|---------------------|
| CTL4 (ENSANGG00000018677) | 49 483 (211) | 13 | 15 | 0.24        | 2.11        | 1.30                |
| Mt. Cameroon         | 20 483 (211) | 5  | 10 | 0.16        | 2.10        | 1.16                |
| Mbita                | 21 483 (214) | 4  | 5  | 0.24        | 1.77        | 1.13                |
| CTLMA2 (ENSANGG00000018421) | 42 455 (230) | 15 | 12 | 0.22        | 1.87        | 2.51                |
| Mt. Cameroon         | 20 454 (230) | 2  | 6  | 0.08        | 1.30        | 0.99                |
| Mbita                | 19 473 (230) | 7  | 6  | 0.49        | 2.46        | 2.76                |
| ENSANGG00000019219   | 38 519 (117) | 0  | 18 | 0.00        | 2.79        | 3.91                |
| Mt. Cameroon         | 20 519 (117) | 0  | 12 | 0.00        | 2.70        | 2.47                |
| Mbita                | 18 531 (127) | 0  | 15 | 0.00        | 2.68        | 4.22                |
| APL2 (ENSANGG00000019333) | 40 750     | 14 | 26 | 0.47        | 3.27        | ..                  |
| Mt. Cameroon         | 22 750     | 10 | 20 | 0.36        | 2.53        | ..                  |
| Mbita                | 16 750     | 6  | 14 | 0.50        | 3.36        | ..                  |
| ENSANGG00000008286   | 44 411     | 4  | 18 | 0.34        | 4.26        | ..                  |
| Mt. Cameroon         | 24 444     | 5  | 14 | 0.46        | 3.26        | ..                  |
| Mbita                | 20 411     | 2  | 11 | 0.20        | 4.81        | ..                  |
| LRIM1 (ENSANGG00000010552) | 44 1344 (52) | 23 | 47 | 0.39        | 3.05        | 1.35                |
| Mt. Cameroon         | 22 1344 (76) | 6 | 22 | 0.18        | 2.49        | 0.35                |
| Mbita                | 17 1455 (57) | 22 | 33 | 0.54        | 3.01        | 1.30                |

\( n \) is the number of haplotypes sampled; \( N \) the number of nonsynonymous segregating sites; \( S \) the number of synonymous segregating sites; \( \pi \) is the average number of pairwise differences per site, reported as a percentage. Note that all analysed Mt. Cameroon samples were \( M \) molecular form and all Mbita samples were \( S \)-form.
PLASMODIUM RESISTANCE GENES IN ANOPHELES GAMBIAE

Table 4 HKA tests

|            | sites | n  | S  | K  | S/K |
|------------|-------|----|----|----|-----|
| An. gambiae (all) |       |    |    |    |     |
| CTL4       | 111   | 41 | 14 | 11 | 1.27|
| CTLMA2     | 99    | 39 | 12 | 11 | 1.09|
| ENSANGG..19219 | 120   | 38 | 18 | 5  | 3.60|
| APL2       | 188   | 38 | 26 | 14 | 1.86|
| ENSANGG..8286 | 94    | 44 | 18 | 6  | 3.00|
| LRIM1      | 319   | 39 | 47 | 16 | 2.94|
| An. gambiae (Mt. Cameroon) |       |    |    |    |     |
| CTL4       | 116   | 20 | 10 | 12 | 0.83|
| CTLMA2     | 99    | 20 | 6  | 10 | 0.60|
| ENSANGG..19219 | 120   | 20 | 12 | 4  | 3.00|
| APL2       | 188   | 22 | 20 | 14 | 1.43|
| ENSANGG..8286 | 103   | 24 | 14 | 6  | 2.33|
| LRIM1      | 321   | 22 | 22 | 17 | 1.29|
| An. gambiae (Mbита) |       |    |    |    |     |
| CTL4       | 116   | 21 | 5  | 11 | 0.45|
| CTLMA2     | 104   | 19 | 6  | 12 | 0.50|
| ENSANGG..19219 | 123   | 18 | 15 | 5  | 3.00|
| APL2       | 189   | 16 | 14 | 14 | 1.00|
| ENSANGG..8286 | 94    | 20 | 11 | 7  | 1.57|
| LRIM1      | 333   | 17 | 33 | 19 | 1.74|

‘sites’ is the number of (synonymous) sites analysed, n the number of haplotypes, S the number of synonymous segregating sites, K the number of synonymous substitutions between species. An. merus was used as the outgroup for all tests.

(Mbита, all ‘S’ molecular form) and the Mount Cameroon population (all ‘M’ form; \( P < 0.05 \) for each locus, based on permutation tests). However, differentiation did not differ markedly between loci (highest, ENSAGG00000019219 \( F_{ST} = 0.22 \); lowest, CTL4 \( F_{ST} = 0.13 \)).

Statistics that measure a departure from the expected allele frequency distribution can be sensitive to population structure (e.g. Tajima 1989a); therefore, we calculated these statistics separately for the Mount Cameroon and Mbitа (Kenya) populations. All statistics were calculated for synonymous sites only. For Mbitа, Tajima’s \( D \) statistic was generally positive, which is indicative of a high proportion of intermediate-frequency alleles, while for Mount Cameroon Tajima’s \( D \) was generally negative, indicating an excess of high and low frequency alleles (Fig. 3a). This deviation in Tajima’s \( D \) was only significant for APL2 in the Mbitа population (no correction for multiple tests). However, across all loci the difference in Tajima’s \( D \) between locations was marginally significant (\( P = 0.04 \), Mann–Whitney test). Fu and Li’s \( D \) statistic was similarly more positive for the Mbitа population than the Mount Cameroon population (\( P = 0.03 \), Mann–Whitney), but again only differed significantly from neutrality for one locus (CTL4 in Mbitа, no correction for multiple tests; Fig. 3b). Fu and Li’s \( D^* \) statistic, which does not require an outgroup, was highly correlated with Fu and Li’s \( D \) (\( r^2 = 0.97 \), data not shown). Fay and Wu’s \( H \) statistic, which is sensitive to an excess of high-frequency derived alleles, did not differ from the neutral expectation for any loci (Fig. 3c).

© 2007 The Authors
Journal compilation © 2007 Blackwell Publishing Ltd
A priori, it might be expected that the 2La inversion would affect patterns of genetic diversity in LRIM1 and ENSANGG00000008286. This could either be through fluctuations in inversion frequency leading to reduced polymorphism, or balancing selection acting on the inversion as a whole incidentally maintaining divergent haplotypes in these genes. However, we saw no evidence of this. Despite the Mt. Cameroon population being fixed for 2L+, and the Mbita population being polymorphic for the inversion, diversity was not appreciably elevated or reduced in the region of the inversion in the Mbita sample relative to the Mount Camaroon sample (Table 3), and deviations from the neutral allele frequency spectrum (e.g. as measured by Tajima’s D or Fu & Li’s D) appear to be no larger for LRIM1 and ENSANGG00000008286 than for other loci in the Mbita population (Fig. 3). Moreover, based on karyotype-assayed mosquitoes homozygous for either the 2La and 2L+ inversion type, we could detect only one fixed difference across LRIM1 and ENSANGG00000008286, out of 115 segregating sites in total (only 3 homozygous mosquitoes genotyped for LRIM1 and ENSANGG00000008286 were available from the Mbita population for the 2La assay, making this a very conservative test for the effect of 2La).

Discussion

Synonymous site divergence, introgression, and outgroup choice

This study used DNA polymorphism and divergence data to test whether mosquito immunity genes thought to be important for Plasmodium development might also be subject to host-pathogen arms races or balancing selection. For some of the tests, in particular MK and HKA tests, an accurate estimate of the number of synonymous substitutions between species is needed. This requires an outgroup for which synonymous substitution does not approach saturation. For many well-studied taxa this is a not difficult (e.g. Drosophila melanogaster — D. simulans KS = 0.12), but our data suggest that it may be hard to meet this requirement for Anopheles gambiae (e.g. Fig. 2). In particular, although previous morphology-based phylogenetic analyses suggested that An. vagus (an oriental member of the Pyrethorophorus series) and An. christyi (thought to be the basal taxon of the Pyrethorophorus series), were amongst the most closely related taxa to the An. gambiae complex (Anthony et al. 1999), for the genes we analysed KS was approaching 1 between these lineages (Fig. 2, Table 1), making them potentially inappropriate for outgroup-based tests. Although further systematic work may reveal that other taxa (unavailable to us) constitute a more appropriate outgroup, given the level of divergence in the genes analysed here it seems probable that future comparative work of this type will be restricted to species within the gambiae complex.

Unfortunately, using a species from within the complex as an outgroup presents its own problems. Members of the An. gambiae complex (including An. arabiensis, An. bwambae, An. merus and An. quadriannulatus A) are closely related, with KS highly variable between loci, but of the order of 3%–10% (Table 2, see also Besansky et al. 2003b) — and if a correction is made for the level of diversity within lineages (KS – πS, p220. Nei 1987) then this becomes 0%–10%. Although the close relationships and low level of diversity within the An. gambiae complex make MK and HKA analyses viable in principle, in a given length of sequence the number of fixed differences is small, reducing statistical power. Furthermore, for some loci, sequences from other members of the gambiae complex nest within the An. gambiae s.s. clade (e.g. Fig. 2). In fact FST (which measures genetic differentiation as the proportion of total diversity that is due to between-group differences) can be calculated between members of the An. gambiae complex as if they were partially isolated populations (Besansky et al. 2003b), and does not approach one (FST is also highly variable between loci, Table 5; FST range 0.3–0.9 here and 0.1–0.8 in Besansky et al. 2003b). This may result either from introgression between species, or from the continued segregation of inherited ancestral polymorphisms, and may have implications for the inference of selection. Specifically, this may mean that, in effect, we are using an allele from within the population as an arbitrary outgroup for our analyses, based only on divergence. Moreover, introgressed alleles will not be differentiated at all between the hybridizing species — preventing comparative analysis on those loci — and this may even be compounded if introgression is more common for strongly selected loci.

Evidence for selection

Despite the potential difficulties in selecting a suitable outgroup, we were able to apply several population-genetic methods for detecting selection. Amino acid conservation between An. gambiae and other species was lower for APL2, CTL4 and CTLMA2 than for the ‘housekeeping’ locus ENSANGG00000008286 (Table 1), and our repeated failure to amplify LRIM1 outside of the An. gambiae complex was also suggestive of reduced sequence conservation. Moreover, within the gambiae complex KS/KS ratios were generally higher for the four putative Plasmodium-associated genes than for the two ‘housekeeping’ loci; for example, across all within-complex comparisons the average KS/KS for candidate resistance/susceptibility-associated genes was 0.14, but only 0.04 for the ‘housekeeping’ loci. An elevated rate of amino-acid evolution is in-line with what is known for immunity genes in other taxa (Hurst & Smith 1999; Schlenke & Begun 2003). However, the KS/KS ratio per se...
provides no evidence for adaptive change, as the observed reduction in amino-acid conservation may equally be due to lower selective constraints in these immunity-related genes relative to our house-keeping genes.

McDonald-Kreitman tests can identify positive selection by comparing the relative numbers of synonymous and nonsynonymous within-species polymorphisms and between-species substitutions (McDonald & Kreitman 1991). If there are an excess of amino-acid substitutions between species, this can be ascribed to selection fixing new amino-acid variants in each species. For most of the loci we analysed we found an excess of amino-acid variation, rather than an excess of amino acid substitutions (NI > 1, Table 2). This suggests that these none of the analysed loci (APL2, LRIM1, CTL4 or CTLMA2) have been under strong or consistent directional selection since the common ancestor of extant *An. gambiae* and other members of the *gambiae* complex. Although it may be argued that this is not a robust result, as the relatively close relationship (K<sub>S</sub> ~ 3–11%) between species leads to low power in MK tests, subject to the caveats regarding introgression given above, it does at least suggest that if directional selection is acting it is likely to be weak. For example, it is possible to find positive MK tests in comparisons between very closely related species, such as *Drosophila yakuba*—*D. santomea* (K<sub>S</sub> ~3%) (antiviral genes R2D2 and Ago2, Obbard et al. 2006).

Recent selective sweeps associated with the fixation of a single new allele can be detected through their effect on genetic diversity and the allele frequency spectrum (reviewed in Nielsen 2005). Although these statistics can be sensitive to demographic processes (such as population growth), when compared between loci they provide a useful method for detecting selection in the absence of an outgroup. However, none of the loci we analysed showed an extreme level of synonymous site diversity, and all were close to the level found previously for other genes in *An. gambiae* (1.9%–4.3% here vs. 1%–4.9% in Besansky et al. 2003a). Genetic diversity was slightly higher for the house-keeping genes than for the putative *Plasmodium*-resistance genes (average 3.5% vs. 2.6% at synonymous sites), but HKA tests found that this effect was not significant (Table 4). This may be due to the low power we have to detect small deviations, given the sample sizes (of the order of 30 polymorphic sites and 20 fixed differences for each test, Table 4), but again at least suggests that any reduction in diversity compared to housekeeping loci is small.

Although skews in the allele-frequency distribution were not significantly different from a neutral model for individual loci, suggesting that none have undergone a recent selective sweep, overall there was a slight but significant difference between the Mbita (East Africa) and Mount Cameroon (West Africa) populations. Tajima’s D differed between populations, being generally positive in the Kenya population (a slight nonsignificant excess of intermediate frequency alleles), and negative in the Mount Cameroon population (a slight nonsignificant excess of extreme-frequency alleles). Fu and Li’s D (and D*) showed a similar pattern, with a greater proportion of new mutations in the Cameroon population than the Kenya population. This may indicate a difference in demographic history between populations, e.g. population growth or a recent bottleneck in the Mount Cameroon population, and/or some form of substructuring within in the Mbita population. However, because all Mount Cameroon individuals analysed in this study were M form, and all Mbita individuals were S form (as assessed by PCR-RFLP) it is impossible to say whether this difference in the allele frequency spectrum, and the level of population differentiation, is due to demographic factors associated with geographical region, or with the M and S molecular forms, or both. Perhaps surprisingly, we also found the 2La

---

**Table 5 Differentiation within the *An. gambiae* complex**

|                  | *An. gambiae*† | *An. merus*‡ | *An. arabiensis*§ |
|------------------|---------------|---------------|------------------|
| **K<sub>S</sub><sup>T</sup>** | **F<sub>S</sub><sup>T</sup>** | **F<sub>S</sub><sup>T</sup>** | **F<sub>S</sub><sup>T</sup>** |
| CTL4             | 0.05          | 0.13          | 0.89             | 0.30             |
| CTLMA2           | 0.08          | 0.20          | 0.80<sup>¶</sup>  | 0.34             |
| ENSANGG.19219    | 0.09          | 0.22          | 0.69<sup>¶</sup>  | 0.63<sup>¶</sup>  |
| APL2             | 0.07          | 0.21          | 0.59<sup>¶</sup>  | 0.29<sup>¶</sup>  |
| ENSANGG.8286     | 0.05          | 0.14          | 0.49             | 0.68<sup>¶</sup>  |
| LRIM1            | 0.05          | 0.18          | 0.65             | 0.72<sup>¶</sup>  |

†Differentiation between Mbita (all S molecular form) and Mount Cameroon (all M molecular form) populations of *An. gambiae s.s*. ‡Differentiation between *An. gambiae s.s* and *An. merus*. §Differentiation *An. gambiae s.s* and *An. arabiensis*. ¶Less than six haplotypes available for the non-*gambiae* species (sample sizes as given in Table 2) so that these values should be treated with caution. All K<sub>S</sub><sup>T</sup> statistics are significantly greater than zero, by 1000 permutations.
inversion had little or no impact on the two loci it spans (LRIM1 and ENSANGG0000008286). However, it is known that recombination is most strongly suppressed near inversion breakpoints (these genes are c. 9.8 Mbp and 3.5 Mbp from the inversion ends, respectively) and gene conversion can homogenize genetic diversity between inversions over time (e.g. Schaeffer & Anderson 2005).

Understanding Plasmodium resistance in Anopheles mosquitoes

Despite being previously implicated in Plasmodium resistance or susceptibility, none of the genes we analysed showed any evidence of strong or recent adaptive evolution in An. gambiae, either under directional or balancing selection. It is therefore possible that that there is little or no selective pressure for adaptive change exerted on these genes. Indeed, contrary to the study that identified LRIM1, CTL4 and CTLMA2 (which used the non-natural host-parasite combination of An. gambiae and P. berghei; Osta et al. 2004), and in broad agreement with our findings, recent work indicates that these genes may not be important in resistance to P. falciparum, a natural parasite of An. gambiae (Cohuet et al. 2006). However, this does not necessarily detract from the wider host-parasite implications of our work, as genes involved in the encapsulation process are likely to engage with many parasites other than P. falciparum.

In addition, many Anopheles loci other than those analysed here have been identified as putative immunity genes (Christophides et al. 2002), and a few are good candidates for mediating levels of Plasmodium infection. These include APL1 (Riehle et al. 2006), which was identified in the same screen as APL2 but which we were unable to reliably amplify by PCR, serine protease inhibitors and TEPs (Blandin et al. 2004; Abraham et al. 2005; Michel et al. 2005), which are the subject to ongoing population-genetic work in our lab. We believe studies which can uncover the rate and mode of evolution in genes mediating vector competence are essential if we are to gain a full understanding of insect-vectorized diseases such as malaria. It is perhaps too often overlooked that Plasmodium evolution may be shaped by interaction with the vector’s immune system, as well as that of the vertebrate host.

Acknowledgements

The authors wish to thank the following people for providing samples used in this study: W. Takken (An. quadrimaculatus A), C. Curtis (An albimanus, An. stephensi, An. Arabiensis) R. Maharaj (An. merus), H. M. Fergussen (An. gambiae) and S. Wanji (An. gambiae). We also wish to thank C. Walton and D. Callister for discussion, and C. Walton, R. Butlin, and three anonymous reviewers for valuable comments on an earlier version of this manuscript. Funding was provided through Wellcome Trust Grant 073210 to T.J.L.

References

Abraham EG, Pinto SB, Ghosh A et al. (2005) An immune-responsive serpin, SRPN6, mediates mosquito defense against malaria parasites. Proceedings of the National Academy of Science, USA, 102, 16327–16332.

Andolfatto P, Depaulis F, Navarro A (2001) Inversion polymorphisms and nucleotide variability in Drosophila. Genetical Research, 77, 1–8.

Anthony TG, Harbach RE, Kitching J (1999) Phylogeny of the Pyretophorus Series of Anopheles subgenus Cellia (Diptera: Culicidae). Systematic Entomology, 24, 193–205.

Ayala FJ, Coluzzi M (2005) Chromosome speciation: humans, Drosophila, and mosquitoes. Proceedings of the National Academy of Sciences of the USA, 102, 6535–6542.

Beaumont MA (2005) Adaptation and speciation: what can FST tell us? Trends in Ecology and Evolution, 20, 435.

Besansky NJ, Powell JR, Caccione A et al. (1994) Molecular phylogeny of the Anopheles gambiae complex suggests genetic introgression between principal malaria vectors. Proceedings of the National Academy of Science, USA, 91, 6885–6888.

Besansky NJ, Krzywinski J, Lehmann T et al. (2003a) Semipermeable species boundaries between Anopheles gambiae and Anopheles arabiensis: evidence from multilocus DNA sequence variation. Proceedings of the National Academy of Science, USA, 100, 10818–10823.

Besansky NJ, Severson DW, Ferdig MT (2003b) DNA barcoding of parasites and invertebrate disease vectors: what you don’t know can hurt you. Trends in Parasitology, 19, 545–546.

Blandin S, Shiao S-H, Moita LF et al. (2004) Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector Anopheles gambiae. Cell, 116, 661.

Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W (1995) The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics, 140, 785–796.

Brooke BD, Hunt RH, Coetsee M (2000) Resistance to dieldrin; fipronil assorts with chromosome inversion 2La in the malaria vector Anopheles gambiae. Medical and Veterinary Entomology, 14, 190–194.

Brooke BD, Hunt RH, Matambo TS et al. (2006) Dieldrin resistance in the malaria vector Anopheles gambiae in Ghana. Medical and Veterinary Entomology, 20, 294–299.

Butlin RK (2005) Recombination and speciation. Molecular Ecology, 14, 2621–2635.

Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome regions. Plos Genetics, 2, 379–384.

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

Cohuet A, Osta MA, Morlais I et al. (2006) Anopheles and Plasmodium: from laboratory models to natural systems in the field. Embo Reports, 7, 1285–1289.

Coluzzi M (1992) Malaria vector analysis and control. Parasitology Today, 8, 113–118.

Coluzzi M, Sabatini A, della Torre A, Di Deco MA, Petrarca V (2002) A polytene chromosome analysis of the Anopheles gambiae species complex. Science, 298, 1415–1418.

della Torre A, Costantini C, Besansky NJ et al. (2002) Speciation within Anopheles gambiae—the glass is half full. Science, 298, 115–117.
Dong Y, Aguilar R, Xi Z et al. (2006a) Anopheles gambiae immune responses to human and rodent Plasmodium parasite species. PLoS Pathogens, 2, e52.

Dong YM, Taylor HE, Dimopoulos G (2006b) AgDscam, a hypervariable immunoglobulin domain-containing receptor of the Anopheles gambiae innate immune system. PLoS Biology, 4, 1137–1146.

Donnelly MJ, Pinto J, Giord I, Besansky NJ, Lehmann T (2004) Revisiting the role of introgression vs. shared ancestral polymorphisms as key processes shaping genetic diversity in the recently separated sibling species of the Anopheles gambiae complex. Heredity, 92, 61–68.

Favia G, dellaTorre A, Bagayoko M et al. (1997) Molecular identification of sympatric chromosomal forms of Anopheles gambiae and further evidence of their reproductive isolation. Insect Molecular Biology, 6, 377–383.

Fay JC, Wu C-I (2000) Hitchhiking under positive Darwinian selection. Genetics, 155, 1405–1194.

Foley DH, Bryan JH, Yeates D, Saul A (1998) Evolution and systematics of Anopheles: insights from a molecular phylogeny of Australasian mosquitoes. Molecular Phylogenetics and Evolution, 9, 262.

Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. Genetics, 133, 693–709.

Harbach RE (2004) The classification of genus Anopheles (Diptera: Culicidae): a working hypothesis of phylogenetic relationships. Bulletin of Entomological Research, 94, 537–553.

Hey J (2004) HKA – A computer program for tests of natural selection. (URL http://lifesci.rutgers.edu/~heylab/HeylabSoftware.htm#HKA) [accessed on 14 August 2006].

Holt RA, Subramanian GM, Halpern A et al. (2002) The genome sequence of the malaria mosquito Anopheles gambiae. Science, 298, 129–149.

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

Hunt RH, Coetze M, Fetene M (1998) The Anopheles gambiae complex: a new species from Ethiopia. Transactions of the Royal Society of Tropical Medicine and Hygiene, 92, 231–235.

Hurst LD, Smith NGC (1999) Do essential genes evolve slowly? Current Biology, 9, 747–750.

Kaplan NL, Hudson RR, Langley CH (1989) The ‘hitchhiking effect’ revisited. Genetics, 123, 887–899.

Kim Y, Stephan W (2000) Joint effects of genetic hitchhiking and background selection on neutral variation. Genetics, 155, 1415–1427.

Kirkpatrick M, Barton N (2006) Chromosome inversions, local adaptation and speciation. Genetics, 173, 419–434.

Kumar S, Tamura K, Nei M (2004) MEGA 3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Briefings in Bioinformatics, 5, 150–163.

Little TJ (2006) Immune system polymorphism: implications for genetic engineering. In: Genetically Modified Mosquitoes for Malaria Control (ed. Boete C), pp. 36–59. Landes Bioscience, Georgetown, Texas.

Little TJ, Cobbe N (2005) The evolution of immune-related genes from disease carrying mosquitoes: diversity in a peptidoglycan-and a thioester-recognizing protein. Insect Molecular Biology, 14, 599–605.

Maynard Smith J, Haigh J (1974) The hitchhiking effect of a favourable gene. Genetical Research, Cambridge, 219, 23–35.

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

Michel K, Kafatos FC (2005) Mosquito immunity against Plasmodium. Insect Biochemistry and Molecular Biology, 35, 677–689.

Michel K, Budd A, Pinto S, Gibson TJ, Kafatos FC (2005) Anopheles gambiae SRPN2 facilitates midgut invasion by the malaria parasite Plasmodium berghei. EMBO Reports, 6, 891–897.

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

Nielsen R (2005) Molecular signatures of natural selection. Annual Review of Genetics, 39, 197–218.

Obbard DJ, Jiggins FM, Halligan DL, Little TJ (2006) Natural selection drives extremely rapid evolution in antiviral RNAi genes. Current Biology, 16, 580–585.

Osta MA, Christophides GK, Kafatos FC (2004) Effects of mosquito genes on Plasmodium development. Science, 303, 2030–2032.

Petracca V, Beier JC (1992) Intraspecific chromosomal polymorphism in the Anopheles gambiae complex as a factor affecting malaria transmission in the Kisumu area of Kenya. American Journal of Tropical Medicine and Hygiene, 46, 229–237.

Rand DM, Kann LM (1996) Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. Molecular Biology and Evolution, 13, 735–748.

Riehle MM, Markianos K, Niare O et al. (2006) Natural malaria infection in Anopheles gambiae is regulated by a single genomic control region. Science, 312, 577–579.

Rozas J, Sanchez-Del Barrio JC, Meseguer X, Rozas R (2003) DNAsp, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics, 19, 2496–2497.

Scheaffer SW, Anderson WW (2005) Mechanisms of genetic exchange within the chromosomal inversions of Drosophila pseudoobscura. Genetics, 171, 1729–1739.

Schenke TA, Begun DJ (2003) Natural selection drives Drosophila immune system evolution. Genetics, 164, 1471–1480.

Scott JA, Brogdon WG, Collins FH (1993) Identification of single specimens of the Anopheles gambiae complex by the polymerase chain reaction. American Journal of Tropical Medicine and Hygiene, 49, 520–529.

Seger J (1988) Dynamics of some simple host-parasite models with more than two genotypes in each species. Philosophical Transactions of The Royal Society of London Series B-Biological Sciences, 319, 541–555.

Sharakhov IV, White BJ, Sharakhova MV et al. (2006) Breakpoint structure reveals the unique origin of an interspecific chromosomal inversion (2Lo) in the Anopheles gambiae complex. PNAS, 103, 6258–6262.

Slater J (2005) Molecular evolution of the sheep prion protein gene. Proceedings of the Royal Society B-Biological Sciences, 272, 2371–2377.

Slotman MA, Della Torre A, Calzetta M, Powell JR (2005) Differential introgression of chromosomal regions between Anopheles gambiae and An. arabiensis. American Journal of Tropical Medicine and Hygiene, 73, 326–335.

Slotman MA, Tripet F, Cornel AJ et al. (2007) Evidence for subdivision within the M molecular form of Anopheles gambiae. Molecular Ecology, 16, 639–649.

Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J (1999) Dynamics of disease resistance polymorphism at the Rpml locus of Arabidopsis. Nature, 400, 667–671.

Tajima F (1989a) The effect of change in population size on DNA polymorphism. Genetics, 123, 597–601.

Tajima F (1989b) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics, 123, 585–595.
Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in Anopheles gambiae. PLoS Biology, 3, e285.
Vernick KD, Collins FH (1989) Association of a Plasmodium refractory phenotype with an esterase locus in Anopheles gambiae. American Journal of Tropical Medicine and Hygiene, 40, 593–597.
Vlachou D, Kafatos FC (2005) The complex interplay between mosquito positive and negative regulators of Plasmodium development. Current Opinion in Microbiology, 8, 415–421.
Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theoretical Population Biology, 7, 256–276.
White GB (1971) Chromosomal evidence for natural interspecific hybridization by mosquitoes of anopheles-gambiae complex. Nature, 231, 184–&.
White GB (1974) Anopheles-gambiae complex and disease transmission in Africa. Transactions of the Royal Society of Tropical Medicine and Hygiene, 68, 278–302.
White BJ, Santolamazza F, Kamau L et al. (2007) Molecular karyotyping of the 2La inversion in Anopheles gambiae. American Journal of Tropical Medicine and Hygiene, 76, 334–339.
WHO (2005) World Malaria Report. www.rbm.who.International/wmr2005/.
Yang ZH (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. Computer Applications in the Biosciences, 13, 555–556.
Yang Z, Bielawski JP (2000) Statistical methods for detecting molecular adaptation. Trends in Ecology and Evolution, 15, 496.
Yawson AE, Weetman D, Wilson MD, Donnelly MJ (2007) Ecological Zones Rather Than Molecular Forms Predict Genetic Differentiation in the Malaria Vector Anopheles gambiae s.s. in Ghana. Genetics, 175, 751–761.

Zheng LB, Cornel AJ, Wang R et al. (1997) Quantitative trait loci for refractoriness of Anopheles gambiae to Plasmodium cynomolgi B. Science, 276, 423–428.

DJO is a postdoctoral researcher working for TJL on the evolutionary genetics of dipteran immunity genes. YML is a molecular systematist with a special interest in Anopheles species complexes. FMJ is a Wellcome Trust Research Career Development Fellow, interested in the evolution of insect innate immune systems. GY is interested in the ecology and evolutionary genetics of host–parasite interactions. TJL is a Wellcome Trust Senior Research Fellow in Basic Biomedical Sciences, with broad interests in host-parasite coevolution and ecology.

Supplementary material
The following supplementary material is available for this article:

Table S1 PCR primer sequences.

This material is available as part of the online article from: http://www.blackwell-synergy.com/doi/abs/10.1111/j.1365-294X.2007.03395.x
(This link will take you to the article abstract).

Please note: Blackwell Publishing are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.