Am. J. Trop. Med. Hyg., 71(6), 2004, pp. 795–802

INTRASPECIFIC NUCLEOTIDE VARIATION IN *ANOPHELES GAMBIAE*: NEW INSIGHTS INTO THE BIOLOGY OF MALARIA VECTORS

ISABELLE MORLAIS, NICOLAS PONÇON, FRÉDÉRIC SIMARD, ANNA COHUET, AND DIDIER FONTENILLE

Laboratoire de Lutte Contre les Insectes Nuisibles, Institut de Recherche pour le Développement, Montpellier, France; Organisation de Coodination pour la lutte Contre les Endémise en Afrique Centrale, Yaounde, Cameroon; European Molecular Biology Laboratory, Heidelberg, Germany

Abstract. The *Anopheles gambiae* genome sequence, together with the recent development of molecular tools for genome-wide analysis, promises new insights into the biology of the malaria vector. These insights should help define the best possible breakdown point for interrupting transmission in the mosquito vector. A survey of the intraspecific nucleotide diversity in coding regions of three different mosquito strains showed an average of one single nucleotide polymorphism (SNP) every 125 coding base pairs. High levels of nucleotide polymorphism were observed in mosquito immune-related genes and pathogen recognition receptors harbored higher replacement substitutions. Genotyping at SNP loci in natural populations of *An. gambiae* from three malaria foci showed contrasting patterns. The distribution of mutation Y443H in the thioester-containing protein 3 (TEP3) gene suggested this mutational event has occurred under selective constraints. Our results show that SNP-based studies will be valuable in identifying the sequence variation associated with phenotypic traits shaping vector competence.

INTRODUCTION

Anopheles gambiae is the major mosquito vector of malaria in Africa, one of the most devastating diseases that affects 300–500 million people and causes 2–3 million deaths per year. Infection of the mosquito vector takes place while blood-feeding on an infected human host. During development within the mosquito, malaria parasites undergo large numerical losses, and it was shown that vector immune mechanisms attenuate the efficiency of parasite development.[1–3] Several approaches to identify key mosquito genes interacting with *Plasmodium* parasites during their development have been developed and the number of putative antimalarial immune defense molecules has increased rapidly.[4–7] Germline transformation of mosquitoes has now been achieved in the laboratory and the principal remaining goal for producing transgenic mosquitoes, with altered vector competence, is to identify suitable targets to be driven into the genome.[8,9] Thus, discovering mutations underlying specific traits that could be used for developing new malaria control strategies has become a great challenge. Different ways are currently envisioned, including reduction of the population density and/or longevity of the mosquito, alteration of their anthropophily level, and disruption of the malaria parasite cycle in the insect vector. Thus, investigations of sequence variation between *An. gambiae* populations of divergent phenotype are needed.

The effort to identify single nucleotide polymorphisms (SNPs) has recently been accelerated by the availability of whole genome sequences and the emergence of high-throughput genotyping tools.[10] The SNPs are the most frequent sequence variations found in vertebrate and invertebrate genomes. They are extremely abundant, co-dominant and easy to score. Thus, SNPs have become genetic markers of choice for high-resolution mapping, population genomics, and allelic association studies.[11–13] The SNPs located in noncoding regions of the genome and SNPs at synonymous sites in coding regions (sSNPs), with no direct impact on the phenotype, provide useful markers in population genetics and evolutionary studies; whereas SNPs at nonsynonymous sites (nsSNPs), which alter the structure and potentially the function of the encoded proteins, represent target markers to identify genetic variations associated with complex phenotypic traits. In the *An. gambiae* genome assembly released in public databases at the end of 2002, an overall SNP frequency of 1.6×10^{-3} was found throughout the 278 Mb of the mosquito genome[14]; this large amount of polymorphisms raises the possibility of undertaking SNP-based studies.

We examined nucleotide variation at coding regions of mosquito nuclear genes by sequence comparison of individual mosquitoes that originated from three different laboratory strains: Yaounde, L35, and 4arr. These three mosquito strains belong to the M molecular form,[15] which allows measurement of intraspecific genetic variation. Genes to be compared were chosen for two purposes: 1) they were spread out over the whole genome so that it will be possible to draw a SNP map for *An. gambiae*, and 2) most of them were genes involved in the immune system, in parasite-vector interactions, or in sexual selection, which will facilitate association studies. To assess the usefulness of SNP markers in population genetics, genotyping was investigated at three SNP loci in natural populations of *An. gambiae* from three different locations.

MATERIALS AND METHODS

Mosquitoes. All mosquitoes used in this study were adult female *An. gambiae* s.s. Mosquitoes used to assess the nucleotide diversity were from three laboratory strains: Yaounde, L35, and 4arr. The three mosquito strains belong to the M molecular form. The Yaounde strain, which is permissive to *Plasmodium falciparum,* was colonized in Cameroon in 1988 and was selected for membrane feeding.[16] The L35 and 4arr strains were genetically selected for *Plasmodium* refractoriness and permissiveness, respectively; both strains were derived from the G3 strain, colonized in The Gambia in 1975, following several selection processes.[17,18] Mosquitoes used for SNP genotyping were collected in three different field areas where both M and S molecular forms are sympatric: Dielmo in Senegal, Vallée du Kou in Burkina Faso, and Simbock in Cameroon.

Extraction of DNA, primer design, polymerase chain reaction amplification, and sequencing. Genomic DNA was extracted from whole mosquitoes by grinding tissues with a micropestle in an extraction buffer (0.1 M Tris HCl, pH 8.0, 0.01

M EDTA, 1.4 M NaCl, 2% cetylltrimethyl ammonium bromide). The mixture was incubated at 65°C for ten min. Total DNA was extracted with chloroform, precipitated in isopropanol, washed in 70% ethanol, and resuspended in sterile water.

Target genes for this study were selected based on putative role in pathogen-vector interactions or in mosquito behavior. They were also chosen to be dispersed throughout the genome (Table 1 and Figure 1). Three genes of redox processes that were up-regulated upon malaria challenge or associated with oocyst development were screened.[4,19] Two housekeeping genes were used as controls, BolA, which plays a role in regulating transcription, and β-tubulin, which is involved in the cytoskeleton formation. Target genes sequences were extracted from the PEST strain genome database, non-coding regions were identified, and primer sets were designed to allow amplification of 600–1,000-basepair DNA fragments using Primer3 software.[20] Regions containing repetitive sequences, introns and untranslated regions were avoided. Primer sequences are available from the corresponding author (IM). For each primer set, the optimum annealing temperature was determined using an Eppendorf MasterCycler Gradient (Eppendorf, Hamburg, Germany).

Polymerase chain reactions (PCRs) were carried out with 5 ng of template DNA in a final volume of 50 μL containing 50 mM Tris-HCl, pH 8.3, 2.5 mM $MgCl_2$, 50 mM KCl, 200 μM of each dNTP (Eurogentec, Herstal, Belgium), 0.25 units of *Taq* polymerase (Qiagen, Courtaboeuf, France), and 10 pmoles of each primer. The PCRs included an initial denaturation step at 95°C for three minutes, followed by 30 cycles at 95°C for 30 seconds, at the specific annealing temperature for 30 seconds, and at 72°C for one minute, and a final extension step at 72°C for six minutes. Five microliters of PCR products were resolved by electrophoresis on a 1% agarose gel, stained with ethidium bromide, and visualized under ultraviolet light (UV). The remaining PCR products were used for sequencing, using reverse- and forward-specific primers.

Electrophoregrams were inspected visually and heterozygous sites were scored according to Black and others.[11] Sequences and position of polymorphisms were deposited in GeneBank. Non-coding regions (3,538 basepairs) were removed for sequence analysis. Sequences from 2–3 individuals from each strain were aligned using ClustalX.[21] Nucleotide variation was analyzed using MEGA version 2.1.[22] The average number of nucleotide substitutions per site π was calculated for the 35 genes for all sequenced individual mosquitoes and for each strain (Table 1). Estimates of synonymous and nonsynonymous substitution rates Ks and Ka were computed between sequences using the approximate methods of Kumar, which corrects for transition/transversion bias and handles degenerated sites.[23]

Single nucleotide polymorphism genotyping. The SNP genotyping assay was conducted on field-caught females (see mosquitoes section). Species and molecular form identification was conducted according to Fanello and others[24]; only *An. gambiae* s.s. mosquitoes were used. Polymorphism in the three natural populations was screened at three SNP loci. The SNPs were chosen in coding regions FBN9_Y185H and TEP3_Y443H, corresponding to nonsynonymous substitutions, and PGRP-LP, which was a 27-basepair insertion/deletion (indel) polymorphism (Table 2).

Specific primers flanking the SNP or the indel were defined using Primer3 software,[20] and the optimum annealing temperature was determined using an Eppendorf MasterCycler Gradient. The PCRs were carried out in a 25-μL reaction mixture as described earlier, but the PCR elongation time was reduced to 30 seconds. Ten microliters of PCR products were used for digestion in a final volume of 25 μL containing 1× reaction buffer, 1× bovine serum albumin, and two units of the appropriate enzyme (Qbiogene, Carlsbad, CA). Digestions were performed overnight at 37°C. Fifteen microliters of the reaction products were resolved by electrophoresis on a 2% agarose gel, photographed under UV light, and individual genotypes were scored. Genotypes for each SNP marker are shown in Table 2.

RESULTS

Intraspecific nucleotide diversity. The intraspecific diversity of *An. gambiae* was assessed at coding regions of 35 nuclear genes. Table 1 details the characteristic features of these genes and Figure 1 shows their genome location. Sequence comparison identified 460 SNPs and 9 indels over 22,680 basepairs examined. All genes contained at least one SNP. All indels were base triplets and did not cause frame shifting; 30% (140 of 460) of the SNPs were nonsynonymous substitutions and thus induced an amino acid change, only one (in the agCP14332 sequence) was a nonsense SNP.

The average nucleotide diversity per gene was 7.9×10^{-3}, ranging from 0.8×10^{-3} to 26.6×10^{-3}. Nucleotide diversity was not statistically different between chromosomes 2 and 3 (8.9×10^{-3} and 6.8×10^{-3}; $P = 0.16$, by Student's *t*-test). The Yaounde strain exhibited higher intra-strain variability ($P < 0.001$, by Student's *t*-test). This could be due to lower selection procedures that have maintained a certain extent of genetic polymorphism, whereas multiple selections of the L3-5 and 4aRR strains might have drastically reduced their genetic heterogeneity. Interestingly, pairwise comparison of nucleotide diversity between the three strains did not yield significant difference ($\pi_{Yd/4arr} = 9.8 \times 10^{-3}$, $\pi_{Yd/L35} = 11.9 \times 10^{-3}$, $\pi_{L35/4arr} = 8.9 \times 10^{-3}$, $P > 0.1$); L35 and 4arr strains were as divergent to each other as with the Yaounde strain. We also observed, depending on genes, haplotypes conserved between both Yaounde and 4arr (CP14332), Yaounde and L35 (Gambif), or 4arr and L35 (SRPN10, LRR-Toll). These findings may reflect bottlenecks induced by selection procedures and that have preserved different ancestral alleles.

Nucleotide diversity was significantly reduced in behavior genes (Arrestin, Period, and Orr1) ($P < 0.05$), and no nonsynonymous substitutions were found in this group (Table 1). These genes are associated with sexual selection or host-feeding recognition and, as for the housekeeping genes, they are likely under selective constraints for sequence conservation.

The average Ka was significantly higher for genes putatively involved in the recognition of parasites (i.e., the pattern recognition receptors and membrane-bound receptors) than for others immune-related genes ($P < 0.04$). This finding could indicate a diversity-enhancing selection on receptor genes, which supports the hypothesis that high levels of polymorphism are maintained in genes involved in parasite recognition because heterozygotes can recognize a greater diversity of antigens.[25,26]

Genotyping natural populations with SNP markers. Allele

TABLE 1
Nucleotide diversity in coding regions of 35 genes of *Anopheles gambiae**

Gene name	Gene abbreviation	Acc. no.	Fc†	Ch
An. gambiae G12	ANG12	EAA11870	1	2
Peritrophin A	Aper1	EAA04177	1	2
Aminopeptidase N	APN	EAA13235	1	2
Gelsolin	Gelsolin	EAA13073	1	3
Intestinal chitin binding protein	ICHIT	EAA11412	1	2
Integrin	Integrin	EAA13939	1	3
Laminin	Laminin	EAA04042	1	2
Transferin-like	TSF-like	EAA12967	1	3
Fibrinogen 23	FBN23	EAA05439	2	3
Fibrinogen 9	FBN-9	EAA05102	2	3
Gram negative binding protein-A1	GNBP-A1	EAA04713	2	2
Gram negative binding protein-B1	GNBP-B1	EAA09116	2	2
Peptidoglycan recognition protein LB	PGRP-LB	EAA01800	2	2
Thioester-containing protein 3	TEP3	EAA10529	2	3
Thioester-containing protein 4	TEP4	EAA13702	2	3
Serine protease 14A	SP14A	EAA08418	3	2
Serine protease D1	SP14D1	EAA45573	3	2
Serine protease 22D	SP22D	EAA11708	3	2
Serpin 10	SRPN10	EAA09492	3	2
Serpin 9	SRPN9	EAA08448	3	2
Leucine-rich repeats toll-like receptor	LRR-Toll	EAA11514	4	2
Nitric oxide synthase	NOS	EAA12335	4	3
Toll receptor 9	TOLL9	EAA04650	4	2
Phagocytic adaptor CED-6	CED-6	EAA04093	5	2
Gambif	Gambif	EAA05931	5	3
Cytochrome P450 9K1	CYP9K1	EAA12048	6	X
Ferredoxin reductase	Ferr-Red	EAA03876	6	2
Vesicular ATPase	vATPase	EAA08175	6	2
Arrestin 1	Arrestin1	EAA13874	7	3
Odorant receptor Orr1	Orr1	EAA13838	7	3
Period	Period	EAA01734	7	2
Unknown gene agCP14332	agCP14332	EAA03115	8	2
Unknown gene agCP5039	agCP5039	EAA10648	8	3
Beta tubulin	b tubulin	EAA09971	9	3
Morphoprotein BolA	BolA	EAA08757	9	2

								Diversity		
Gene	N	L	cSNPs	nsSNPs	indels	Ks	Ka	π	Yd	4arr	L35
ANG12	8	584	18	2		0.0291	0.0007	0.0073	0.0162	0.0000	0.0000
Apert	8	462	26	8		0.0467	0.0072	0.0185	0.0111	0.0000	0.0023
APN	8	491	19	10		0.0195	0.0090	0.0113	0.0159	0.0000	0.0044
Gelsolin	9	765	10	4		0.0096	0.0021	0.0042	0.0032	0.0000	0.0000
ICHIT	7	1,003	26	15	1	0.0157	0.0089	0.0108	0.0038	0.0027	0.0027
Integrin	8	583	3	0		0.0098	0.0000	0.0022	0.0023	0.0000	0.0000
Laminin	7	881	16	3		0.0104	0.0009	0.0034	0.0103	0.0000	0.0000
TSF-like	7	588	18	5		0.0403	0.0038	0.0129	0.0092	0.0000	0.0000
FBN23	7	806	5	3		0.0038	0.0023	0.0026	0.0027	0.0000	0.0000
FBN-9	9	841	25	1		0.0347	0.0008	0.0114	0.0091	0.0078	0.0000
GNBP-A1	9	567	8	2	1	0.0238	0.0013	0.0061	0.0012	0.0000	0.0000
GNBP-B1	8	574	3	2		0.0021	0.0012	0.0016	0.0000	0.0000	0.0006
PGRP-LB	8	539	16	11	2	0.0208	0.0132	0.0149	0.0000	0.0000	0.0000
TEP3	9	567	4	3		0.0108	0.0274	0.0024	0.0019	0.0011	0.0000
TEP4	9	664	28	10		0.0466	0.0091	0.0181	0.0164	0.0000	0.0000
SP14A	7	715	19	3		0.0325	0.0021	0.0111	0.0073	0.0030	0.0080
SP14D1	7	591	7	2		0.0110	0.0018	0.0046	0.0023	0.0000	0.0000
SP22D	8	733	23	7	3	0.0340	0.0047	0.0124	0.0116	0.0000	0.0019
SRPN10	7	603	10	5		0.0129	0.0042	0.0063	0.0081	0.0000	0.0000
SRPN9	9	780	10	1		0.0197	0.0007	0.0056	0.0027	0.0048	0.0000
LRR-Toll	8	858	25	7		0.0281	0.0025	0.0104	0.0104	0.0000	0.0008
NOS	7	463	9	0		0.0281	0.0000	0.0077	0.0077	0.0000	0.0044
TOLL9	9	860	12	0		0.0212	0.0000	0.0053	0.0031	0.0006	0.0000
CED-6	8	478	3	2		0.0030	0.0023	0.0026	0.0000	0.0000	0.0011
Gambif	7	574	5	1		0.0101	0.0011	0.0038	0.0009	0.0000	0.0000
CYP9K1	9	580	3	3		0.0000	0.0014	0.0011	0.0018	0.0000	0.0009
Ferr-Red	9	696	14	3		0.0365	0.0027	0.0093	0.0000	0.0008	0.0000

TABLE 1 (continued)
Nucleotide diversity in coding regions of 35 genes of *Anopheles gambiae**

Gene	N	L	cSNPs	nsSNPs	indels	Ks	Ka	Diversity			
								π	Yd	4arr	L35
vATPase	8	693	18	6		0.0183	0.0017	0.0061	0.0066	0.0029	0.0041
Arrestin1	9	734	8	0		0.0128	0.0000	0.0033	0.0025	0.0020	0.0000
Orr1	9	606	1	0		0.0042	0.0000	0.0008	0.0000	0.0000	0.0000
Period	9	856	8	0		0.0131	0.0000	0.0031	0.0027	0.0000	0.0000
agCP14332	7	312	19	10	2	0.0397	0.0174	0.0266	0.0118	0.0054	0.0101
agCP5039	8	772	31	11		0.0442	0.0082	0.0168	0.0126	0.0009	0.0000
b tubulin	8	564	2	0		0.0053	0.0000	0.0019	0.0024	0.0000	0.0000
BolA	9	297	8	0		0.0325	0.0019	0.0095	0.0109	0.0000	0.0000
Total		22,680	460	140	9						
Average						0.0209	0.0040	0.0079	0.0060	0.0009	0.0012

*Acc. no. = GeneBank peptide accession number; Ch = chromosome location; N = number of individuals screened; L = length of aligned sequences; cSNPs = coding single nucleotide polymorphisms; nsSNPs = nonsynonymous SNPs; indels = insertion/deletion polymorphisms; Ks = number of substitutions per nonsynonymous site; Ka = number of substitutions per synonymous site; Diversity = average number of nucleotide substitutions per site; π = diversity for all strains; Yd, 4arr, and L35 = diversity within the corresponding laboratory strain. Diversity and estimation of substitution rates Ka and Ks were computed using MEGA version 2.1.

† Fc = functional class; 1 = receptor; 2–5 = immune-related (2 = recognition; 3 = modulation; 4 = transduction; 5 = effector peptide synthesis); 6 = redox processes and ion transport; 7 = host seeking or sexual selection; 8 = unknown; 9 = housekeeping.

frequencies for the three SNP markers are shown in Figure 2. Allelic distributions differed from one sampling site to another and between molecular forms. For example, the deleted fragment (allele C) of PGRP-LP marker had a higher frequency in the S molecular form from Dielmo, whereas allele H of FBN9_Y185H marker was more frequent in the Simbock mosquito population of molecular form M.

Mutation Y443H in TEP3 gene showed marked differences in its distribution pattern between sampling sites (Figures 2 and 3). Allele H was fixed in all mosquitoes of molecular form S, but allele frequencies differed from one location to another for individuals of molecular form M. In Simbock, only allele H was found, whereas in Vallée du Kou all mosquitoes but one had allele Y, and the mosquito harboring allele H was heterozygote. In contrast, in Dielmo both alleles Y and H were found at almost equal frequency.

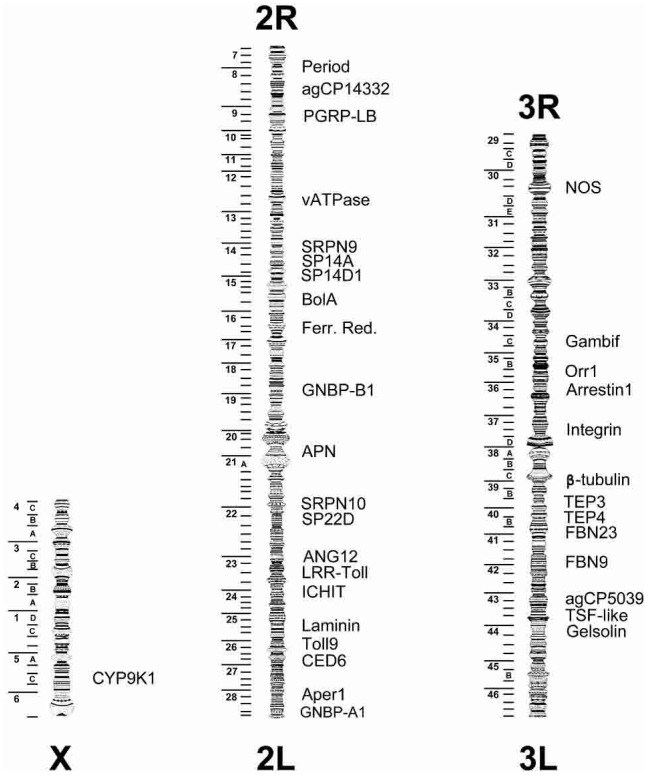

FIGURE 1. Relative location of target genes on the *Anopheles gambiae* genome. For definitions of genes, see Table 1.

DISCUSSION

Genome sequencing projects offer the possibility to undertake studies at a whole-genome scale, and the availability of a growing number of genome sequences facilitates the discovery of massive numbers of genetic markers. In this study, we investigated the genetic variation in coding regions of 35 nuclear genes of *An. gambiae* and our data, showing an overall average of one SNP every 125 coding basepairs in the mosquito nuclear genes, confirmed the high nucleotide diversity observed in insect genomes.[14,27,28]

Anopheles gambiae s.s. is characterized by two distinct molecular forms, M and S.[15] The molecular forms have maintained their ability to cross-mate, yielding fertile progeny in the laboratory, but the M and S forms exist as isolated populations in sympatric areas, which is indicative of incipient speciation.[29–31] The *An. gambiae* genome sequence was obtained from the PEST strain, which was produced by cross-mating a laboratory strain originating in Nigeria and harboring the Mopti cytotype (molecular form M) with the progeny of field-caught mosquitoes from western Kenya that carried the Savanna cytotype (molecular form S). The genetic composition of the outbred colony was predominantly derived from the Savanna form, but the assembly of the mosquito genome was nonetheless hampered by the presence of dual haplotypes and the PEST strain appeared to have a mosaic genome structure.[14] The SNP distribution on a genome-wide scale was heterogeneous across the genome, with SNP densities varying approximately 500-fold according to genomic regions.[14] This heterogeneity in the PEST strain is probably due in part to genetic introgression between molecular forms. However, nucleotide diversity is usually positively correlated with recombination rates and a higher level of polymorphism is generally observed at the telomere regions.[27,32] The SNP distribution on genes in *P. falciparum* chromosome 2 was shown to be regional, with most polymorphisms occurring in genes encoding antigens and membrane-associated proteins.[33] Most of

TABLE 2
Characteristics of the three SNP loci screened*

SNP	Location	Screening	Primer sequence	Genotype size
PGRP-LB_LP	Coding	LP, 27 bp	For TAC GTT GGC AAA CAG CTG AC Rev TGG CAG TAG GCT CTA AAT ATGC	L: 120 C: 93
FBN9_Y185H	Coding, ns	RFLP-*Rsa*I	For GCG GTC GGC AAT AAT CTA AC Rev AAA CTC CTG ATC GAC GTT CC	H: 355/35 Y: 265/90/35
TEP3_Y443H	Coding, ns	RFLP-*Rsa*I	For GGC AAA CTG ATA CGC CTC AT Rev CGT CCC TTC AGC AGT AGC TC	H: 400 Y: 234/166

* SNP = single nucleotide polymorphism; LP = length polymorphism; genotype size gives the product size (basepairs [bp]) for each allele by agarose gel electrophoresis after a polymerase chain reaction (PCR) or PCR–restriction fragment length polymorphism (RFLP); For = forward; Rev = reverse; ns = nonsynonymous. For definitions of genes, see Table 1.

these proteins interact with the host immune system and, as for the acquired immune system of mammals, genes involved in parasite recognition are likely subject to selective pressures. In *Drosophila simulans*, DNA variation indicated that adaptive selection has an important role in driving immunity gene evolution.[34] The higher rate of replacement substitutions we observed in *An. gambiae* genes that have a putative receptor function would then suggest that mosquito genes

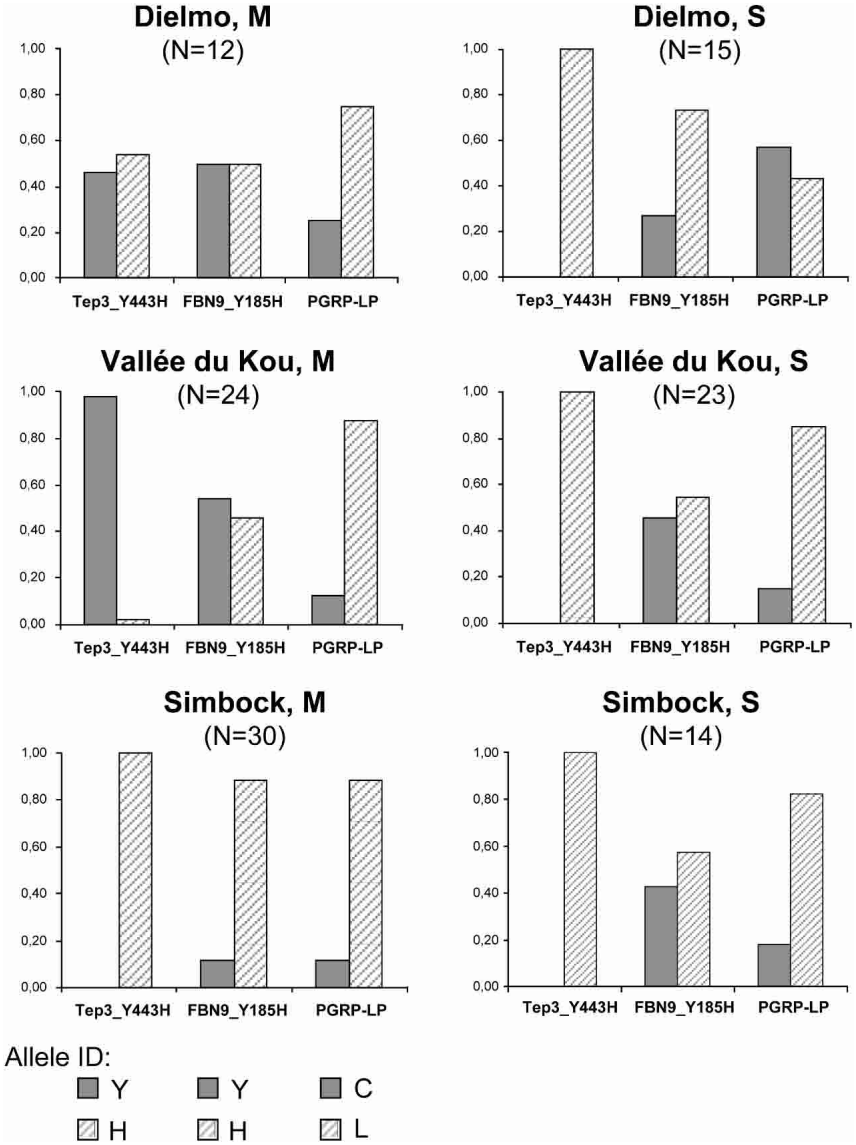

FIGURE 2. Allele frequencies of three single nucleotide polymorphisms in *Anopheles gambiae* natural populations from three different sampling sites (Dielmo in Senegal, Vallée du Kou in Burkina Faso, and Simbock in Cameroon) where molecular forms M and S are sympatric. The numbers of mosquitoes screened are shown in parentheses. For definitions of genes, see Table 1.

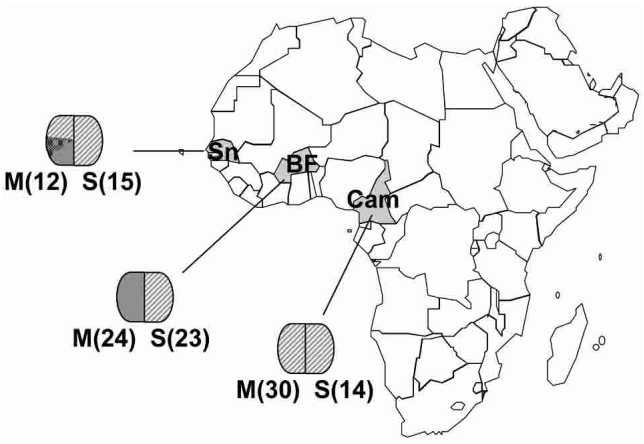

FIGURE 3. Genotypic distribution of the TEP3_Y443H single nucleotide polymorphism in *Anopheles gambiae* natural populations from three different sampling sites. Sn = Senegal; BF = Burkina Faso; Cam = Cameroon. Numbers in parentheses indicate the sampling size. TEP3 = thioester-containing protein 3.

associated in host-pathogen interactions may be well specialized.

Genes exhibiting large number of nonsynonymous substitutions are generally fast-evolving.[35] These genes can evolve rapidly because of a lack of strong functional or structural constraints: they evolve in a neutral fashion or because of a positive diversifying selection, with sequence polymorphism giving a selective advantage under a variety of conditions. Thus, to verify adaptive molecular evolution putatively acting on the mosquito receptor genes, it would be interesting to investigate nucleotide divergence in closely related species such as *An. arabiensis* and *An. melas,* other efficient malaria vectors, and *An. quadriannulatus*, which does not naturally transmit *P. falciparum* parasites because of its zoophilic feeding preferences.[36] Similarly, the high sequence diversity we observed in agCP14332 and agCP5039, the function of which is still unknown, is indicative they represent fast-evolving genes. In *Drosophila*, it has been estimated that 20% of genes represent fast-evolving genes and this proportion is approximately 11% in the mosquito genome.[37,38] A large number of these genes do not have homologs in other species, which is the case here for agCP14332 and agCP5039, and it will be of critical importance to determine the function of the encoded proteins since they could be involved in specific adaptations.

The distribution of mutation Y443H in the TEP3 gene is likely due to selective constraints, but this particular allelic distribution could also be linked to chromosomal forms. Indeed, in Simbock, only chromosomal form Forest occurs[31] and in Vallée du Kou, mosquitoes of molecular form M carry chromosomal form Mopti, whereas molecular form S has the Savana cytotype.[29] In Dielmo, we do not know the relationship between molecular and chromosomal forms, but mosquitoes of molecular form S are harboring the Savana karyotype throughout Senegal whereas molecular form M can carry Savana, Mopti, or Bissau karyotypes.[29] However, three of the 12 individuals of the M form screened in Dielmo were heterozygotes and heterocaryotypes are rare in nature.[39] Nonetheless mutation Y443H in the TEP3 gene can be associated with some ecologic adaptations and chromosomal forms reflect adaptations to environmental constraints.[30] In Vallée du Kou, *An. gambiae* M form occurs in irrigated zones of rice cultivation, whereas the S form develops in rain-dependent breeding sites.[40] At Simbock, both forms are found in the same breeding sites that are characterized by temporary puddles. No information, to our knowledge, is available concerning breeding sites of the different molecular forms in Dielmo.

TEP3 is an immune gene, a member of a family of thioester-containing proteins (TEPs) that are parasite recognition molecules sharing similarities with vertebrates complement factors and α_2-macroglobulins.[41] TEP3 is up-regulated upon bacterial challenge and parasite infection.[7,41] Mutation Y443H in the TEP3 gene is unlikely related to *Plasmodium* infection because both molecular forms transmit malaria parasites in these areas. Conversely, the presence of bacteria in permanent breeding sites such as irrigated zones could have exerted a selection pressure on the larval stage and selected this mutation. Characterization of breeding habitats for the M and S molecular forms of *An. gambiae* in Vallée du Kou would allow confirmation of this assumption, and genotyping mutation Y443H in mosquitoes of known chromosomal form in Dielmo would help decipher the correlation between the mutation and mosquito cytotypes. Functional mutations have already been characterized in genes associated to insecticide resistance, and it was shown that such mutations with strong adaptive value were passed through introgression between incipient species.[42,43] Introgressive hybridization is thought to represent a mechanism for generating new adaptations, and monitoring such adaptive genetic changes in natural populations could improve vector control operations implementation and management.[30]

Population genetics approaches are based on the distribution of genetic variability within and among populations of mosquitoes and SNP markers will provide powerful tools to analyze variation throughout the entire genome. The three SNPs we screened here gave promising results for such population genetics analysis. Genotyping large numbers of SNP markers should facilitate the study of genetic structure of natural populations, which is necessary to predict the spread of genes of interest, such as insecticide resistance genes or malaria-associated genes.

Development of new malaria control measures necessitates the identification of candidate gene(s) that affects phenotypes such as anthropophily, fertility, or parasite permissiveness. Functional genomics tools will help to identify these genes, for example, by comparing genome-wide expression profiles between mosquito populations exhibiting contrasting phenotypes and comparative genomics of functional regions of genes between closely related species (e.g., *An. arabiensis* and *An. quadriannulatus*) would help identifying phenotypic markers linked to vectorial capacity and competence. Whole-genome analysis promises new insights into the biology of malaria vectors, and the mosquito genetic variation offers the possibility of designing genotype-phenotype association studies that will aid improving or developing strategies to reduce malaria transmission. The numerous polymorphisms we found in the mosquito genes will be useful for drawing a preliminary SNP map for *An. gambiae* and to undertake association studies. Genome-wide analysis using SNPs will allow discovering genetic factors associated with complex traits such as permissiveness to *Plasmodium*.

Received April 7, 2004. Accepted for publication June 29, 2004.

Acknowledgments: We thank Mark Benedict (Malaria Research Activities, National Institute of Allergy and Infections Diseases, National Institutes of Health, Bethesda, MD) for providing L35 and 4arr mosquito strains and Myriam Collin for help in mosquito maintenance in the insectary. We are grateful to Abdoulaye Diabaté (Institut de Recherche en Sciences de la Santé, Bobo-Dioulasso, Burkina Faso) and Ibrahima Dia (Institut Pasteur, Dakar, Senegal) for kindly providing DNA samples from Burkina Faso and Senegal, respectively.

Financial support: This work was supported by the Centre National de la Recherche Scientifique. Isabelle Morlais was supported by the Fondation pour la Recherche Médicale.

Authors' addresses: Isabelle Morlais, Nicolas Ponçon, and Didier Fontenille, Laboratoire de Lutte Contre les Insectes Nuisibles, Institut De Recherche Pour Le Développement, 911 Avenue Agropolis, BP 64501, 34 394 Montpellier Cedex 5, France, Telephone: 33-467-043-222, Fax: 33-467-542-044, E-mails: isabelle.morlais@ird.fr, poncon@mpl.ird.fr, and fontenil@mpl.ird.fr. Frédéric Simard, Organisation de Coordination pour la Lutte Contre les Endémies en Afrique Centrale, BP 288, Yaounde, Cameroon, Telephone: 237-223-2232, Fax: 237-223-0061, out_of_cameroun@yahoo.fr. Anna Cohuet, European Molecular Biology Laboratory, Meyerhofstrasse 1, 69117 Heidelberg, Germany, Telephone: 49-6221-387-288, Fax: 49-6221-387-211, cohuet@embl.de.

Reprint requests: Isabelle Morlais, Laboratoire de Lutte Contre Les Insectes Nuisibles, Institut de Recherche pour le Développement, 911 Avenue Agropolis, BP 64501, 34 394 Montpellier Cedex 5, France.

REFERENCES

1. Gouagna LC, Mulder B, Noubissi E, Tchuinkam T, Verhave JP, Boudin C, 1998. The early sporogonic cycle of *Plasmodium falciparum* in laboratory-infected *Anopheles gambiae*: an estimation of parasite efficacy. *Trop Med Int Health 3:* 21–28.
2. Dimopoulos G, Muller HM, Levashina EA, Kafatos FC, 2001. Innate immune defense against malaria infection in the mosquito. *Curr Opin Immunol 13:* 79–88.
3. Barillas-Mury C, Wizel B, Han YS, 2000. Mosquito immune responses and malaria transmission: lessons from insect model systems and implications for vertebrate innate immunity and vaccine development. *Insect Biochem Mol Biol 30:* 429–442.
4. Dimopoulos G, Christophides GK, Meister S, Schultz J, White KP, Barillas-Mury C, Kafatos FC, 2002. Genome expression analysis of *Anopheles gambiae*: responses to injury, bacterial challenge, and malaria infection. *Proc Natl Acad Sci USA 99:* 8814–8819.
5. Ito J, Ghosh A, Moreira LA, Wimmer EA, Jacobs-Lorena M, 2002. Transgenic anopheline mosquitoes impaired in transmission of a malaria parasite. *Nature 417:* 452–455.
6. Morlais I, Severson DW, 2001. Identification of a polymorphic mucin-like gene expressed in the midgut of the mosquito, *Aedes aegypti*, using an integrated bulked segregant and differential display analysis. *Genetics 158:* 1125–1136.
7. Christophides GK, Zdobnov E, Barillas-Mury C, Birney E, Blandin S, Blass C, Brey PT, Collins FH, Danielli A, Dimopoulos G, Hetru C, Hoa NT, Hoffmann JA, Kanzok SM, Letunic I, Levashina EA, Loukeris TG, Lycett G, Meister S, Michel K, Moita LF, Muller HM, Osta MA, Paskewitz SM, Reichhart JM, Rzhetsky A, Troxler L, Vernick KD, Vlachou D, Volz J, von Mering C, Xu J, Zheng L, Bork P, Kafatos FC, 2002. Immunity-related genes and gene families in *Anopheles gambiae. Science 298:* 159–165.
8. Catteruccia F, Nolan T, Loukeris TG, Blass C, Savakis C, Kafatos FC, Crisanti A, 2000. Stable germline transformation of the malaria mosquito *Anopheles stephensi. Nature 405:* 959–962.
9. Grossman GL, Rafferty CS, Clayton JR, Stevens TK, Mukabayire O, Benedict MQ, 2001. Germline transformation of the malaria vector, *Anopheles gambiae*, with the *piggyBac* transposable element. *Insect Mol Biol 10:* 597–604.
10. Syvanen AC, 2001. Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nat Rev Genet 2:* 930–942.
11. Black WC IV, Baer CF, Antolin MF, DuTeau NM, 2001. Population genomics: genome-wide sampling of insect populations. *Annu Rev Entomol 46:* 441–469.
12. Berger J, Suzuki T, Senti KA, Stubbs J, Schaffner G, Dickson BJ, 2001. Genetic mapping with SNP markers in *Drosophila. Nat Genet 29:* 475–481.
13. Botstein D, Risch N, 2003. Discovering genotypes underlying human phenotypes: past successes for Mendelian disease, future approaches for complex disease. *Nat Genet 33 (Suppl):* 228–237.
14. Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark AG, Ribeiro JM, Wides R, Salzberg SL, Loftus B, Yandell M, Majoros WH, Rusch DB, Lai Z, Kraft CL, Abril JF, Anthouard V, Arensburger P, Atkinson PW, Baden H, de Berardinis V, Baldwin D, Benes V, Biedler J, Blass C, Bolanos R, Boscus D, Barnstead M, Cai S, Center A, Chatuverdi K, Christophides GK, Chrystal MA, Clamp M, Cravchik A, Curwen V, Dana A, Delcher A, Dew I, Evans CA, Flanigan M, Grundschober-Freimoser A, Friedli L, Gu Z, Guan P, Guigo R, Hillenmeyer ME, Hladun SL, Hogan JR, Hong YS, Hoover J, Jaillon O, Ke Z, Kodira C, Kokoza E, Koutsos A, Letunic I, Levitsky A, Liang Y, Lin JJ, Lobo NF, Lopez JR, Malek JA, McIntosh TC, Meister S, Miller J, Mobarry C, Mongin E, Murphy SD, O'Brochta DA, Pfannkoch C, Qi R, Regier MA, Remington K, Shao H, Sharakhova MV, Sitter CD, Shetty J, Smith TJ, Strong R, Sun J, Thomasova D, Ton LQ, Topalis P, Tu Z, Unger MF, Walenz B, Wang A, Wang J, Wang M, Wang X, Woodford KJ, Wortman JR, Wu M, Yao A, Zdobnov EM, Zhang H, Zhao Q, Zhao S, Zhu SC, Zhimulev I, Coluzzi M, della Torre A, Roth CW, Louis C, Kalush F, Mural RJ, Myers EW, Adams MD, Smith HO, Broder S, Gardner MJ, Fraser CM, Birney E, Bork P, Brey PT, Venter JC, Weissenbach J, Kafatos FC, Collins FH, Hoffman SL, 2002. The genome sequence of the malaria mosquito *Anopheles gambiae. Science 298:* 129–149.
15. della Torre A, Fanello C, Akogbeto M, Dossou-yovo J, Favia G, Petrarca V, Coluzzi M, 2001. Molecular evidence of incipient speciation within *Anopheles gambiae* s.s. in West Africa. *Insect Mol Biol 10:* 9–18.
16. Tchuinkam T, Mulder B, Dechering K, Stoffels H, Verhave JP, Cot M, Carnevale P, Meuwissen JH, Robert V, 1993. Experimental infections of *Anopheles gambiae* with *Plasmodium falciparum* of naturally infected gametocyte carriers in Cameroon: factors influencing the infectivity to mosquitoes. *Trop Med Parasitol 44:* 271–276.
17. Collins FH, Sakai RK, Vernick KD, Paskewitz S, Seeley DC, Miller LH, Collins WE, Campbell CC, Gwadz RW, 1986. Genetic selection of a *Plasmodium*–refractory strain of the malaria vector *Anopheles gambiae. Science 234:* 607–610.
18. Zheng L, Cornel AJ, Wang R, Erfle H, Voss H, Ansorge W, Kafatos FC, Collins FH, 1997. Quantitative trait loci for refractoriness of *Anopheles gambiae* to *Plasmodium cynomolgi* B. *Science 276:* 425–428.
19. Cociancich SO, Park SS, Fidock DA, Shahabuddin M, 1999. Vesicular ATPase-overexpressing cells determine the distribution of malaria parasite oocysts on the midguts of mosquitoes. *J Biol Chem 274:* 12650–12655.
20. Rozen S, Skaletsky H, 2000. Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol 132:* 365–386.
21. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG, 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res 25:* 4876–4882.
22. Kumar S, Tamura K, Jakobsen IB, Nei M, 2001. MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics 17:* 1244–1245.
23. Nei M, Kumar S, 2000. *Molecular Evolution and Phylogenetics*. New York: Oxford University Press.
24. Fanello C, Santolamazza F, della Torre A, 2002. Simultaneous identification of species and molecular forms of the *Anopheles gambiae* complex by PCR-RFLP. *Med Vet Entomol 16:* 461–464.
25. Wang HY, Tang H, Shen CK, Wu CI, 2003. Rapidly evolving genes in human. I. The glycophorins and their possible role in evading malaria parasites. *Mol Biol Evol 20:* 1795–1804.

26. Hughes AL, 1997. Rapid evolution of immunoglobulin superfamily C2 domains expressed in immune system cells. *Mol Biol Evol 14:* 1–5.
27. Aquadro CF, Bauer DuMont V, Reed FA, 2001. Genome-wide variation in the human and fruitfly: a comparison. *Curr Opin Genet Dev 11:* 627–634.
28. Morlais I, Severson DW, 2003. Intraspecific DNA variation in nuclear genes of the mosquito *Aedes aegypti. Insect Mol Biol 12:* 631–639.
29. della Torre A, Costantini C, Besansky NJ, Caccone A, Petrarca V, Powell JR, Coluzzi M, 2002. Speciation within *Anopheles gambiae*–the glass is half full. *Science 298:* 115–117.
30. 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.
31. Wondji C, Simard F, Fontenille D, 2002. Evidence for genetic differentiation between the molecular forms M and S within the Forest chromosomal form of *Anopheles gambiae* in an area of sympatry. *Insect Mol Biol 11:* 11–19.
32. Lercher MJ, Hurst LD, 2002. Human SNP variability and mutation rate are higher in regions of high recombination. *Trends Genet 18:* 337–340.
33. Volkman SK, Hartl DL, Wirth DF, Nielsen KM, Choi M, Batalov S, Zhou Y, Plouffe D, Le Roch KG, Abagyan R, Winzeler EA, 2002. Excess polymorphisms in genes for membrane proteins in *Plasmodium falciparum. Science 298:* 216–218.
34. Schlenke TA, Begun DJ, 2003. Natural selection drives *Drosophila* immune system evolution. *Genetics 164:* 1471–1480.
35. Schmid KJ, Tautz D, 1997. A screen for fast evolving genes from Drosophila. *Proc Natl Acad Sci USA 94:* 9746–9750.
36. Takken W, Eling W, Hooghof J, Dekker T, Hunt R, Coetzee M, 1999. Susceptibility of *Anopheles quadriannulatus* Theobald (Diptera: Culicidae) to *Plasmodium falciparum. Trans R Soc Trop Med Hyg 93:* 578–580.
37. Zdobnov EM, von Mering C, Letunic I, Torrents D, Suyama M, Copley RR, Christophides GK, Thomasova D, Holt RA, Subramanian GM, Mueller HM, Dimopoulos G, Law JH, Wells MA, Birney E, Charlab R, Halpern AL, Kokoza E, Kraft CL, Lai Z, Lewis S, Louis C, Barillas-Mury C, Nusskern D, Rubin GM, Salzberg SL, Sutton GG, Topalis P, Wides R, Wincker P, Yandell M, Collins FH, Ribeiro J, Gelbart WM, Kafatos FC, Bork P, 2002. Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster. Science 298:* 149–159.
38. Schmid KJ, Nigro L, Aquadro CF, Tautz D, 1999. Large number of replacement polymorphisms in rapidly evolving genes of *Drosophila*. Implications for genome-wide surveys of DNA polymorphism. *Genetics 153:* 1717–1729.
39. Gentile G, Slotman M, Ketmaier V, Powell JR, Caccone A, 2001. Attempts to molecularly distinguish cryptic taxa in *Anopheles gambiae* s.s. *Insect Mol Biol 10:* 25–32.
40. Diabate A, Baldet T, Chandre C, Dabire KR, Kengne P, Guiguemde TR, Simard F, Guillet P, Hemingway J, Hougard JM, 2003. KDR mutation, a genetic marker to assess events of introgression between the molecular M and S forms of *Anopheles gambiae* (Diptera: Culicidae) in the tropical savannah area of west Africa. *J Med Entomol 40:* 195–198.
41. Blandin S, Levashina EA, 2004. Thioester-containing proteins and insect immunity. *Mol Immunol 40:* 903–908.
42. Martinez-Torres D, Chandre F, Williamson MS, Darriet F, Berge JB, Devonshire AL, Guillet P, Pasteur N, Pauron D, 1998. Molecular characterization of pyrethroid knockdown resistance (kdr) in the major malaria vector *Anopheles gambiae* s.s. *Insect Mol Biol 7:* 179–184.
43. Weill M, Lutfalla G, Mogensen K, Chandre F, Berthomieu A, Berticat C, Pasteur N, Philips A, Fort P, Raymond M, 2003. Comparative genomics: insecticide resistance in mosquito vectors. *Nature 423:* 136–137.