Insect Molecular Biology (2004) **13**(4), 371–377

Variation in an intron sequence of the voltage-gated sodium channel gene correlates with genetic differentiation between *Anopheles gambiae s.s.* molecular forms

G. Gentile*†, F. Santolamazza*, C. Fanello*‡, V. Petrarca*§, A. Caccone¶ and A. della Torre*

**Sezione di Parassitologia, Dipartimento di Scienze di Sanità Pubblica, and §Dipartimento di Genetica e Biologia Molecolare, Università di Roma 'La Sapienza', Rome, Italy; †Dipartimento di Biologia, II Università di Roma 'Tor Vergata', Rome, Italy; ‡Prince Leopold Institut of Tropical Medicine, Antwerp, Belgium; ¶Yale Institute for Biospheric Studies and Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT 06520–8106, USA*

Abstract

We present the results of a geographical survey of genetic variation in *Anopheles gambiae* M and S molecular forms from ten African countries at Intron I of the voltage-gated sodium channel gene. We found two major haplotypes separated by a single mutational step, which cosegregate almost completely with the rDNA sites that identify M and S, consistent with previous estimates of strong reductions of gene flow between the two forms. We also report ten additional haplotypes stemming from the two major haplotypes, mostly present in single localities. The low levels of genetic variation found in this intronic region are discussed in light of a possible selective sweep. These findings offer additional elements to the ongoing debate on the amount of genetic differentiation and isolation between the two molecular forms and on their taxonomic status.

Keywords: *Anopheles gambiae*, molecular forms, voltage-gated sodium channel gene.

Received 21 January 2003; accepted after revision 25 March 2004. Correspondence: Dr Alessandra della Torre, Sezione di Parassitologia, Dipartimento di Scienze di Sanità Pubblica, Università di Roma 'La Sapienza', Piazzale Aldo Moro, 5, 00185 Rome, Italy. Tel.: +39 06 49914932; fax: +39 06 49914653; e-mail: ale.dellatorre@uniroma1.it

Introduction

Anopheles gambiae s.s. (hereafter simply *A. gambiae*) is the species most specifically adapted to humans within the *A. gambiae* complex and shows the highest malaria vectorial capacity in the whole Afrotropical region. Genotyping X-linked rDNA of *A. gambiae* populations from West Africa led to the characterization of two molecular forms (provisionally named M and S), which differ in both the transcribed and the nontranscribed spacers in the rDNA repeat unit (della Torre *et al*., 2001; Gentile *et al*., 2001, 2002). The relationship between molecular and chromosomal forms, described on the basis of analysis of the distribution of fixed and polymorphic paracentric inversions on chromosome-2 (Coluzzi *et al*., 1985), varies in different ecogeographical conditions (della Torre *et al*., 2002). Although both indirect and direct genetic evidence indicate significant barriers to gene flow among different *A. gambiae* forms, these barriers do not appear to be complete throughout the entire range of distribution (Touré *et al*., 1998; Taylor *et al*., 2001; Tripet *et al*., 2001, 2003; Wondji *et al*., 2002). Moreover, analyses of several allozyme loci (Cianchi *et al*., 1983), coding and noncoding DNA regions scattered throughout the genome revealed no differentiation between forms, in contrast to results obtained from the analysis of more rapidly evolving DNA regions, such as rDNA and microsatellites (Black & Lanzaro, 2001). On the whole, molecular and chromosomal data suggest an ongoing speciation process, with the persistence of variation shared because of recent common ancestry and low levels of gene flow continuing to homogenize regions of the genome not directly involved in the speciation process; in other words, M and S appear to have mosaic genomes consisting of parts that are completely differentiated and between which gene flow is limited, whereas it occurs among other parts of the genome (Gentile *et al*., 2002; Besansky *et al*., 2003). Whether this picture justifies raising M and S molecular forms to the rank of species is still a matter of discussion.

Apart from the taxonomic debate, the most epidemiologically relevant issue is how the observed restrictions to gene

flow affect bionomics of the molecular forms, and to what extent and under what circumstances do they hinder the circulation of genes associated with vector capacity or insecticide resistance. Analysis of the voltage-gated sodium channel gene has already shown a non-homogeneous distribution of the *knock-down resistance* (*kdr*) allele in the two molecular forms, which accounts for differences in susceptibility to pyrethroid insecticides and DDT (Chandre *et al*., 1999; Diabate *et al*., 2003; Fanello *et al*., 2003).

Preliminary sequence analysis of the Intron I sequence upstream of the *kdr* mutation also showed two nucleotide differences associated with M and S in populations from Burkina Faso and Ivory Coast (Weill *et al*., 2000). These data also provided support for the hypothesis that in Benin, where the first M population carrying the *kdr* allele has been described (Fanello *et al*., 2000), this allele was acquired by M from S via introgressive hybridization. We tested these preliminary observations by increasing the sample size and enlarging the sampling area to ten West and Central African countries, with the aim to validate the possible role of the above nucleotide differences in Intron I as markers of genetic differentiation between M and S.

Results and discussion

We sequenced a 576 bp fragment of Intron I of the voltage-gated sodium channel gene in fifty-seven and seventy-one *A. gambiae* M and S specimens, respectively. Each specimen has been identified as M or S by PCR-RFLP, characterized for the *kdr* genotype and karyotyped in half-gravid females. Specimens were collected in twenty-one localities from ten West and Central African countries (Fig. 1). Both molecular forms were present at all localities, with the exception of Sombili (Guinea) and the three Angolan sites, where only one form was found. Because in several samples one form predominates, we have not been able to analyse a balanced number of specimens of the two forms for each locality.

We found twelve Intron I haplotypes, six and six in the M and S samples (Table 1). The genealogical relationships between these haplotypes are shown in Fig. 2 as a TCS network (Clement *et al*., 2000). Two haplotypes (S1 and M1) occurred in very high frequencies over the entire geographical area studied. They are separated by a single mutational step, a substitution of a cytosine (C) with a thymine (T) at site 702 (numbering follows Weill *et al*., 2000), which distinguishes two main clusters that almost completely correspond to the S and M molecular forms.

The third most common haploype (M5) was found in M individuals only, and it is characterized by an adenine/cytosine (A/C) polymorphism at position 896, consistent with the preliminary observations by Weill *et al*. (2000). The other nine haplotypes are geographical variants of either the S or the M forms, mostly linked to single localities. It is interesting to note that S samples from N'Gabakoro (Mali) included both Savanna and Bamako specimens and that one haplotype (S3) was found in the latter chromosomal form only. A large screening of sympatric Savanna and Bamako populations is under way to evaluate the possible significance of this mutation as a marker of genetic differentiation between these chromosomal forms.

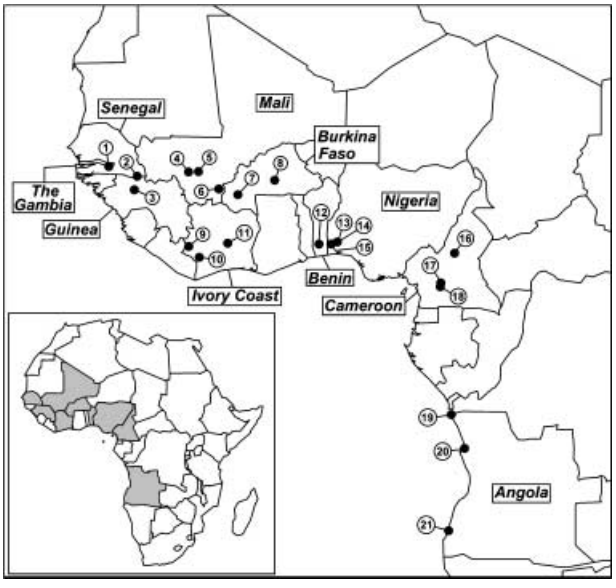

Figure 1. Map of the twenty-one localities (numbered as in Table 1) in the ten West and Central African countries where the *Anopheles gambiae* samples were collected.

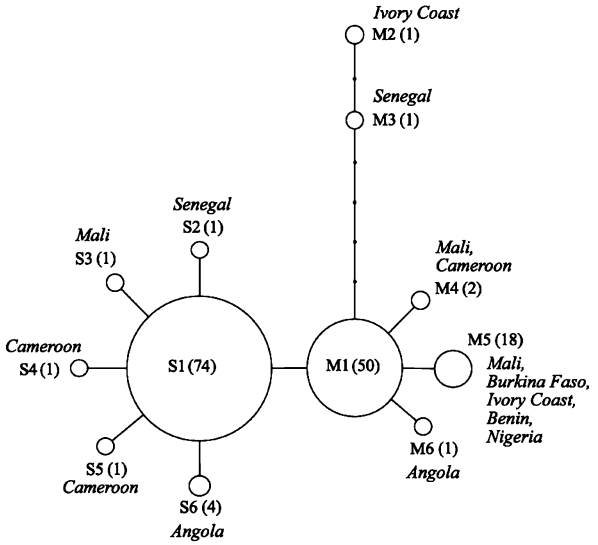

Figure 2. Genealogical relationships between haplotypes in the Intron I region of the voltage-gated sodium channel gene as revealed by statistical parsimony (Clement *et al*., 2000). Circles indicate the haplotypes (Table 1) with the size of the circle roughly proportional to the number of individuals carrying that haplotype (in parentheses). Branches connecting circles and/or dots represent single mutational steps between haplotypes. Countries where haplotypes were found are also reported. Haplotypes S1 (n = 74) and M1 (n = 50) were found in all the S and M form samples, respectively.

Table 1. Genotype distribution for the Intron I region of the voltage-gated sodium channel gene ('Intron I'), for the 702 site within this Intron ('Intron I 702') and for the *kdr* locus in populations of *Anopheles gambiae* collected in twenty-one localities (numbered as in Fig. 1) from ten African countries. The chromosomal forms given for each specimen are based on either direct chromosomal analysis or ecogeographical considerations (della Torre *et al*., 2001)

(Number of) Locality (longitude, latitude)	Sample size	Chromosomal form	Molecular form	Intron I	Intron I 702	*kdr*
Senegal						
(1) Kédougou	3	nr*	S	S1/S1	T/T	*s/s*
(12°10′W, 12°33′N)	1	Savanna	S	S1/S2	T/T	*s/s*
	1	Savanna	M	M1/M3	C/C	*s/s*
The Gambia						
(2) Basse	1	Savanna	S	S1/S1	T/T	*s/s*
(14°15 W, 13°13′N)	3	Savanna	M	M1/M1	C/C	*s/s*
Guinea						
(3) Sombili	2	Savanna	S	S1/S1	T/T	*s/s*
(12°16′W, 11°24′N)	1	Mopti	S	S1/S1	T/T	*s/s*
Mali						
(4) N'Gabakoro	4	Bamako	S	S1/S1	T/T	*s/s*
(07°01′W, 12°48′N)	1	Sav/Bam†	S	S1/S1	T/T	*s/s*
	1	Bamako	S	S1/S3	T/T	*s/s*
	2	Savanna	S	S1/S1	T/T	*r/r*
	1	Mopti	M	M4/M5	C/C	*s/s*
	1	Mopti	M	M1/M1	C/C	*s/s*
	1	Mopti	M	M1/M5	C/C	*s/s*
(5) Banambani	1	Bamako	S	S1/S1	T/T	*s/s*
(08°03′W, 12°48′N)	1	Bamako	S	S1/S1	T/T	*s/s*
	1	Savanna	S	S1/S1	T/T	*r/r*
	3	Mopti	M	M1/M1	C/C	*s/s*
(6) Pimperena	2	Savanna	S	S1/S1	T/T	*r/r*
(05°42′W, 11°28′N)	1	Savanna	S	S1/S1	T/T	*s/r*
	1	Savanna	S	S1/S1	T/T	*s/s*
	1	Savanna	M	M1/M5	C/C	*s/s*
	1	Mop/Sav [3]	M	M1/M1	C/C	*s/s*
	1	Mopti	M	M5/M5	C/C	*s/s*
Burkina Faso						
(7) Koubri	1	nr	S	S1/S1	T/T	*r/r*
(01°22′W, 12°09′N)	1	nr	S	S1/S1	T/T	*s/r*
	1	nr	M	M5/M5	C/C	*s/s*
	1	nr	M	M1/M1	C/C	*s/s*
(8) Dioulassouba	3	Savanna	S	S1/S1	T/T	*r/r*
(04°13′W, 11°02′N)	1	Savanna	M	M1/M5	C/C	*s/s*
	1	Mop/Sav‡	S	S1/S1	T/T	*r/r*
	2	Mop/Sav‡	M	M1/M5	C/C	*s/s*
	1	Mop/Sav‡	M	M5/M5	C/C	*s/s*
	1	Mopti	M	M1/M5	C/C	*s/s*
	1	Mopti	M	M1/M1	C/C	*s/s*
Ivory Coast						
(9) Danta	2	Forest	S	S1/S1	T/T	*s/s*
(08°02′W, 07°00′N)	2	Forest	M	M1/M1	C/C	*s/s*
	1	Forest	M	M2/M2	C/C	*s/s*
(10) Ziglo	4	Forest	S	S1/S1	T/T	*s/s*
(07°13′W, 06°09′N)	1	Forest	S	S1/S1	T/T	*s/r*
	2	Forest	M	M1/M1	C/C	*s/s*
(11) M'bé	2	Savanna	S	S1/S1	T/T	*r/r*
(05°01′W, 07°14′N)	1	Savanna	S	S1/S1	T/T	*s/r*
	1	Forest	S	S1/S1	T/T	*s/r*
	1	Forest	M	M1/M5	C/C	*s/s*
	1	Savanna	M	M5/M5	C/C	*s/s*
	1	Mopti	M	M1/M5	C/C	*s/s*
	1	Savanna	M	M1/M1	C/C	*s/s*
Benin						
(12) Bohicon	3	nr	S	S1/S1	T/T	*s/s*
(02°05E, 07°10′N)	2	nr	S	S1/S1	T/T	*r/r*
	2	nr	M	S1/M5	**C/T**	*s/r*
	1	nr	M	S1/M1	**C/T**	*s/r*
	1	nr	M	M1/M5	C/C	*s/s*

Table 1. (*Continued*)

(Number of) Locality (longitude, latitude)	Sample size	Chromosomal form	Molecular form	Intron I	Intron I 702	*kdr*
Nigeria						
(13) Olugbo	3	nr	S	S1/S1	T/T	*s/s*
(03°30′E, 07°20′N)	2	nr	M	M1/M1	C/C	*s/s*
	1	nr	M	M1/M5	C/C	*s/s*
(14) Kobape	1	nr	S	S1/S1	T/T	*s/s*
(03°30′E, 07°20′N)	1	nr	M	M1/M1	C/C	*s/s*
(15) Aiyetoro	3	nr	S	S1/S1	T/T	*s/s*
(03°01′E, 07°11′N)	1	nr	M	M1/M5	C/C	*s/s*
	1	nr	M	M1/M1	C/C	*s/s*
Cameroon						
(16) Simbock	1	Forest	S	S1/S1	T/T	*s/s*
(11°30′E, 03°51′N)	5	Forest	M	M1/M1	C/C	*s/s*
(17) Tibati	1	Forest	S	S1/S1	T/T	*s/s*
(12°37′E, 06°28′N)	4	Savanna	S	S1/S1	T/T	*s/s*
	1	Forest	S	S1/S5	T/T	*s/s*
	2	Mopti	M	M1/M1	C/C	*s/s*
	1	Forest	M	M1/M1	C/C	*s/s*
	1	Forest	M	S1/M1	**C/T**	*s/s*
	1	Mopti	M	M1/M4	C/C	*s/s*
(18) Obala	3	Forest	S	S1/S1	T/T	*s/s*
(11°33′E, 04°09′N)	1	Forest	S	S1/S4	T/T	*s/s*
	1	Forest	M	M1/M1	C/C	*s/s*
Angola						
(19) Namibe (12°09′E, 15°10′S)	3	Forest	M	M1/M1	C/C	*s/s*
(20) Luanda	2	For/Sav§	M	M1/M1	C/C	*s/s*
(13°23′E, 08°45′S)	1	For/Sav§	M	M1/M6	C/C	*s/s*
(21) Kikudo	4	For/Sav§	S	S1/S1	T/T	*s/s*
(12°22′E, 06°07′S)	2	For/Sav§	S	S1/S6	T/T	*s/s*
	1	For/Sav§	S	S6/S6	T/T	*s/s*
	1	For/Sav§	S	M1/S6	**C/T**	s/s

*nr = no chromosomal preparation was obtained.
†Karyotype with characteristics of both Savanna and Bamako: usually interpreted as a putative hybrid between the two forms.
‡Karyotypes with characteristics of both Mopti and Savanna and/or inversion 2Rd carriers: usually interpreted either as putative hybrids between the two forms, or rare variants of either the two forms (Coluzzi *et al*., 2002).
§Karyotypes monomorphic standard for 2R inversions (typical of the Forest chromosomal form), and polymorphic for 2La inversion (typical of the Savanna form).

The star-like shape of the TCS network suggests recent dynamics of genetic differentiation combined with recent population expansion, as proposed for other sites of the Afrotropical belt (e.g. São Tomé: Pinto *et al*., 2003; Kenya: Donnelly *et al*., 2001). Interestingly, the distribution of the M form haplotypes suggests a differentiation faster and/or earlier than within the S haplotypes: in fact, the M cluster includes haplotypes that are separated by more than one mutational step, i.e. M2 and M3, and are distributed over a wide geographical range, i.e. M5 and M4 (Fig. 2); by contrast, within the S cluster all the low-frequency haplotypes are only one mutational step from the most abundant one (S1), regardless of geographical location.

As mentioned above, most individuals can be assigned to the correct molecular form on the basis of the substitution at site 702. Only five exceptions contradict this general trend: three M specimens from Benin, one M from Cameroon and one S from Angola (indicated in bold type in Table 1). The three M specimens from Benin are *kds*/*kdr* heterozygotes and thus they are expected to be heterozygotes C/T at position 702, as a consequence of the introgression event that allowed the transfer of the *kdr* allele from S to M (Weill *et al*., 2000). The remaining two C/T heterozygotes are homozygous for the *kds* allele. These genotypes could result from past introgression events between S and M.

It is worth noting that the *kdr* allele was found in nineteen S form specimens from Mali, Burkina Faso, Ivory Coast and Benin in both homozygous and heterozygous states; all the M samples were *kds* homozygous, with the single exception of the Bohicon (Benin) population. As already pointed out, three out of four M individuals analysed from this village are heterozygotes C/T for the mutation at position 702. By contrast, the single M specimen from the same sample that does not carry the *kdr* allele is a C/C homozygous (see Table 1), as are all M specimens collected from other localities. This is entirely consistent with the hypothesis that segregating sites in the *kdr* and the Intron I region (separated only by a few tens of bases) have been both introgressed from S to M (Weill *et al*., 2000).

We confirm a low level of variation in Intron I, as reported by Weill *et al*. (2000). Interestingly, no variation was found in the coding region of the voltage-gated sodium channel gene in twenty M and S individuals sequenced from Senegal, Cameroon, Nigeria and Angola. Weill *et al*. (2000) hypothesized that this low level of variation could have resulted from a local selective sweep. This hypothesis is based on the assumption that a selective sweep, which occurs when a favourable mutation rapidly spreads through a population, tends strongly to reduce or eliminate existing variability (Harvey & Steers, 1999). Thus, the variation found in these regions should have arisen after the sweep. In fact, the percentage of variable sites that we found in Intron I was quite low (2.4%) compared with that found in different intronic regions of other *A. gambiae* populations (7.5–18.6%, Gentile *et al*., 2001). However, at present, we do not know to which gene this region is linked. The voltage-gated sodium channel gene is not a likely candidate, because the strong selective pressure that acts on this gene is related to the resistance to insecticide conferred by the *kdr* allele. If the voltage-gated sodium channel gene was responsible for the selective sweep, we would expect that most individuals would have the *kdr* allele. Interestingly, this allele occurred only in nineteen S form individuals from Mali, Burkina Faso, Ivory Coast and Benin. All were homozygous for the same haplotype S1. In these individuals the percentage of variable sites is zero, suggesting a genetic sweep directly related to the voltage-gated sodium channel gene, in which the *kdr* allele could have arisen only once, before spreading across West Africa. We are currently investigating this hypothesis on a larger sample. Following the same line of reasoning, the variation pattern at site 702 is not correlated with the voltage-gated sodium channel gene that carries the *kdr* allele, with the only obvious exception of the three M individuals from Bohicon, which are *kds*/*kdr* heterozygotes and also heterozygous at position 702.

Could the variation pattern observed in the sequence of Intron I be due to linkage and cosegregation with other markers that show similar patterns, such as rDNA region and the chromosomal inversions? This is unlikely, because the voltage-gated sodium channel gene maps in Div. 20C on the centromeric end of arm L of chromosome-2 (Ranson *et al*., 2000), which is not linked to the rDNA cluster (chromosome-X) or close to the paracentric inversions of chromosome-2 that are used to define the forms (Coluzzi *et al*., 1985; Touré *et al*., 1998).

In conclusion, our study describes the first marker found in an Intron DNA region that consistently correlates over a wide geographical range with the rDNA markers used to define *A. gambiae* molecular forms. Because the sodium channel gene maps on arm L of chromosome-2, this marker is completely independent from those described so far. The results provide further data supporting the genetic differentiation between *A. gambiae* molecular forms and corroborate the hypothesis that M and S have mosaic genomes of exchangeable and not-exchangeable regions. In fact, other empirical evidence suggests that reproductive isolation may not be a genome-wide phenomenon, especially at the early stage of the speciation process (Noor *et al*., 2001; Wu, 2001). Our results offer additional elements in the debate about the specific status of the M and S forms. We are now extending the sequence analysis of Intron I both to samples of *A. gambiae s.s*. molecular forms from a wider geographical range (from West to East Africa and Madagascar) and to other species of the *A. gambiae* complex.

Experimental procedures

The 128 specimens of *A. gambiae* used in this study were collected in twenty-one localities from ten African countries, which are shown in Fig. 1. Karyotypes of collected samples were assessed as in Coluzzi *et al*. (1979) and della Torre (1997). Genomic DNA was extracted from each individual following Favia *et al*. (1994). Specimens were assigned to a given taxon *a posteriori*, using PCR-RFLP (Favia *et al*., 1997; Fanello *et al*., 2002).

All individuals were scored for the occurrence of the *kdr* allele in the voltage-gated sodium channel gene either by using the protocol reported in Fanello *et al*. (1999) and Kolaczinski *et al*. (2000) or by direct sequencing.

A 576 bp long fragment of Intron I of the voltage-gated sodium channel gene was PCR-amplified. Primers were I1rev (Weill *et al*., 2000), and Ganest (5-CATACAT TGCT TAAAGCTCTAAT TATC-3), a new primer located from position 388 to 414 of the sequence published by Weill *et al*. (2000). In a few individuals the occurrence of the *kdr* allele was determined by sequencing. In these cases the primer Ganest was used in combination with a new reverse primer, 359–378R (5′-TGGTGCAGACAAGGATGATG-3′), located within the coding region downstream of Intron II, at position 359–378 in the GenBank sequence AGY13592. By using this primer pair we PCR-amplified a 794 bp long fragment that included the partial sequence of Intron I, the entire 188 bp sequence of the coding region downstream of Intron I, the entire 57 bp sequence of Intron II and the partial sequence of the coding region downstream of Intron II. Details of the PCR conditions are available upon request. PCR fragments were gel purified using the QIAquick Gel Extraction Kit (Qiagen). Double-stranded DNA fragments were sequenced on an ABI 310 automated sequencer (Applied Biosystems) using protocols recommended by the manufacturer. Both strands were sequenced. Sequences were analysed using the program Sequencer 3.1.1 (Gene Codes Corp.), and aligned by eye. All new Intron I haplotypes have been deposited in GenBank under accession number AY615526–AY615653.

The gametic phase of the heterozygotes at more than one site was determined either by cloning, using the PGEM–Tvector System I cloning kit (Promega Corp., Madison, WI, USA) and sequencing multiple clones from each individual, or by using the statistical method described in Stephens *et al*. (2001). This is a Bayesian method based on *a priori* predictions from the coalescent theory regarding the patterns of haplotypes to be expected in natural populations. This method is implemented in the software PHASE v.1.0.1 (Stephens *et al*., 2001). In order to test the reliability of the method, those individuals whose gametic phase was

established by cloning were included in the dataset for statistical estimation. For all these individuals, the statistical method provided the same gametic phase as that detected by cloning.

Variability estimates were obtained using DNASP v.3 (Rozas & Rozas, 1999).

We used the TCS software (Clement *et al*., 2000) to investigate genealogical relationships among haplotypes. This software is based on the method of statistical parsimony described by Templeton *et al*. (1992), which uses parsimony in a pair-wise calculation of the number of mutational steps between haplotypes until the probability threshold of 95% is exceeded. The software represents the relationships between haplotypes in the form of a network in which branches represent mutational steps between different haplotypes.

Acknowledgements

We are grateful to all the scientists and entomology teams who made this study possible by providing us with field samples. We especially wish to thank: M. Akogbeto, A. Bosman, J. H. Bryan, P. J. Cani, C. Carrara, C. Costantini, C. F. Curtis, I. Dia, J. Dossou-Yovo, D. Fontenille, M. Kristan, N'F. Sagnon, F. Simard, Y. T. Touré and C. Wondji. We acknowledge the technical assistance of M. Calzetta and G. Petrangeli and the support and advice of Professor Mario Coluzzi. Financial support was provided by the UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR) and the MIUR/COFIN funds to G.G., A.d.T., F.S. and V.P. and YIBS funds to A.C.

References

Besansky, N.J., Krzywinski, J., Lehmann, T., Simard, F., Kern, M., Mukabayire, O., Fontenille, D., Toure, Y. and Sagnon, N. (2003) Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: evidence from multilocus DNA sequence variation. *Proc Natl Acad Sci USA* **100**: 10818–10823.

Black, W.C. IV and Lanzaro, G.C. (2001) Distribution of genetic variation among chromosomal forms of *Anopheles gambiae s.s*. introgressive hybridization, adaptive inversions, or recent reproductive isolation? *Insect Mol Biol* **10**: 3–7.

Chandre, F., Manguin, S., Brengues, C., Dossou Yovo, J., Darriet, F., Diabate, A., Carnevale, P. and Guillet, P. (1999) Current distribution of pyrethroid resistance gene (*kdr*) in *Anopheles gambiae* complex from West Africa and further evidence for reproductive isolation of the Mopti form. *Parassitologia* **41**: 319–322.

Cianchi, R., Villani, F., Touré, Y.T., Petrarca, V. and Bullini, L. (1983) Electrophoretic study of different chromosomal forms within *Anopheles gambiae s.s*. *Parassitologia* **25**: 239–241.

Clement, M., Posada, D. and Crandall, K.A. (2000) TCS: a computer program to estimate gene genealogies. *Mol Ecol* **9**: 1657–1659.

Coluzzi, M., Petrarca, V. and Di Deco, M.A. (1985) Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae*. *Boll Zool* **52**: 45–63.

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

Coluzzi, M., Sabatini, A., Petrarca, V. and Di Deco, M.A. (1979) Chromosomal differentiation and adaptation to human environment in the *Anopheles gambiae* complex. *Trans R Soc Trop Med Hyg* **73**: 483–497.

Diabate, A., Baldet, T., Chandre, R., Dabire Koumbor, R., Kengne, P., Guiguemde, T., Simard, F., Guillet, P., Hemingway, J. and Hougard, J. (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.

Donnelly, M.J., Licht, M.C. and Lehmann, T. (2001) Evidence for recent population expansion in the evolutionary history of the malaria vectors *Anopheles arabiensis* and *Anopheles gambiae*. *Mol Biol Evol* **8**: 1353–1364.

Fanello, C., Akogbeto, M. and della Torre, A. (2000) Distribution of the pyrethroid knockdown resistance gene (*kdr*) in *Anopheles gambiae s.l*. from Benin. *Trans R Soc Trop Med Hyg* **94**: 132.

Fanello, C., Kolaczinski, J.H., Conway, D.J., Carnevale, P. and Curtis, C. (1999) The *kdr* pyrethroid resistance gene in *Anopheles gambiae*: tests of non-pyrethroid insecticides and a new detection method for the gene. *Parassitologia* **41**: 323–326.

Fanello, C., Petrarca, V., della Torre, A., Santolamazza, F., Alloueche, A., Coulibaly, M., Dolo, G., Curtis, C., Touré, Y.T. and Coluzzi, M. (2003) The pyrethrooid *knock-down* resistance gene in *Anopheles gambiae* complex in Mali and further evidence of incipient speciation within *An. gambiae s.s*. *Insect Mol Biol* **12**: 241–245.

Fanello, C., Santolamazza, F. and della Torre, A. (2002) PCR-RFLP methods for simultaneos detection of molecular forms of *Anopheles gambiae s.s*. and other species of the complex. *Med Vet Entomol* **16**: 461–464.

Favia, G., della Torre, A., Bagayoko, M., Lanfrancotti, A., Sagnon N'F., Touré, Y.T. and Coluzzi, M. (1997) Molecular identification of sympatric chromosomal forms of *Anopheles gambiae* and further evidence of their reproductive isolation. *Insect Mol Biol* **6**: 377–383.

Favia, G., Dimopoulos, G., della Torre, A., Touré, Y.T., Coluzzi, M. and Louis, C. (1994) Polymorphisms detected by random PCR distinguish between different chromosomal forms of *Anopheles gambiae*. *Proc Natl Acad Sci USA* **91**: 10315–10319.

Gentile, G., della Torre, A., Maegga, B., Powell, J.R. and Caccone, A. (2002) Genetic differentiation in the African malaria vector, *Anopheles gambiae s.s*. and the problem of taxonomic status. *Genetics* **161**: 1561–1578.

Gentile, G., Slotman, M., Ketmaier, V., Powell, J.R. and Caccone, A. (2001) Attempts to molecularly distinguish cryptic taxa in *Anopheles gambiae s.s*. *Insect Mol Biol* **10**: 25–32.

Harvey, P.H. and Steers, H. (1999) One use of phylogenies for conservation biologists: inferring population history from gene sequences. In *Genetics and the Extinction of Species* (Landweber, L.F. and Dobson, A.P., eds), pp. 101–120. Princeton University Press, Princeton.

Kolaczinski, J.H., Fanello, C., Hervé, J.-P., Conway, D.J., Carnevale, P. and Curtis, C. (2000) Experimental and molecular genetic analysis of the impact of pyrethroid and non-pyrethroid insecticide impregnated bednets for mosquito control in an area of pyrethroid resistance. *Bull Entomol Res* **90**: 125–132.

Noor, M.A.F., Grams, K.L., Bertucci, L.A. and Reiland, J. (2001) Chromosomal inversions and the reproductive isolation of species. *Proc Natl Acad Sci USA* **98**: 12084–12088.

Pinto, J., Donnelly, M.J., Sousa, C.A., Malta-Vacas, J., Gil, V., Ferreira, C., Petrarca, V., Do Rosario, V.E. and Charlwood, J.D. (2003) An island within an island: genetic differentiation of *Anopheles gambiae* in Sao Tome, West Africa, and its relevance to malaria vector control. *Heredity* **91**: 407–414.

Ranson, H., Jensen, B., Vulule, J.M., Wang, X., Hemingway, J. and Collins, F.H. (2000) Identification of a point mutation in the voltage-gated sodium channel gene of Kenyan *Anopheles gambiae* associated with resistance to DDT and pyrethroids. *Insect Mol Biol* **9**: 491–497.

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

Stephens, M., Smith, N. and Donnelly, P. (2001) A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet* **68**: 978–989.

Taylor, C.E., Touré, Y.T., Carnahan, J., Norris, D.E., Dolo, G., Traoré, S.F., Edilla, F.E. and Lanzaro, G.C. (2001) Gene flow among populations of the malaria vector, *Anopheles gambiae*, in Mali, West Africa. *Genetics* **157**: 743–750.

Templeton, A.R., Crandall, K.A. and Sing, C.F. (1992) A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* **132**: 619–633.

della Torre, A. (1997) Polytene chromosome preparation from Anopheline mosquitoes. In *Molecular Biology of Insect Disease Vectors: a Methods Manual* (Crampton, J.M., Beard, C.B. and Louis, C., eds), pp. 329–336. Chapman & Hall, London.

della Torre, A., Costantini, C., Besansky, N.J., Caccone, A., Petrarca, V., Powell, J.R. and Coluzzi, M. (2002) Incipient speciation in *Anopheles gambiae*: the glass is half-full. *Science* **298**: 115–117.

della Torre, A., Fanello, C., Akogbeto, M., Dossou-yovo, J., Favia, F., Petrarca, V. and Coluzzi, M. (2001) Molecular evidence of incipient speciation within *Anopheles gambiae s.s.* in West Africa *Insect Mol Biol* **10**: 9–18.

Touré, Y.T., Petrarca, V., Traoré, S.F., Coulibaly, A., Maiga, H.M., Sankaré, O., Sow, M., Di Deco, M.A. and Coluzzi, M. (1998) The distribution and inversion polymorphism of chromosomally recognized taxa of the *Anopheles gambiae* complex in Mali, West Africa. *Parassitologia* **40**: 477–511.

Tripet, F., Toure, Y.T., Dolo, G. and Lanzaro, G.C. (2003) Frequency of multiple inseminations in field-collected *Anopheles gambiae* females revealed by DNA analysis of transferred sperm. *Am J Trop Med Hyg* **68**: 1–5.

Tripet, F., Touré, Y.T., Taylor, C.E., Norris, D.E., Dolom, G. and Lanzaro, G.C. (2001) DNA analysis of transferred sperm reveals significant levels of gene flow between molecular forms of *Anopheles gambiae*. *Mol Ecol* **10**: 1725–1732.

Weill, M., Chandre, F., Berngues, C., Manguin, S., Akogbeto, M., Pasteur, N., Guillet, P. and Raymond, M. (2000) The *kdr* mutation occurs in the Mopti form of *Anopheles gambiae s.s.* through introgression. *Insect Mol Biol* **9**: 451–455.

Wondji, C., Simard, F. and 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.

Wu, C.I. (2001) The genic view of the process of speciation. *J Evol Biol* **14**: 851–865.