Medical and Veterinary Entomology (2007) **21**, 358–369

Evidence for late Pleistocene population expansion of the malarial mosquitoes, *Anopheles arabiensis* and *Anopheles gambiae* in Nigeria

S. D. MATTHEWS[1,2], L. J. MEEHAN[2], D. Y. ONYABE[3], J. VINEIS[2], I. NOCK[4], I. NDAMS[4] and J. E. CONN[1,2]

[1]Department of Biomedical Sciences, School of Public Health, University at Albany, State University of New York, New York, U.S.A., [2]Wadsworth Center, Griffin Laboratory, New York State Department of Health, Slingerlands, New York, U.S.A., [3]Bacilligen, Inc., Rockville, Maryland, U.S.A. and [4]Department of Biological Sciences, Ahmadu Bello University, Zaria, Nigeria

Abstract. *Anopheles gambiae* Giles *s.s.* and *Anopheles arabiensis* Patton (Diptera: Culicidae) are major vectors of malaria in Nigeria. We used 1115 bp of the mitochondrial *COI* gene to assess their population genetic structures based on samples from across Nigeria ($n = 199$). The mtDNA neighbour-joining tree, based on F_{ST} estimates, separated *An. gambiae* M and S forms, except that samples of *An. gambiae* M from Calabar clustered with all the *An. gambiae* S form. *Anopheles arabiensis* and *An. gambiae* could be combined into a single star-shaped, parsimonious haplotype network, and shared three haplotypes. Haplotype diversity values were high in *An. arabiensis* and *An. gambiae* S, and intermediate in *An. gambiae* M; all nucleotide diversities were relatively low. Taken together, patterns of haplotype diversity, the star-like genealogy of haplotypes, five of seven significant neutrality tests, and the violation of the isolation-by-distance model indicate population expansion in *An. arabiensis* and *An. gambiae* S, but the signal was weak in *An. gambiae* M. Selection is supported as an important factor shaping genetic structure in *An. gambiae* in Nigeria. There were two geographical subdivisions in *An. arabiensis*: one included all southern localities and all but two central localities; the other included all northern and two central localities. Re-analysing an earlier microsatellite dataset of *An. arabiensis* using a Bayesian method determined that there were two distinctive clusters, northern and southern, that were fairly congruent with the mtDNA subdivisions. There was a trend towards decreasing genetic diversity in *An. arabiensis* from the northern savannah to the southern rainforest that corroborated previous data from microsatellites and polytene chromosomes.

Key words. *Anopheles arabiensis*, *Anopheles gambiae*, cytochrome oxidase I, mitochondrial DNA, molecular forms, population expansion, population structure, Nigeria.

Introduction

The *Anopheles gambiae* Giles complex is comprised of seven species, including *Anopheles arabiensis* Patton and *An. gambiae s.s.* (Hunt *et al.*, 1998), two of the most important malaria vectors in Africa. *Anopheles arabiensis* and *An. gambiae* are found sympatrically throughout their common ranges in sub-Saharan Africa (Besansky *et al.*, 1997), although *An. arabiensis* is more widespread (Coetzee *et al.*, 2000). This species is also more tolerant of dry environments and is the primary vector in Sahelian savannah (Gillies & Coetzee, 1987). *Anopheles gambiae* has been subdivided into five chromosomal forms (Bryan *et al.*, 1982; Coluzzi *et al.*, 1985; Touré *et al.*, 1998), and two distinct genotypes, designated as molecular forms M and S (Favia *et al.*,

Correspondence: Jan E. Conn, Griffin Laboratory, New York State Department of Health, 5668 State Farm Road, Slingerlands, New York 12159, U.S.A. Tel.: +1 518 869 4575; Fax: +1 518 869 6487; E-mail: jconn@wadsworth.org

1997). The S form appears to be very widespread, whereas the M form is mostly restricted to West and Central Africa (Onyabe *et al.*, 2003; Masendu *et al.*, 2004; della Torre *et al.*, 2005). Genetic differentiation between *An. gambiae* M and S suggest assortative mating and incipient speciation (Taylor *et al.*, 2001; della Torre *et al.*, 2001; Wondji *et al.*, 2002; Lehmann *et al.*, 2003) resulting, at least in part, from selection (Stump *et al.*, 2005a). Reproductive isolation between M and S is proposed to be the result of the low recombination rate associated with the centromere on the X chromosome (Stump *et al.*, 2005b).

The population structures of *An. arabiensis* and *An. gambiae* in Africa have been investigated in numerous studies. The major geographic barrier to gene flow is the arid Great Rift Valley in East Africa. With microsatellite markers, *An. gambiae* demonstrates restricted gene flow across the Rift Valley (Kamau *et al.*, 1999; Lehmann *et al.*, 2003). By contrast, *An. arabiensis* data are conflicting, with Kamau *et al.* (1999) finding no restriction using microsatellite loci, and Temu & Yan (2005) detecting population differentiation using microsatellites inside inversions, but little differentiation with microsatellites outside inversions or with an mtDNA *ND5* fragment (in a comparison of samples from western Kenya, the Rift Valley and coastal Kenya). On a much larger geographic scale in East Africa, from Sudan to Mozambique, Donnelly & Townson (2000) uncovered strong evidence of population subdivision and recent range expansion in *An. arabiensis*, based on microsatellite data. Both species also exhibit minimal isolation by distance (IBD) across the entire continent (Besansky *et al.*, 1997), and population expansions that were detected in Kenya (Donnelly *et al.*, 2001) are assumed to have spread across the continent (Lehman *et al.*, 2003; Michel *et al.*, 2005) at about the time agriculture arose (Coluzzi *et al.*, 2002). One result of recent population expansion is the shallow population structure of *An. gambiae* across Africa (Lehmann *et al.*, 2003).

Densely populated Nigeria, in western Africa, is devoid of major geographical features that provide obvious barriers to gene flow. It is drier in the north and wetter in the south, with a cline of rainfall between the two extremes of 500 mm and 4000 mm annually (Davies, 1977; Coluzzi *et al.*, 1979). Malaria transmission is subcontinuous in northern Nigeria, occurring largely during the rainy season, and continuous in southern Nigeria (Taylor *et al.*, 1993; Charlwood *et al.*, 1995). Both *An. arabiensis* and *An. gambiae* are involved in malaria transmission and are found sympatrically throughout the country (Lindsay & Martens, 1998; Coetzee *et al.*, 2000; Onyabe & Conn, 2001a). Using microsatellite loci located both within and outside inversions and sampled from across the genome, Onyabe & Conn (2001b) concluded that there is extensive gene flow in *An. gambiae* across Nigeria but that selection counters the homogenizing effects of gene flow. By contrast, Onyabe & Conn (2001c) found severely restricted gene flow with IBD in *An. arabiensis* in Nigeria using microsatellite loci, and they observed a north–south reduction of heterozygosity and number of alleles per locus and concluded that this species had expanded its range from northern to southern Nigeria.

Detailed knowledge of the population structure of vectors, such as the geographical limits of populations, estimates of gene flow useful in the prediction of the spread of genes associated with insecticide resistance, the determination and timing of bottlenecks or expansions, and the use of effective population size (N_e) and genetic diversity of microsatellites to evaluate the effectiveness of insecticide treatment, can provide more effective control of malaria transmission (Lehmann *et al.*, 2003; Coetzee & Fontenille, 2004; Wondji *et al.*, 2005). The possibility of releasing transgenic mosquitoes also requires understanding of gene flow to assess the potential for hybridization with or replacement of the natural population (reviewed in Ashburner *et al.*, 1998; Collins *et al.*, 2000; Benedict & Robinson, 2003; Moreira *et al.*, 2004; Morlais *et al.*, 2005).

Rubinoff & Holland (2005) emphasize the utility of comparing data from mtDNA and nuclear DNA to further understand the complexities of population structure. Because there have been no previous studies using mtDNA to assess population structure across Nigeria, we chose the mitochondrial *COI* gene as an independent marker to investigate *An. arabiensis* and *An. gambiae* in comparison with studies that focused on nuclear DNA (Coluzzi *et al.*, 1979; Onyabe & Conn, 2001b, 2001c). With this focus on mtDNA sequences, we addressed the following questions. Is there population genetic evidence of subdivision and restricted gene flow within *An. arabiensis?* Is there evidence of reduced genetic diversity in *An. arabiensis* from the northern savannah to the southern rainforest? Is the assumption of a recent population expansion in *An. arabiensis* and *An. gambiae* in Nigeria valid? Do both species follow the IBD model?

Materials and methods

Mosquito collection and identification

Mosquito collection and identification procedures are described in Onyabe & Conn (2001a). Larvae and pupae were collected in 1997 and 1999 from at least seven pools within a 1-km radius at each locality (Fig. 1). Larvae were reared on site until emergence and identified as adults to species complex using the keys of Gillies & Coetzee (1987). All samples (adult mosquitoes from all localities except Garki and Sokoto, where larvae were used) were stored in 95% ethanol. We analysed 115 *An. arabiensis*, 58 *An. gambiae* S and 26 *An. gambiae* M from multiple localities in Nigeria (Fig. 1, Table 1). Sokoto and Garki are in Sudan savannah; Giwa and Kaduna are in northern Guinea savannah; Abuja, Jebba, Bida, Lafia, Kwenev and Otukpo are in southern Guinea savannah, Okigwe and Ogbomoso are in savannah–forest transition, and Benin, Sapele and Calabar are in rainforest (Coluzzi *et al.*, 1979).

DNA extraction and sequencing

DNA was extracted using the method of Collins *et al.* (1987) and identification to species was conducted by polymerase chain reaction (PCR) (Scott *et al.*, 1993). Identification of *An. gambiae* M and S was carried out using both restriction fragment length polymorphism (RFLP)-PCR and diagnostic PCR (Paskewitz & Collins, 1990; Favia *et al.*, 1997, 2001). A

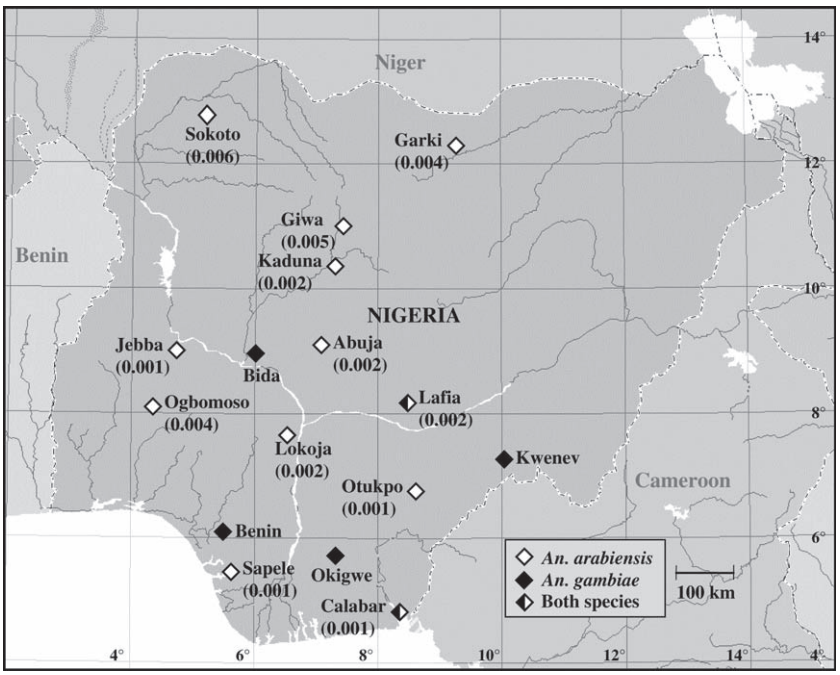

Fig. 1. Map showing collection sites of *Anopheles arabiensis* and *Anopheles gambiae* in Nigeria. Numbers in parentheses are nucleotide diversity values (π; Tajima, 1983) for *An. arabiensis*.

1300-bp region of the mitochondrial *COI* gene was amplified, using primers from UEA3: TAT AGC ATT CCC ACG AAT AAA TAA, and UEA10: TCC AAT GCA CTA ATC TGC CAT ATT A (Lunt *et al.*, 1996). Polymerase chain reactions were carried out in 25 μL reactions using 1 PuReTaq Ready-To-Go PCR Bead (Amersham Biosciences, Piscataway, NJ, U.S.A.), (~2.5 units of PuReTaq DNA polymerase, 10 mM Tris-HCl, [pH 9.0 at room temperature], 50 mM KCl, 1.5 mM $MgCl_2$, 200 μM of each dNTP, stabilizers, and bovine serum albumin), ~1/100 the DNA of the mosquito, 2 μM of each primer. A 4-min initial denaturation at 95 °C was followed by 45 cycles of amplification (95 °C for 40 s, 42 °C for 1 min, 72 °C for 1 min) (PTC 100

Table 1. Distribution of haplotypes across Nigeria. Haplotypes sampled in both species are in bold type; numbers in parentheses indicate the frequency if > 1.

	Anopheles arabiensis		*Anopheles gambiae*			
			M form		S form	
Site	*n*	Haplotypes	*n*	Haplotypes	*n*	Haplotypes
Sokoto	10	1(2),2,3,4,5,6,7,8,9				
Garki	10	2,10,11,12,13,14,**15**,16,17,**27**				
Giwa	10	3,14,18,**19**,20,21,22,23,24,25				
Kaduna	3	26,**27**,28				
Abuja	10	**27**(5)**,**28(2),29,30,31				
Bida			6	**27**,**15,**55(4)	11	48,49(2),50,51(2),52,53,54,55(2)
Jebba	10	**27**(5),28(5)				
Lafia	11	**27**(3),32,33(2),34(5)			11	**27**(3),54,55(3),70,71,72,73
Ogbomoso	10	**27**,33,35(2),36,37(3),38,39				
Lokoja	10	28,33(4),40(5)				
Kwenev			8	**15**(2),55(6)	8	**19**,52,65,66,67,68,69(2)
Otukpo	10	26(2),**27**(7),41				
Benin			10	55(7),81,82,83	6	43,44,45(2),46,47
Okigwe					11	54,55(3),74,75,76,77,78,79,80
Sapele	10	**27**(3),28(6),42				
Calabar	11	26,**27**(9),28	2	55,84	11	**27**,55,56,57,58,59,60,61,62,63,64

n, number of individuals.

and PTC 200, MJ Research/BioRad Laboratories, Inc., Waltham, MA, U.S.A.). The PCR product was visualized under ultraviolet (UV) light and purified using Centri-Spin 40 Columns (Princeton Separations, Inc., Adelphia, NJ, U.S.A.) and sequenced in both directions using Applied Biosystems model 3700 DNA analyser or Applied Biosystems model 3100 genetic analyser. Sequences were deposited into Genbank (accession numbers DQ465250–DQ465336). SEQUENCHER 4.1 was used to compile, edit and align sequences (Gene Codes Corp., Ann Arbor, MI, U.S.A.).

Phylogenetic relatedness

Once aligned, sequences were imported into PAUP* 4.0b10 (Swofford, 2002) and haplotypes were condensed in MACCLADE 3.04 (Maddison & Maddison, 1992). MEGA 2.1 (Kumar *et al.*, 2001) was used to construct maximum parsimony and minimum evolution trees. Nodal support was determined using non-parametric bootstrap proportion (Felsenstein, 1985). F_{ST} estimates of genetic distances between populations (Wright, 1951) were calculated using ARLEQUIN 2.000 (Schneider *et al*., 2000), and used as distance measures to conduct a neighbour-joining (NJ) analysis with MEGA 2.1 (Kumar *et al.*, 2001). The program TCS was employed to create a minimum-spanning haplotype network (Clement *et al.*, 2000) and the probability of parsimonious connections between haplotypes was calculated as in Templeton *et al.* (1992). Homoplasies were removed from the haplotype networks using the rules presented by Crandall & Templeton (1993) and Posada & Crandall (2001), summarized in Uthicke & Benzie (2003), for resolving undetermined pathways.

Population analyses

The organization of the genetic variation in *An. arabiensis* was initially assessed with the program STRUCTURE Version 2.0 (Pritchard *et al.*, 2000), which assigns individuals from all populations to a predetermined number of clusters (*k*). We used microsatellite data from Onyabe & Conn (2001c) to estimate the number of populations/clusters represented by the sample localities in the present study (Fig. 1). All samples of *An. arabiensis* in Fig. 1 except those from Lafia, Lokoja and Calabar (excluded because they were not analysed in the microsatellite study by Onyabe & Conn [2001c]) were analysed with STRUCTURE. The origin of each specimen was unknown, but *k*, the number of clusters, was determined a priori for each run. A series of independent runs was conducted using *k* from 1 to 8, with a burn-in period of 100 000 Markov Chain Monte Carlo (MCMC) iterations, and a data collection period of 200 000 MCMC iterations. We performed 10 runs at each *k*-value. Log-likelihood scores were averaged across runs and compared to determine the posterior probability of each *k*. At $k = 2$ the likelihood reached an obvious plateau. The two clusters consisted of northern (Garki, Giwa and Sokoto) and southern (Ogbomoso, Otukpo and Sapele), and a group of admixed populations (Abuja, Jebba and Kaduna; Fig. 2) that are located approximately midway geographically between the northern and southern clusters. The northern *An. arabiensis* were assigned at rates of 90.5–95.6%

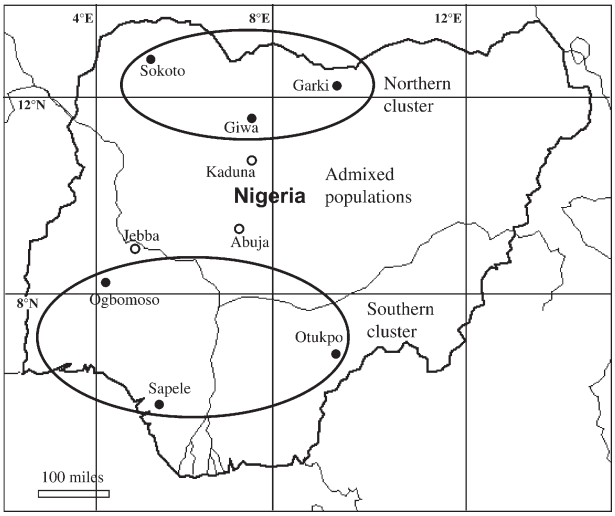

Fig. 2. Results of Bayesian analysis of *Anopheles arabiensis* microsatellite data from Onyabe & Conn (2001c) depicting the northern and southern clusters. Admixed populations, shown by ○, are those with some individuals assigned to each of the northern and southern clusters.

and the southern *An. arabiensis* were assigned at rates of 93.6–97.6%. Assignment rates for admixed populations were as follows: Kaduna, 50.1% N, 49.9% S; Abuja,– 38.8% N, 61.2% S, and Jebba, 36.2% N, 63.8% S. The three populations in the mtDNA analysis (Calabar, Lafia and Lokoja; Fig. 1) that were not represented in the STRUCTURE analysis, fall geographically either between the admixed populations and the southern cluster (Lokoja, Lafia) or south of the southern cluster (Calabar). Therefore, we created a non-northern cluster (based on geography) consisting of Ogbomoso, Otukpo, Sapele, Abuja, Jebba, Kaduna, Calabar, Lafia and Lokoja, and compared it with the northern cluster to determine whether the mtDNA data would detect the signature of a population expansion consistent with the finding of Onyabe & Conn (2001c).

Historical demography

For both *An. arabiensis* and *An. gambiae*, mtDNA datasets for haplotype diversity (Nei, 1987) and nucleotide diversity (Tajima, 1983) were calculated using DNASP 4.0 (Rozas *et al.*, 2003). The hypothesis that all mutations are selectively neutral, and the null hypothesis of longterm stable population size were examined using Tajima's D_T, Fu & Li's D^* and F^* statistics (Tajima, 1989; Fu & Li, 1993), Fu's F_S (Fu, 1997), R^2 (Ramos-Onsins & Rozas, 2002) and Strobeck's S statistic (Strobeck, 1987), calculated in DNASP 4.0. Tajima's D_T is estimated by comparing the differences between the number of segregating sites and the average number of nucleotide differences. Fu & Li's D^* and F^* statistics compare estimates of theta based on mutations in internal and external branches of a genealogy. Fu's F_S and Strobeck's S were used to assess the haplotype structure on the basis of the haplotype frequency distribution. The mismatch distribution, a frequency distribution of the observed number of

pairwise sequence differences, was used to distinguish between a smooth unimodal distribution and a ragged or multimodal distribution (Rogers & Harpending, 1992; Rogers, 1995) in ARLEQUIN 2.000 using the generalized non-linear least-square approach of Schneider & Excoffier (1999). The raggedness statistic (r) to quantify the smoothness of the mismatch distribution (Harpending, 1994) was calculated in DNASP 4.0. To test for positive correlation between geographic and genetic distance, Mantel tests (Mantel, 1967) were conducted in ARLEQUIN 2.000 with 1000 permutations.

Results

Genetic variation

A 1115-bp region of the *COI* mitochondrial gene was sequenced from each end, with a region of 716-bp overlap, for 199 individuals (Table 1). Mutations were all transitions or transversions, resulting in a total of 89 synonymous changes and six amino acid changes. Forty-four *COI* haplotypes were detected in 115 individual *An. arabiensis* from 12 populations (Fig. 1). The number of haplotypes per population ranged from two to 10 (Table 1). The sequences of *An. arabiensis* included 27 singleton variable sites and 31 parsimony informative sites. Haplotype (27) was widespread across Nigeria, found in nine of 12 populations, and three haplotypes (15, 19 and 27) were common to both *An. arabiensis* and *An. gambiae* (Table 1). Although *An. gambiae* is found throughout Nigeria, for this study we analysed samples from six central and southern populations (Fig. 1), and identified 58 S form and 26 M form. Forty-five *COI* haplotypes were detected in 84 *An. gambiae*, and the number per population ranged from two to 11. The sequences of *An. gambiae* included 27 singleton variable sites and 34 parsimony informative sites. One haplotype (55) was found in all six populations in central and southern Nigeria (Table 1). High haplotype diversity relative to nucleotide diversity was the common pattern detected (Table 2), suggesting a population bottleneck followed by a rapid expansion and build-up of mutations (Avise, 2000). Only *An. gambiae* M form, which had a moderate haplotype diversity (0.520), does not fit this pattern. There was also a general reduction of π values in *An. arabiensis* from the northern Sudan savannah to the southern rainforest localities, with a slight increase in Ogbomoso, possibly because it is a transition zone (Fig. 1). The haplotype diversities followed a similar north–south trend (data not shown).

Phylogeography

We inferred phylogenetic relationships using the NJ analysis for each of *An. gambiae* and *An. arabiensis*. We detected two subdivisions in *An. arabiensis*: one included all southern localities plus all but two central localities; the other included all northern plus two central localities (Lafia and Lokojo; Fig. 3). The two *An. gambiae* forms were clearly separated except that samples of M form from Calabar in southern Nigeria clustered with *An. gambiae* S form. In both NJ analyses the sample size per locality is typically $n = 10$, which may weaken our conclusions of subdivision/division. However, it should be mentioned that the *An. arabiensis* subdivision is fairly well supported by the Bayesian clustering analysis of microsatellite data, and only one sample of *An. gambiae* M (from Calabar) clustered with *An. gambiae* S in the analysis shown in Fig. 3. Both maximum parsimony and minimum evolution analyses of species relationships were unresolved, with low bootstrap support values, probably reflecting a paucity of phylogenetic signal (data not shown).

The networks constructed in TCS 1.21 (Clement *et al.*, 2000) were parsimoniously connected at the 95% probability level with less than 14 mutational steps between haplotypes for *An. arabiensis* and for *An. gambiae* M and S forms. *Anopheles arabiensis* and *An. gambiae* combined may also be connected into a single parsimonious network (Fig. 4). Many haplotypes in both species differed by only one or two mutational steps, suggesting a demographic expansion (Slatkin & Hudson, 1991; Fu, 1997). Haplotype 27, the most interior and therefore theoretically the most ancestral (Crandall & Templeton, 1993), included 20% of individuals sequenced, and contained representatives from *An. arabiensis* and both the M and S forms of *An. gambiae*

Table 2. Summary statistics, diversity measures, and neutrality tests for Nigerian *Anopheles arabiensis* and *Anopheles gambiae.*

	N/U	*h*(SD)	π(SD)	D_T	*D**	*F**	*F*s	R^2	*S*
Anopheles arabiensis									
All	115/42	0.883 (0.024)	0.003 (0.0002)	−2.298†	−3.873†	−3.854†	−36.179‡	0.027‡	1.000
North	30/26	0.991 (0.011)	0.005 (0.0004)	−2.032*	−2.310	−2.619*	−20.843‡	0.0491‡	1.000
Admixed	54/6	0.672 (0.070)	0.001 (0.0002)	−1.043	−2.201	−2.163	−0.903	0.103	0.869
South	31/11	0.828 (0.053)	0.002 (0.0004)	−1.647	−1.490	−1.810	−2.828	0.0698	0.979
Non-North	44/17	0.795 (0.035)	0.002 (0.0002)	−1.841*	−2.979*	−3.044*	−5.898	0.0413*	0.999
Anopheles gambiae									
All	84/45	0.893 (0.031)	0.005 (0.0003)	−1.939*	−2.816*	−2.954*	−35.425‡	0.0382*	1.000
M form	26/7	0.520 (0.115)	0.002 (0.0007)	−1.563	−2.007	−2.189	−0.237	0.0792	0.745
S form	58/40	0.970 (0.014)	0.005 (0.0003)	−1.967*	−2.482*	−2.735*	−33.457‡	0.0412*	1.000

*$P < 0.05$; †$P < 0.02$; ‡$P < 0.001$.

n, number of individuals; U, number of unique haplotypes; *h*, haplotype diversity; π, nucleotide diversity; SD, standard deviation; D_T = Tajima's D; *D** = Fu & Li's *D** test; *F** = Fu & Li's *F** test; F_S = Fu's F_S statistic; R^2 = Ramos-Onsins & Rozas' R^2 test; *S* = Strobeck's *S* statistic. See text for definitions of north, admixed, south and non-north groupings of *An. arabiensis*.

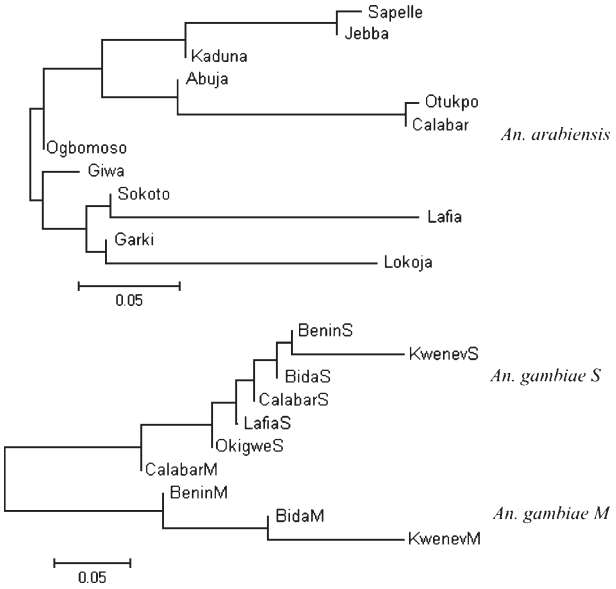

Fig. 3. Neighbour-joining tree for populations of *Anopheles arabiensis*, *Anopheles gambiae* M form, and *An. gambiae* S form based on F_{ST} estimates of pairwise genetic differences. Branch lengths are proportional to F_{ST} estimates.

(Fig. 4, Table 1). Most haplotypes were tip alleles, which are considered by Crandall & Templeton (1993) and Castelloe & Templeton (1994) to be both more recently derived and geographically restricted. For *An. gambiae* analysed alone, haplotype 55 is the theoretical ancestral haplotype, as it is both interior and widespread (i.e. found in all M form localities and four of six S form localities; Table 1).

Population genetic structure

In general, F_{ST} estimates suggested somewhat greater genetic distance between *An. arabiensis* and *An. gambiae* (−0.09489 to 0.87673), with 75% (90/120) significant pairwise comparisons, vs. a range of −0.24091 to 0.63612 between *An. gambiae* M and S forms, with 67% (16/24) of the F_{ST} values significant. Between the two subdivisions in *An. arabiensis* (Fig. 3), F_{ST} estimates ranged from −0.11897 to 0.43581, and 66% (23/35) of these values were significant (see supplementary material). These ranges of F_{ST} estimates are much greater than those found in comparable mtDNA studies of *An. arabiensis* and *An. gambiae* (e.g. Besansky *et al.*, 1997), and higher than expected between *An. gambiae* M and S, considering that one of the hallmarks of these two forms is their low degree of genetic differentiation (della Torre *et al.*, 2005).

Under the IBD model, the greatest genetic distance would be found between the localities separated by the greatest geographic distance. In *An. arabiensis*, the maximum geographic distance (957 km) is between Sokoto in the Sudan savannah and Calabar in the southern rainforest (Fig. 1), and the genetic distance, although relatively small (0.05655), is statistically significant ($P < 0.05$, SM 1). For *An. gambiae* M, the greatest genetic distance would be expected between Bida and Calabar (552 km), but this estimate is 0.04280 ($P > 0.05$); in fact, none of the pairwise F_{ST} estimates among *An. gambiae* M from the four localities (Benin, Bida, Calabar, Kwenev; Fig. 1) is significant (see supplementary material). The estimates within *An. gambiae* S appear to violate the IBD model. There are only two significant F_{ST} estimates: between Kwenev and Calabar (338 km), and between Kwenev and Lafia (194 km). Localities at greater geographic distances (e.g. Calabar and Bida, 552 km; Benin and Kwenev, 499 km) were not significant ($P > 0.05$). Furthermore,

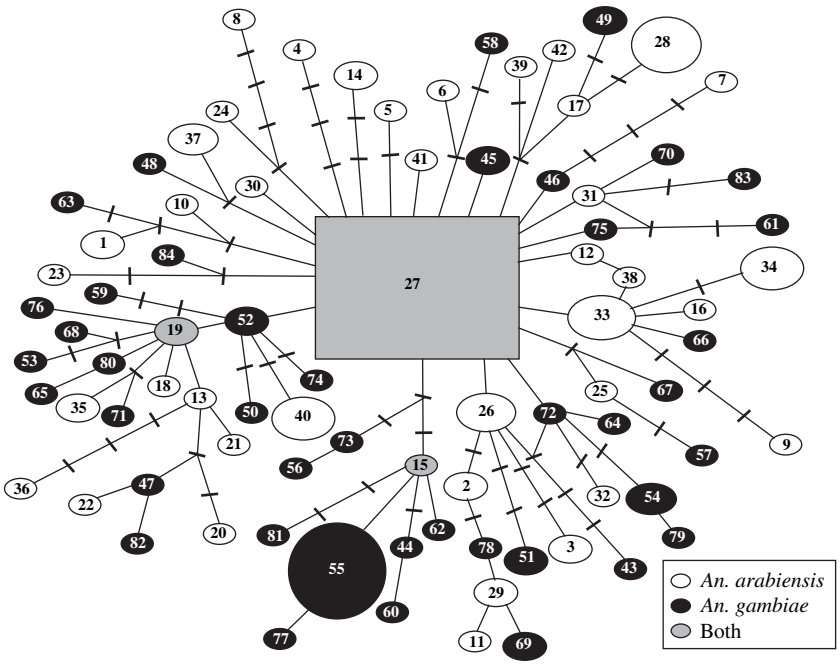

Fig. 4. Minimum spanning network of *Anopheles arabiensis* and *Anopheles gambiae*. The inferred ancestral haplotype is rectangular; dashes represent presumed haplotypes.

Mantel tests indicated no significant correlation between pairwise F_{ST} and distance (*An. arabiensis*, $R^2 = 0.0346$; *An. gambiae*, $R^2 = 0.0207$; S form alone, $R^2 = 0.0173$; M form alone, $R^2 = 0.1125$). The scatterplot of *An. arabiensis* demonstrates that many of the greatest distances represent very low genetic differentiation (Fig. 5). These data together suggest that distance alone does not explain either the subdivision in *An. arabiensis*, or the deeper division between the two forms of *An. gambiae* (Fig. 3), and the pattern that has emerged is more consistent with a recent demographic expansion. By contrast, the relatively small sample sizes per locality may have limited the ability to detect IBD.

Historical demography

Tajima's (1989) D_T and Fu & Li's (1993) F^* and D^* neutrality tests found that all *An. arabiensis*, northern *An. arabiensis*, non-northern *An. arabiensis*, all *An. gambiae* and *An. gambiae* S form have significant negative values (except the D value for northern *An. arabiensis*, which is negative but insignificant) (Table 2), indicating an excess of recently derived haplotypes. Our results allow rejection of the neutral model in these regions of Nigeria, resulting from either purifying selection or a recent population expansion.

Fu's (1997) F_S and Ramos-Onsins & Rozas' (2002) R^2-tests, considered to be the most powerful for the detection of population growth (Michel *et al.*, 2005), were both highly significant for all *An. arabiensis* and northern *An. arabiensis*. F_S test results were also very significant for all *An. gambiae* and S form alone, but R^2 results were somewhat less so (Table 2). Significant negative F_S values occur when an excess of rare haplotypes is present, and suggest either population expansion or genetic hitchhiking (Fu, 1997). None of the groups tested was significant for Strobeck's (1987) S-test.

Mismatch distribution of all *An. arabiensis* individuals analysed together, and nearly all of the other groups analysed (admixed, south, non-north), each resulted in a multimodal distribution consistent with equilibrium. However, analysis of the northern *An. arabiensis* and of the M and S forms of *An. gambiae* suggested a unimodal pattern following a Poisson distribution (Fig. 6A–C), consistent with an historical expansion. Although none of the P-values for the raggedness indices were significant, the raggedness statistics for each of northern *An. arabiensis* and *An. gambiae* S were quite small (Fig. 6), and also suggestive of a population expansion.

Time since expansion was estimated using the equation $t = \tau/2\mu$, in which μ is the mutation rate per site per generation (Slatkin & Hudson, 1991). The mutation rate of Drosophila (10^{-8}/site/year; Powell *et al.*, 1986) and 10 generations/year

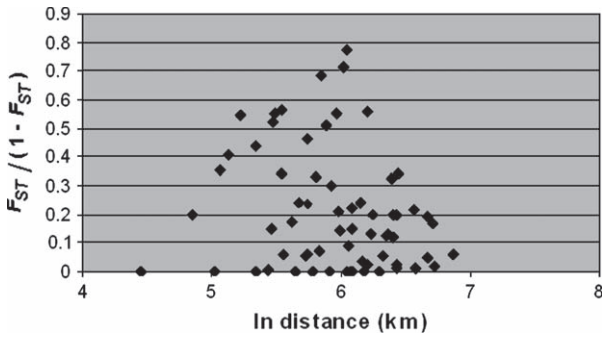

Fig. 5. Scatterplot of pairwise *Anopheles arabiensis* F_{ST} values against geographic distance separating pairs of localities in Nigeria.

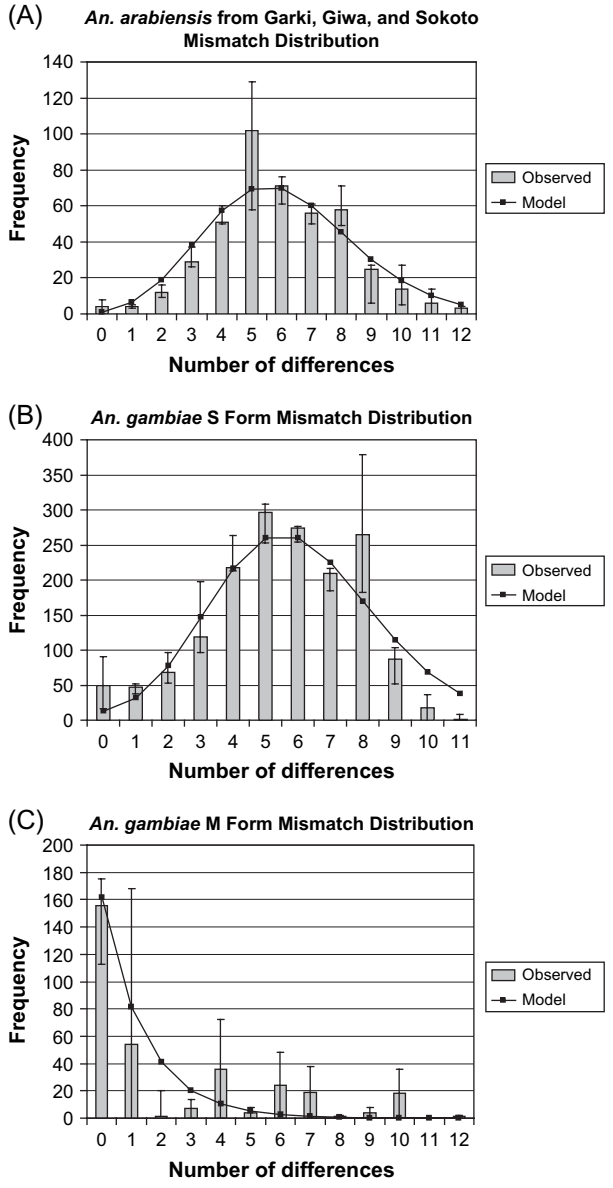

Fig. 6. Frequency for *COI* sequences showing distribution of segregating sites (=number of differences). The bar graph is the observed mismatch distribution for each taxon, and the line graph is the expected distribution of a recently expanded population. The error bars are 95% confidence interval estimates of the observed values. (A) Northern cluster of *Anopheles arabiensis* including samples from Garki, Giwa and Sokoto (raggedness statistic [r] = 0.0313; $P = 0.04$). (B) *Anopheles gambiae* S form ($r = 0.02316$; $P = 0.08$). (C) *An. gambiae* M form ($r = 0.1552$; $P = 0.55$).

(Walton *et al.*, 2000) were used. Estimates of τ came from the raggedness calculation. These expansions started approximately 27 085 years ago (95% confidence interval [CI] 16 157–31 897 years) for northern *An. arabiensis*; 27 144 years ago (95% CI 17 471–32 242 years) for *An. gambiae* S, and 32 372 years ago (95% CI 7480–54 233 years) for *An. gambiae* M. All of these expansions occurred during the late Pleistocene.

Discussion

Anopheles arabiensis

Our study detected some variation across Nigeria, and about two-thirds of the pairwise population samples were significantly differentiated. The mtDNA diversity values we detected are similar to those found using *ND5* sequences in Senegal, Kenya, Malawi and Ethiopia (Besansky *et al.*, 1997; Donnelly *et al.*, 2001; Temu & Yan, 2005). Both the mtDNA NJ analysis we performed and the Bayesian clustering analysis (STRUCTURE) of the microsatellite loci (from Onyabe & Conn, 2001c) suggest that there are two geographic subdivisions in *An. arabiensis* within Nigeria, one predominantly northern and the other predominantly southern. The Bayesian clusters include (a) northern localities Sokoto, Garki and Giwa, and (b) southern localities Ogbomoso, Sapele and Otukpo. Samples from central Nigeria (Kaduna, Jebba and Abuja) were not strongly assigned to either cluster. In the mtDNA clusters (Fig. 3), these three central localities are most closely aligned with the southern ones. Although the geographic boundaries are not identical, their detection using two independent markers strengthens the conclusions of a geographical subdivision within Nigeria.

By contrast with the study by Onyabe & Conn (2001c), who determined that all 10 microsatellite loci and the 12 localities each contributed to the finding of highly significant IBD in *An. arabiensis*, the mtDNA sequences in our study provided no convincing pattern of IBD; in fact, variation in levels of differentiation was unrelated to geographic distance and does not explain the subdivision we detected in *An. arabiensis*. These findings in *An. arabiensis* are consistent with a regional study, using microsatellite loci, which found no IBD among geographically diverse samples from the northwestern highlands and the Rift Valley in Ethiopia and Eritrea (Nyanjom *et al.*, 2003).

In the microsatellite study (Onyabe & Conn, 2001c), the genetic structure of *An. arabiensis* was characterized by reductions in heterozygosity and number of alleles per locus from savannah (north) to forest (south). A similar finding of a lower diversity of inverted arrangements in Nigerian forest samples compared with savannah samples was reported when three chromosome inversions were examined (Coluzzi *et al.*, 1979). These congruent results led Onyabe & Conn (2001c) to hypothesize that *An. arabiensis* extended its range southward, from savannah into forest habitat, with the new forest populations founded by individuals that originated in savannah zones. The mitochondrial data show the same trend, although the current study lacks adequate samples from the forest zone to rigorously test the founder effect. Given that the vegetation of tropical West Africa south of the Sahara varies mostly across latitudes (White, 1983), that sub-Sahelian Africa was in a severe drought (21% decline in precipitation) during 1900−1994 (Hulme, 1996), and that *An. arabiensis* is more tolerant of dry environments, we believe our data reinforce the suggestion made by Lindsay & Martens (1998) that *An. arabiensis* may be in the process of expanding its range from savannah southward into areas affected by drought, including forest, possibly replacing *An. gambiae s.s.* as the major vector. It would be of interest to know whether this is occurring in other West African countries and whether it might constitute a regional phenomenon.

Our study provides evidence that *An. arabiensis*, particularly in northern Nigeria, is not in equilibrium, most likely as a result of a population expansion, and corroborates findings from a study using microsatellite data in Nigeria (Onyabe & Conn, 2001c). Because this pattern was detected in two independent datasets, it is probably the result of a past demographic change, rather than selection (Donnelly *et al.*, 2001). The Nigerian expansion in *An. arabiensis* is more strongly supported than in reports using mtDNA conducted in other regions of Africa. Besansky *et al.* (1997), who examined samples from Senegal, South Africa and Kenya, found neither Tajima's *D* nor Fu & Li's *D* to be significant, although both values were negative for Senegal, which, like Nigeria, is in West Africa. Furthermore, the Senegal mismatch distribution was unimodal. Interestingly, in a study of *An. arabiensis* from Malawi and Ethiopia and *An. gambiae* from eastern and western Kenya, Donnelly *et al.* (2001) detected less evidence for an expansion using microsatellite data in *An. arabiensis* compared with *An. gambiae*, and no support for an expansion using mtDNA data in *An. arabiensis*, whereas there was considerable evidence in *An. gambiae*. Possible explanations for the weaker signal in *An. arabiensis* included an earlier expansion, a smaller change between pre- and post-expansion populations and/o) smaller current population size (Donnelly *et al.*, 2001). Perhaps the difference in signal strength of expansions detected with mtDNA in *An. arabiensis* between western (Nigeria) and eastern (Malawi, Ethiopia) Africa indicates distinctive demographic histories.

The disequilibrium may have resulted in an overestimation of rates of gene flow between the *An. arabiensis* subdivisions ($N_m > 1$ for 85% of pairwise estimates; data not shown). The potential exaggeration of gene flow estimates as a result of population or range expansion, noted in *Anopheles dirus* (Walton *et al.*, 2000) has been observed in several anopheline disease vectors (e.g. *An. gambiae* and *An. arabiensis* [Donnelly *et al.*, 2001]; *Anopheles funestus* [Michel *et al.*, 2005]; *Anopheles darlingi* [Mirabello & Conn, 2006]), and in other mosquito genera (e.g. *Culex tarsalis* [Venkatesan *et al.*, 2007]). Donnelly *et al.* (2001) note that historical expansions in both *An. arabiensis* and *An. gambiae* may be linked with the rise of agriculture in sub-Saharan Africa (4000–10 000 years ago), providing significant new breeding sites and sources of bloodmeals from both humans and domestic animals (Coluzzi, 1982). Our estimate for expansion of *An. arabiensis* from northern Nigeria is within the same time-frame, albeit slightly earlier (approximately 27 085 years ago). African anopheline expansions could also have been influenced by the significant continental human expansions that are estimated to have begun between 70 000 and 50 000 years ago (Rogers & Harpending, 1992; Rogers, 1995;

Lahr & Foley, 1998). Recent expansions in a variety of disease vectors, including triatomines (Monteiro *et al.*, 2003; reviewed in Conn & Mirabello, 2007), and black flies (Pramual *et al.*, 2005) appear to be tied to human population expansions, Pleistocene climate changes, or possibly interactions between the two, depending on local circumstances, population size, ecology and host range.

Anopheles gambiae

In Nigeria, the inference of expansion was well supported in *An. gambiae* S by the star-shaped genealogy, five of seven neutrality tests and the diversity pattern, but extremely weakly in the M form (only by the unimodal distribution, which was not significant). Despite this poor support, one interpretation of the lack of IBD detected in *An. gambiae* M is that it is consistent with a relatively recent demographic expansion (Michel *et al.*, 2005). Another interpretation is that extensive genetic exchange occurs within each of the two forms in Nigeria. The difference in evidence of equilibrium for the M and S forms, from several localities where they are sympatric (Bida and Kwenev in savannah, Benin and Calabar in rainforest; Fig. 1), and where the larvae were collected in the same or several adjacent pools within a 1-km radius, is puzzling. It is hard to imagine conditions in both savannah and rainforest that would have resulted in such a difference between forms, even allowing for the fact that during the late Pleistocene, the area of forest was greatly reduced, and open vegetation covered much of equatorial West Africa (Dupont *et al.*, 2000). However, our sampling strategy (collecting larvae from several pools within a 1-km radius) may have inadvertently resulted in the collection of multiple disparate populations, leading to a false conclusion of population expansion. This seems unlikely considering previously published results of levels of gene flow in *An. gambiae* (Besansky *et al.*, 1997). There was no evidence for an expansion in *An. gambiae* in Nigeria based on microsatellite markers (Onyabe & Conn, 2001b); therefore the expansion detected by mtDNA data in *An. gambiae* S may be the result of selection (Donnelly *et al.*, 2001). Another possibility, however, is that the expansion was not detected in the previous *An. gambiae* study using microsatellite markers because of rapid mutation rates (Cornuet & Luikart, 1996).

The results of IBD tests based on $F_{ST}/(1-F_{ST})$, using the microsatellite dataset of Onyabe & Conn (2001b), were equivocal. The test conducted using F_{ST} estimates was highly significant, but not when using R_{ST}-based estimates. Furthermore, when the rainforest localities of Benin and Lagos were removed, and when three loci, *AG2H26*, *AG2H637* and *AG2H79*, located within chromosomal inversions *2Rb*, *2La* and *2Ra*, respectively, were removed, the F_{ST}-based estimates were no longer significant. The conclusion of this study was that selection of genes located within inversions on chromosome II appears to counter the homogenizing effect of gene flow. Even when the IBD tests were rerun on the microsatellite data after the M and S forms of *An. gambiae* had been identified, genetic distance was not significantly correlated with geographical distance in either form (Onyabe *et al.*, 2003). The insignificant results of the IBD tests using mtDNA data may also be interpreted in support of selection as a substantial factor influencing the genetic differentiation of *An. gambiae* in Nigeria.

The three haplotypes common to *An. gambiae* and *An. arabiensis* in Nigeria lend support to earlier studies that demonstrated intertwined sequences in mtDNA trees (Besansky *et al.*, 1997; Caccone *et al.*, 1998), recent evidence of semipermeable boundaries between these closely related species (Besansky *et al.*, 2003), and/or the retention of ancestral haplotypes in both species (Donnelly *et al.*, 2004). Further, when analysed in an NJ tree together, *An. arabiensis* and *An. gambiae* did not cluster into two species. Instead, the M molecular form of *An. gambiae* clustered separately, and the S form of *An. gambiae* clustered with *An. arabiensis* (data not shown). This could be indicative of shared ancestral polymorphisms or contemporary gene flow between the S form and *An. arabiensis*, with the M form diverging under selective pressure (Wondji *et al.*, 2002). If introgressive hybridization is occurring, these findings are congruent with the mosaic model of speciation in which genes are exchanged at some loci, such as *COI*, but not at others, such as microsatellites inside inversions (Krzywinski & Besansky, 2003), and also with the model of incipient speciation as proposed by Lehmann *et al.* (2003).

In conclusion, we detected evidence from two independent datasets for population subdivision in *An. arabiensis* and corroborated earlier findings of a north–south range expansion, and a corresponding reduction in genetic diversity in the southern rainforest. We suggest this southward expansion may be a regional trend in West African *An. arabiensis*. The relatively strong mtDNA support for expansion in Nigerian *An. arabiensis* may indicate different demographic histories for this species in West Africa vs. East Africa. Distance does not appear to be a significant factor influencing mtDNA genetic structure in Nigerian *An. arabiensis*, by contrast with findings from microsatellite loci. Evidence for selection shaping genetic structure in *An. gambiae* was particularly noteworthy in the S form.

Supplementary material

The following material is available at www.Blackwell-Synergy.com under the DOI reference doi:10.1111/j.1365-2915.2007.00703.x

Table S1. Pairwise F_{ST} values between each locality for *Anopheles arabiensis* and *An. gambiae* M and S.

Acknowledgements

We are grateful to Dr C. Vajime, of Ahmadu Belo University, Zaria, Nigeria for logistical support in acquiring field samples. We are very grateful to two astute reviewers whose comments improved this manuscript significantly. This study was funded by National Institutes of Health AI R01 54139 to JEC. We thank Lisa Mirabello and Kathryn Powers of the Conn Lab (Griffin Laboratory, Albany, NY, U.S.A.), for their interest and input. All sequencing was carried out at the Wadsworth Center Molecular Genetics Core Facility (Albany, NY, U.S.A.).

References

Ashburner, M., Hoy, M.A. & Peloquin, J.J. (1998) Prospects for the genetic transformation of arthropods. *Insect Molecular Biology*, **7**, 201–213.

Avise, J.C. (2000) *Phylogeography. The History and Formation of Species*, pp. 37−108. Cambridge University Press, Harvard, MA.

Benedict, M.Q. & Robinson, A.S. (2003) The first release of transgenic mosquitoes: an argument for the sterile insect technique. *Trends in Parasitology*, **19**, 349–355.

Besansky, N.J., Krzywinski, J., Lehmann, T. *et al*. (2003) 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, N.J., Lehmann, T., Fahey, G.T., Fontenille, D., Braack, L.E.O., Hawley, W.A. & Collins, F. (1997) Patterns of mitochondrial variation within and between African malaria vectors, *Anopheles gambiae* and *An. arabiensis*, suggest extensive gene flow. *Genetics*, **147**, 1817–1828.

Bryan, J.H., Di Deco, M.A., Petrarca, V. & Coluzzi, M. (1982) Inversion polymorphism and incipient speciation in *Anopheles gambiae s.s*. in the Gambia, West Africa. *Genetica*, **59**, 167–176.

Caccone, A. Min, G.S. & Powell, J.R. (1998) Multiple origins of cytologically identical chromosome inversion in the *Anopheles gambiae* complex. *Genetics*, **150**, 807–814.

Castelloe, J. & Templeton, A.R. (1994) Root probabilities for intraspecific gene trees under neutral coalescent theory. *Molecular Phylogenetics and Evolution*, **3**, 102–113.

Charlwood, J.D., Billingsley, P.F. & Hoc, T.Q. (1995) Mosquito-mediated attraction of female European but not African mosquitoes to hosts. *Annals of Tropical Medicine and Parasitology*, **89**, 327–329.

Clement, M., Posada, D. & Crandall, K.A. (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1660.

Coetzee, M., Craig, M. & le Sueur, D. (2000) Distribution of African malaria mosquitoes belonging to the *Anopheles gambiae* complex. *Parasitology Today*, **16**, 74–77.

Coetzee, M. & Fontenille, D. (2004) Advances in the study of *Anopheles funestus*, a major vector of malaria in Africa. *Insect Biochemistry and Molecular Biology*, **4**, 599–605.

Collins, F.H., Kamau, L., Ranson, H.A. & Vulule, J.M. (2000) Molecular entomology and prospects for malaria control. *Bulletin of the World Health Organization*, **78**, 1412–1423.

Collins, F.H., Mendez, M.A., Rasmussen, M.O., Mehaffey, P.C. & Besansky, N.J. (1987) A ribosomal RNA gene probe differentiates member species of the *Anopheles gambiae* complex. *American Journal of Tropical Medicine and Hygiene*, **37**, 37–41.

Coluzzi, M. (1982) Spatial distribution of chromosomal inversions and speciation in the Anopheline mosquitoes. *Progress in Clinical and Biological Research*, **96**, 143–153.

Coluzzi, M., Petrarca, V. & Di Deco, M. (1985) Chromosomal inversion intergradation and incipient speciation in *Anopheles gambiae*. *Bolletino di Zoologia*, **52**, 45–63.

Coluzzi, M., Sabatini, A., Petrarca, V. & Di Deco, M.A. (1979) Chromosomal differentiation and adaptation to human environments in the *Anopheles gambiae* complex. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **73**, 483–497.

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

Conn, J.E. & Mirabello, L. (2007) The biogeography and population genetics of neotropical vector species. *Heredity*, **99**, 245–256.

Cornuet, J.M. & Luikart, G. (1996) Description and power analysis of two tests for detecting population bottlenecks from allele frequency data. *Genetics*, **144**, 2001–2014.

Crandall, K.A. & Templeton, A.R. (1993) Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics*, **134**, 959–969.

Davies, H. (1977) *Tsetse Flies in Nigeria: a Handbook for Junior Control Staff*, pp. 7−42. Oxford University Press, Ibadan.

Donnelly, M.J., Licht, M.C. & Lehmann, T. (2001) Evidence for recent population expansion in the evolutionary history of the malaria vectors *Anopheles arabiensis* and *Anopheles gambiae*. *Molecular Biology and Evolution*, **18**, 1353–1364.

Donnelly, M.J., Pinto, J., Girod, R., Besansky, N.J. & 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**, 61–88.

Donnelly, M.J. & Townson, H. (2000) Evidence for extensive genetic differentiation among populations of the malaria vector *Anopheles arabiensis* in Eastern Africa. *Insect Molecular Biology*, **9**, 357–367.

Dupont, L.M., Jahns, S., Marret, F. & Ning, S. (2000) Vegetation change in equatorial West Africa: time-slices for the last 150 ka. *Palaeogeography, Palaeoclimatology, Palaeoecology*, **155**, 95–122.

Favia, G., Lanfrancotti, A., Spanos, L., Siden-Kiamos, I. & Louis, C. (2001) Molecular characterization of ribosomal DNA polymorphisms discriminating among chromosomal forms of *Anopheles gambiae s.s*. *Insect Molecular Biology*, **10**, 19–23.

Favia, G., della Torre, A., Bagayoko, M., Lanfrancotti, A., Sagnon, N. & Touré, Y.T. (1997) Molecular identification of sympatric chromosomal forms of *Anopheles gambiae* and further evidence of their reproductive isolation. *Insect Molecular Biology*, **6**, 377–383.

Felsenstein, J. (1985) Confidence limits of phylogeny: an approach using bootstrap. *Evolution*, **39**, 783–791.

Fu, Y.X. (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, **147**, 915–925.

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

Gillies, M.T. & Coetzee, M. (1987) *A Supplement to the Anophelinae of Africa South of the Sahara*. Publications of the South African Institute for Medical Research No. 55, pp. 1−143. SAIMR, Johannesburg.

Harpending, H.C. (1994) Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology*, **66**, 591–600.

Hulme, M. (1996) *Climate Change and Southern Africa: An Exploration of some Potential Impacts and Implications in the SADC Region*, pp. 104. Climatic Research Unit, University of East Anglia and World Wide Fund International, .

Hunt, R.H., Coetzee, M. & Fettene, M. (1998) The *Anopheles gambiae* complex: a new species from Ethiopia. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **92**, 231–235.

Kamau, L., Mukabana, W.R., Hawley, W.A., Lehmann, T., Irungu, L.W., Orago, A.A.S. & Collins, F.H. (1999) Analysis of genetic variability in *Anopheles arabiensis* and *Anopheles gambiae* using microsatellite loci. *Insect Molecular Biology*, **8**, 287–297.

Krzywinski, J. & Besansky, N.J. (2003) Molecular systematics of *Anopheles*: from subgenera to subpopulations. *Annual Review of Entomology*, **48**, 111–139.

Kumar, S., Tamura, K., Jakobsen, B. & Nei, M. (2001) MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics*, **17**, 1244–1245.

Lahr, M.M. & Foley, R.A. (1998) Towards a theory of modern human origins: geography, demography and diversity in recent human evolution. *Yearbook of Physical Anthropology*, **41**, 137–176.

Lehmann, T., Licht, M., Elissa, N. *et al*. (2003) Population structure of *Anopheles gambiae* in Africa. *Journal of Heredity*, **94**, 133–147.

Lindsay, S.W. & Martens, W.J. (1998) Malaria in the African highlands: past, present and future. *Bulletin of the World Health Organization*, **76**, 33–45.

Lunt, D.H., Zhang, D.X., Szymura, J.M. & Hewitt, G.M. (1996) The insect cytochrome oxidase I gene: evolutionary patterns and conserved primers for phylogenetic studies. *Insect Molecular Biology*, **5**, 153–165.

Maddison, W.P. & Maddison, D.R. (1992) *MacClade*. Sinauer Associates, Inc., Sunderland, MA.

Mantel, N. (1967) The detection of disease clustering and a generalized regression approach. *Cancer Research*, **27**, 209–220.

Masendu, H.T., Hunt, R.H., Govere, J., Brooke, B.D., Awolola, T.S. & Coetzee, M. (2004) The sympatric occurrence of two molecular forms of the malaria vector *Anopheles gambiae* Giles *sensu stricto* in Kanyemba, in the Zambezi Valley, Zimbabwe. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **98**, 393–396.

Michel, A.P., Ingrasci, M.J., Schemerhorn, B.J. *et al*. (2005) Rangewide population genetic structure of the African malaria vector *Anopheles funestus*. *Molecular Ecology*, **14**, 4235–4248.

Mirabello, L. & Conn, J.E. (2006) Molecular population genetics of the malaria vector *Anopheles darlingi* in Central and South America. *Heredity*, **96**, 311–321.

Monteiro, F.A., Barrett, T.V., Fitzpatrick, S., Cordon-Rosales, C., Feliciangeli, D. & Beard, C.B. (2003) Molecular phylogeography of the Amazonian Chagas disease vectors *Rhodnius prolixus* and *R. robustus*. *Molecular Ecology*, **12**, 997–1006.

Moreira, L.A., Wang, J., Collins, F.H. & Jacobs-Lorena, M. (2004) Fitness of anopheline mosquitoes expressing transgenes that inhibit *Plasmodium* development. *Genetics*, **166**, 1337–1341.

Morlais, I., Girod, R., Hunt, R., Simard, F. & Fontenille, D. (2005) Population structure of *Anopheles arabiensis* on La Réunion Island, Indian Ocean. *American Journal of Tropical Medicine and Hygiene*, **73**, 1077–1082.

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

Nyanjom, S.R., Chen, H., Gebre-Michael, T. *et al*. (2003) Population genetic structure of *Anopheles arabiensis* mosquitoes in Ethiopia and Eritrea. *Journal of Heredity*, **94**, 457–463.

Onyabe, D.Y. & Conn, J.E. (2001a) The distribution of two major malaria vectors, *Anopheles gambiae* and *Anopheles arabiensis*, in Nigeria. *Memorias do Instituto Oswaldo Cruz*, **96**, 1081–1084.

Onyabe, D.Y. & Conn, J.E. (2001b) Genetic differentiation of the malaria vector *Anopheles gambiae* across Nigeria suggests that selection limits gene flow. *Heredity*, **87**, 647–658.

Onyabe, D.Y. & Conn, J.E. (2001c) Population genetic structure of the malaria mosquito *Anopheles arabiensis* across Nigeria suggests range expansion. *Molecular Ecology*, **10**, 2577–2591.

Onyabe, D.Y., Vajime, C.G., Nock, I.H., Ndams, I.S., Akpa, A.U., Alaribe, A.A. & Conn, J.E. (2003) The distribution of M and S molecular forms of *Anopheles gambiae* in Nigeria. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **97**, 1–4.

Paskewitz, S.M. & Collins, F.H. (1990) Use of the polymerase chain reaction to identify mosquito species of the *Anopheles gambiae* complex. *Medical and Veterinary Entomology*, **4**, 367–373.

Posada, D. & Crandall, K.A. (2001) Intraspecific gene genealogies: trees grafting into networks. *Trends in Ecology and Evolution*, **16**, 37–45.

Powell, J.R., Caccone, A., Amato, G.D. & Yoon, C. (1986) Rate of nucleotide substitution in Drosophila mitochondrial DNA and nuclear DNA are similar. *Proceedings of the National Academy of Sciences USA*, **83**, 9090–9093.

Pramual, P., Kuvangkadilok, C., Baimai, V. & Walton, C. (2005) Phylogeography of the black fly *Simulium tani* (Diptera: Simuliidae) from Thailand as inferred from mtDNA sequences. *Molecular Ecology*, **14**, 3989–4001.

Pritchard, J.K., Stephens, M. & Donnelly, P. (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Ramos-Onsins, S.E. & Rozas, J. (2002) Statistical properties of new neutrality tests against population growth. *Molecular Biology and Evolution*, **19**, 2092–2100.

Rogers, A.R. (1995) Genetic evidence for a Pleistocene population explosion. *Evolution*, **49**, 608–615.

Rogers, A.R. & Harpending, H. (1992) Population growth makes waves in the distribution of pairwise genetic differences. *Molecular Biology and Evolution*, **9**, 552–569.

Rozas, J., Sanchez-DelBarrio, J.C., Messenguer, X. & Rozas, R. (2003) DNASP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.

Rubinoff, D. & Holland, B.S. (2005) Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. *Systematic Biology*, **54**, 952–961.

Schneider, S. & Excoffier, L. (1999) Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA. *Genetics*, **152**, 1079–1089.

Schnieder, S., Roessli, D. & Excoffier, L. (2000) *ARLEQUIN Version 2.000: a Software for Population Genetics Data Analysis*. Genetics and Biometry Laboratory, University of Geneva, Geneva.

Scott, J.A., Brogdon, W.G. & Collins, F.H. (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.

Slatkin, M. & Hudson, R.R. (1991) Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics*, **129**, 555–562.

Strobeck, C. (1987) Average number of nucleotide differences in a sample from a single subpopulation: a test for population subdivision. *Genetics*, **117**, 149–153.

Stump, A.D., Fitzpatrick, M.C., Lobo, N.F. *et al*. (2005b) Centromere-proximal differentiation and speciation in *Anopheles gambiae*. *Proceedings of the National Academy of Science USA*, **102**, 15930–15935.

Stump, A.D., Shoener, J.A., Costantini, C., Sagnon, N. & Besanasky, N. (2005a) Sex-linked differentiation between incipient species of *Anopheles gambiae*. *Genetics*, **169**, 1509–1519.

Swofford, D.L. (2002) *paup*. Phylogenetic Analysis Using Parsimony (*and Other Methods)*, Version 4.0b. Sinauer Associates, Inc., Sunderland, MA.

Tajima, F. (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437–460.

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

Taylor, C., Touré, Y.T., Carnahan, J. *et al*. (2001) Gene flow among populations of the malaria vector, *Anopheles gambiae*, in Mali, West Africa. *Genetics*, **157**, 743–750.

Taylor, C.E., Touré, Y.T., Coluzzi, M. & Petrarca, V. (1993) Effective population size and persistence of *Anopheles arabiensis* during the dry season in West Africa. *Medical and Veterinary Entomology*, **7**, 351–357.

Templeton, A.R., Crandall, K.A. & 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.

Temu, E.A. & Yan, G. (2005) Microsatellite and mitochondrial genetic differentiation of *Anopheles arabiensis* (Diptera: Culicidae) from

western Kenya, the Great Rift Valley, and coastal Kenya. *American Journal of Tropical Medicine and Hygiene*, **73**, 726–733.

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 Molecular Biology*, **10**, 9–18.

della Torre, A., Tu, Z. & Petrarca, V. (2005) On the distribution and genetic differentiation of *Anopheles gambiae s.s.* molecular forms. *Insect Biochemistry and Molecular Biology*, **35**, 755–769.

Touré, Y.T., Petrarca, V., Traore, S.F. *et al.* (1998) The distribution and inversion polymorphism of chromosomally recognized taxa of the *Anopheles gambiae* complex in Mali, West Africa. *Parassitologia*, **40**, 477–511.

Uthicke, S. & Benzie, J.A.H. (2003) Gene flow and population history in high dispersal marine invertebrates: mitochondrial DNA analysis of *Holothuria nobilis* (Echinodermata: Holothuroidea) populations from the Indo-Pacific. *Molecular Ecology*, **12**, 2635–2648.

Venkatesan, M., Westbrook, C.J., Hauer, M.C. & Rasgon, J.L. (2007) Evidence for a population expansion in the West Nile virus vector *Culex tarsalis*. *Molecular Biology & Evolution*, **24**, 1208–1218.

Walton, C., Handley, J.M., Willoughby, T.-L., Collins, F.H., Harbach, R.E., Baimai, V. & Butlin, R.K. (2000) Population structure and population history of *Anopheles dirus* mosquitoes in southeast Asia. *Molecular Biology and Evolution*, **17**, 962–974.

White, F. (1983) *The Vegetation of Africa*. UNESCO Natural Resources Research XX, pp. 356. UNESCO, Paris.

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 Molecular Biology*, **11**, 11–19.

Wondji, C., Simard, F., Petrarca, V., Etang, J., Santolamazza, F., della Torre, A. & Fontenille, D. (2005) Species and populations of the *Anopheles gambiae* complex in Cameroon with special emphasis on chromosomal and molecular forms of *Anopheles gambiae s.s.* *Journal of Medical Entomology*, **42**, 998–1005.

Wright, S. (1951) The genetical structure of populations. *Annals of Eugenics*, **15**, 323–354.

Accepted 1 October 2007

