Molecular Ecology (2005) **14**, 4235–4248 doi: 10.1111/j.1365-294X.2005.02754.x

Rangewide population genetic structure of the African malaria vector *Anopheles funestus*

A. P. MICHEL,* M. J. INGRASCI,* B. J. SCHEMERHORN,* M. KERN,* G. LE GOFF,† M. COETZEE,‡ N. ELISSA,§ D. FONTENILLE,¶ J. VULULE,** T. LEHMANN,†† N'F. SAGNON,‡‡ C. COSTANTINI¶¶ and N. J. BESANSKY*

**Center for Tropical Disease Research and Training, Department of Biological Sciences, University of Notre Dame, Notre Dame, IN, 46556, USA, †Service d'Entomologie Medicale, Groupe de Recherche sur le Paludisme, Institut Pasteur de Madagascar, BP 1274, Antananarivo 101, Madagascar, ‡Vector Control Reference Unit, National Institute for Communicable Diseases, Private Bag X4, Sandringham 2131, South Africa, and Medical Entomology, Division of Clinical Microbiology & Infectious Diseases, School of Pathology of the National Health Laboratory Service and the University of Witwatersrand, Johannesburg, South Africa, §Unité d'Entomologie Médicale, CIRMF, BP 769, Franceville, Gabon, ¶Laboratoire de Lutte Contre les Insectes Nuisibles, Institut de Recherche pour le Developpement (LIN-IRD), 911 Avenue Agropolis, BP 64501, 34394 Montpellier cedex 5, France, **Center for Vector Biology Control Research, Kenya Medical Research Institute, PO Box 1578, Kisumu, Kenya, ††Laboratory of Malaria and Vector Research, National Institutes of Health, NIAID, Rockville, MD 20852, USA, ‡‡Centre National de Recherche et de Formation sur le Paludisme, 01 BP 2208 Ouagadougou 01, Burkina Faso, ¶¶Institut de Recherche pour le Développement (IRD), 01 BP 182, Ouagadougou 01, Burkina Faso*

Abstract

***Anopheles funestus* is a primary vector of malaria in Africa south of the Sahara. We assessed its rangewide population genetic structure based on samples from 11 countries, using 10 physically mapped microsatellite loci, two per autosome arm and the X ($N = 548$), and 834 bp of the mitochondrial *ND5* gene ($N = 470$). On the basis of microsatellite allele frequencies, we found three subdivisions: eastern (coastal Tanzania, Malawi, Mozambique and Madagascar), western (Burkina Faso, Mali, Nigeria and western Kenya), and central (Gabon, coastal Angola). *A. funestus* from the southwest of Uganda had affinities to all three subdivisions. Mitochondrial DNA (mtDNA) corroborated this structure, although mtDNA gene trees showed less resolution. The eastern subdivision had significantly lower diversity, similar to the pattern found in the codistributed malaria vector *Anopheles gambiae*. This suggests that both species have responded to common geographic and/or climatic constraints. The western division showed signatures of population expansion encompassing Kenya west of the Rift Valley through Burkina Faso and Mali. This pattern also bears similarity to *A. gambiae*, and may reflect a common response to expanding human populations following the development of agriculture. Due to the presumed recent population expansion, the correlation between genetic and geographic distance was weak. Mitochondrial DNA revealed further cryptic subdivision in *A. funestus*, not detected in the nuclear genome. Mozambique and Madagascar samples contained two mtDNA lineages, designated clade I and clade II, that were separated by two fixed differences and an average of 2% divergence, which implies that they have evolved independently for ~1 million years. Clade I was found in all 11 locations, whereas clade II was sampled only on Madagascar and Mozambique. We suggest that the latter clade may represent mtDNA capture by *A. funestus*, resulting from historical gene flow either among previously isolated and divergent populations or with a related species.**

Keywords: Africa, *Anopheles funestus*, malaria vector, microsatellites, mitochondrial DNA, population genetics

Received 15 June 2005; revision accepted 24 August 2005

Correspondence: Nora J. Besansky, Fax: 574-631-3996; E-mail: nbesansk@nd.edu

Introduction

Malaria is a leading cause of child mortality in Africa, claiming at least 1 million lives each year (WHO 2003). High transmission intensity is enforced by two primary mosquito vectors that feed preferentially on humans: *Anopheles gambiae* and *Anopheles funestus*. These two species are widely codistributed across much of tropical Africa, and frequently occupy the same habitat in close proximity (Gillies & De Meillon 1968). They are also united by synanthropic tendencies. Thus, a marked preference for human blood is accompanied by domestic habits typical of pest species — breeding in the vicinity of human settlements and resting inside houses for considerable portions of their adult lives. The better studied of these vectors, *A. gambiae*, has a history shaped by human activity. The dependence of *A. gambiae* upon humans is thought to date from the beginnings of agriculture in Africa (Coluzzi *et al*. 2002). Human population expansion associated with agricultural development seems to have triggered parallel expansions of *Plasmodium* parasites and *Anopheles* ectoparasites (Hume *et al*. 2003; Joy *et al*. 2003). A consequence of recent population expansion by *A. gambiae* is the shallow population structure recorded for this species across Africa, with only one major subdivision into western and eastern lineages, probably resulting from the Rift Valley barrier as well as periods of extreme drought in East Africa (Lehmann *et al*. 2003). Given their common anthropophilic tendencies and species range, it is possible that *A. funestus* has been exposed to the same demographic opportunities and severe climatic pressures as *A. gambiae*. If so, it is expected that these species should share a similar pattern of subdivision into western and eastern lineages overlaid upon shallow rangewide population structure.

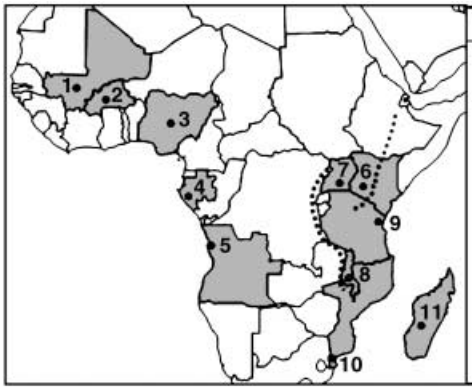
Before the recent application of molecular markers to the genetics of *A. funestus* populations, several observations led to a contrary expectation of extensive substructure. First, unlike *A. gambiae* whose temporary breeding sites are ubiquitous throughout the rainy season, *A. funestus* breeds in semipermanent or permanent sites whose distribution is patchy, such as seepages, rice fields and marshes. Second, in limited areas of East Africa where indoor residual application of insecticide was used to control malaria, this species was eliminated from many areas and re-infestation took up to 3 years after cessation of spraying (Gillies & De Meillon 1968). Finally, characterization of chromosomal inversions from different parts of Africa revealed not only significant frequency differences between locales, but also uncovered inverted arrangements that were apparently endemic to particular regions (Green & Hunt 1980; Dia *et al*. 2000). In contrast, microsatellite markers generally have not revealed genetically isolated populations, despite the patchy distribution of *A. funestus*. Similar to the picture that has emerged from indirect genetic studies of *A. gambiae*, microsatellite-based studies of *A. funestus* population genetic structure found only weak isolation by distance: no significant differentiation between populations within 200 km of each other, and only moderate differentiation ($F_{ST} < 0.07$) between populations separated by up to 3000 km (Braginets *et al*. 2003; Cohuet *et al*. 2004; Temu *et al*. 2004). The difference between indirect molecular estimates indicating shallow population structure and direct field-based observations suggesting otherwise is consistent with a system that has not yet reached migration–drift equilibrium, potentially because of a demographic history that included population expansion. A striking exception to the pattern of shallow genetic divergence, also mirrored in *A. gambiae*, is the large amount of differentiation ($F_{ST} = 0.2$) recorded across the Great Rift Valley in Kenya (Braginets *et al*. 2003). In addition, sequence differences in another type of molecular marker, the ITS2 and D3 regions of ribosomal DNA, suggested the existence of at least one main division between populations of *A. funestus* from West-Central and East Africa (Garros *et al*. 2004).

An understanding of the patterns and processes responsible for the distribution of variation in anopheline mosquito vectors is relevant to vector-based malaria control programmes (Coetzee & Fontenille 2004). However, population genetic studies of *A. funestus* to date have been few in number and relatively limited in geographic scope, leaving open to question the extent to which their conclusions can be generalized to different ecological settings across its nearly pan-African species range. To provide a more comprehensive framework for the interpretation of *A. funestus* population structure, we sampled > 500 specimens from 11 countries spanning much of the length and breadth of its range: Mali, Burkina Faso, Nigeria, Gabon, Angola, Kenya, Uganda, Tanzania, Malawi, Mozambique and Madagascar. We used mitochondrial DNA (mtDNA) sequences from the *ND5* gene and 10 microsatellite loci distributed on all five chromosome arms to address the following questions: Is there population genetic evidence for subdivision within *A. funestus*? What is the maximum amount of differentiation observed within *A. funestus*, and what is the main force responsible — geographic distance, physical barrier(s), or demographic history? Is the suggestion of a recent population expansion across Africa plausible for *A. funestus*? Finally, does the population history of this species reflect the influence of humans?

Materials and methods

Mosquitoes

Collections were made from 11 localities as indicated in Fig. 1. Adult female mosquitoes morphologically identified as *A. funestus s.l.* (Gillies & De Meillon 1968) were stored at room temperature in labelled tubes with



Map Reference	Country	Village	Latitude	Longitude	Collection Date MM/YY	Collection Method	Sample Size (STR/mtDNA)
1	Mali	Niono	14°15′N	06°00′W	04/01	IR	40/42
2	Burkina Faso	Koubri-Kuiti	12°10′N	01°21′W	12/01 (STR) 09/99 (mtDNA)	IR	Folonzo 50/23 Kiribina 50/47
3	Nigeria	Bakin-Kogi	10°04′N	10°20′E	06/99	IR	29/28
4	Gabon	Bengui	01°53′S	10°09′E	12/99	HL	50/26
5	Angola	Bairro Pinto	08°50′S	13°33′E	04/02	IR	47/52
6	Kenya	Rota	00°05′S	35°40′E	03/98	IR	47/45
7	Uganda	Rakai	00°43′S	31°23′E	07-09/99	IR	29/28
8	Malawi	Kela	16°05′S	34°48′E	02/01	IR	54/53
9	Tanzania	Kibanda	05°10′S	38°46′E	12/97	CDC	58/40
10	Mozambique	Maputo	29°39′S	32°52′E	02-04/00	IR	43/36
11	Madagascar	Miandrivazo	19°31′S	45°28′E	02/00	HL	51/50
							548/470

Fig. 1 *A. funestus* sampling sites in Africa. Dashed lines indicate approximate location of the eastern and western arms of the Great Rift Valley. Collection methods included indoor resting (IR), human landing (HL), and CDC light traps (CDC). STR, simple tandem repeat (microsatellite); mtDNA, mitochondrial DNA.

desiccant. Identification was verified using genomic DNA extracted from individual specimens (DNeasy Tissue Kit, QIAGEN, Inc), diluted 1:10 in H_2O (~5 ng/μL) prior to application of an rDNA-based polymerase chain reaction (PCR) diagnostic assay for members of the Funestus Group (Koekemoer *et al*. 2002 as modified by Michel *et al*. 2005). In addition to *A. funestus*, collections from Mali and Tanzania contained *Anopheles leesoni*, and those from Kenya contained *Anopheles rivulorum*. Only specimens verified as *A. funestus* were analysed further.

Sequencing and analysis

Mitochondrial *ND5* sequences from Burkina Faso were previously published (Michel *et al*. 2005); all others were determined using the primer pair and methods from Michel *et al*. (2005). In rare cases where this primer pair yielded no product, other combinations of primers were designed (sequences available upon request). Microsatellite flanking sequences were determined from apparent homozygotes at loci AFND19, AFND40 and FunL using the same primers (unlabelled) as were used for genotyping. Both strands were directly sequenced using ABI PRISM BigDye protocols on an ABI 377 or ABI 3700 per manufacturer's instructions (Applied Biosystems). Mitochondrial sequences were trimmed to 834 bp and aligned using either SEQUENCE NAVIGATOR (Applied Biosystems) or LASERGENE (DNASTAR, Inc.) software; polymorphic sites were verified against the chromatograms. Mitochondrial sequences were deposited in GenBank under accession numbers DQ102854–DQ103253.

Sequence polymorphism statistics, including θ and π estimators of $4N_e\mu$, were calculated using DNASP 4.0 (Rozas *et al*. 2003). Initial estimates of phylogenetic relationships were reconstructed using the neighbour-joining (NJ) algorithm as implemented in MEGA2 (Kumar *et al*. 2001). Further analysis involved construction of a haplotype network, estimated using the statistical parsimony method implemented in TCS version 1.13 (Clement *et al*. 2000) with reticulations resolved according to the predictions of coalescence theory (Crandall & Templeton 1993). Genetic differentiation (F_{ST}) based on the average number of pairwise nucleotide differences within and between populations was computed in ARLEQUIN 2.0 (Schneider *et al*. 2000), and tested for significance by permuting haplotypes between samples 10 000 times. In addition, ARLEQUIN was used to perform an AMOVA. The relationship between genetic and linear geographic distance was explored using two methods. The program ISOLDE as implemented in GENEPOP 3.3 (Raymond & Rousset 1995) was used to compute a regression of $F_{ST}/(1 - F_{ST})$ on geographic distance for all mainland samples (i.e. excluding Madagascar), and to test for positive correlation using a Mantel test with 1000 permutations. Tests for population growth were implemented in DNASP 4.0 using the statistics Fu's F_S (Fu 1997) and R_2 (Ramos-Onsins & Rozas 2002), with significance assessed through 10 000 coalescent simulations.

Microsatellite genotyping and analysis

Ten microsatellite loci, two per autosome arm and the X, were chosen for this study based on their location outside of polymorphic chromosomal inversions (Fig. S1, Supplementary material, modified from Sharakhov *et al*. 2004). Genotyping methods followed Michel *et al*. (2005). Briefly, after PCR amplification, products were diluted, pool-plexed, and genotyped on a Beckman-Coulter CEQ8000 per manufacturer's instructions, using CEQ8000 software to size alleles.

Diversity was analysed per locus and per population on the basis of allelic richness (R_S, number of alleles independent of sample size), observed heterozygosity (H_O), the

inbreeding coefficient (F_{IS}), and linkage disequilibrium. All calculations were performed in FSTAT 2.9.3.2 (Goudet 2001), with significance based on Bonferroni-corrected *P* values after 10 000 random permutations. Suspected null allele frequencies per locus and per population were generated using the Brookfield 2 estimate (Brookfield 1996), and alleles and genotypes were modified in MICRO-CHECKER (van Oosterhout *et al*. 2004) to create a null allele adjusted data set, used to explore the impact of the null alleles on analyses.

Relationships among populations were examined in several ways. We reconstructed NJ trees based on Nei *et al*. (1983) D_A distance and Cavalli-Sforza & Edwards (1967) chord (D_C) distance as implemented in POPULATIONS 1.2.28 (www.pge.cnrs-gif.fr/bioinfo/populations/index.php). Significance of the nodes was assessed by bootstrapping individuals 1000 times. Bayesian clustering analysis was performed using STRUCTURE 2.0 (Pritchard *et al*. 2000) with a burn-in length of 750 000 and 1 000 000 Markov chain Monte Carlo replications for each of $K = 1$ to 8. Three replicate runs for each *K* were compared to verify the consistency of the estimated likelihood of the data [log Pr($X \mid K$); Pritchard *et al*. 2000]. Genetic differentiation between populations was estimated using F_{ST} calculated in MSA software (Dieringer & Schlotterer 2003) and R_{ST} calculated in ARLEQUIN 2.0, with *P* values based on 10 000 random permutations of genotypes. The isolation by distance model was tested using the same procedure as for mtDNA F_{ST} values (see above). Additionally, within the western and eastern population clusters inferred by STRUCTURE, the program SPAIDA (Palsson 2004) was used to calculate two estimates of spatial autocorrelation based on frequencies of different alleles (Moran's *I*) or differences in the numbers of repeats (Geary's *c*). The significance of the correlation for each distance class was evaluated by 1000 permutations of individuals among locations. Inference of past population growth was made using the β-imbalance index (Kimmel *et al*. 1998), with the β_1 estimator. We generated 95% confidence intervals by bootstrapping over loci using a program written by T.L. (Donnelly *et al*. 2001).

Folonzo and Kiribina from Burkina Faso

Codistributed in Burkina Faso, West Africa, are chromosomally recognized incipient species of *A. funestus* designated Folonzo and Kiribina (Costantini *et al*. 1999). Based on variation at 16 microsatellite loci distributed genomewide and mtDNA *ND5* sequences, molecular differentiation between Folonzo and Kiribina was significant but very slight (mtDNA $F_{ST} = 0.023$, $P < 0.001$; microsatellite $F_{ST} = 0.004$, $P < 0.001$; Michel *et al*. 2005). The distribution of these taxa outside of Burkina Faso has not been established. Because the specimens available for this study were not karyotyped with the exception of those from Burkina Faso, and due to the lack of any molecular diagnostic tool, it was not possible to assign the remaining *A. funestus* collections to Folonzo or Kiribina. Given the small divergence between emerging species for both marker classes, relative to levels of divergence between genetic clusters identified in this study (e.g. microsatellite $F_{ST} > 0.2$; see Results), any potential pooling of these taxa within and/or between localities in this or other studies should not have a significant impact on the results or conclusions. As an illustration, samples identified as Folonzo and Kiribina from Burkina Faso (Fig. 1) were incorporated into both the Bayesian clustering analysis performed with microsatellite data, and the statistical parsimony analysis used to construct a haplotype network from mtDNA sequences; in neither case was any structure detected in relation to these taxa (see Figs 2 and 3). All other analyses were conducted using Burkina Faso Folonzo specimens exclusively.

Results

mtDNA polymorphism

We analysed sequence variation at 834 bp of the *ND5* gene from a total of 423 individual mosquitoes (excluding 47 Kiribina from Burkina Faso). Unambiguous electropherograms and the predominance of synonymous mutations support the absence of nuclear-integrated copies of mtDNA. Across the total data set, there were 146 variable sites, of which half were singletons (substitutions present in only one sampled sequence). The large number of singleton mutations contributed both to a high haplotype diversity (0.951 for 190 haplotypes) and to a marked difference between two estimates of the population mutation parameter $4N_e\mu$, based on the average pairwise nucleotide diversity ($\pi = 0.007$ per site) or the number of segregating sites ($\theta = 0.026$ per site). This trend is not a trivial result of pooling geographically separated populations, as it was consistently found within samples from single locales (Table 1). Not surprisingly, the sample from the island of Madagascar had much lower haplotype diversity, but the relative difference between π and θ persisted.

Microsatellite polymorphism

All 10 microsatellite loci amplified reliably from a total of 548 individual mosquitoes, with no evidence of nonspecific or selective amplification. Similar to mtDNA, levels of microsatellite polymorphism across loci and samples were moderate to high (mean allelic richness, 6.0; mean heterozygosity, 0.59; Table 1). However, polymorphism was notably lower not only on the island of Madagascar, as expected, but also in Mozambique. This was especially pronounced for the two loci on the X chromosome, AFND12 and FUNQ.

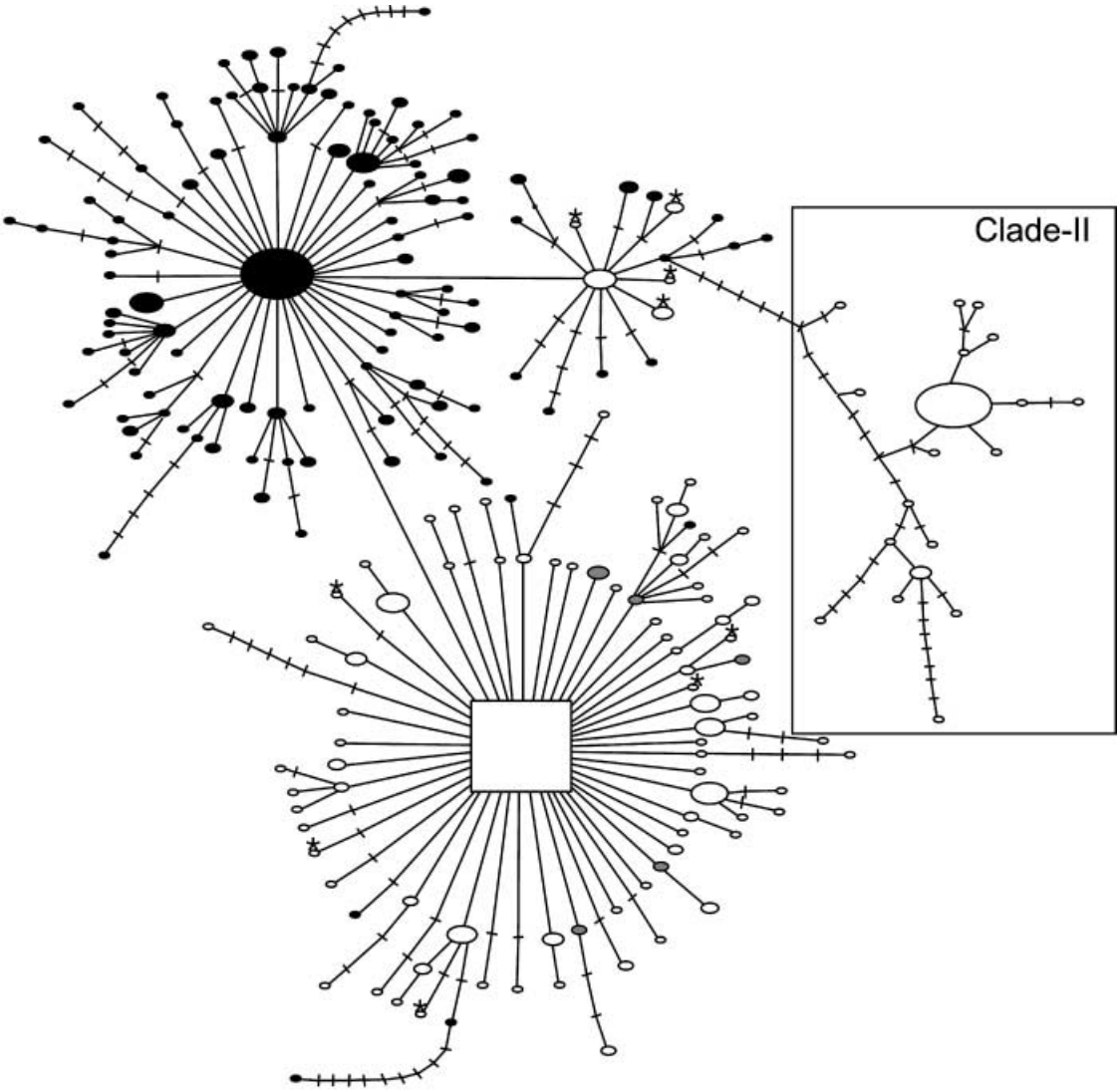


Fig. 2 Haplotype network of mtDNA *ND5* sequences. Haplotypes are represented as ovals, scaled to reflect frequency. In addition, the most frequent haplotype ($N = 53$), inferred as ancestral, is represented by a square. Lines connecting haplotypes and hatch marks (missing intermediate haplotypes) indicate one mutational step. Regions are denoted by colour: black, West Africa + Kenya; white, East Africa + Gabon + Angola; grey, shared between regions. Asterisks mark haplotypes unique to Uganda. Box encloses haplotypes belonging to clade II (see Results).

After correction for multiple tests, significant deviations from Hardy–Weinberg equilibrium were found in 7 out of 130 tests (5%), owing to heterozygote deficits. The deviations were not limited to particular loci, nor were they clustered by geographic sample. Taken together with the absence of linkage disequilibrium in these samples, we attribute the heterozygote deficits to null alleles rather than assortative mating or Wahlund effect. Assuming random mating, suspected null allele frequencies were estimated for each locus and population (Table S1, Supplementary material), yielding a mean frequency of 6%. This is in accord with the mean frequency of null alleles (8%) estimated for *A. funestus* from Burkina Faso in a previous study (Michel *et al*. 2005). These estimates are also consistent with direct measurements of null alleles based on X-linked loci in *A. gambiae* males, which suggest that most sampled loci are affected at a mean frequency of 5–8% (Barnes *et al*. 2005; Stump *et al*. 2005). Null alleles had a negligible effect on our conclusions, as the output of analyses run before and after adjustment of allele and genotype frequencies for null alleles were indistinguishable.

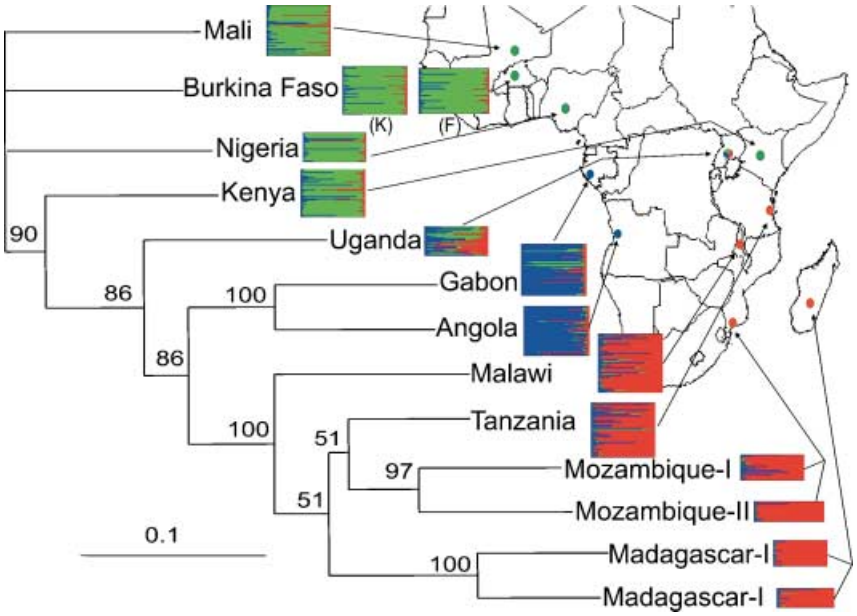


Fig. 3 Population structure of *A. funestus* inferred from microsatellites. Neighbour-joining tree based on Nei's D_A (at left), shown with corresponding membership coefficients for $K = 3$ inferred by the model-based clustering method implemented in STRUCTURE (Pritchard *et al*. 2000). Green, K_1 (mainly West Africa); blue, K_2 (mainly Gabon + Angola); red, K_3 (mainly East Africa). Folonzo (F) and Kiribina (K) from Burkina Faso were assigned with virtually identical membership coefficients to K_1. Bootstrap values are based on 100 replicates.

Homoplasy

In the analysis of microsatellite data, electromorphs (alleles of the same length as assessed by electrophoresis) are assumed to be identical by descent, not only identical in state. Given microsatellite mutation rates that are often at least three orders of magnitude greater than nucleotide substitution rates in flanking DNA, it is expected that electromorphs have identical flanking sequences unless there has been convergent evolution to the same allele size. This assumption does not always hold between closely related species, between geographically distant populations of the same species, or even within a single large population that has expanded historically (Mank & Avise 2003). Indeed, the phenomenon of null alleles, presumably caused by a nucleotide substitution in the primer-binding site, is a manifestation of homoplasy. If present, homoplasy results in the underestimation of diversity and divergence. To obtain a rough gauge of homoplasy in our data, we determined flanking sequence from electromorphs at three loci (AFND19, AFND40, FUNL) in population subsamples from geographically distant locations: Burkina Faso, Gabon, Kenya and Malawi. Outside of the microsatellite itself, no insertions/deletions were detected, but nucleotide substitutions indicated some homoplasy. Across all loci (398 bp of aligned sequence from 30 amplicons), eight polymorphic sites were found, of which seven were due to singletons (Table S2, Supplementary material). The remaining substitution, a G↔A transition in AFND19, was polymorphic in three distant samples (Burkina Faso, Gabon and Malawi).

Phylogeography

As a first step in the inference of geographic population structure, we applied clustering methods that do not rely on a priori definitions of populations based on collection locales (Mank & Avise 2004). Beginning with mtDNA, phylogenetic relationships were inferred initially from an NJ tree. This tree lacked resolution, with the exception of one clade branching off the star phylogeny, comprised of a subsample of mosquitoes from both Mozambique and Madagascar (data not shown). Examining this pattern in more detail, we built a tree based only on the total samples from Mozambique and Madagascar. As before, this resulted in one major subdivision into two clades differing by an average of 18 mutational steps including two fixed differences (~2% sequence divergence), compared to <0.4% divergence within clades (data not shown). Surprisingly, each clade contained mosquitoes collected from both locales. Based on this unexpected result, the mtDNA samples from Mozambique and Madagascar were no longer treated as homogeneous. Those sequences that clustered with, or apart from, samples from the other nine mainland locales are hereafter referred to as belonging to clade I (designated Mozambique-I, $n = 22$ and Madagascar-I, $n = 29$) or clade II (Mozambique-II, $n = 14$ and Madagascar-II, $n = 21$), respectively. Clade I represents the majority of *A. funestus* analysed in this study; clade II contains individuals that, to our knowledge, have only been sampled from Mozambique and Madagascar.

For better resolution of intraspecific phylogeny, we implemented a statistical parsimony method (Clement

Table 1 Summary of variation at 10 microsatellite loci and mtDNA *ND5* by collection site. Mozambique and Madagascar have been partitioned into putative clades I and II (See Results)

Locus	MAL	BFA	NIG	GAB	ANG	UGA	KEN	MAW	TAN	MOZ-I	MAD-I	MOZ-II	MAD-II
AFND12 (X)													
R_S	9.0	11.0	8.0	4.0	5.0	7.0	9.9	2.0	3.0	1.7	1.0	1.0	1.0
H_O	0.61	0.63	0.50	0.31	0.28	0.37	0.64	0.0	0.02	0.04	0	0	0
F_{IS}	0.27	**0.26**	0.39	0.27	0.03	0.15	0.23	1.0	0.66	0	NA	NA	NA
FUNQ (X)													
R_S	7.8	6.0	7.0	7.9	7.9	5.0	6.8	6.0	6.8	1.0	1.6	1.0	3.8
H_O	0.53	0.60	0.66	0.54	0.60	0.45	0.55	0.17	0.37	0	0.03	0	0.19
F_{IS}	0.30	0.15	0.16	0.08	–0.01	–0.09	**0.27**	–0.05	0.24	NA	0	NA	–0.05
AFND40 (2R)													
R_S	5.0	6.0	8.9	5.0	5.0	6.0	5.0	4.0	5.0	4.0	4.0	4.0	4.0
H_O	0.72	0.73	0.72	0.55	0.60	0.55	0.57	0.57	0.5	0.48	0.17	0.33	0.24
F_{IS}	0.01	0.01	0.05	0.10	–0.05	0.23	0.17	0.14	0.08	0.33	**0.72**	0.53	0.53
FUNO(2R)													
R_S	7.9	11.0	7.9	6.0	6.0	6.0	8.8	5.0	5.9	4.4	3.6	3.9	5.0
H_O	0.63	0.69	0.55	0.66	0.69	0.66	0.68	0.50	0.55	0.52	0.56	0.39	0.62
F_{IS}	0.13	0.10	0.27	0.06	0.01	–0.14	–0.07	0.17	0.11	0.15	–0.23	0.38	0.09
FUNL (2L)													
R_S	10.9	15.0	10.8	6.0	7.8	14.0	12.9	9.0	7.8	3.7	5.3	4.9	4.7
H_O	0.56	0.65	0.66	0.31	0.38	0.66	0.55	0.70	0.52	0.49	0.50	0.67	0.55
F_{IS}	**0.34**	0.20	0.17	0.34	0.17	0.21	**0.29**	0.05	0.29	0.28	0.18	–0.08	0.11
AFUB11 (2L)													
R_S	7.9	9.0	8.0	7.0	6.0	6.0	6.9	6.0	3.9	3.7	3.0	3.0	3.0
H_O	0.63	0.54	0.69	0.62	0.60	0.38	0.51	0.54	0.31	0.32	0.50	0.39	0.57
F_{IS}	0.11	0.17	0.06	0.03	0.03	0.10	–0.02	0.01	0.18	0.16	0.24	0.37	0.11
AFND7 (3R)													
R_S	6.9	9.0	7.0	6.0	6.0	7.0	7.9	9.0	6.9	3.9	5.5	5.0	6.5
H_O	0.45	0.71	0.73	0.8	0.81	0.72	0.66	0.69	0.67	0.52	0.70	0.71	0.71
F_{IS}	**0.37**	0.12	0.09	–0.06	–0.07	0.01	0.16	0.06	0.10	0	–0.08	–0.33	–0.10
AFND19 (3R)													
R_S	11.8	11.0	8.9	7.0	7.9	7.0	9.0	7.0	6.9	4.0	3.0	4.9	4.0
H_O	0.65	0.79	0.73	0.78	0.74	0.59	0.70	0.65	0.62	0.48	0.60	0.56	0.57
F_{IS}	0.12	0.03	0.01	–0.16	–0.07	0.10	0.11	–0.03	–0.07	0.09	–0.06	0.10	0.08
AFUB12 (3L)													
R_S	6.8	4.0	4.0	4.0	4.0	4.0	3.0	4.0	3.0	3.0	2.7	2.9	2.8
H_O	0.4	0.43	0.59	0.34	0.68	0.55	0.40	0.56	0.55	0.72	0.17	0.33	0.14
F_{IS}	0.34	0.11	0.10	**0.47**	–0.01	0.18	0.29	0.11	0.12	–0.09	–0.05	0.33	0.36
FUNF (3L)													
R_S	7.9	7.0	6.9	5.0	4.0	6.0	6.0	8.0	6.0	4.6	5.5	6.0	4.8
H_O	0.75	0.69	0.79	0.65	0.85	0.83	0.52	0.76	0.64	0.75	0.60	0.83	0.57
F_{IS}	0.04	0.10	–0.06	0.01	–0.25	–0.16	0.24	–0.07	–0.05	–0.21	0	–0.10	–0.13
Average R_S	8.19	8.9	7.74	5.79	5.96	6.8	7.62	6	5.52	3.4	3.52	3.66	3.96
Average H_O	0.593	0.646	0.662	0.556	0.623	0.576	0.578	0.514	0.475	0.432	0.383	0.421	0.416
mtDNA *ND5*													
Hd	0.974	0.992	0.906	0.803	0.908	0.907	0.978	0.800	0.862	0.900	0.621	0.956	0.500
π	0.004	0.006	0.004	0.003	0.003	0.003	0.005	0.002	0.002	0.004	0.001	0.005	0.001
θ	0.010	0.013	0.007	0.004	0.009	0.006	0.012	0.006	0.003	0.008	0.003	0.006	0.002

Country abbreviations: MAL, Mali; BFA, Burkina Faso; NIG, Nigeria; GAB, Gabon; ANG, Angola; UGA, Uganda; KEN, Kenya; MAW, Malawi; TAN, Tanzania; MOZ, Mozambique; MAD, Madagascar. R_S, allelic richness; H_O, observed heterozygosity; F_{IS}, inbreeding coefficient (significant values after Bonferroni correction given in bold); Hd, haplotype diversity; π, average pairwise nucleotide difference per site; θ, the population mutation rate (= $4N\mu$) based on the number of segregating sites.

Table 2 Proportion of membership of each predefined population in each of the three clusters inferred by STRUCTURE

Population	West	Central	East
Burkina Faso	0.861	0.081	0.057
Mali	0.857	0.097	0.046
Nigeria	0.805	0.130	0.065
Kenya	0.779	0.140	0.081
Uganda	0.337	0.316	0.348
Gabon	0.119	0.789	0.091
Angola	0.042	0.857	0.101
Tanzania	0.046	0.162	0.792
Malawi	0.048	0.184	0.767
Mozambique-I	0.021	0.168	0.811
Mozambique-II	0.012	0.062	0.925
Madagascar-I	0.014	0.051	0.935
Madagascar-II	0.032	0.087	0.881

et al. 2000) to construct an mtDNA haplotype network representing the complete data set. All haplotypes could be connected within the limits of parsimony (12 steps). After resolving reticulations using predictions of coalescent theory (Crandall & Templeton 1993), four main mtDNA clusters emerged (Fig. 2). One that was separated from the others by multiple mutational steps contained all mosquitoes from clade II. The hubs of the other three clusters were within one mutational step of each other. The first of these three clusters was composed largely of haplotypes from East Africa, although it did not contain the majority of Kenyan and Ugandan haplotypes. In addition, haplotypes from Gabon and Angola were included. The second consisted of haplotypes from the West African locales, as well as most haplotypes from Kenya. The third, a small group of haplotypes found only in Uganda and Kenya, connected directly to clade II.

Turning to the microsatellite data, we implemented an unsupervised model-based clustering method (without predefined population information based on geographic sampling location). The program STRUCTURE (Pritchard *et al*. 2000) assigns individuals to K clusters, where K represents the number of populations with distinctive multilocus allele frequencies. The most likely fit to the data on each of three replicate runs was $K = 3$, where K_1 = West Africa (Mali, Burkina Faso, Nigeria) + Kenya west of the Rift Valley, K_2 = Gabon + coastal Angola, and K_3 = East Africa (coastal Tanzania, Malawi, Mozambique, Madagascar). The membership coefficients of geographic samples to their respective clusters were relatively high, ranging from 0.77 to 0.94, except for Uganda whose membership was evenly divided among the three clusters (Table 2). Of note is the absence of any nuclear subdivision that corresponds to mtDNA clade II, based on microsatellite evidence. All individuals from Madagascar and Mozambique were strongly assigned to K_3 based on microsatellites. We compared the results of unsupervised clustering to NJ trees based on genetic distances between geographic samples, using the measures D_A (Nei *et al*. 1983) and D_C (Cavalli-Sforza & Edwards 1967). Both distance measures gave the same topology and degree of resolution. Figure 3 illustrates the good agreement between the NJ tree constructed with D_A and model-based clustering.

Overall, although both marker classes revealed the same subdivision between East and West Africa, microsatellites provided better resolution, indicating the distinctive grouping of Gabon and coastal Angola. On the other hand, the cryptic subdivision revealed by mtDNA as clade II apparently does not extend to the nuclear genome.

Genetic differentiation

Genetic differentiation between all pairs of geographic samples and across loci was estimated, based on allele frequency differences at microsatellite loci and average nucleotide differences between mtDNA sequences. Instead of presenting the complete matrix (available upon request), for clarity these individual pairwise estimates were grouped according to four types of comparisons: (i) within each of the mainland clusters — i.e. the three clusters inferred by STRUCTURE but excluding Madagascar, (ii) among the three mainland clusters, (iii) between mainland and Madagascar, and (iv) between clade I and clade II (Table 3). Within mainland clusters, although all or nearly all pairwise estimates of differentiation (F_{ST} and R_{ST}) were significantly different from zero, overall levels of differentiation were relatively small, especially in the west. At this level of comparison, there was general agreement between F_{ST} and R_{ST} as well as between different markers, although the magnitude of mtDNA estimates exceeded those based on microsatellites.

Overall differentiation among mainland clusters was as much as an order of magnitude higher than differentiation within clusters, a trend that was consistent between F_{ST}/R_{ST} measures and different marker classes. However, global comparisons among clusters mask the much higher affinity of the east and central clusters (average microsatellite F_{ST} = 0.069) relative to the west (east-west, microsatellite F_{ST} = 0.238; central-west, microsatellite F_{ST} = 0.268).

Because of the physical ocean barrier, it was expected that levels of differentiation between mainland and island samples would be relatively high. To avoid confounding genetic distance between mtDNA clades and genetic distance due to the ocean barrier, comparisons between mainland and island were made only within each mtDNA clade, between Mozambique and Madagascar samples. For both clades and both marker classes, genetic differentiation was as high or higher (by as much as twofold) across the Mozambique Channel as it was between mainland regions, with the exception of the mtDNA estimate for clade I.

Table 3 Genetic differentiation in *A. funestus*

	Within mainland clusters*			Between mainland clusters	Mainland vs. island†		Clade I vs. Clade II	
	West	Central	East		Clade I	Clade II	Mozambique	Madagascar
Microsatellites								
Ave. pairwise F_{ST}	0.010	0.020	0.039	0.110	0.111	0.130	0.018	0.000
Range	0.004–0.020	NA	0.018–0.065	0.056–0.158	NA	NA	NA	NA
Significant pairwise	5/6	1/1	6/6	32/32	1/1	1/1	1/1	0/1
Ave. pairwise R_{ST}	0.016	0.018	0.059	0.106	0.143	0.190	0.000	0.015
Range	0.000–0.042	NA	0.015–0.114	0.051–0.182	NA	NA	NA	NA
Significant pairwise	2/6	1/1	6/6	32/32	1/1	1/1	0/1	0/1
mtDNA								
Ave. pairwise F_{ST}	0.015	0.081	‡0.061	‡0.212	0.031	0.336	0.753	0.951
Range	0.000–0.023	NA	0.042–0.075	0.027–0.323	NA	NA	NA	NA
Significant pairwise	4/6	1/1	3/3	26/26	1/1	1/1	1/1	1/1

*Mainland Clusters: West — Mali, Burkina Faso, Nigeria, western Kenya; Central — Gabon, Angola; East — Tanzania, Malawi, Mozambique-I and II.
†Mainland represented by Mozambique; Island — Madagascar.
‡Excludes Mozambique-II.

Clades I and II were defined based on mtDNA sequence differences. Not surprisingly, when mtDNA sequence comparisons were made between sympatric samples of these clades in Mozambique or Madagascar, F_{ST} estimates were extremely high (0.753 and 0.951, respectively). By contrast, corresponding microsatellite differentiation was very slight (0.018) or nonsignificant.

For additional insight into how variation was partitioned at different hierarchical levels across the continent, we performed an AMOVA on the mtDNA and microsatellite data. After excluding samples from Madagascar (and clade II from Mozambique for the mtDNA AMOVA), we compared populations within and among the three clusters inferred by STRUCTURE. The AMOVAs on both data sets revealed similar patterns. The vast majority of variation was explained at the intrapopulation level (78% and 89% for mtDNA and microsatellites, respectively). The bulk of the remaining variation was explained by differences among groups (19%, mtDNA; 8%, microsatellites). The small amount of remaining variation was due to differences between populations within groups (3%, mtDNA; 2% microsatellites).

Isolation by distance

Neighbouring populations often resemble one another more than geographically distant ones. Under the isolation-by-distance model, this is assumed to result from ongoing genetic mixing occurring more frequently across smaller distances. The geographic composition of the western cluster, which contains samples from Kenya as well as from West Africa, already suggests that distance alone is inadequate to explain *A. funestus* population structure. However, formal analysis of the correlation of genetic and geographic distance can lend insight about the relative importance of ongoing gene flow and population history. Initially, we examined this relationship for all mainland samples (excluding Madagascar), for microsatellite and mtDNA markers. Clade II was excluded from the mtDNA analysis. All pairwise genetic distance measures were plotted against geographic distance separating the populations, and $F_{ST}/(1 - F_{ST})$ was regressed against linear distance (Fig. 4). Mantel tests indicated a weak but

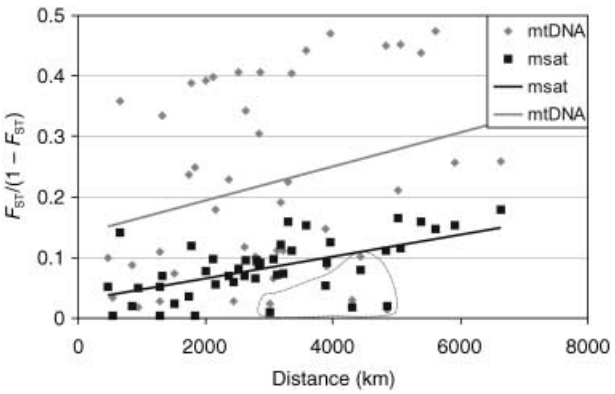


Fig. 4 Scatterplots of pairwise F_{ST} values against geographic distance separating pairs of locations, shown with regression line through the points. Grey, mtDNA; black, microsatellites. Points enclosed by dashed line are comparisons between Kenya or Uganda and West African locations.

significantly positive association between pairwise F_{ST} and distance (mtDNA, $r^2 = 0.076$, $P < 0.049$; microsatellites, $r^2 = 0.336$, $P < 0.003$). Detailed analysis of the scatterplot revealed that some of the 'noise' in the data — the points in the lower right quadrant of Fig. 4, representing relatively low genetic differentiation given large geographic distances — corresponded to comparisons between Kenya or Uganda and West African locations. Separate tests of isolation by distance within either the western or eastern clusters, whether conducted by regression of F_{ST} against distance or by spatial autocorrelation, showed no significant relationship. Because the present study lacks adequate samples within each cluster to provide a robust test, we do not suggest that there is no effect of distance on differentiation. However, it appears that the positive correlation indicated by the analysis in Fig. 4 may be a misleading effect of having combined diverse genetic clusters that are separated spatially. At best, large geographic distances have only a weak relationship with genetic differentiation in this species. Instead, the pattern is consistent with a relatively recent demographic expansion spanning West Africa to western Kenya. The following analysis lends some support for this hypothesis.

Deviation from equilibrium

The relationship between gene flow and F_{ST} expressed as $F_{ST} \cong 1/(4Nm + 1)$ (Wright 1931) assumes an equilibrium between migration and drift. The large number of singleton mtDNA mutations, and the discrepancy between π and θ estimates of diversity, hint that this assumption may be violated. Demographic events such as a recent population expansion disrupt equilibrium and can inflate the apparent rate of gene flow estimated by indirect genetic methods. Several tests have been developed for mtDNA sequence and microsatellites to detect departures from equilibrium. For mtDNA sequences, we performed the two most powerful tests for detecting population growth, the R_2 and F_S tests (Fu 1997; Ramos-Onsins & Rozas 2002), based on the segregating site frequency and the haplotype distribution, respectively. Both approaches are based on the expectation that in an expanding population, there will be an excess of singleton mutations, but they use different aspects of the data. For microsatellites, we used a statistic based on the imbalance between allele size variance and heterozygosity at a locus, the β imbalance index (Kimmel *et al*. 1998), as it has been shown to be the most powerful of the microsatellite-based statistics against population growth (King *et al*. 2000). At equilibrium, the ratio between size variance and heterozygosity at a locus should equal 1. The test for population expansion is premised on the expectation that recovery of variance in allele size will lag behind recovery of heterozygosity at a locus after an expansion, causing the β index to fall below 1. Results of the two mtDNA-based tests concurred in showing that most populations significantly deviated from equilibrium (Table 4). Results from the microsatellite-based test also were consistent with a population expansion ($\beta < 1$) for all locations in West Africa and Kenya. Although β-values in Burkina Faso and Kenya were not significantly less than 1 at the 5% level, only 5–10% of the bootstrap replicates produced $\beta > 1$, indicating that these results are marginally significant.

Table 4 Tests of population growth vs. constancy

	mtDNA		Microsatellites
	R_2	F_S	β_1‡ (95% CI)
West†			
Mali	0.04**	–29.8***	0.56* (0.80–0.44)
Burkina Faso	0.05**	–16.6***	0.75ns (1.22–0.42)
Nigeria	0.10ns	–2.6ns	0.56* (0.71–0.40)
Kenya	0.04***	–21.5***	0.75ns (1.10–0.47)
Central			
Gabon	0.08ns	–4.2*	1.14ns (2.11–0.65)
Angola	0.02***	–31.2***	1.15ns (1.30–0.58)
East			
Uganda	0.07*	–5.0*	1.13ns (1.58–0.55)
Tanzania	0.09ns	–2.7ns	1.01ns (1.50–0.61)
Malawi	0.03**	–18.3***	1.35ns (2.19–0.73)
Mozambique-I	0.05***	–8.9*	1.40ns (1.70–0.39)
Mozambique-II	0.12ns	–3.9*	1.01ns (1.65–0.4)
Island			
Madagascar-I	0.05***	–9.8***	1.40ns (2.25–0.69)
Madagascar-II	0.08*	–4.15**	1.38ns (2.20–0.64)

†Population membership as defined by STRUCTURE analysis (see Results).
‡β_1 Imbalance index and 95% confidence intervals (CI) generated by bootstrapping, to assess whether the index deviates significantly from 1.
*$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

Discussion

A. funestus is a species characterized by abundant levels of molecular and chromosomal polymorphism across its range. Most of this variation can be explained by differences among individuals within population samples, rather than between samples from the same geographic region or between different geographic regions. Nevertheless, we found that nearly all pairs of population samples were significantly differentiated, and Bayesian clustering (STRUCTURE) analysis of microsatellite data suggested that there are at least two main divisions within *A. funestus*. Although the mtDNA data did not conflict with this outcome, they were generally less informative (with the important exception of uncovering clades I and II; see below). The

two divisions revealed by Bayesian clustering define three clusters that correspond to (i) Africa east of the Great Rift Valley (coastal Tanzania, Malawi, Mozambique, Madagascar), (ii) western Africa (Mali, Burkina Faso, Nigeria) + Kenya west of the Rift Valley, and (iii) central Africa (Gabon, coastal Angola). Samples from the southwestern part of Uganda near the western shore of Lake Victoria were not strongly assigned to any of the three clusters. A previous rDNA-based study (Garros *et al*. 2004) that included samples of *A. funestus* from across Africa also found evidence for two genetic subdivisions similar to the ones proposed here: an eastern cluster represented by coastal Kenya (east of the Rift Valley) and Madagascar, presumably equivalent to our eastern cluster; and a west-central cluster represented by Burkina Faso, Senegal and Cameroon, possibly corresponding to what we have called the western cluster. Both clusters were defined based on fixed differences in rDNA haplotypes (ribotypes). Garros *et al*. (2004) described a third cluster represented by southern Angola, Mozambique and South Africa which contains both ribotypes within individual genomes, and thus may represent admixture between the east and west-central clusters. (Retained ancestral polymorphism is a less-likely explanation, given that the tandem rDNA gene arrays evolve through a process of concerted evolution and molecular drive). Although the geographic boundaries of these genetic subdivisions do not precisely correspond, they are not contradictory and their appearance in independent studies and samples based on three different marker systems reinforces the conclusion of a major east–west subdivision.

This overall pattern bears a striking resemblance to rangewide population structure in the codistributed malaria vector *Anopheles gambiae*, in which a main continental split has been found between the east (coastal Kenya, Tanzania, Malawi and Zambia) and populations to the northwest, extending from western Kenya to Senegal (Lehmann *et al*. 2003). Thus, in *A. gambiae* as in *A. funestus*, samples from western Kenya are genetically more closely related to those from West Africa than to samples from the coast of Kenya or Tanzania. Phylogeographic structure of *A. gambiae* shows further similarities to *A. funestus*, namely a series of nested population clusters in the eastern region, and the placement of *A. gambiae* from Gabon as relatively distant and basal to the main northwestern cluster that included western Kenya and Nigeria (Angola was not sampled in the *A. gambiae* study of Lehmann *et al*. (2003). Available evidence suggests that at the continental scale, the population structure of *A. gambiae* and *A. funestus* is similar and likely has been shaped by the same geographic and climatic forces.

The Rift Valley represents a barrier to gene flow for *A. gambiae* (Lehmann *et al*. 1999, 2000), and was invoked as a main contributor to its subdivision (Lehmann *et al*. 2003); the same has been suggested for *A. funestus* (Braginets *et al*. 2003). In addition to the Rift Valley as a physical barrier, another factor hypothesized as contributing to the subdivision in *A. gambiae* into western and eastern clusters is a bottleneck in the eastern division (Donnelly *et al*. 2001), possibly due to very severe droughts (Lehmann *et al*. 2003). In the present study, we found evidence of reduced heterozygosity in *A. funestus* from the corresponding geographic region (our eastern cluster), in agreement with Braginets *et al*. (2003), and consistent with predictions for a codistributed drought-sensitive species. (Despite its association with 'permanent' breeding sites, there are records of *A. funestus* having been eliminated from parts of Africa due to prolonged severe drought; e.g. Mouchet *et al*. 1996). Average heterozygosity (H_O) and allelic richness (R_S) across all microsatellite loci were not only lower for samples from Madagascar, as might be expected for an island population, but these parameters also were markedly lower for the entire eastern cluster than for the rest of Africa. Whereas values of R_S and H_O in the west + central clusters averaged 6.0 and 0.606, respectively, the corresponding values for the east were 3.9 and 0.455, significantly lower (one sided *P* values obtained after 10 000 permutations: R_S, 0.0013; H_O, 0.0004). The same was true for mtDNA haplotype diversity between the east (mean, 0.773) and the rest of Africa (mean, 0.924). This trend remained significant when Madagascar was omitted from the comparisons (data not shown). Although the reduction in diversity was particularly striking for both loci on the X chromosome, suggesting that selection could be involved, the trend of lower diversity in the east was genome-wide and therefore is consistent with lower effective population size due to historical drought or other unknown factors.

Although tests for population expansion in *A. gambiae* have only been applied in Kenya, it has been assumed that the shallow population structure in *A. gambiae* across Africa, particularly in the western division, is a reflection of an expansion that spread across the continent in response to human population expansion (Donnelly *et al*. 2001; Lehmann *et al*. 2003). There is evidence for a similar history in *A. funestus*. In the western cluster, both mtDNA and microsatellite markers rejected mutation–migration–drift equilibrium in most localities, indicating a genome-wide effect most likely due to population expansion rather than selection. Our data do not support other explanations, such as population structure within localities or very high levels of homoplasy. Outside of the western cluster, mtDNA but not microsatellites rejected equilibrium. If this result is interpreted to mean that microsatellite variation is at equilibrium, the mtDNA signal might reflect selection on mtDNA. In fact, the extraordinarily low observed heterozygosity for the two microsatellite loci on the X chromosome across the entire eastern cluster may suggest a selective interaction between mtDNA and the X chromosome.

Alternatively, the significantly reduced genetic diversity in the east suggests the possibility of a bottleneck. If this preceded an expansion, a long transient period (several thousand generations) is expected during which the β imbalance index for microsatellites should exceed 1, before eventually falling below 1 (Kimmel *et al*. 1998). Further studies will be required to improve inference of population demography in the east and to test these hypotheses. Based on the absence of a strong genome-wide signal elsewhere, we propose that the western cluster of *A. funestus* experienced population growth across a region spanning West Africa to western Kenya. Southwestern Uganda does not appear to have been involved in the expansion, nor is it associated strongly with any of the three clusters, showing roughly equal membership to each. It is possible that this region of Africa may represent part of a corridor of migration between the clusters.

There was only a weak positive correlation between geographic and genetic distance in *A. funestus*, which appeared to be dictated by genetic differences between clusters from the West vs. East and Central Africa rather than physical distance per se. Moreover, the magnitude of genetic differentiation was very low when examined only within mainland clusters. Samples from the western cluster separated by 1500–3000 km were characterized by mean F_{ST} estimates of 0.004 and 0.020 for microsatellites and mtDNA, respectively. For the same geographic range, samples from the eastern cluster had slightly larger but nevertheless low F_{ST} estimates (microsatellites, 0.043; mtDNA, 0.081). Although both marker classes were largely in agreement, mtDNA reflected at least twofold greater levels of differentiation than the microsatellites. This may result from a smaller effective population size for mtDNA relative to the autosomes or X chromosome. Additionally, because microsatellites evolve at a rate that is at least three orders of magnitude in excess of mtDNA sequence, shallower microsatellite divergence estimates may reflect the contribution of allelic homoplasy in the microsatellite data, a phenomenon for which we have both indirect and direct evidence (Tables S1 and S2). Nevertheless, homoplasy was not rampant and neither mutational saturation nor constraint is adequate to explain such shallow differentiation, given the much higher levels of microsatellite differentiation across the Great Rift Valley (Braginets *et al*. 2003), between genetic clusters, and across the Mozambique Channel. Again, there are strong parallels to the situation in *A. gambiae*, perhaps a surprising outcome considering the significant bionomic differences between the two species (Gillies & De Meillon 1968). A likely explanation is that indirect genetic methods are reflecting a history of recent population growth linked to a common trigger, human population expansion. Because molecular markers of population structure have not regained equilibrium, they are insensitive to present-day dynamics.

Microsatellite and mtDNA markers generally revealed concordant patterns within and among different geographic samples of *A. funestus*, with an unexpected exception. Within single collections from Maputo, Mozambique, and from Miandrivazo, Madagascar, were highly differentiated mtDNA lineages, separated by two fixed nucleotide substitutions and differing on average by 18.5 nt, in contrast to the average pairwise nucleotide difference within lineages of only 2–3 nt. As each mosquito from both samples had been classified on the basis of adult morphological characters as belonging to the *A. funestus* subgroup (Gillies & Coetzee 1987), and had been identified as *A. funestus s.s.* by an rDNA-based PCR assay, such a deep mtDNA divergence within a collection made from the same time and place was unanticipated. One of the mtDNA lineages, designated clade I, encompassed the mtDNA sequences from all other African samples of *A. funestus*. The second, clade II, was found only in Mozambique and Madagascar (Fig. 2). The deep mtDNA split, ~2.2% sequence divergence (~1.9% after accounting for within-clade polymorphism), may correspond to ~850 000 years divergence, assuming a rate of mtDNA evolution of 1.1–1.2% per million years per lineage (Brower 1994). Yet, despite deep mtDNA divergence, there was no corresponding nuclear divergence (Table 2).

In considering plausible explanations for the origin of mtDNA clade II, vicariance biogeography involving Madagascar can be excluded, because its separation from the African mainland began ~165 million years ago (Rabinowitz *et al*. 1983). Moreover, a Malagasy origin of clade II owing to human transport also can be excluded, as Madagascar was devoid of native anthropoids of any kind until only 2000 years ago, being one of the last great landmasses to be colonized by humans (Burney *et al*. 2004). An origin of this clade on Madagascar through aerial invasion from the west across the Mozambique Channel is not consistent with typical weather patterns, although this possibility should not be dismissed; there is precedent for transoceanic dispersal of Malagasy insects (Monaghan *et al*. 2005). A more plausible explanation for the origin of clade II ~850 000 years ago involves climatic fluctuations in rainfall and temperature during the Pleistocene on the southern African mainland (Cooke 1964). Decreased annual rainfall during nonpluvial periods would have restricted suitable habitat into isolated islands, facilitating differentiation and speciation (Brain & Meester 1964). Later habitat expansion during pluvial periods would have renewed contact between diverging populations or emergent species. Because mtDNA is haploid, maternally transmitted, and nonrecombining, admixed populations could retain both distinct mtDNA clades, while their nuclear genomes would become homogenized. Likewise, hybridization between species could have resulted in the introgression of the mtDNA genome characteristic of one species into

the nuclear background of another, a process known as mtDNA capture that appears to be widespread in both plants and animals (Avise 2004). Subsequent spread of clades I and II to Madagascar can be envisioned in association with the historical slave trade and/or more recent commerce between island and mainland.

Complicating the possibility of testing these hypotheses is a highly successful malaria control campaign conducted via indoor residual spraying of insecticides in the highlands of Madagascar beginning in 1949, and extending to the whole island until 1960 (Zahar 1985). The control campaign was reported to have either eliminated *A. funestus* from the central plateau or to have reduced its population below detectable levels in this region. The sudden re-emergence of malaria epidemics in Madagascar in the 1980s signalled the return of *A. funestus* to epidemiologically significant numbers (Fontenille & Rakotoarivony 1988), presumably through recolonization from one or more refugia on the island. Further investigation of the distribution of mtDNA clades I and II within the framework of *A. funestus* population structure, on Madagascar and the mainland, will be facilitated by an mtDNA ND5-based PCR diagnostic assay, whose ability to consistently distinguish the two clades is being evaluated in our laboratories. Among other potential applications of this discovery is the opportunity to identify the source of clade II and to track the extent, direction and rate of its spread beyond Mozambique, allowing inference of aspects of *A. funestus* population structure otherwise beyond our grasp using indirect genetic approaches.

Acknowledgements

We thank A. Spiers, M. Donnelly, C. Curtis, and G. Carrara for providing *A. funestus* specimens from Malawi, Tanzania and Angola. We are grateful to J. Romero-Severson for many helpful discussions and four anonymous reviewers for constructive comments. This project was funded by a grant from the National Institutes of Health R01-AI48842 to NJB. APM and BJS were supported by University of Notre Dame Graduate Fellowships from Arthur J. Schmidt and Clare Boothe Luce, respectively.

Supplementary material

The supplementary material is available from http://www.blackwellpublishing.com/products/journals/suppmat/MEC/MEC2754/MEC2754sm.htm

Table S1 Estimated null allele frequencies

Table S2 Allelic homoplasy in electromorphs of FUNL, AFND19 and AFND40

Fig. S1 Physical map of microsatellite loci in relation to common inversions found throughout Africa (modified from Sharakhov *et al*. 2004).

References

Avise JC (2004) *Molecular Markers, Natural History, and Evolution*, 2nd edn. Sinauer Associates, Sunderland, Massachusetts.

Barnes MJ, Lobo NF, Coulibaly MB *et al.* (2005) SINE insertion polymorphism on the X chromosome differentiates *Anopheles gambiae* molecular forms. *Insect Molecular Biology*, **14**, 353–363.

Braginets OP, Minakawa N, Mbogo CM, Yan G (2003) Population genetic structure of the African malaria mosquito *Anopheles funestus* in Kenya. *American Journal of Tropical Medicine and Hygiene*, **69**, 303–308.

Brain CK, Meester J (1964) Past climatic changes as biological isolating mechanisms in southern Africa. In: *Ecological Studies in Southern Africa* (ed. Davis DHS), pp. 332–340. Dr W. Junk Publishers, The Hague, The Netherlands.

Brookfield JF (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. *Molecular Ecology*, **5**, 453–455.

Brower AV (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences, USA*, **91**, 6491–6495.

Burney DA, Burney LP, Godfrey LR *et al.* (2004) A chronology for late prehistoric Madagascar. *Journal of Human Evolution*, **47**, 25–63.

Cavalli-Sforza LL, Edwards AW (1967) Phylogenetic analysis. Models and estimation procedures. *American Journal of Human Genetics*, **19**, 233–257.

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. *Molecular Ecology*, **9**, 1657–1659.

Coetzee M, Fontenille D (2004) Advances in the study of *Anopheles funestus*, a major vector of malaria in Africa. *Insect Biochemistry and Molecular Biology*, **34**, 599–605.

Cohuet A, Dia I, Simard F, Raymond M, Fontenille D (2004) Population structure of the malaria vector *Anopheles funestus* in Senegal based on microsatellite and cytogenetic data. *Insect Molecular Biology*, **13**, 251–258.

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

Cooke HBS (1964) The pleistocene environment in southern Africa. In: *Ecological Studies in Southern Africa* (ed. Davis DHS), pp. 1–23. Dr W. Junk Publishers, The Hague, The Netherlands.

Costantini C, Sagnon NF, Ilboudo-Sanogo E, Coluzzi M, Boccolini D (1999) Chromosomal and bionomic heterogeneities suggest incipient speciation in *Anopheles funestus* from Burkina Faso. *Parassitologia*, **41**, 595–611.

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

Dia I, Lochouarn L, Boccolini D, Costantini C, Fontenille D (2000) Spatial and temporal variations of the chromosomal inversion polymorphism of *Anopheles funestus* in Senegal. *Parasite*, **7**, 179–184.

Dieringer D, Schlotterer C (2003) MICROSATELLITE ANALYSER (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes*, **3**, 167–169.

Donnelly MJ, Licht MC, 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.

Fontenille D, Rakotoarivony I (1988) Reappearance of *Anopheles funestus* as a malaria vector in the Antananarivo region, Madagascar. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, **82**, 644–645.

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

Garros C, Koekemoer LL, Kamau L *et al.* (2004) Restriction fragment length polymorphism method for the identification of major African and Asian malaria vectors within the *Anopheles funestus* and *An. minimus* groups. *American Journal of Tropical Medicine and Hygiene*, **70**, 260–265.

Gillies MT, Coetzee M (1987) *A Supplement to the Anophelinae of Africa South of the Sahara*. South African Institute for Medical Research, Johannesburg, South Africa.

Gillies MT, De Meillon B (1968) *The Anophelinae of Africa South of the Sahara*, 2nd edn. South African Institute for Medical Research, Johannesburg, South Africa.

Goudet J (2001) FSTAT, *Version 2.9.3, A program to estimate and test gene diversities and fixation indices*. Available from www.unil.ch/izea/softwares/fstat.html.

Green CA, Hunt RH (1980) Interpretation of variation in ovarian polytene chromosomes of *Anopheles funestus* Giles, *A. parensis* Gillies, and *A. aruni? Genetica*, **51**, 187–195.

Hume JC, Lyons EJ, Day KP (2003) Human migration, mosquitoes and the evolution of *Plasmodium falciparum*. *Trends in Parasitology*, **19**, 144–149.

Joy DA, Feng X, Mu J *et al.* (2003) Early origin and recent expansion of *Plasmodium falciparum*. *Science*, **300**, 318–321.

Kimmel M, Chakraborty R, King JP *et al.* (1998) Signatures of population expansion in microsatellite repeat data. *Genetics*, **148**, 1921–1930.

King JP, Kimmel M, Chakraborty R (2000) A power analysis of microsatellite-based statistics for inferring past population growth. *Molecular Biology and Evolution*, **17**, 1859–1868.

Koekemoer L, Weeto MM, Kamau L, Hunt RH, Coetzee M (2002) A cocktail polymerase chain reaction (PCR) assay to identify members of the *Anopheles funestus* (Diptera: Culicidae) group. *American Journal of Tropical Medicine and Hygiene*, **66**, 804–811.

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

Lehmann T, Hawley WA, Grebert H *et al.* (1999) The Rift Valley complex as a barrier to gene flow for *Anopheles gambiae* in Kenya. *Journal of Heredity*, **90**, 613–621.

Lehmann T, Blackston CR, Besansky NJ *et al.* (2000) The Rift Valley complex as a barrier to gene flow for *Anopheles gambiae* in Kenya: the mtDNA perspective. *Journal of Heredity*, **91**, 165–168.

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

Mank JE, Avise JC (2003) Microsatellite variation and differentiation in North Atlantic eels. *Journal of Heredity*, **94**, 310–314.

Mank JE, Avise JC (2004) Individual organisms as units of analysis: Bayesian-clustering alternatives in population genetics. *Genetical Research*, **84**, 135–143.

Michel AP, Guelbeogo WM, Grushko O *et al.* (2005) Molecular differentiation between chromosomally defined incipient species of *Anopheles funestus*. *Insect Molecular Biology*, **14**, 375–387.

Monaghan MT, Gattolliat JL, Sartori M *et al.* (2005) Trans-oceanic and endemic origins of the small minnow mayflies (Ephemeroptera, Baetidae) of Madagascar. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **272**, 1829–1836.

Mouchet J, Faye O, Juivez J, Manguin S (1996) Drought and malaria retreat in the Sahel, West Africa. *Lancet*, **348**, 1735–1736.

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *Journal of Molecular Evolution*, **19**, 153–170.

Palsson S (2004) Isolation by distance, based on microsatellite data, tested with spatial autocorrelation (SPAIDA) and assignment test (SPASSIGN). *Molecular Ecology Notes*, **4**, 143–145.

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

Rabinowitz PD, Coffin MF, Falvey D (1983) The separation of Madagascar and Africa. *Science*, **220**, 67–69.

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

Raymond M, Rousset F (1995) GENEPOP (version 1.2): a population genetics software for exact tests and ecumenicism. *Journal of Heredity*, **86**, 248–249.

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

Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN: *A software for population genetics data analysis, Version 2.000*. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Switzerland.

Sharakhov I, Braginets O, Grushko O *et al.* (2004) A microsatellite map of the African human malaria vector *Anopheles funestus*. *Journal of Heredity*, **95**, 29–34.

Stump AD, Shoener JA, Costantini C, Sagnon N, Besansky NJ (2005) Sex-linked differentiation between incipient species of *Anopheles gambiae*. *Genetics*, **169**, 1509–1519.

Temu EA, Hunt RH, Coetzee M (2004) Microsatellite DNA polymorphism and heterozygosity in the malaria vector mosquito *Anopheles funestus* (Diptera: Culicidae) in east and southern Africa. *Acta Tropica*, **90**, 39–49.

van Oosterhout C, Hutchinson WF, Wills DPM, Shipley PF (2004) MICROCHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, **4**, 535–538.

WHO (2003) *World Health Report 2003*. World Health Organization, Geneva, Switzerland.

Wright S (1931) Evolution in Mendelian populations. *Genetics*, **16**, 97–159.

Zahar AR (1985) *Vector Bionomics in the Epidemiology and Control of Malaria*. World Health Organization, Geneva, Switzerland.

This study was part of Andrew Michel's PhD research project in the laboratory of NJB; whose group studies evolutionary genetics and genomics of mosquitoes at various levels. GLG, MC, NE, DF, JV, TL, N'F and CC study population genetics in relation to vector control in Africa.
