Insect Molecular Biology (2005) **14**(4), 375–387

Molecular differentiation between chromosomally defined incipient species of *Anopheles funestus*

A. P. Michel*, W. M. Guelbeogo†‡, O. Grushko*, B. J. Schemerhorn*§, M. Kern*, M. B. Willard*, 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, USA;* †*Université de Ouagadougou, Ouagadougou, Burkina Faso;* ‡*Centre National de Recherche et Formation sur le Paludisme, Ouagadougou, Burkina Faso; and* ¶*Parasitology Unit, Department of Public Health, University of Rome 'La Sapienza', Rome, Italy*

Abstract

Anopheles funestus **Giles is one of the most important vectors of malaria in sub-Saharan Africa. The population structure of this mosquito in Burkina Faso, West Africa based on chromosomal inversion data led to the description of two chromosomal forms, Kiribina and Folonzo. Because both forms co-occur in the same locales yet differ significantly, both in the frequency of inverted arrangements on chromosome arms 3R and 2R and in vectorial capacity, they were hypothesized to be emerging species with at least partial barriers to gene flow. This hypothesis would be strengthened by molecular evidence of differentiation between Kiribina and Folonzo at loci outside chromosomal inversions. We surveyed molecular variation in sympatric populations of the two forms using sequences from the mitochondrial *ND5* gene and genotypes at sixteen microsatellite loci distributed across the genome. Both classes of marker revealed slight but significant differentiation between the two forms (mtDNA $F_{ST} = 0.023$, $P < 0.001$; microsatellite $F_{ST} = 0.004$, $P < 0.001$; $R_{st} = 0.009$, $P = 0.002$). Locus-by-locus analysis of the microsatellite data showed that significant differentiation was not genome-wide, but could be attributed to five loci on chromosome 3R ($F_{ST} = 0.010$, $P < 0.001$; $R_{st} = 0.016$, $P = 0.002$). Importantly, three of these loci are outside of, and in linkage equilibrium with, chromosomal inversions, suggesting that differentiation between chromosomal forms extends beyond the inversions themselves. The slight overall degree of differentiation indicated by both marker classes is likely an underestimate because of recent population expansion inferred for both Folonzo and Kiribina. The molecular evidence from this study is consistent with the hypothesis of incipient speciation between Kiribina and Folonzo.**

Keywords: ***Anopheles funestus*****, chromosomal forms, incipient species, malaria vector, molecular markers, null alleles.**

Introduction

Malaria remains the leading pathogen-specific cause of morbidity and mortality in children from tropical Africa, taking at least one million lives each year (WHO, 2003). The most important vectors of malaria in this region are members of the *Anopheles gambiae* complex (*A. gambiae*, *Anopheles arabiensis*) and *Anopheles funestus*, all highly anthropophilic mosquitoes that are specifically adapted to humans and human habitats for larval development, bloodmeals and shelter. The widespread distribution of these species in contemporary Africa is limited only by desert, saltwater (coastal mangrove), high elevation, and most importantly, dense vegetation of rainforest or humid savannah that shades potential breeding sites (Coluzzi, 1994). Characteristic of all three species are breeding sites often found in association with agriculture and created by such activities as deforestation, irrigation, dam construction and rice cultivation. Because such man-made sites would have been rare before the establishment of large permanent agricultural settlements within the past 10 000 years, adaptations to the anthropogenic environment must be recent (Powell *et al.*, 1999). Rapid adaptation to changing environments implies a genetic flexibility, in evidence even now (della Torre *et al.*, 2002), which can lead to genetic divergence,

doi: 10.1111/j.1365-2583.2005.00568.x

Received 22 November 2004; accepted after revision 8 February 2005. Correspondence: Nora J. Besansky, Department of Biological Sciences, 317 Galvin Life Sciences Bldg, University of Notre Dame, Notre Dame, IN 46556–0369, USA; Tel.: +1 574 631 9321; fax: +1 574 631 3996; e-mail, nbesansk@nd.edu

§Present address: USDA-ARS, Department of Entomology, Purdue University, West Lafayette, IN 47907 USA.

reproductive isolation and speciation (Powell *et al*., 1999). Realistic epidemiological modelling of disease transmission and efficient implementation of new and existing vector control strategies will require a comprehensive understanding of vector population structure and dynamics, which may vary across Africa according to ecoclimatic zone, ongoing anthropogenic modifications to the environment, and cryptic population structure between emerging species (Coluzzi, 1992, 1994; della Torre *et al*., 2002). Until recently, insufficient attention and the paucity of molecular markers hindered progress toward an understanding of *A. funestus* population structure and the processes that shape it, with the majority of data stemming from a small number of chromosomal inversion studies (Green & Hunt, 1980; Lochouarn *et al*., 1998; Dia *et al*., 2000a,b; Kamau *et al*., 2002).

Evidence suggests that polymorphic chromosomal rearrangements play a role in rapid adaptation to heterogeneous environments, and that alternative rearrangements are associated with divergent ecophenotypes and possibly speciation (Coluzzi, 1982; Coluzzi *et al*., 1985; Noor *et al*., 2001; Rieseberg, 2001). If they are under selection, chromosomal inversions (or molecular markers within them) will yield misleading estimates of migration rates between populations. However, the distribution of inversion polymorphism within population samples remains a valuable indicator of population structure potentially due to assortative mating. For example, in some West African populations of *A. gambiae*, a temporally stable situation exists in which the observed proportion of inversion heterozygotes is significantly less than expected under random mating. Long-term genetic and ecological measurements in these areas led to the description of different reproductive units – termed chromosomal forms – defined by non-random associations of chromosomal inversions, characteristic breeding habitats, and significant premating barriers to gene flow (Coluzzi *et al*., 1985; Toure *et al*., 1998; Tripet *et al*., 2001). These chromosomal forms of *A. gambiae* have been considered to represent speciation in its earliest stages (della Torre *et al*., 2002). There are indications that this phenomenon is also occurring within *A. funestus* from West Africa.

Cytogenetic studies of *A. funestus* conducted on samples from southern Africa, Kenya, Nigeria and Senegal have found no compelling evidence for substructure within populations (Green & Hunt, 1980; Lochouarn *et al*., 1998; Dia *et al*., 2000b; Kamau *et al*., 2002). Similarly, available molecular evidence based on microsatellites has supported the view that *A. funestus* populations are panmictic at a local scale (Braginets *et al*., 2003; Cohuet *et al*., 2004; Temu *et al*., 2004). In contrast to this general picture, chromosomal inversion studies of *A. funestus* conducted in Burkina Faso, West Africa revealed evidence for chromosomally defined incipient species with striking parallels to the situation in *A. gambiae* (Costantini *et al*., 1999; Guelbeogo *et al*., 2005). These studies were carried out on more than 6000 karyotyped females collected from nine villages along a south-west-north-east cline of increasing aridity. They included a longitudinal investigation across three consecutive breeding seasons in two of the villages. Regardless of locale, climatic zone or breeding season, each geographical sample showed significant heterokaryotype deficits and linkage disequilibrium between inversions on chromosome arms 2R and 3R, consistent with assortative mating. Hardy–Weinberg and linkage equilibria were re-established under the hypothesis of two strictly sympatric reproductively isolated chromosomal forms differing in levels of inversion polymorphism. According to this hypothesis, one chromosomal form, designated Folonzo, is characterized by high frequencies of inversions 3Ra, 3Rb and 2Ra. The second, Kiribina, is characterized by high frequencies of the corresponding uninverted arrangements. There was also an indication of behavioural differences impacting vectorial capacity, as the Kiribina form was less likely to rest indoors, feed on humans, and carry malaria parasites (Costantini *et al*., 1999). A recent study in Cameroon also demonstrated high levels of chromosomal heterogeneity in *A. funestus* consistent with the coexistence of Kiribina and Folonzo chromosomal forms (Cohuet *et al*., 2005). The fact that this situation has been recorded so far only from Burkina Faso and Cameroon is not without precedent; only in West Africa are *A. gambiae* populations strongly structured based on inversion polymorphism. Precisely the same factors invoked to explain the apparent contrast in population structure between East and West Africa in *A. gambiae* apply to *A. funestus*: there is richer inversion polymorphism (Kamau *et al*., 2002) and greater environmental heterogeneity in West Africa due to anthropogenic modifications and fluctuations in the rain forest belt (Powell *et al*., 1999).

While the case for population structuring in Burkina Faso based on chromosomal inversions appears robust, the question remains whether genetic differentiation extends to molecular markers outside of inversions in a pattern suggestive of incipient speciation. If complete reproductive isolation has evolved between chromosomal forms, genetic divergence should be found genome-wide, given sufficient time since lineage splitting. However, a source of confusion and contention in previous studies that have examined chromosomal forms of *A. funestus* or *A. gambiae* is the assumption that evolutionary divergence in sympatry requires complete reproductive isolation beforehand; this is not the case (Coyne & Orr, 2004). What is required is that natural selection bars from introgression those genes that directly contribute to functional divergence, owing to their role in ecological and behavioural adaptations – a process that leads to reproductive isolation. This view of speciation in the face of gene flow, termed the genic view (Wu, 2001; Wu & Ting, 2004), predicts the pattern of variation actually observed for *A. gambiae* incipient species: not genome-wide

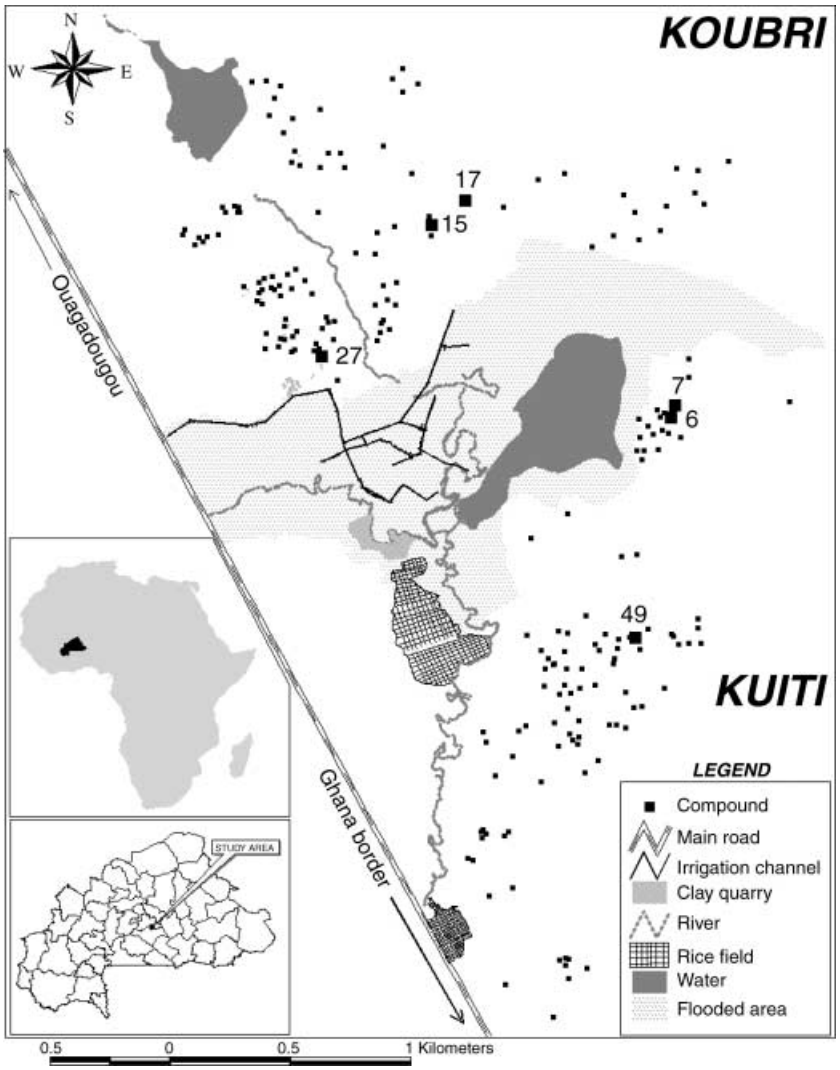

Figure 1. Map of collection sites in Burkina Faso, West Africa.

divergence, but a mosaic genome architecture of differentiated islands in a sea of shared variation (della Torre *et al*., 2002; Barnes *et al*., 2005; Stump *et al*., 2005). Because these islands can be overlooked in a typical coarse and random screen of the genome, even the complete absence of genetic differentiation is inconclusive if other lines of evidence (e.g. chromosomal, behavioural and ecological) suggest persistent heterogeneities between sympatric populations. In an effort to explore the pattern and magnitude of genetic differentiation outside of inversions as it relates to population structure in *A. funestus*, we performed a genome-wide survey of chromosomally defined Folonzo and Kiribina forms from Burkina Faso, using molecular markers that included sixteen microsatellite loci and sequence from the mitochondrial *ND5* gene. Almost 600 karyotyped *A. funestus* were analysed from the same pair of adjacent villages that hosted the longitudinal chromosomal investigation (Guelbeogo *et al*., 2005), in a nested spatial study design that allowed us to discount geographical distance as a source of genetic differentiation.

Results

We made simultaneous collections of the Folonzo and Kiribina chromosomal forms at the smallest possible spatial scale, inside homesteads (compounds) and outside in pit shelters, in two adjacent villages (Fig. 1). As is typical for *A. funestus*, outdoor resting densities were low, and Folonzo was present in markedly lower abundance than Kiribina, particularly in outdoor collections.

Microsatellite polymorphism

From the initial set of eighteen microsatellite loci, two were discarded due to non-specific amplification (AFUB3) or an error in primer synthesis (AFND1). The remaining sixteen loci were amplified from 573 chromosomally identified specimens, comprising 418 Kiribina and 155 Folonzo. Polymorphism was high in all samples (Table 1). Allelic richness, a sample-size-adjusted measure of the number of alleles, averaged 8.2 alleles per locus (range, 3–21). FUNG and FUND, located within inversion 3Rb, were at the upper end of this range;

Table 1. Summary of microsatellite variation at sixteen loci for two chromosomal forms of *Anopheles funestus* in Burkina Faso, West Africa

	Folonzo chromosomal form							Kiribina chromosomal form						
	Koubri compound			Kuiti compound				Koubri compound			Kuiti compound			
Locus	C15 (N = 31)	C17 (N = 28)	C27 (N = 18)	C49 (N = 18)	C6 (N = 28)	C7 (N = 21)	OUT (N = 11)	C15 (N = 76)	C17 (N = 90)	C27 (N = 54)	C49 (N = 57)	C6 (N = 56)	C7 (N = 54)	OUT (N = 31)
AFND12 (X)														
R_s	8.8	9.5	7.9	7.9	7.4	7.6	7.0	10.5	9.5	10.9	9.0	9.8	8.9	9.0
H_o	0.600	0.556	0.778	0.667	0.815	0.571	0.677	0.684	0.633	0.667	0.685	0.517	0.685	0.455
F_{IS}	**0.305**	**0.357**	0.048	0.203	0.018	0.316	0.459	**0.193**	**0.243**	0.222	0.190	**0.367**	0.179	0.186
FUNQ (X)														
R_s	5.7	4.6	5.9	5.9	4.0	5.0	5.0	12.0	9.2	7.9	7.8	6.9	7.9	8.0
H_o	0.452	0.741	0.611	0.722	0.643	0.476	0.742	0.605	0.674	0.722	0.737	0.571	0.698	0.636
F_{IS}	**0.387**	–0.068	0.174	0.058	0.125	0.372	0.097	**0.263**	**0.217**	**0.314**	**0.253**	0.348	**0.271**	0.055
AFND40 (2R)														
R_s	5.8	5.6	4.9	4.9	5.8	4.8	6.0	6.7	8.4	7.9	5.9	7.8	8.9	6.0
H_o	0.714	0.654	0.611	0.889	0.714	0.571	0.613	0.647	0.647	0.685	0.727	0.554	0.704	0.636
F_{IS}	0.038	0.121	0.097	–0.234	0.030	0.191	0.021	0.124	0.128	0.054	–0.011	0.251	0.038	0.124
AFND32 (2R)														
R_s	8.0	7.9	6.8	7.8	8.4	6.9	4.0	10.1	9.9	7.0	9.9	10.6	8.9	9.0
H_o	0.645	0.821	0.444	0.889	0.714	0.571	0.742	0.684	0.700	0.792	0.754	0.679	0.667	0.727
F_{IS}	0.244	–0.004	0.424	0.098	0.236	0.157	0.042	0.147	0.152	0.016	0.070	0.176	0.163	0.054
FUNO (2R)														
R_s	6.5	7.0	9.8	7.0	6.9	7.6	4.0	11.1	11.0	12.8	10.0	11.8	10.9	9.0
H_o	0.645	0.714	0.611	0.722	0.643	0.476	0.645	0.764	0.767	0.759	0.825	0.804	0.759	0.364
F_{IS}	0.100	0.039	0.253	–0.169	0.318	0.145	0.512	0.023	0.012	–0.023	–0.078	–0.041	–0.019	0.124
AFUB10 (2L)														
R_s	6.9	6.4	5.9	7.9	6.5	6.8	7.0	9.8	8.2	8.0	10.0	8.9	7.0	9.0
H_o	0.645	0.750	0.833	0.722	0.679	0.905	0.710	0.658	0.648	0.722	0.789	0.709	0.685	0.455
F_{IS}	0.202	0.006	–0.056	0.160	0.157	–0.087	0.394	0.153	**0.189**	0.068	0.051	0.121	0.156	0.087
FUNL (2L)														
R_s	10.7	10.5	8.8	9.8	10.0	11.0	5.0	12.2	12.6	10.9	14.6	13.9	15.0	7.0
H_o	0.742	0.679	0.667	0.722	0.714	0.714	0.581	0.776	0.633	0.618	0.702	0.661	0.734	0.727
F_{IS}	0.099	0.188	0.109	0.116	0.144	0.161	0.018	0.075	**0.267**	**0.270**	0.163	**0.234**	0.134	0.238
AFND23 (2L)														
R_s	8.6	8.9	7.0	9.8	8.8	7.6	8.0	9.9	9.6	8.9	8.0	9.9	9.0	8.0
H_o	0.679	0.720	0.667	0.886	0.714	0.667	0.677	0.731	0.854	0.855	0.717	0.714	0.774	0.909
F_{IS}	0.197	0.146	0.174	–0.129	0.155	0.213	–0.042	0.143	–0.010	–0.012	0.092	0.166	0.093	0.212
AFUB11 (2L)														
R_s	6.5	7.1	6.9	5.0	5.4	7.4	5.0	8.5	10.2	10.8	8.9	7.0	10.8	9.0
H_o	0.613	0.788	0.667	0.667	0.571	0.714	0.548	0.600	0.605	0.6363	0.589	0.538	0.667	0.545
F_{IS}	0.101	–0.062	–0.046	0.073	0.086	0.080	0.055	0.129	0.108	0.111	0.058	0.225	0.016	0.231
AFND20 (3R)														
R_s	8.1	5.6	7.9	8.8	6.9	7.2	8.0	9.4	8.7	8.9	8.0	9.9	9.9	8.0
H_o	0.871	0.630	0.556	0.889	0.714	0.714	0.581	0.813	0.681	0.618	0.654	0.636	0.704	0.727
F_{IS}	–0.052	0.184	0.344	–0.077	0.067	0.080	0.171	–0.032	0.137	0.197	0.175	0.212	0.128	0.220
AFND7 (3R)														
R_s	6.4	6.5	5.9	7.9	6.2	6.6	7.0	7.8	7.8	8.0	7.8	8.0	8.0	6.0
H_o	0.677	0.643	0.944	0.556	0.643	0.667	0.419	0.605	0.674	0.722	0.737	0.571	0.698	0.818
F_{IS}	0.155	0.216	–0.180	0.319	0.205	0.138	–0.059	0.247	0.181	0.125	0.073	0.273	0.160	**0.459**
AFND19 (3R)														
R_s	8.3	8.5	8.0	8.9	9.7	9.4	9.0	10.6	11.8	11.8	11.8	11.8	9.0	11.0
H_o	0.742	0.786	0.833	0.778	0.704	0.857	0.742	0.707	0.711	0.727	0.750	0.750	0.722	0.909
F_{IS}	0.050	–0.054	–0.008	–0.028	0.194	–0.010	–0.117	0.066	0.061	0.068	0.012	0.043	0.008	0.087
FUNG (3R)														
R_s	10.1	12.0	11.7	8.9	11.2	8.6	10.0	11.9	12.3	12.0	12.5	12.8	13.8	9.0
H_o	0.828	0.846	0.722	0.778	0.821	0.857	0.871	0.724	0.730	0.774	0.825	0.804	0.889	0.909
F_{IS}	0.063	0.054	0.201	0.120	0.071	0.015	–0.010	0.135	0.133	0.109	0.003	0.058	–0.046	–0.061
FUND (3R)														
R_s	12.9	14.6	13.6	12.6	15.8	16.2	10.0	20.6	20.9	16.9	19.5	19.7	18.8	13.0
H_o	0.548	0.654	0.722	0.722	0.630	0.810	0.677	0.750	0.789	0.648	0.684	0.709	0.685	0.455
F_{IS}	**0.396**	**0.293**	0.205	0.171	**0.310**	0.128	**0.479**	**0.161**	0.118	**0.237**	**0.240**	**0.208**	0.188	0.207
AFUB12 (3L)														
R_s	4.3	3.8	3.0	4.0	3.6	4.0	2.0	5.0	5.0	4.0	5.0	5.0	5.0	4.0
H_o	0.323	0.357	0.388	0.444	0.393	0.429	0.452	0.355	0.450	0.352	0.600	0.304	0.302	0.518
F_{IS}	**0.528**	0.344	0.168	0.364	0.282	0.089	0.762	**0.396**	0.178	**0.412**	0.062	**0.439**	**0.488**	0.128
FUNF (3L)														
R_s	6.5	5.8	7.8	6.9	5.8	5.8	7.0	7.0	9.4	5.0	7.0	6.9	7.9	6.0
H_o	0.645	0.571	0.778	0.889	0.750	0.857	0.839	0.730	0.586	0.667	0.815	0.655	0.811	0.818
F_{IS}	0.216	0.247	0.052	–0.108	0.017	–0.118	–0.065	0.096	**0.283**	0.108	–0.060	0.156	–0.061	–0.103

Samples were collected indoors from compounds (C) or outdoors from pit shelters (OUT); sample sizes are given in parentheses. R_s, allelic richness; H_o, observed heterozygosity; F_{IS}, inbreeding coefficient. Significant F_{IS} values ($P < 0.05$ after Bonferroni correction) are given in bold.

at the lower end was trinucleotide repeat AFUB12. Average allelic richness was significantly higher in Kiribina than Folonzo (9.7 vs. 7.5; Wilcoxon signed rank test, $N = 16$, $P < 0.001$).

Observed heterozygosity values, ranging from 0.30 to 0.91, were relatively high at most loci (Table 1). Deviations from Hardy–Weinberg equilibrium, as measured by the inbreeding coefficient F_{IS}, were detected within both Kiribina and Folonzo chromosomal forms. Significantly positive F_{IS} values in 21 out of 112 tests for Kiribina and 8 out of 112 tests for Folonzo after Bonferroni corrections indicated heterozygote deficiency. Heterozygote deficiency is consistent with biological processes, such as inbreeding or the Wahlund effect, as well as technical factors: non-amplifying (null) alleles that cause true heterozygotes to be scored as homozygotes. Inbreeding and a Wahlund effect due to population substructure should produce genome-wide heterozygote deficits and linkage disequilibrium. Yet, most of the significant heterozygote deficits were measured only at AFND12, FUNQ, FUND and AFUB12, and no significant linkage disequilibrium was detected between any pair of loci. It is therefore likely that positive values of F_{IS} were caused by null alleles, as commonly observed for microsatellites (Dakin & Avise, 2004). Under this assumption, we estimated null allele frequencies based on the observed heterozygote deficits (Brookfield, 1996). Suspected null allele frequencies averaged 0.08 in both forms (range, 0.01–0.16; Table 2), in line with frequencies reported from other studies (Dakin & Avise, 2004; Stump *et al.*, 2005). Because of the possibility that null alleles could impact our population genetic analyses, we performed these analyses both before and after adjusting allele and genotype frequencies based on the estimated null allele frequency. Additionally, we performed analyses before and after excluding the four loci displaying the largest heterozygote deficits (highest suspected null allele frequencies). Neither of these treatments significantly altered the magnitude or the statistical significance of the parameters estimated (see below). Accordingly, we considered that null alleles did not introduce any bias of consequence into the population genetic analyses.

Table 2. Estimated null allele frequencies within Kiribina and Folonzo chromosomal forms of *Anopheles funestus* (averaged among compounds and outdoor collection)

Locus	Folonzo	Kiribina
AFND12 (X)	0.133	0.125
FUNQ (X)	0.058	0.137
AFND40 (2R)	0.083	0.099
AFND32 (2R)	0.077	0.066
FUNO (2R)	0.071	0.012
AFUB10 (2L)	0.045	0.077
FUNL (2L)	0.049	0.091
AFND23 (2L)	0.110	0.125
AFUB11 (2L)	0.017	0.102
AFND20 (3R)	0.057	0.095
AFND19 (3R)	0.058	0.092
AFND7 (3R)	0.028	0.033
FUNG (3R)	0.074	0.061
FUND (3R)	0.137	0.107
AFUB12 (3L)	0.102	0.157
FUNF(3L)	0.033	0.110

mtDNA polymorphism

An 834 bp sequence was determined from the mtDNA *ND5* gene of chromosomally identified specimens, forty-seven Kiribina and twenty-three Folonzo. Although it has not been used previously in studies of *A. funestus*, this gene was chosen because of its utility in population surveys of the malaria vectors *A. gambiae* and *A. arabiensis* across Africa (Besansky *et al.*, 1997). The fact that electropherograms were unambiguous and that most substitutions were silent suggests that the sequences were not derived from nuclear-integrated copies. Table 3 summarizes the polymorphism statistics. The haplotype diversity was very high; the number of haplotypes observed in a sample nearly approached the sample size. However, the haplotypes were closely related. The average number of pair-wise differences (π) for Kiribina was 0.43% per site, or 3.6 nt per sequence. Corresponding values for Folonzo were slightly higher, 0.63% and 5.3 nt. Singleton sites (sampled only once) were the most common type of polymorphic site; compared to π-values, estimates of the population mutation parameter $4N_e\mu$ based on the number of segregating sites (θ) were significantly higher (see below).

Population structure inferred from microsatellite variation

Flight range for *A. funestus* is reported as up to 7 km (Gillies & De Meillon, 1968), but the physical distance corresponding to the smallest level of population structure (the deme) has

Table 3. Polymorphism summary statistics for mtDNA *ND5* of *Anopheles funestus* chromosomal forms

Chromosomal form	N	len	S	S_s	N_s	N_a	h	H_d	π	θ
Kiribina (all)	47	834	41	23	38	3	32	0.969 ± 0.015	0.0043 ± 0.0005	0.0111 ± 0.0035
Compound2	19	834	24	18	22	2	17	0.988 ± 0.021	0.0039 ± 0.0005	0.0082 ± 0.0032
Compound7	14	834	16	12	14	2	11	0.956 ± 0.045	0.0034 ± 0.0006	0.0060 ± 0.0026
Compound12	14	834	21	12	21	0	10	0.945 ± 0.045	0.0056 ± 0.0010	0.0079 ± 0.0033
Folonzo (all)	23	834	40	28	40	2	21	0.992 ± 0.015	0.0063 ± 0.0012	0.0130 ± 0.0046

N, sample size; len, length in bp; S, number of segregating sites; S_s, number of singleton segregating sites; N_s and N_a, number of synonymous and replacement changes, respectively; h, number of haplotypes; H_d, haplotype diversity; π (pi, per site), and θ (theta, per site).

Table 4. Microsatellite differentiation between *Anopheles funestus* Kiribina and Folonzo forms

Locus	F_{ST}	*P*-value
AFND12 (X)	0.003	0.09
FUNQ (X)	0	0.39
AFND40 (2R)	0	0.25
AFND32 (2R)	0.002	0.12
FUNO (2R)	0	0.51
AFUB10 (2L)	0	0.43
FUNL (2L)	0.001	0.29
AFND23 (2L)	0.001	0.32
AFUB11 (2L)	0	0.81
AFND20 (3R)	**0.012**	**< 0.001**
AFND19 (3R)	**0.006**	**0.007**
AFND7 (3R)	**0.009**	**< 0.001**
FUNG (3R)	**0.010**	**< 0.001**
FUND (3R)	**0.010**	**< 0.001**
AFUB12 (3L)	0.004	0.10
FUNF(3L)	0	0.67
All loci (16 loci)	**0.004**	**< 0.001**
Excluding 3R (11 loci)	0.001	0.15
Only 3R (5 loci)	**0.010**	**< 0.001**

Significant *P*-values are given in bold.

not been well-studied. Therefore, taking the most conservative approach possible, we began by testing for population structure within each chromosomal form using a nested spatial design. Pair-wise comparisons were performed between samples from three compounds in each village. Multilocus F_{ST} values were not significantly different from zero between compounds in the same village for either chromosomal form (data not shown). Therefore, indoor samples of each form were pooled by village and compared (Folonzo form: 77 mosquitoes in Koubri village vs. 67 in Kuiti village; Kiribina form: 220 in Koubri village vs. 167 in Kuiti village). Again, multilocus F_{ST} values were not significant (data not shown). Thus, at this small geographical scale (< 2 km), no population structure was detected within either the Kiribina or Folonzo chromosomal form. Therefore, all 387 (220 + 167) indoor samples of the Kiribina form were compared to all 144 (77 + 67) indoor samples of the Folonzo form. The F_{ST} value across all sixteen loci was very small, but highly significant (0.004, $P < 0.001$; Table 4). Using the same nested approach for Rst statistics, the outcome was the same; only the comparison between chromosomal forms was significant ($R_{st} = 0.009$, $P = 0.002$). However, examination of F_{ST} values for individual loci revealed that this was not a genome-wide effect (Table 4). Only five loci had values significantly different from zero, and all five are located on chromosome 3R either between inversions 3Ra and 3Rb (AFND19, AFND7 and AFND20) or within inversion 3Rb (FUNG and FUND) (Fig. 2). Re-analysis of differentiation excluding these five markers, followed by an analysis based only on these markers, showed that all significant differentiation could be explained by loci on 3R (Table 4).

It is unlikely that this result is a consequence of null alleles. We adjusted all observed allele and genotype frequencies based on the estimated frequency of null alleles (Table 2), and repeated the analysis on this adjusted data set. F_{ST} values for loci on 3R remained significant, and all other values were zero except for loci FUNF ($F_{ST} = 0.006$, $P < 0.01$) and AFUB11 ($F_{ST} = 0.007$, $P < 0.01$), which had the largest discrepancy in estimated null allele frequencies between the two forms.

Population structure inferred from mtDNA variation: In the total set of seventy sequences, there were sixty-one polymorphisms segregating at fifty-nine sites. Between Folonzo and Kiribina chromosomal forms, there were no fixed differences and twenty-one shared polymorphisms. The remaining polymorphisms were segregating only within one form. Beyond sharing individual mutations, Folonzo and Kiribina shared two haplotypes (Fig. 3).

We estimated genealogical relationships among all mtDNA sequences using a network approach that allows for multifurcations and for the persistence of ancestral haplotypes in the present-day population (Posada & Crandall, 2001). The shape of the network resembled a star phylogeny, with sequence variants from both Folonzo and Kiribina radiating from a single common ancestor (Fig. 3). There was no clear-cut separation of these forms based on visual inspection of the network. Aside from its basic star shape, the most obvious feature of the network was the relatively large number of missing (unobserved) haplotypes inferred to connect those that were sampled. Indeed, one haplotype (Folonzo, GenBank #AY727743) could not be connected to the network within the 95% limits of parsimony (Clement *et al*., 2000 and references therein).

We tested the null hypothesis that Folonzo and Kiribina are not genetically different, using test statistics for DNA sequences based on the number of nucleotide differences or on haplotype frequencies (Hudson *et al*., 1992a,b; Hudson, 2000) (Table 5). The null hypothesis was rejected by all nucleotide difference statistics. The haplotype statistic H_s was marginally significant, but because haplotype diversity was so high (> 0.97), this statistic is expected to lack power (Hudson *et al*., 1992a).

Evidence for past population growth

The neutral equilibrium assumptions of the usual infinite island model of population structure include absence of

Table 5. mtDNA differentiation between *Anopheles funestus* Kiribina and Folonzo forms

H_{st} (*P*-value)	K^*_{st} (*P*-value)	Z^*_s (*P*-value)	S_{nn} (*P*-value)	F_{ST} (*P*-value)
0.0046 (0.065)	0.0087 (0.013)	6.787 (0.023)	0.655 (0.027)	0.023 (0.009)

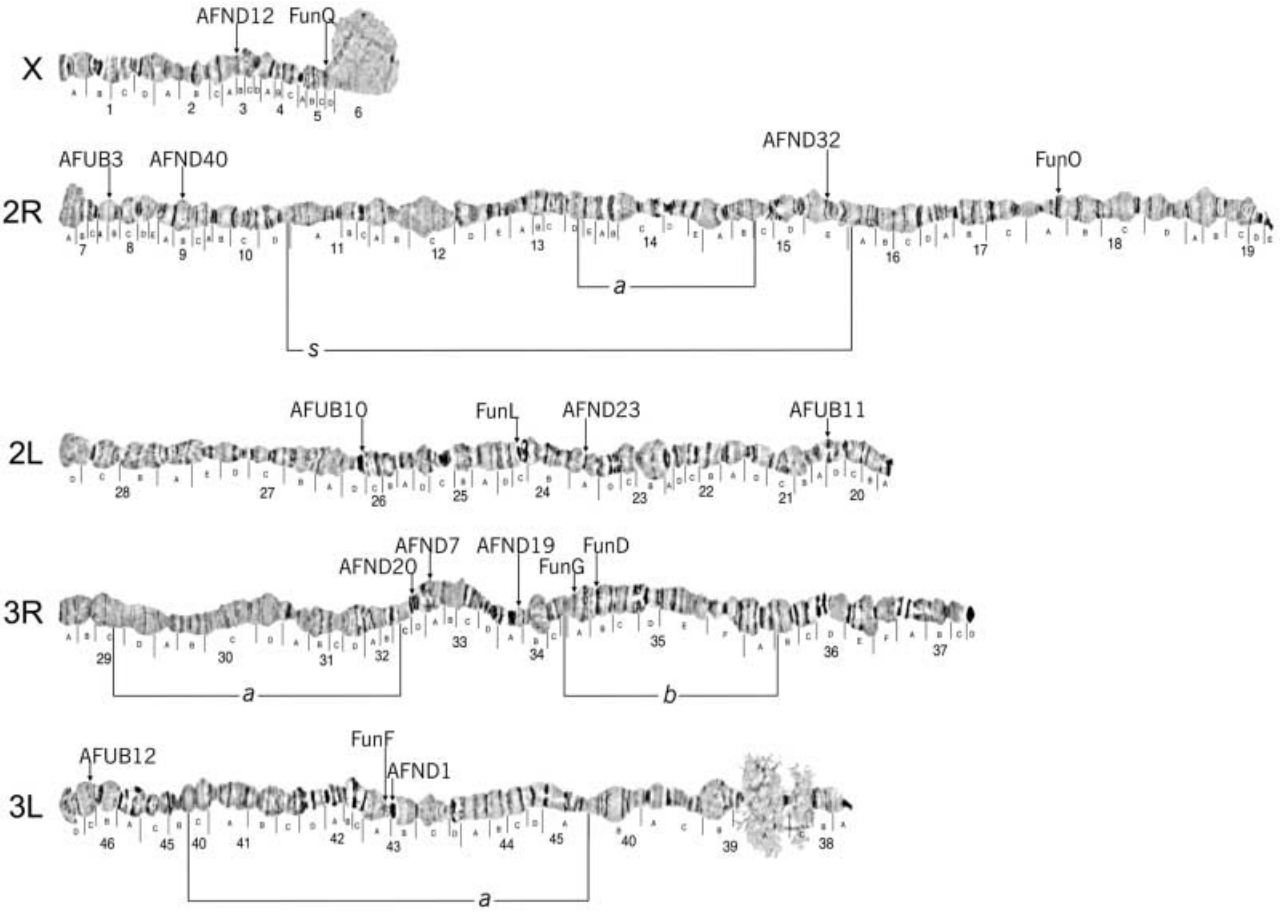

Figure 2. Physical map of microsatellite loci in relation to chromosomal inversions common in Burkina Faso (modified from Sharakhov *et al.*, 2004).

selection, equilibrium with respect to mutation, migration and genetic drift and constant population size. Violations of these assumptions have important consequences for indirect estimates of gene flow. Tests applied to mtDNA or microsatellite data indicated significant deviation from this model for both Kiribina and Folonzo, in the direction of a recent population expansion (Table 6).

For mtDNA, three different classes of tests were performed on Folonzo and Kiribina samples to detect population growth. These tests are based on the frequencies of segregating sites (Tajima's *D*, Fu and Li's *D** and *F**, and R_2), the haplotype distribution (Fu's F_s) or the mismatch distribution (the raggedness statistic *rg*) (Tajima, 1989; Fu & Li, 1993; Harpending, 1994; Fu, 1997; Ramos-Onsins & Rozas, 2002). These tests are premised on the expectation that in an expanded population, mutations are more likely to be recent. Therefore segregating sites occur predominantly in external branches of a genealogy; these sites – and corresponding sequence haplotypes – are more likely to be singletons, and a histogram of the number of pairwise sequence differences (mismatch distribution) is more likely to be smooth. All tests rejected the null hypothesis of

Table 6. Tests of population expansion in *Anopheles funestus*

Test	Kiribina	Folonzo
mtDNA		
Tajima's *D*	−2.120**	−2.087**
Fu & Li's *D**	−2.721*	−2.492*
Fu & Li's *F**	−2.983*	−2.771*
R_2	0.037***	0.052***
Fu's Fs	−32.266***	−16.631***
rg	0.031^{ns}	0.016^{ns}
Microsatellites		
Homozygosity†		
IAM	$0/16^{ns}$	$0/16^{ns}$
SMM	15/16**	16/16**
TPM-90%	14/16**	12/16**
TPM-80%	13/16*	8/16
TPM-70%	$8/16^{ns}$	$6/16^{ns}$
k‡	$0/16^{ns}$	$0/16^{ns}$
g	0.71^{ns}	1.32^{ns}
β§ (95% CI)	0.67** (0.49–0.84)	0.69** (0.45–0.90)

**P* < 0.05, ***P* < 0.01, ****P* < 0.001.
†number of loci out of total showing significant heterozygosity deficit by a Wilcoxon sign-rank test.
‡number of loci out of total with positive *k*-value.
§Significance of β-value is based on 95% confidence interval of bootstrapped β-values.

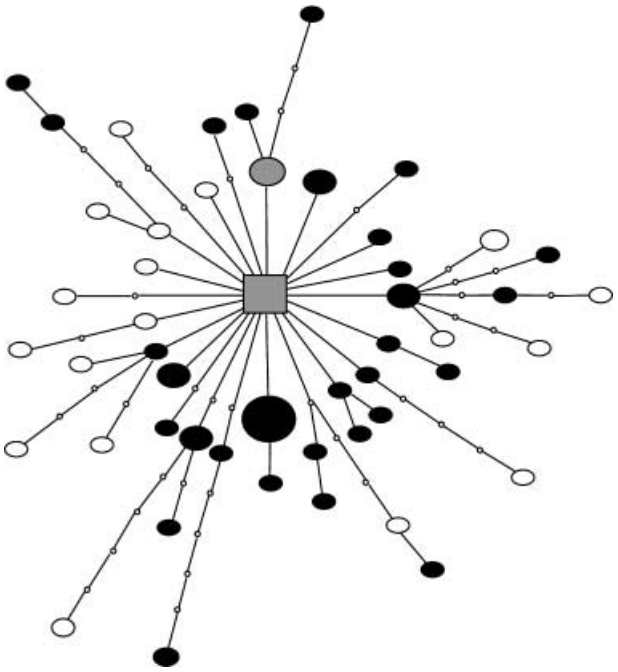

Figure 3. Haplotype network of mtDNA *ND5* sequences. Haplotypes are represented by ovals whose dimensions are scaled to sample size; circles represent inferred haplotypes that were not sampled. The haplotype displayed as a square is inferred as ancestral. Colours indicate source of haplotype: black, Kiribina; white, Folonzo; grey, both. Connecting lines represent a single mutational step.

constant population size except *rg* (Table 6), which has been shown to have little power against population growth (Ramos-Onsins & Rozas, 2002).

For microsatellites, we calculated four statistics designed to detect population expansion, based on: allelic diversity vs. heterozygosity (homozygosity test, Cornuet & Luikart, 1996), between-locus variance in allele size (*g* statistic, Reich & Goldstein, 1998), within-locus shape of the allele size distribution (*k* statistic, Reich & Goldstein, 1998), and variance in allele size vs. heterozygosity (imbalance index, Kimmel *et al*., 1998). Similar to the nucleotide statistics, these tests assume that in an expanded population most mutations are recent, and because of the nature of microsatellites, most recent mutations only differ in size by single repeat units. Results are given in Table 6. The homozygosity test was performed under three different models of microsatellite evolution: the infinite allele model (IAM), the step-wise mutation model (SMM) and the two-phase mutation model (TPM) given that one-step mutations comprise 90, 80 or 70% of the total. A signature of expansion – significantly higher heterozygosity based on allelic diversity – was detected for Kiribina and Folonzo under the most realistic models, the SMM and the TPM-90% (Ellegren, 2000). Neither the *g* statistic nor the *k* statistic detected the respective reduction in observed variance or unimodal allele size distribution expected for expanded populations. However, the β imbalance index was significantly < 1 for Kiribina and Folonzo. This indicates that allele size variance was less than heterozygosity, as expected after population expansion. Because β is the most powerful statistic for inferring past population growth (King *et al*., 2000), this result should carry most weight.

Taken together, the results of several tests which examine different aspects of both microsatellite and mtDNA data suggest that Kiribina and Folonzo populations are not at neutral equilibrium, due to population growth.

Discussion

At the microspatial scale of this study, there was no evidence for population structure within either of the two sympatric chromosomal forms of *A. funestus*. However, between the Folonzo and Kiribina forms slight but highly significant differentiation was detected using two different marker systems, mtDNA and multiple independent microsatellite loci distributed genome-wide. Thus, the genetic structure is due to inherent differences between the two forms and not to physical distance. These molecular data are consistent with the hypothesis initially proposed based on chromosomal inversion evidence (Costantini *et al*., 1999), that Folonzo and Kiribina are incipient species in Burkina Faso.

The population genetic factors responsible for the inherent differences between Folonzo and Kiribina are apparently complex. Across the set of sixteen microsatellite loci, there was significant differentiation. However, locus-by-locus analysis revealed that genetic differentiation was not genome-wide, nor was it randomly distributed. Instead, it derived from five loci on 3R, the arm responsible for the main chromosomal inversion differences between Folonzo and Kiribina. This pattern strongly suggests that selection is responsible, consistent with an adaptive role of inversions. Importantly, only two of these loci were inside inversion 3Rb (FunG, FunD). The remaining three (AFND7, AFND19, AFND20) were outside both inversions on 3R, and in linkage equilibrium with them (data not shown). Thus, while genetic divergence based on nuclear markers is not genome-wide, it extends beyond the inversions themselves. Moreover, selection alone may be insufficient to explain the data, given divergence based on a cytoplasmic marker. The slight but significant mtDNA differentiation is due to frequency differences rather than private mutations, and is most simply explained by stochastic factors arising through genetic drift or experimental sampling. In our samples, the Kiribina form had lower haplotype diversity than Folonzo, and thus had a greater number of identical sequences (Table 3, Fig. 3). After collapsing the entire mtDNA data set down to only one copy of each haplotype and repeating estimates of differentiation between Kiribina and Folonzo, the F_{ST} value decreased from 0.023 to 0.012 and was no longer significant ($P = 0.09$). The present data

do not allow us to firmly exclude experimental sampling as the cause of mtDNA differentiation. However, mtDNA data from an additional 190 Kiribina and 302 Folonzo specimens sampled from a larger region of Burkina Faso confirmed significant differentiation between but not within forms (A.P. Michel, W.M. Guelbeogo, O. Grushko, M.B. Willard, N'F. Sagnon, C. Costantini, N.J. Besansky, unpubl. data), suggesting that interform differentiation is not an artefact of experimental sampling. If the difference in the frequency of mtDNA haplotypes between Kiribina and Folonzo indeed reflects underlying population structure, it is tempting to speculate that the greater sensitivity of mtDNA as compared to nuclear microsatellite loci (outside chromosome 3R) is due to lower 'swarm-fidelity' of males. That is, interform mating may occur because males occasionally join the 'wrong' mating swarm, while females always choose the correct swarm (not always the correct partner). Improved insight into the nature of Kiribina and Folonzo can be gleaned from integrated genetic and behavioural studies in the field, but such studies depend on a resource currently lacking for this species – additional informative markers.

Recently, Cohuet *et al*. (2005) studied *A. funestus* population structure in Cameroon, using chromosomal inversions and ten microsatellite markers. Similar to results from Burkina Faso (Costantini *et al*., 1999; Guelbeogo *et al*., 2005), they found high levels of chromosomal heterogeneity within or between ten sampling locales. However, in contradiction to the present study, they reported that their microsatellite markers revealed no departure from panmixia within populations and that markers within inversion breakpoints were not differentiated. How can the results of these two studies be reconciled, particularly as seven microsatellite loci were in common (FunO, FunL, FunG, AFND19, AFND20, FunD, FunF)? One possibility is that *A. funestus* population structure is truly different between West and Central Africa. A second possibility is that microsatellite heterogeneity is actually present in the data, but went undetected because only 78 of 583 genotyped mosquitoes were also karyotyped. Accordingly, microsatellite F_{ST} values were estimated based on comparisons between the total *A. funestus* sample from each locale, without knowledge of the chromosomal form composition within or between locales. Based on chromosomal data from the subsample of seventy-eight mosquitoes (fig. 1 and table 1 of Cohuet *et al*., 2005), only one village (Tibati) showed a significant deficit of heterokaryotypes suggestive of the presence in sympatry of both chromosomal forms; other villages showed no deviation from Hardy–Weinberg equilibrium. Villages to the north were mainly monomorphic standard (uninverted) and those to the south were highly polymorphic, suggestive of the exclusive presence of Kiribina or Folonzo, respectively, in different halves of a north–south cline of decreasing aridity, with Tibati at its centre. The mainly allopatric distribution of chromosomal forms in the sampled villages potentially explains the finding of panmixia within locales. If Tibati is the only village representative of an area of sympatry of both chromosomal forms, the isolation by distance phenomenon reported by Cohuet *et al*. (2005) is confounded by the differential distribution of chromosomal forms along this cline. It is therefore premature to conclude that these studies disagree; further investigation is warranted.

Analysis of mtDNA in Folonzo and Kiribina samples revealed a significantly negative Tajima's *D* statistic. This and four out of five other tests exploiting different aspects of the mtDNA data significantly rejected neutral evolution, results that are in theory consistent with either selection or population expansion. We were able to discount selection as the likely cause by applying microsatellite-based tests of population expansion to multiple loci. Two out of four tests rejected the null hypothesis of mutation-drift equilibrium, in the direction of population expansion. Moreover, the null hypothesis was still significantly rejected in favour of population expansion when the two tests were repeated on two modified datasets: excluding the five loci on chromosome 3R, and excluding the four loci with the highest levels of null alleles (data not shown). Thus, as expected for a demographic process, the effect is genome-wide, and is not due to selection on 3R or to null alleles. In this context, it might be significant that *A. funestus* has recently re-colonized some arid Sahel areas in West Africa in the wake of development projects (dams, irrigation systems) after a 20 year absence of this species triggered by severe droughts during the 1970s through early 1990s (Mouchet *et al*., 1996; Konate *et al*., 2001). More generally, there is evidence for population expansion in other synanthropic African vectors such as *A. gambiae* and *A. arabiensis* (Donnelly *et al*., 2001), and as suggested by Donnelly *et al*. (2002), this may be a phenomenon characteristic of all pest species whose demographics are entwined with that of our own species.

The case has been made that indices of differentiation such as F_{ST} do not translate into meaningful rates of gene flow (Nm) if the populations under study are not stationary (Whitlock & McCauley, 1999). If these populations have experienced recent expansion, migration rates will be overestimated by F_{ST}. Looked at another way, differentiation between recently expanded populations will be underestimated relative to neutral equilibrium expectations. Thus, the relatively low level of differentiation measured between Folonzo and Kiribina (mtDNA F_{ST}, 0.023; microsatellite F_{ST}, 0.004) may not reflect the current degree of reproductive isolation between these chromosomal forms. Moreover, such low differentiation makes it difficult to detect significant differences between forms. In the present study, sample sizes were well above 100 when pooled by chromosomal form, which gave statistical power not routinely available in other studies. Under these circumstances, the failure to detect significant microsatellite-based genetic differentiation between chromosomally differentiated *A. funestus*

populations elsewhere in West Africa (e.g. Cohuet *et al.*, 2004) should be interpreted cautiously.

Preliminary data have suggested that differences between Folonzo and Kiribina are not merely academic, but are manifested in behaviours that could impact malaria transmission and vector control measures (Costantini *et al.*, 1999). Thus, although both chromosomal forms were found resting indoors significantly more often than outdoors, this tendency was more pronounced in Folonzo (seen also in the present study), implying that Folonzo has a higher potential for contact with insecticide residual sprays applied indoors, and for contact with humans at night when blood meals are taken. Indeed, Folonzo had a higher human blood index, and was more likely to be infected with *Plasmodium falciparum*, suggesting that Folonzo is the superior vector. In this study, of thirty-six *A. funestus* found infected with *P. falciparum*, we found a higher infection rate in Folonzo ($N = 12$, 7.7% of the indoor Folonzo sample) vs. Kiribina ($N = 24$, 5.7% of the indoor Kiribina sample). Although these results apparently follow the trend noted by Costantini *et al.* (1999), the difference was not statistically significant possibly because of small sample size ($\chi^2 = 0.80$, $P > 0.10$, df = 1). Although we found no significant differences in infection rates, these values convey a similarly high intensity of transmission imposed by both chromosomal forms of *A. funestus* in Burkina Faso. Apparently, these chromosomal forms represent incipient speciation of an already highly efficient vector. Perhaps more importantly, quantitative differences between forms in resting behaviour (based on more extensive data to be presented elsewhere) may account for the variable degree of success at controlling *A. funestus* in West Africa by indoor residual spraying with insecticide. Most notably, lower success was encountered in savannahs like those of Burkina Faso relative to forest areas (reviewed in Zahar, 1985). Similarly, behavioural heterogeneities associated with chromosomal polymorphism in vector populations were invoked to explain the failure to interrupt transmission in the West African savannah under extensive state-of-the-art malaria control programs such as the Garki Project (Molineaux & Grammicia, 1980; Coluzzi, 1982). Thus, a more detailed knowledge of the distribution and bionomics of the *A. funestus* chromosomal forms, as well as of their population structures and residual gene flow, will be relevant to the understanding of their response to environmental changes and to vector control measures.

Experimental procedures

Mosquitoes

The study sites were two villages in Burkina Faso, Koubri (12°11′54 N; 1°23′43 W) and Kuiti (12°11′36 N; 1°23′11 W), which lie ~1 km apart on opposite margins of a permanent swamp (Fig. 1). Because of the relatively low adult resting densities of *A. funestus*, the smallest possible spatial unit for sampling was the family compound, which consists of one to thirteen closely spaced huts enclosed by a fence. Collections were made indoors by insecticide spray-sheet catches in September 1999 (for the mtDNA analysis) and December 2001 (for microsatellite analysis). Outdoor collections were made from pit shelters. Mosquitoes were sorted in the field under a dissecting scope, using morphological keys (Gillies & Coetzee, 1987). Ovaries from females at the appropriate gonotrophic stage were dissected into individually labelled tubes containing Carnoy's solution and stored at −20 °C for later polytene chromosome analysis. Mosquito carcasses were stored individually at room temperature in correspondingly labelled tubes with desiccant.

Chromosome preparation followed Sharakhov *et al.* (2001). Assignment of individual females to the Folonzo or Kiribina form was performed according to Guelbeogo *et al.* (2005). Only those specimens that could be unambiguously assigned were analysed further.

After removal of abdomens for later blood meal analysis (whose results will be presented elsewhere), genomic DNA was extracted from single mosquito carcasses in individual tubes (DNeasy Tissue Kit, Qiagen, Inc, Valencia, CA), or in 96-well plates (Wizard SV-96 Genomic DNA Purification System, Promega Corporation, Madison, WI) with the aid of a Biomek FX workstation (Beckman Coulter, Fullerton, CA). Prior to analysis by PCR, genomic DNA was diluted 1 : 10 in H_2O (~5 ng/μl). Morphological identification of *A. funestus* was verified for each specimen by an rDNA-based PCR assay (Koekemoer *et al.*, 2002), modified with the inclusion of a primer specific to the '*A. rivulorum-like*' taxon (Hackett *et al.*, 2000; Cohuet *et al.*, 2003) and an alternative primer specific for *A. rivulorum* (RL, 5′-TTTGACCCTGATTCAGTAGG and RIV2, 5′-AAGTTCGCAACACGCAGCGT, respectively). Sporozoite infection was assayed by nested PCR targeting the *Msp*I gene (Ranford-Cartwright *et al.*, 1993).

Mitochondrial DNA sequencing and analysis

A 984 bp fragment of the mitochondrial *ND5* gene spanning positions 6950–7933 in the *A. gambiae* reference sequence (GenBank L20934) was PCR amplified in 50 μl reactions containing 20 mM Tris-HCl pH 8.3, 50 mM KCl, 0.2 mM each dNTP, 1.5 mM MgCl, 2.5 U *Taq* polymerase, 1 μl (~5 ng) diluted DNA, and 10 pmol of each primer (6950: 5′-TCCTTTGAATAAAACCCAGC; 7933: 5′-TCAGTGAGAAGTATTAAGCC), in a GeneAmp 9600 thermocycler (Perkin-Elmer). Cycling conditions were: 94 °C for 5 min, thirty-five cycles of 94 °C for 40 s, 58 °C for 44 s, 72 °C for 1 min, and final extension at 72 °C for 10 min. Both strands of the amplified products were directly sequenced on a Beckman-Coulter CEQ8000.

Sequences were aligned, trimmed to a common 834 bp segment, and edited using Lasergene software (DNASTAR, Inc, Madison, WI). Sequences were deposited in GenBank under accession numbers AY727671–AY727744. *DnaSP* 4.0 (Rozas *et al.*, 2003) was used to compute summary statistics for mtDNA polymorphism of Kiribina and Folonzo samples. Tests of neutral evolution: Tajima's *D* (Tajima, 1989), Fu & Li's *D** & *F** (Fu & Li, 1993), R2 (Ramos-Onsins & Rozas, 2002), Fu's Fs (Fu, 1997) and *rg* (Harpending, 1994), also were computed using *DnaSP*. Statistical significance was evaluated by coalescent simulations (10 000 replicates) in *DnaSP*. *Arlequin* 2.0 (Schneider *et al.*, 2000) was used to calculate F_{ST} between samples based on pair-wise nucleotide distance, and to test for significance by 10 000 permutations of haplotypes between samples. Other indices of differentiation, H_{st}, K^*_{st}, Z^*_s and S_{nn} (Hudson *et al.*, 1992a; Hudson, 2000), were computed in *DnaSP* and tested for significance by permutation with 10 000 replicates. A haplotype network was created using the program

TCS v. 1.13 (Clement *et al.*, 2000). Reticulations in the network were resolved following the methods of Crandall & Templeton (1993).

Microsatellite genotyping and analysis

Physically mapped microsatellite loci were chosen from a reference set supplemented by two additional loci located within a key inversion (3Rb) distinguishing Kiribina and Folonzo forms (Sharakhov *et al.*, 2004; Fig. 2). Forward primers for four loci were redesigned either to facilitate pool-plexing of genotyping reactions (FUNG: 5′-GCAAGCAGCTTACTGCACTG; FUND: 5′-TCGCTTCCAATGCACAATAC) or to improve the reliability of PCR amplification (AFND7: 5′-TTCGACTCGGAAGCTTTACC; AFND20: 5′-CGAGAACCACATTAGGGAACA). PCR was performed for each locus individually in 25 μl reactions containing 20 mM Tris-HCl pH 8.3, 50 mM KCl, 0.2 mM dNTPs, 2 mM $MgCl_2$, 15 pmol of each primer, 1.25 U *Taq* polymerase, and 1 μl (~5 ng) diluted DNA. Forward primers were labelled with one of three Beckman-Coulter dyes (Invitrogen, Carlsbad, CA). Cycling conditions in a GeneAmp 9700 thermocycler (Applied Biosystems) were: 94 °C for 5 min, thirty-eight cycles of 94 °C for 20 s, 55 °C for 15 s, 72 °C for 30 s, and a final extension at 72 °C for 10 min. Amplified products were diluted and pool-plexed (three groups of six loci each). Pools consisted of 1 μl aliquots of each of six PCR reactions, 0.5 μl of a 400 bp size standard (Beckman-Coulter) and 30 μl SLS buffer (Beckman-Coulter). Alleles were resolved on a Beckman-Coulter CEQ8000 per manufacturer's instructions, and sized with CEQ8000 software.

For each locus and population sample, a sample-size adjusted measure of number of alleles (allelic richness, Rs) and observed heterozygosity were computed using Fstat 2.9.3.2 (Goudet, 2001). Fstat was also used to assess deviation from Hardy–Weinberg equilibrium at each locus and population sample, as indicated by the inbreeding coefficient F_{IS}, as well as linkage disequilibrium between pairs of loci within each chromosomal form. Significance was tested using the randomization approach implemented in *Fstat*, with Bonferroni-adjusted *P*-values. At the compound level for each chromosomal form, the frequency of suspected null alleles was calculated using the Brookfield 2 estimate (Brookfield, 1996); allele and genotype frequencies were modified accordingly in MICRO-CHECKER (van Oosterhout *et al.*, 2004). The null allele-adjusted data set was used to explore the impact of null alleles on the analysis of genetic differentiation.

Deviation from neutral equilibrium in Folonzo and Kiribina samples was investigated using four tests: the homozygosity test of Cornuet & Luikart as implemented in BOTTLENECK (Cornuet & Luikart, 1996); the *k*-test and *g*-test (Reich & Goldstein, 1998); and Kimmel's β-imbalance index, using the β_1 estimator of β(t) (Kimmel *et al.*, 1998). For the imbalance index, 95% confidence intervals were estimated by bootstrapping over loci, using a SAS program kindly provided by T. Lehmann (Donnelly *et al.*, 2001). Significance of the *k*-test and *g*-test were determined according to (Reich *et al.*, 1999), and that of the homozygosity test by simulations implemented in BOTTLENECK.

Genetic differentiation between samples was measured using F_{ST} as computed in MSA (Dieringer & Schlotterer, 2003) and R_{st} as computed in Arlequin, with significance assessed by 10 000 random permutations.

Acknowledgements

For assistance in the laboratory, we thank B. Harker and M. Ingrasci. For fieldwork, we thank the entomological technicians at CNRFP and the villagers of Koubri and Kuiti. B. Hutton and M. Chrystal helped with programming. Tovi Lehmann kindly provided SAS code for bootstrapping β-values. We thank Jeanne Romero-Severson, Jeff Feder, Hope Hollocher and the Evolutionary Discussion Group at the University of Notre Dame for helpful discussion and advice. 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.

References

Barnes, M.J., Lobo, N.F., Coulibaly, M.B., Sagnon, N'F., Costantini, C. and Besansky, N.J. (2005) SINE insertion polymorphism on the X chromosome differentiates *Anopheles gambiae* molecular forms. *Insect Mol Biol* doi: 10.1111/j.1365-2583.2005.00566.x.

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

Braginets, O.P., Minakawa, N., Mbogo, C.M. and Yan, G. (2003) Population genetic structure of the African malaria mosquito *Anopheles funestus* in Kenya. *Am J Trop Med Hyg* **69**: 303–308.

Brookfield, J.F. (1996) A simple new method for estimating null allele frequency from heterozygote deficiency. *Mol Ecol* **5**: 453–455.

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

Cohuet, A., Simard, F., Toto, J.-C., Kengne, P., Coetzee, M. and Fontenille, D. (2003) Species identification within the *Anopheles funestus* group of malaria vectors in Cameroon and evidence for a new species. *Am J Trop Med Hyg* **69**: 200–205.

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

Cohuet, A., Dia, I., Simard, F., Raymond, M., Rousset, F., Antonio-Nkondjio, C., Awono-Ambene, P., Wondji, C. and Fontenille, D. (2005) Gene flow between chromosomal forms of the malaria vector *Anopheles funestus* in Cameroon, Central Africa, and its relevance in malaria fighting. *Genetics* **169**: 301–311.

Coluzzi, M. (1982) Spatial distribution of chromosomal inversions and speciation in anopheline mosquitoes. In *Mechanisms of Speciation* (Barigozzi, C., ed.), pp.143–153. Alan R. Liss Inc., New York.

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

Coluzzi, M. (1994) Malaria and the Afrotropical ecosystems: impact of man-made environmental changes. *Parassitologia* **36**: 223–227.

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

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

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

Coyne, J.A. and Orr, H.A. (2004) *Speciation*. Sinauer Associates Inc., Sunderland, MA.

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

Dakin, E.E. and Avise, J.C. (2004) Microsatellite null alleles in parentage analysis. *Heredity* **95**: 504–509.

Dia, I., Boccolini, D., Antonio-Nkondjio, C., Costantini, C. and Fontenille, D. (2000a) Chromosomal inversion polymorphism of *Anopheles funestus* from forest villages of South Cameroon. *Parassitologia* **42**: 227–229.

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

Dieringer, D. and Schlotterer, C. (2003) Microsatellite analyzer (MSA): a platform independent analysis tool for large microsatellite data sets. *Mol Ecol Notes* **3**: 167–169.

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

Donnelly, M.J., Simard, F. and Lehmann, T. (2002) Evolutionary studies of malaria vectors. *Trends Parasitol* **18**: 75–80.

Ellegren, H. (2000) Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet* **16**: 551–558.

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

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

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

Gillies, M.T. and De Meillon, B. (1968) *The Anophelinae of Africa South of the Sahara*. South African Institute for Medical Research, Johannesburg.

Goudet, J. (2001) *FSTAT, a Program to Estimate and Test Gene Diversities and Fixation Indices* Version 2.9.3. Available from http://www.unil.ch/izea/softwares/fstat.html

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

Guelbeogo, W.M., Grushko, O., Boccolini, D., Ouédraogo, P.A., Besansky, N.J., Sagnon, N'F. and Costantini, C. (2005) Chromosomal evidence of incipient speciation in the Afrotropical malaria mosquito *Anopheles funestus*. *Med Vet Entomol* in press.

Hackett, B.J., Gimnig, J., Guelbeogo, W., Costantini, C., Koekemoer, L.L., Coetzee, M., Collins, F.H. and Besansky, N.J. (2000) Ribosomal DNA internal transcribed spacer (ITS2) sequences differentiate *Anopheles funestus* and *An. rivulorum*, and uncover a cryptic taxon. *Insect Mol Biol* **9**: 369–374.

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

Hudson, R.R. (2000) A new statistic for detecting genetic differentiation. *Genetics* **155**: 2011–2014.

Hudson, R.R., Boos, D.D. and Kaplan, N.L. (1992a) A statistical test for detecting geographic subdivision. *Mol Biol Evol* **9**: 138–151.

Hudson, R.R., Slatkin, M. and Maddison, W.P. (1992b) Estimation of levels of gene flow from DNA sequence data. *Genetics* **132**: 583–589.

Kamau, L., Hunt, R. and Coetzee, M. (2002) Analysis of the population structure of *Anopheles funestus* (Diptera: Culicidae) from western and coastal Kenya using paracentric chromosomal inversion frequencies. *J Med Entomol* **39**: 78–83.

Kimmel, M., Chakraborty, R., King, J.P., Bamshad, M., Watkins, W.S. and Jorde, L.B. (1998) Signatures of population expansion in microsatellite repeat data. *Genetics* **148**: 1921–1930.

King, J.P., Kimmel, M. and Chakraborty, R. (2000) A power analysis of microsatellite-based statistics for inferring past population growth. *Mol Biol Evol* **17**: 1859–1868.

Koekemoer, L., Weeto, M.M., Kamau, L., Hunt, R.H. and Coetzee, M. (2002) A cocktail polymerase chain reaction (PCR) assay to identify members of the *Anopheles funestus* (Diptera: Culicidae) group. *Am J Trop Med Hyg* **66**: 804–811.

Konate, L., Diop, A., Sy, N., Faye, M.N., Deng, Y., Izri, A., Faye, O. and Mouchet, J. (2001) Comeback of *Anopheles funestus* in Sahelian Senegal. *Lancet* **358**: 336.

Lochouarn, L., Dia, I., Boccolini, D., Coluzzi, M. and Fontenille, D. (1998) Bionomical and cytogenetic heterogeneities of *Anopheles funestus* in Senegal. *Trans R Soc Trop Med Hyg* **92**: 607–612.

Molineaux, L. and Grammicia, G. (1980) *The Garki Project. Research on the Epidemiology and Control of Malaria in the Sudan Savanna of West Africa*. World Health Organization, Geneva, Switzerland.

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

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

van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M. and Shipley, P.F. (2004) Micro-Checker: Software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes* **4**: 535–538.

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

Powell J.R., Petrarca V., della Torre A., Caccone A. and Coluzzi M. (1999) Population structure, speciation, and introgression in the *Anopheles gambiae* complex. *Parassitologia* **41**: 101–113.

Ramos-Onsins, S.E. and Rozas, J. (2002) Statistical properties of new neutrality tests against population growth. *Mol Biol Evol* **19**: 2092–2100.

Ranford-Cartwright, L.C., Balfe, P., Carter, R. and Walliker, D. (1993) Frequency of cross-fertilization in the human malaria parasite *Plasmodium falciparum*. *Parasitology* **107**: 11–18.

Reich, D.E. and Goldstein, D.B. (1998) Genetic evidence for a Paleolithic human population expansion in Africa. *Proc Natl Acad Sci USA* **95**: 8119–8123.

Reich, D.E., Feldman, M.W. and Goldstein, D.B. (1999) Statistical properties of two tests that use multilocus data sets to detect population expansions. *Mol Biol Evol* **16**: 453–466.

Rieseberg, L.H. (2001) Chromosomal rearrangements and speciation. *Trends Ecol Evol* **16**: 351–358.

Rozas, J., Sanchez-Delbarrio, J.C., Messeguer, X. and Rozas, R. (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496–2497.

Schneider, S., Roessli, D. and Excoffier, L. (2000) Arlequin: A software for population genetics data analysis. 2.0 ed.

Sharakhov, I.V., Sharakhova, M.V., Mbogo, C.M., Koekemoer, L.L. and Yan, G. (2001) Linear and spatial organization of polytene chromosomes of the African malaria mosquito *Anopheles funestus*. *Genetics* **159**: 211–218.

Sharakhov, I., Braginets, O., Grushko, O., Cohuet, A., Guelbeogo, W.M., Boccolini, D., Weill, M., Costantini, C., Sagnon, N., Fontenille, D., Yan, G. and Besansky, N.J. (2004) A microsatellite map of the African human malaria vector *Anopheles funestus*. *J Hered* **95**: 29–34.

Stump, A.D., Schoener, J.A., Costantini, C., Sagnon, N'F. and Besansky, N.J. (2005) Sex-linked differentiation between incipient species of *Anopheles gambiae*. *Genetics* **169**: 1509–1519.

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

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

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

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

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

Whitlock, M.C. and Mccauley, D.E. (1999) Indirect measures of gene flow and migration: FST not equal to 1/(4*Nm* + 1). *Heredity* **82**: 117–125.

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

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

Wu C.-I. and Ting C.T. (2004) Genes and speciation. *Nat Rev Genet* **5**: 114–122.

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