

## **Chromosomal Inversions, Natural Selection and Adaptation in the Malaria Vector *Anopheles funestus***

Diego Ayala, Michael C. Fontaine, Anna Cohuet, Didier Fontenille, Renaud Vitalis, Frédéric Simard

► **To cite this version:**

Diego Ayala, Michael C. Fontaine, Anna Cohuet, Didier Fontenille, Renaud Vitalis, et al.. Chromosomal Inversions, Natural Selection and Adaptation in the Malaria Vector *Anopheles funestus*. *Molecular Biology and Evolution*, Oxford University Press (OUP), 2011, 28 (1), pp.745-758. 10.1093/molbev/msq248 . hal-02651247

**HAL Id: hal-02651247**

**<https://hal.inrae.fr/hal-02651247>**

Submitted on 14 Jun 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



# Chromosomal Inversions, Natural Selection and Adaptation in the Malaria Vector *Anopheles funestus*

Diego Ayala,<sup>\*,1</sup> Michael C. Fontaine,<sup>2</sup> Anna Cohuet,<sup>†,1</sup> Didier Fontenille,<sup>1</sup> Renaud Vitalis,<sup>3</sup> and Frédéric Simard<sup>†,1,4</sup>

<sup>1</sup>Institut de Recherche pour le Développement, UR016 CCPV, Montpellier, France

<sup>2</sup>Centre National de la Recherche Scientifique, Université Paris-Sud, UMR 8079, Orsay, France

<sup>3</sup>Centre National de la Recherche Scientifique, Unité Mixte de Recherche Centre de Biologie pour la Gestion des Populations (Institut National de la Recherche Agronomique—Institut de Recherche pour le Développement—Centre de Coopération Internationale en Recherche Agronomique pour le Développement—Montpellier SupAgro), Montpellier sur Lez Cedex, France

<sup>4</sup>Institut de Recherche pour le Développement, Organisation de Coordination pour la Lutte Contre les Endémies en Afrique Centrale, Yaoundé, Cameroon

<sup>†</sup>Present address: Institut de Recherches pour le Développement/Institut de Recherche en Sciences de la Santé—Direction Régionale de l'Ouest, Bobo Dioulasso, Burkina Faso

\*Corresponding author: E-mail: diego.ayala.g@gmail.com.

Associate editor: Manolo Gouy

## Abstract

Chromosomal polymorphisms, such as inversions, are presumably involved in the rapid adaptation of populations to local environmental conditions. Reduced recombination between alternative arrangements in heterozygotes may protect sets of locally adapted genes, promoting ecological divergence and potentially leading to reproductive isolation and speciation. Through a comparative analysis of chromosomal inversions and microsatellite marker polymorphisms, we hereby present biological evidence that strengthens this view in the mosquito *Anopheles funestus* s.s., one of the most important and widespread malaria vectors in Africa. Specimens were collected across a wide range of geographical, ecological, and climatic conditions in Cameroon. We observed a sharp contrast between population structure measured at neutral microsatellite markers and at chromosomal inversions. Microsatellite data detected only a weak signal for population structuring among geographical zones ( $F_{ST} < 0.013$ ,  $P < 0.01$ ). By contrast, strong differentiation among ecological zones was revealed by chromosomal inversions ( $F_{ST} > 0.190$ ,  $P < 0.01$ ). Using standardized estimates of  $F_{ST}$ , we show that inversions behave at odds with neutral expectations strongly suggesting a role of environmental selection in shaping their distribution. We further demonstrate through canonical correspondence analysis that heterogeneity in eco-geographical variables measured at specimen sampling sites explained 89% of chromosomal variance in *A. funestus*. These results are in agreement with a role of chromosomal inversions in ecotypic adaptation in this species. We argue that this widespread mosquito represents an interesting model system for the study of chromosomal speciation mechanisms and should provide ample opportunity for comparative studies on the evolution of reproductive isolation and speciation in major human malaria vectors.

**Key words:** chromosome inversions, *Anopheles funestus*, natural selection, local adaptation, malaria, Cameroon.

## Introduction

Chromosomal polymorphism (such as inversions, robertsonian fusions and fissions, and translocations) has been recognized as a major driving force in local adaptation, speciation processes, and evolution of sex chromosomes (King 1993; Noor et al. 2001; Rieseberg 2001; Hoffmann et al. 2004; van Doorn and Kirkpatrick 2007). The main evolutionary significance of inversions may come from the fact that they reduce recombination in heterozygotes, protecting genomic regions from introgression (Stump et al. 2007; Kulathinal et al. 2009). Chromosomal inversions can favor local adaptation by capturing sets of locally adapted genes, therefore promoting ecological divergence and fostering reproductive isolation between incipient species (Ortiz-Barrientos et al. 2002; Navarro and Barton 2003; Butlin and Roper 2005; Kirkpatrick and Barton 2006). In this sense,

chromosomal inversions have been repeatedly regarded as genetic markers for ecotypic adaptation (Coluzzi et al. 1985; Hoffmann and Willi 2008). Strong and significant correlations between chromosomal inversion frequencies and environmental parameters have been reported in a number of plant and animal species (Coluzzi et al. 1979; Rodriguez-Trelles et al. 1996; Hoffmann et al. 2004). The adaptive significance of chromosomal inversions is further testified by the presence of parallel geographical clines of inversions on different continents, the rapid reestablishment of inversion frequencies after a new habitat is colonized, and documented seasonal shifts in inversion frequencies associated with environmental conditions in fruitflies and anopheline mosquitoes (Coluzzi et al. 1979; Balanya et al. 2003).

Afrotropical anopheline mosquitoes are emerging as suitable biological models for the study of the role of chromosomal inversions in ecological adaptation and speciation

as exemplified by the rapidly growing literature available on this topic in the malaria mosquito *A. gambiae* (Coluzzi et al. 2002; della Torre et al. 2002). A conceptual model for ecological speciation within the *A. gambiae* complex was first proposed by Coluzzi (1982) more than 25 years ago. This “ecotypification” model explored the role of paracentric polymorphic chromosomal inversions in promoting and maintaining ecological divergence between populations in the context of rapidly fluctuating populations subject to recurrent niche expansions and contractions. In this model, sets of locally adapted alleles are selected in peripheral populations as the species expands into marginal habitats during phases of high population growth as typically observed in these mosquito species all through the rainy season. During phases of population contraction, which occur in the dry season, the peripheral populations become geographically and ecologically distinct from the species core. Then, by chance alone, sets of locally adapted alleles may be captured by a chromosomal inversion in peripheral populations, thereby preventing recombination during the next demographic expansion. Once an inversion is established, new locally adapted alleles are expected to accumulate within the inversion, strengthening ecological divergence and promoting assortative mating between ecotypes eventually leading to speciation. Empirical support for the ecotypification model of chromosomal speciation in *A. gambiae* was provided recently (Manoukis et al. 2008; Costantini et al. 2009; Simard et al. 2009).

Beside pure academic interest, a clear understanding of the mechanisms underlying the dynamics of ecological adaptation, ecological niche divergence, and speciation in major disease vectors has relevance for the epidemiology of the diseases being transmitted and has direct implications for vector control (Besansky et al. 2003; White et al. 2007). Together with *A. gambiae*, *A. funestus* is considered a major vector of the deadly malaria parasite *Plasmodium falciparum* in Africa (Gillies and de Meillon 1968). This highly anthropophilic mosquito, widely spread throughout sub-Saharan Africa, populates a wide range of habitats and presents a high level of chromosomal polymorphism (Coetzee and Fontenille 2004). Nonrandom distribution of chromosomal inversions, strong and significant heterozygote deficits, and linkage disequilibrium (LD) between inversions within natural populations of *A. funestus* have been described and interpreted as indicative of environmental adaptation and/or incipient speciation, a situation that is reminiscent of *A. gambiae* (Costantini et al. 1999; Cohuet et al. 2005; Guelbeogo et al. 2005; Michel, Guelbeogo, et al. 2005).

Here, we focus on the vector species *A. funestus* and present original data that strengthen the view that chromosomal inversions are strongly exposed to natural selection in this species. We collected samples of *A. funestus* from different ecological zones in Cameroon, a country in Central Africa with highly heterogeneous landscapes. We first compared the spatial population genetic structure inferred from presumably neutral nuclear microsatellite markers with that inferred from chromosomal inversions. In doing so, we show that the distribution of chromosomal polymorphism is at odds with neutral expectations.

Then, using canonical correspondence analysis (CCA), we relate inversion frequencies to environmental data and elucidate the contribution of a set of eco-geographical variables (EGVs) to the distribution of chromosomal polymorphisms in natural populations of *A. funestus*. Our results strengthen the view that selection due to environmental factors probably is a major architect of the distribution of chromosomal inversion polymorphism in *A. funestus*. However, the shallow population structure revealed by microsatellite markers localized within and outside of the inversions suggests that recombination between chromosomal rearrangements still occurs and allows significant gene flow at least in genomic areas not targeted by natural selection as previously observed in *A. gambiae* (Stump et al. 2007; White et al. 2007, 2009). We highlight the relevance of our findings for comparative genomic studies and discuss their implications for malaria transmission and control.

## Materials and Methods

### Mosquitoes and Sampling Sites

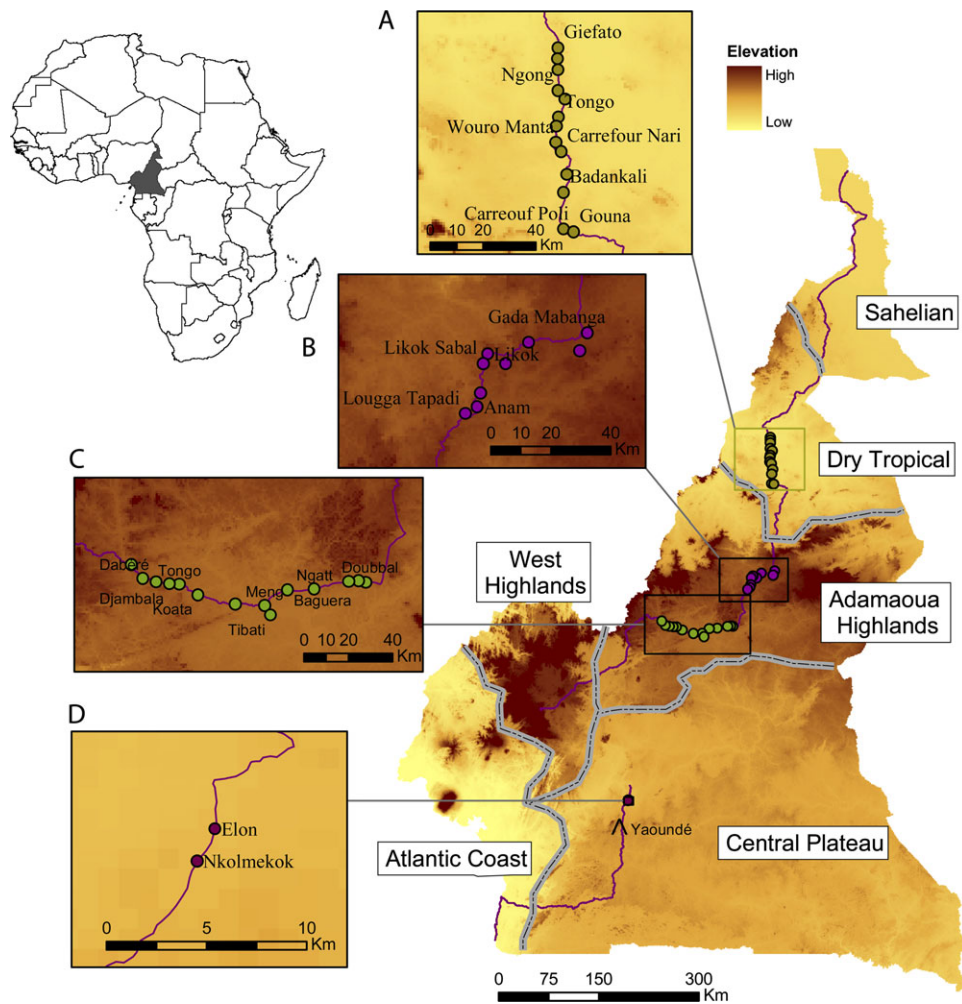
Mosquitoes were sampled in four different ecological zones distributed along a latitudinal transect across Cameroon from August to December 2005, which included a total of 36 villages (fig. 1 and table 1). In each village, mosquito specimens were collected by daytime indoor pyrethroid spraying (Service 1993). Among the 663 *A. funestus* females collected, 313 were at the half-gravid stage. Their ovaries were dissected in the field and placed individually in labeled 1.5 ml tubes containing Carnoy's solution (three parts of 100% ethanol: one part of glacial acetic acid by volume) and stored at  $-20^{\circ}\text{C}$  for cytological analysis (below). Mosquitoes' carcasses were stored individually at room temperature in labeled tubes with desiccant.

### Cytogenetics

Ovaries of half-gravid females were squashed and stained according to standard protocols to reveal the polytene chromosomes (Hunt 1973). The preparations were examined under a phase contrast microscope, and paracentric inversions were scored according to the *A. funestus* chromosome map (Sharakhov et al. 2004). The banding patterns of polytene chromosomes provide an excellent visualization of general chromatin structure, such as inversions or deletions, allowing determination of the three karyotypic states: standard homokaryotypes (encoded as +/+), inverted homokaryotypes (−/−) and heterozygotes for the inversion (+/−). For population genetics analyses, alternative chromosomal arrangements were considered as different alleles at one locus as previously described (Cohuet et al. 2005). Goodness of fit to Hardy–Weinberg expectations, observed ( $H_o$ ) and expected ( $H_e$ ) heterozygosity (Nei 1978, 1987) and LD between inversions were computed using FSTAT 2.9.3.2 (Goudet 2001).

### DNA Extraction and Polymerase Chain Reaction Identification

Genomic DNA was extracted from the body of mosquitoes, which had been successfully karyotyped on all chromosomal



**FIG. 1.** Sampling sites. Topographic Map of Cameroon showing the study area divided into four ecological zones (A–D) and villages sampled in each zone (dots). Dotted lines delimit biogeographic domains as defined in Olivry (1986). Elevation Source: Shuttle Radar Topography Mission, <http://www2.jpl.nasa.gov/srtm>

arms, using the protocol described in Morlais et al. (2004). DNA was then resuspended in sterile water in individual tubes. Morphological identification of *A. funestus* s.s. was confirmed by molecular methods (Koekemoer et al. 2002; Cohuet et al. 2003). Two *A. lesoni* specimens were identified and excluded from the analysis. Finally, a total of 269 *A. funestus* s.s. were used in this study.

### Microsatellite Genotyping and Analysis

Sixteen physically mapped microsatellite loci were selected from published data (Sinkins et al. 2000; Cohuet et al. 2002; Sharakhov et al. 2004; Michel, Guelbeogo, et al.

2005). Microsatellite markers were distributed on all chromosome arms inside and outside common polymorphic chromosomal inversions (supplementary fig. S1, Supplementary Material online). Polymerase chain reaction (PCR) amplifications were performed for each locus individually in a 25  $\mu$ l reaction volume containing 5 $\times$  PCR buffer including 7.5 mM MgCl<sub>2</sub> at pH 8.5 (Promega, France), 200  $\mu$ M of each dNTP (Eurogentec, France), 0.5 units of Taq Polymerase (Promega, France), 10 pmol of each primer, and 5–10 ng of template DNA. The forward primer was 5'-end labeled with PET, VIC, 6FAM, or NED fluorescent dye (Eurogentec, Belgium) to allow for

**Table 1.** Sampling Details.

	Number of Villages	Number of Specimens	Length (km)	Width (km)	Bioclimatic Domain	Elevation (m)	Temp. (°C)	Rainfall (mm)	WVP (%)
Zone A	13	286	70	6	Dry tropical	300	27	1,100	19.9
Zone B	10	117	25	37	Highlands	1,100	22	1,500	17.0
Zone C	12	213	25	106	Highlands	900	23	1,700	20.9
Zone D	2	47	2	2	Rainforest	500	25	1,600	24.3

NOTE.—Bioclimatic domain classification from Olivry (1986); elevation: mean elevation (in meters); rainfall: mean yearly rainfall (in mm); temp.: mean yearly temperature (in °C); WVP: mean yearly water vapor pressure values (in %).

multiplexing prior to electrophoresis. PCR amplifications were carried out in a GeneAmp 9700 thermocycler (Applied Biosystems, France). Cycling conditions were 94 °C for 2 min followed by 36 cycles at 94 °C for 30 s, 54 °C for 30 s, and 72 °C for 30 s and a final extension step at 72 °C for 10 min. PCR products were diluted 1:75—1:100 depending on the fluorescent dye and loaded onto a AB3130sequencer (Applied Biosystems, France) as three multiplex sets. Alleles were scored at each locus using the software GeneMapper v4.0 (Applied Biosystems, France).

Because of low sample sizes of karyotyped females at the village scale (from 2 to 22 specimens per village), mosquitoes were pooled according to their ecological zone of origin to conduct population genetics analyses. Pooling was justified based on 1) the results of a former study of *A. funestus* in Cameroon, which revealed only a weak population structure among villages across the country (Cohuet et al. 2005) and 2) a preliminary population genetics analysis of our data based on a subset of villages with sample sizes above ten specimens that provided no evidence for significant population structuring within each ecological zone. We tested for the presence of null alleles using MICRO-CHECKER software (Van Oosterhout et al. 2004) for all loci. For each microsatellite locus and overall, we estimated  $H_o$ ,  $H_e$ , and allelic richness ( $R_s$ ) (el Mousadik and Petit 1996) using FSTAT 2.9.3.2 (Goudet 2001). Conformity to Hardy–Weinberg equilibrium (HWE) and LD between pairs of loci was tested using FSTAT 2.9.3.2 (Goudet 2001).

Genetic differentiation between ecological zones was examined using exact tests and  $F$ -statistics (Wright 1951) calculated according to Weir and Cockerham (1984) in FSTAT 2.9.3.2 (Goudet 2001). In order to compare levels of differentiation for chromosomal versus microsatellite markers, it is essential to take into account the fundamental differences in the nature and properties of each marker class. Chromosomal inversions are codominant biallelic markers, whereas microsatellite loci are codominant multiallelic markers. They have very different levels of heterozygosity and evolve following distinct mutational models. We used a procedure proposed by Hedrick (2005) and implemented in the RecodeData program (Meirmans 2006) to standardize levels of differentiation according to heterozygosity. This software recodes genetic data by maximizing heterozygosity in the total population, given the observed heterozygosity within subpopulations and derives standardized  $F_{ST}$  estimators, therefore allowing for direct comparisons between estimates obtained from markers with different mutation rates (Meirmans 2006). To compare genetic variation in chromosomal arrangements to that measured with microsatellite markers, we also tested LD between inversions and microsatellite markers using FSTAT 2.9.3.2 (Goudet 2001).

### Homogeneity Across Loci and the Role of Selection

Conformity to expectation under neutral evolution was tested individually for each of the 16 microsatellite loci used in this study through a model-based test directed toward the detection of outlier loci. We used the software

package `FDIST2` (<http://www.rubic.rdg.ac.uk/~mab/software.html>) to identify microsatellite loci showing unusually high  $F_{ST}$  values, taking into account their heterozygosity (Beaumont and Nichols 1996). We ran 500,000 coalescent simulations to characterize the joint distribution of  $F_{ST}$  and heterozygosity, using a 50-demes island model (with 4 sampled demes corresponding to the 4 sampled zones) with an average differentiation equal to that observed across loci ( $F_{ST} = 0.005$ ). Although the true demographic history of *A. funestus* populations is likely to depart from the island model assumptions, the distribution of  $F_{ST}$  estimates so obtained should be robust to the vagaries of demographic history (Beaumont 2005). We used a stepwise mutation model as implemented in `FDIST2`. We obtained a close approximation to the expected joint distribution of the ( $F_{ST}$ ,  $H_e$ ) estimates following (Vitalis et al. 2001), using a 2D histogram of  $100 \times 100$  square cells (see the Appendix in Vitalis et al. 2001). We used the averaged shifted histogram algorithm (Scott 1992) to smooth the simulated data and provide a more continuous region. For each observation ( $F_{ST}$ ,  $H_e$ ) in the sample data set, we computed its associated probability in the simulated data set (i.e., the relative frequency of this observation over the full set of simulated data under the null hypothesis of selective neutrality). Then, we derived the empirical  $P$  value for this observation as one minus the sum of all cell probabilities within the simulated data set that are more probable than the observed data.

Then, we aimed at testing whether chromosomal inversions showed outstanding  $F_{ST}$  values when compared with the microsatellite markers considered here as presumably neutral. Because homoplasy may not be properly accounted for by the above procedure proposed by Hedrick (2005) and Meirmans (2006), we approximated the expected distribution of  $F_{ST}$  conditional upon heterozygosity for chromosomal inversions, given the reference level of differentiation provided by the microsatellite loci. To that end, we modified the software package `DFDIST` (<http://www.rubic.rdg.ac.uk/cgi-bin/MarkBeaumont/dirlist1.cgi>; Beaumont and Balding 2004), which was specifically designed for the analysis of dominant markers (Bonin et al. 2006), in order to simulate codominant biallelic data (code available upon request). We therefore ran 500,000 coalescent simulations of biallelic markers to characterize the joint distribution of  $F_{ST}$  and heterozygosity, using a 50-demes island model (with 4 sampled demes corresponding to the 4 sampled zones) with an average differentiation equal to that observed across microsatellite loci ( $F_{ST} = 0.005$ ). Here, we used the overall heterozygosity of the pooled sample, which makes the conditional density behave better, particularly for biallelic loci. We used the same algorithms as described above to characterize the joint distribution of  $F_{ST}$  and  $H_e$  and to calculate the empirical  $P$  values associated with each studied polymorphism.

### Genetic Cluster Analyses and Isolation by Distance

To infer population structure and to assign specimens to clusters without a priori knowledge of population units, a Bayesian model-based clustering algorithm was used as

implemented in STRUCTURE v2.2 (Pritchard et al. 2000). Individual multilocus genotypic and karyotypic data were used to cluster individuals into  $K$  groups, minimizing HWE and gametic phase disequilibrium between loci within groups (Falush et al. 2003). In STRUCTURE v2.2, the number of distinct genetic clusters in the data set ( $K$ ) was estimated from 1 to 12 by the posterior probability of data under each  $K$ ,  $P(K|X)$  (Pritchard et al. 2000). Each run carried out 500,000 iterations after a burn-in period of 40,000, using the admixture model and correlated allele frequencies. Five replicates for each value of  $K$  were performed to check for the convergence of the Markov Chain Monte Carlo simulations (Fontaine et al. 2007).

Geographic distances can determine patterns of genetic differentiation across populations (Wright 1973). The correlation between genetic differentiation and the logarithm of geographic distances between mosquito collection sites (estimated through Geographical Positioning System, GPS) was tested using the Mantel test available in GENEPOP v4 (Rousset 1997). We used the statistics “ $\hat{e}_r$ ” (Watts et al. 2007), that estimates an analogous of the parameter  $F_{ST}/(1 - F_{ST})$  between pairs of individuals, to compute the regression. Bootstrapped confidence intervals for the slope of the regression (Leblois et al. 2003) were estimated across loci for microsatellites and inversions systems. The inverse of the regression slope provides an estimate of the “neighborhood size,”  $D\sigma^2$ , that is, the product of the effective population density ( $D$ ) by the average squared parent-offspring distance ( $\sigma^2$ ) (Rousset 1997).

### Eco-Geographic Determinants of Chromosomal Inversion Distributions

CCA was carried out to explore the relationships between chromosomal inversions and EGVs using the software CANOCO 4.0 (Ter Braak 1987). CCA shows the main pattern of variation in the distribution of chromosomal inversions as accounted for by the environmental variables and the distribution of the chromosomal inversions along each environmental variable. Thus, this technique allows a quick appraisal of how the distribution of chromosomal inversions varies with the environment. Six different EGVs were considered for each village: latitude (coordinates Universal Transverse Mercator [UTM]), longitude (coordinates UTM), altitude (m), rainfall (mm, yearly means averaged over the last 30 years), water vapor pressure (relative humidity in %, yearly means averaged over the last 30 years), and temperature ( $^{\circ}\text{C}$ , yearly means averaged over the last 30 years). The first three EGVs were obtained directly from the field using portable GPS. The other three EGVs were extracted from the LocClim database developed by the Food and Agriculture Organization ([http://www.fao.org/sd/2002/EN1203a\\_en.htm](http://www.fao.org/sd/2002/EN1203a_en.htm)). Cytogenetic data were encoded as presence (1) or absence (0) of each chromosomal state (standard or inverted) for each inversion. Statistical significance of the canonical axes and environmental variables was tested with a Monte Carlo permutation test using 5,000 permutations, which allowed estimation of the 95% confidence interval around each estimate.

## Results

### Genetic Diversity

The genotypes of 269 *A. funestus* females collected in four distinct ecological zones in Cameroon (fig. 1) were determined at 16 microsatellite loci spread throughout the genome (supplementary fig. S1, Supplementary Material online). Mean allelic richness ( $R_s$ ) was 9.2 alleles per microsatellite locus (table 2) ranging from 4.3 (locus AFUB12) to 18.1 (locus FUND). Mean allelic richness was slightly higher in Zone A ( $R_s = 9.3$ ) as compared with other zones although no significant difference was observed among ecological zones (Wilcoxon signed-rank test,  $P = 0.72$ ). Observed heterozygosity ( $H_o$ ) ranged from 0.28 (AFUB12) to 0.92 (FUNG) and expected heterozygosity ( $H_e$ ) from 0.52 (AFUB12) to 0.94 (FUND). There was significant difference among ecological zones neither in  $H_o$  (Wilcoxon signed-rank test,  $P = 0.71$ ) nor in  $H_e$  (Wilcoxon signed-rank test,  $P = 0.98$ ). No significant LD was found among pairs of microsatellite loci across the country and within zones. Significant heterozygote deficits were detected at some loci in each ecological zone (table 2). Locus FUND consistently showed significant heterozygote deficits across all ecological zones. Only AFND12, FUND, and AFUB12 showed some evidence for the presence of null alleles. However, subsequent analyses did not provide different results with these loci excluded and therefore, the results below are given for the entire set of microsatellite markers (Chapuis and Estoup 2007).

Full karyotypes along the chromosomal complement were scored for all 269 specimens. Seven paracentric inversions out of the 15 polymorphic chromosomal rearrangements known for *A. funestus* (Sharakhov et al. 2004) were found in our samples (table 3). Chromosomal inversions were recorded in all chromosomal arms except in the left arm of chromosome 2 and chromosome X. Inversions frequencies varied from 0.06 (2Rt) to 0.83 (3La). When considering all data together, significant deviation from HWE was detected for all inversion systems (exact test,  $P < 0.001$ ). LD between all pairs of inversion systems was highly significant ( $G$ -test,  $P < 0.001$ ) except between the 2Rh and 2Rt systems. LD between inversions have been reported in a large number of evolutionary contexts (Hoffmann and Rieseberg 2008). Epistatic effects between different inversions, protecting different coadapted genes, may explain LD pattern among inversions.

Although some microsatellite loci mapped within the DNA region encompassed by an inversion (supplementary fig. S1, Supplementary Material online), no significant LD was detected between microsatellite loci and inversion systems within zones and at the country scale ( $G$ -test,  $P > 0.05$ , after Bonferroni correction), which suggests that microsatellites alleles and chromosomal inversions assort independently from each other.

### Population Structure

Genetic differentiation measured across the 16 microsatellite loci was very low among the four ecological zones, with

**Table 2.** Polymorphism at 16 Microsatellite Loci in *Anopheles funestus* Collected in Four Ecological Zones of Cameroon.

Arm	Locus	Zone A <i>n</i> = 114	Zone B <i>n</i> = 37	Zone C <i>n</i> = 83	Zone D <i>n</i> = 35	All <i>n</i> = 269		
X	AFND12	<i>R<sub>s</sub></i>	11.0	8.9	9.1	10.0	14.00	
		<i>F<sub>IS</sub></i>	0.25***	0.27**	0.35***	0.39***	0.30***	
		<i>H<sub>e</sub></i>	0.85	0.86	0.81	0.86	0.84	
		<i>H<sub>o</sub></i>	0.64	0.63	0.53	0.53	0.59	
	FunQ	<i>R<sub>s</sub></i>	6.8	6.8	6.7	4.0	11.00	
		<i>F<sub>IS</sub></i>	0.26***	0.14	0.37***	0.28*	0.29**	
		<i>H<sub>e</sub></i>	0.77	0.69	0.76	0.70	0.75	
		<i>H<sub>o</sub></i>	0.57	0.59	0.48	0.50	0.53	
	2R	AFND40	<i>R<sub>s</sub></i>	7.1	7.6	6.8	5.9	12.00
			<i>F<sub>IS</sub></i>	-0.07	-0.13	0.16*	0.24*	0.03
			<i>H<sub>e</sub></i>	0.71	0.72	0.70	0.70	0.75
			<i>H<sub>o</sub></i>	0.75	0.81	0.59	0.50	0.53
AFND5		<i>R<sub>s</sub></i>	5.7	5.0	5.8	6.0	7.00	
		<i>F<sub>IS</sub></i>	0.14*	-0.02	0.04	0.15	0.08*	
		<i>H<sub>e</sub></i>	0.63	0.66	0.63	0.70	0.64	
		<i>H<sub>o</sub></i>	0.54	0.68	0.60	0.60	0.58	
AFND32		<i>R<sub>s</sub></i>	7.6	9.7	7.7	7.0	12.00	
		<i>F<sub>IS</sub></i>	0.06	0.18*	0.08	0.07	0.09**	
		<i>H<sub>e</sub></i>	0.79	0.83	0.81	0.74	0.79	
		<i>H<sub>o</sub></i>	0.74	0.68	0.74	0.69	0.72	
FunO	<i>R<sub>s</sub></i>	9.3	8.8	9.0	5.9	17.00		
	<i>F<sub>IS</sub></i>	-0.02	-0.13	-0.07	0.04	-0.04		
	<i>H<sub>e</sub></i>	0.77	0.77	0.72	0.72	0.74		
	<i>H<sub>o</sub></i>	0.78	0.86	0.77	0.69	0.77		
2L	AFUB10	<i>R<sub>s</sub></i>	9.2	8.8	8.0	9.0	15.00	
		<i>F<sub>IS</sub></i>	0.23***	0.15	0.30***	0.31**	0.25**	
		<i>H<sub>e</sub></i>	0.81	0.79	0.79	0.78	0.80	
		<i>H<sub>o</sub></i>	0.62	0.67	0.79	0.78	0.80	
	FunL	<i>R<sub>s</sub></i>	12.6	12.5	12.3	13.8	22.00	
		<i>F<sub>IS</sub></i>	0.04	0.02	0.16***	0.03	0.07**	
		<i>H<sub>e</sub></i>	0.84	0.86	0.83	0.86	0.84	
		<i>H<sub>o</sub></i>	0.81	0.84	0.70	0.83	0.78	
	AFUB11	<i>R<sub>s</sub></i>	9.2	9.7	7.3	7.9	15.00	
		<i>F<sub>IS</sub></i>	0.13**	0.17	0.26***	0.10	0.17**	
		<i>H<sub>e</sub></i>	0.71	0.73	0.65	0.75	0.70	
		<i>H<sub>o</sub></i>	0.62	0.61	0.49	0.68	0.58	
AFND20	<i>R<sub>s</sub></i>	9.7	8.7	9.5	9.0	12.00		
	<i>F<sub>IS</sub></i>	0.12**	0.03	0.08	-0.04	0.08**		
	<i>H<sub>e</sub></i>	0.79	0.83	0.81	0.74	0.81		
	<i>H<sub>o</sub></i>	0.74	0.68	0.74	0.69	0.75		
AFND7	<i>R<sub>s</sub></i>	7.9	7.0	5.7	5.0	9.00		
	<i>F<sub>IS</sub></i>	0.04	0.20**	0.08	0.36***	0.13**		
	<i>H<sub>e</sub></i>	0.76	0.80	0.79	0.73	0.79		
	<i>H<sub>o</sub></i>	0.73	0.64	0.73	0.47	0.68		
3R	AFND19	<i>R<sub>s</sub></i>	9.1	9.7	9.4	10.9	13.00	
		<i>F<sub>IS</sub></i>	0.12*	0.10	0.12*	0.12	0.12**	
		<i>H<sub>e</sub></i>	0.76	0.81	0.84	0.81	0.80	
		<i>H<sub>o</sub></i>	0.66	0.73	0.74	0.71	0.70	
	FunG	<i>R<sub>s</sub></i>	12.8	14.5	10.8	13.7	21.00	
		<i>F<sub>IS</sub></i>	-0.02	-0.03	0.02	0.04	0.00	
		<i>H<sub>e</sub></i>	0.86	0.89	0.87	0.89	0.87	
		<i>H<sub>o</sub></i>	0.88	0.92	0.86	0.86	0.87	
	FunD	<i>R<sub>s</sub></i>	19.3	15.7	17.5	19.7	26.00	
		<i>F<sub>IS</sub></i>	0.37***	0.42***	0.34***	0.18***	0.35**	
		<i>H<sub>e</sub></i>	0.93	0.87	0.91	0.94	0.92	
		<i>H<sub>o</sub></i>	0.58	0.51	0.60	0.77	0.60	

Table 2. Continued.

Arm	Locus	Zone A <i>n</i> = 114	Zone B <i>n</i> = 37	Zone C <i>n</i> = 83	Zone D <i>n</i> = 35	All <i>n</i> = 269	
3L	AFUB12	<i>R<sub>s</sub></i>	4.6	3.9	4.6	4.0	7.00
		<i>F<sub>IS</sub></i>	<b>0.50***</b>	<b>0.39**</b>	<b>0.44***</b>	<b>0.32**</b>	<b>0.44**</b>
		<i>H<sub>e</sub></i>	0.56	0.58	0.52	0.55	0.54
		<i>H<sub>o</sub></i>	0.28	0.35	0.29	0.37	0.30
	FunF	<i>R<sub>s</sub></i>	7.4	6.9	6.7	5.9	9.00
		<i>F<sub>IS</sub></i>	-0.02	0.07	-0.02	0.14	0.02
		<i>H<sub>e</sub></i>	0.78	0.75	0.77	0.79	0.77
		<i>H<sub>o</sub></i>	0.80	0.70	0.78	0.69	0.78
	Mean	<i>R<sub>s</sub></i>	9.3	9.0	8.6	8.6	
		<i>F<sub>IS</sub></i>	<b>0.13***</b>	<b>0.11***</b>	<b>0.16***</b>	<b>0.17***</b>	
<i>H<sub>e</sub></i>		0.77	0.76	0.76	0.77		
<i>H<sub>o</sub></i>		0.67	0.69	0.64	0.64		

NOTES.—*n*: sample size (number of specimens); *R<sub>s</sub>*, allelic richness, calculated after (el Mousadik and Petit 1996); *F<sub>IS</sub>* calculated according to Weir and Cockerham (1984); *H<sub>e</sub>* and *H<sub>o</sub>* refer to the observed and expected heterozygosity (Nei 1978), respectively. \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001. In bold italics: significant (*P* < 0.05), after accounting for multiple testing (Holm 1979).

a mean  $F_{ST}$  of 0.005 [95% CI: 0.001–0.010]. Although very low, population differentiation was statistically significant for all population pairs ( $P < 0.05$ ) except between Zone B and Zone C (table 4). Consistent with the weak genetic differentiation observed, the STRUCTURE analysis failed to cluster the multilocus genotypes in more than a single group [ $P(K = 1 | X) = 1.00$ ; see the supplementary fig. S2A, Supplementary Material online]. Nonetheless, we found a significant correlation between the genetic distances measured between pairs of individuals ( $\hat{d}_r$ ) and the logarithm of Euclidian geographic distances between them (one-tailed Mantel test;  $P = 0.009$ ). From the slope of the regression ( $b = 0.002$  [95% CI: 0.001–0.004]), we calculated an overall estimate of  $D\sigma^2$  based on microsatellite data of 46.35 individuals.

In contrast to microsatellite markers, chromosomal inversions exhibited strong and significant genetic differentiation between ecological zones except between Zone B and Zone C (table 4). Pairwise  $F_{ST}$  estimates were much larger for inversion systems than for microsatellites, ranging from 0.021 to 0.483. Using standardized estimates of population differentiation (Hedrick 2005; Meirmans 2006), we found significant differences in  $F_{ST}$  values between chromosomal inversions and microsatellite markers (Wilcoxon signed-rank test,  $P < 0.05$ ). The highest estimate of genetic differentiation was observed between Zone A and Zone D, which are the most geographically distant (fig. 1) and display highly different habitat conditions. From the STRUCTURE clustering analysis, we found the highest posterior probability of the data for  $K = 4$  putative clusters ( $[P(K = 4 | X) = 0.83]$ , supplementary fig. S2B, Supplementary Material online). Two of these clusters segregated the standard and inverted homokaryotypes for all inversions systems, respectively (e.g., yellow and green clusters on supplementary fig. S2C, Supplementary Material online). The other two clusters (in blue and red on supplementary fig. S2C, Supplementary Material online) consisted in a mixture of standard, inverted, and heterozygous karyotypes for all inversions systems. Although standard

homokaryotypes were more prevalent in the northern populations and inverted homokaryotypes in southern ones, no obvious geographical organization of the genetic clusters was revealed. However, we found evidence for isolation by distance across the entire study area with a highly significant correlation between pairwise estimates of genetic distances ( $\hat{d}_r$ ) and the logarithm of Euclidian geographic distances between individuals (one-tailed Mantel test,  $P < 0.001$ ). From the slope of the regression ( $b = 0.224$  [95% CI: 0.107–0.522]) obtained from chromosomal inversions, we estimated a  $D\sigma^2$  at 0.35 individuals. This estimate is two orders of magnitude lower than that obtained from microsatellite markers. Because the effective density  $D$  must be the same for chromosomal inversions and microsatellite markers, this suggests that the effective dispersal of chromosomal inversions is extremely limited.

#### Homogeneity Across Loci and the Role of Selection

Single-locus  $F_{ST}$  estimates obtained from microsatellite loci ranged from 0 to 0.038. We tested for homogeneity in differentiation level, as measured by  $F_{ST}$ , across microsatellite loci among the four sampled areas using the software package FDIST2 (Beaumont and Nichols 1996). When plotted against their respective  $H_e$  value,  $F_{ST}$  estimates obtained from all loci (either within or outside polymorphic chromosomal inversions) mapped within the 99% confidence envelope of  $F_{ST}$  estimates expected under neutrality (fig. 2A). Only locus AFND7 revealed a marginally significant departure from neutral expectations ( $P = 0.018$ ), with an  $F_{ST}$  estimate slightly higher than expected, lying outside the 95% confidence envelope. Discarding AFND7, the overall mean  $F_{ST}$  estimate remained unchanged at 0.005 [CI 95%: 0.001–0.006] across the 15 remaining loci. There was no apparent difference across locus-specific  $F_{ST}$  estimates lying within and outside polymorphic chromosomal inversions (fig. 2A).

We generated the neutral distribution of  $F_{ST}$  estimates for biallelic codominant markers, given the level of differentiation measured at presumably neutral marker loci. To



**Table 3.** Polymorphism at Seven Chromosomal Inversion in *Anopheles funestus* Collected in Four Ecological Zones in Cameroon.

Arm	Locus		Zone A	Zone B	Zone C	Zone D	All
			<i>n</i> = 114	<i>n</i> = 37	<i>n</i> = 83	<i>n</i> = 35	<i>n</i> = 269
2R	a/a	<i>f</i>	0.32	0.55	0.36	0.17	0.35
		<i>F<sub>IS</sub></i>	<b>0.68**</b>	-0.03	<b>0.40**</b>	0.01	<b>0.44***</b>
		<i>H<sub>e</sub></i>	0.44	0.50	0.46	0.29	0.45
		<i>H<sub>o</sub></i>	0.14	0.51	0.28	0.29	0.25
	h/h	<i>f</i>	0.02	0.11	0.23	0.81	0.20
		<i>F<sub>IS</sub></i>	-0.02	0.17	-0.22	0.16	<b>0.35***</b>
		<i>H<sub>e</sub></i>	0.04	0.20	0.36	0.31	0.32
		<i>H<sub>o</sub></i>	0.04	0.16	0.43	0.26	0.21
	ab/ab	<i>f</i>	0.16	0.05	0.07	0.00	0.10
		<i>F<sub>IS</sub></i>	<b>0.28**</b>	-0.04	-0.07	—	<b>0.19***</b>
		<i>H<sub>e</sub></i>	0.27	0.10	0.13	0.00	0.17
		<i>H<sub>o</sub></i>	0.19	0.11	0.14	0.00	0.14
t/t	<i>f</i>	0.02	0.09	0.11	0.00	0.06	
	<i>F<sub>IS</sub></i>	-0.02	0.22	-0.12	—	0.01	
	<i>H<sub>e</sub></i>	0.04	0.17	0.19	0.00	0.10	
	<i>H<sub>o</sub></i>	0.04	0.14	0.22	0.00	0.10	
3R	a/a	<i>f</i>	0.39	0.66	0.66	1.00	0.59
		<i>F<sub>IS</sub></i>	<b>0.82**</b>	-0.38*	0.09	—	<b>0.49***</b>
		<i>H<sub>e</sub></i>	0.48	0.45	0.45	0.00	0.49
		<i>H<sub>o</sub></i>	0.09	0.62	0.41	0.00	0.25
	b/b	<i>f</i>	0.40	0.81	0.72	1.00	0.63
		<i>F<sub>IS</sub></i>	<b>0.95**</b>	-0.04	0.16	—	<b>0.65***</b>
3L	a/a	<i>f</i>	0.40	0.97	0.94	1.00	0.72
		<i>F<sub>IS</sub></i>	<b>0.95**</b>	-0.01	-0.06	—	<b>0.86***</b>
		<i>H<sub>e</sub></i>	0.48	0.05	0.11	0.00	0.40
		<i>H<sub>o</sub></i>	0.03	0.05	0.12	0.00	0.06
	Mean	<i>F<sub>IS</sub></i>	<b>0.75**</b>	-0.07	0.08	0.09	
		<i>H<sub>e</sub></i>	0.32	0.26	0.30	0.09	
		<i>H<sub>o</sub></i>	0.08	0.27	0.28	0.08	

NOTES.—*n*: sample size (number of specimens); *f*: inversion frequency; *F<sub>IS</sub>* calculated according to (Weir and Cockerham 1984); *H<sub>o</sub>* and *H<sub>e</sub>* refer to the observed and expected heterozygosity (Nei 1978), respectively. \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001. In bold italics: significant (*P* < 0.05), after accounting for multiple testing (Holm 1979).

do so, we conducted coalescent simulations using a modified version of the software package *DFDIST* (Beaumont and Balding 2004). We used the multilocus *F<sub>ST</sub>* estimate obtained from the 15 presumably neutral microsatellite loci (*F<sub>ST</sub>* = 0.005, see above) to perform the coalescent simu-

lations. As shown in figure 2B, all chromosomal inversion systems departed significantly from neutral expectations based on microsatellite data (*P* < 2.10<sup>-6</sup>, except for 2Rab and 2Rt for which *P* = 0.003 and *P* = 0.005, respectively). The two inversions 2Rab and 2Rt occur at very low frequency in our data set (table 3), which may reduce statistical power to detect departure from neutrality in this case. The significantly higher differentiation of chromosomal inversions as compared with that expected at migration–drift equilibrium given the microsatellite data, strongly suggests that gene flow is extremely limited for chromosomal inversions systems, and/or that chromosomal inversions are exposed to strong natural selection and local adaptation.

### Inversions and Environment

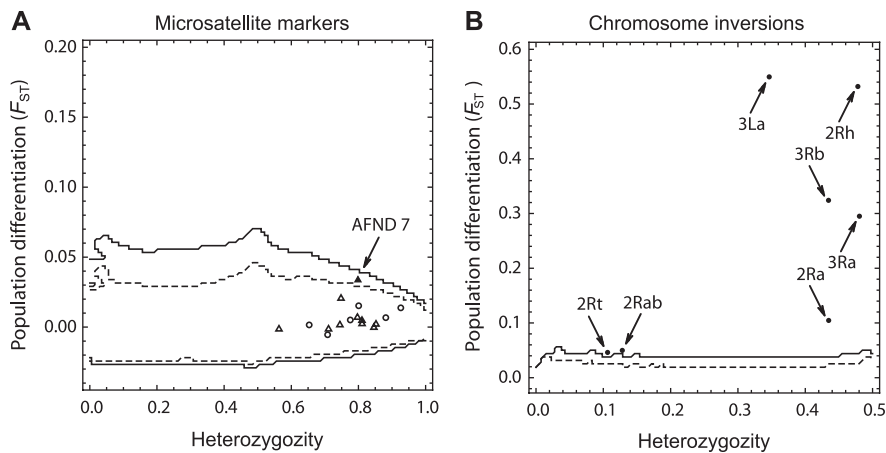
To further test whether chromosomal inversions are exposed to environmentally driven selection, we investigated how their distribution varies with environmental variables (i.e., latitude, longitude, altitude, rainfall, temperature, and water vapor pressure) using CCA. A graphical representation of the contribution of each EGV to the global chromosomal polymorphism distribution is given in figure 3. The first four axes of the CCA were statistically significant (Monte Carlo permutation test, *F* = 2.71, *P* = 0.0176), with the first two axes accounting for 89.7% of the total variance in the data set. The most informative EGV was latitude, explaining 23% of the system variance followed by rainfall and temperature (22% and 19%, respectively). All EGVs, except water vapor pressure, showed a highly significant contribution to the spatial pattern of chromosomal inversion distribution (*P* < 0.05).

The first axis, which explains 66.7% of the total variance in chromosomal polymorphism, can be ascribed to an increasing aridity gradient (from left to right): We observe elevated water vapor pressure and rainfall, which are characteristic of the humid conditions in the rainforest biome on the left-hand side of the axis, whereas high temperatures, which characterize more arid conditions, are found on the right-hand side. The ordination diagram (fig. 3) shows a clear separation among alternative arrangements (i.e., standard vs. inverted) for each inversion system. Standard arrangements are positively associated with warmer and drier environments (except 2Rab ++ and 2Rt ++, which show almost no variance) and map accordingly in the right-hand

**Table 4.** Estimate of Genetic Differentiation (*F<sub>ST</sub>*).

	Zone A <i>n</i> = 114	Zone B <i>n</i> = 37	Zone C <i>n</i> = 83	Zone 4 <i>n</i> = 35
			Microsatellites	
Zone A		0.0067** (0.0296)	0.0038** (0.0163)	0.0047** (0.0206)
Zone B	0.2253** (0.3219)		0.0025* (0.0108)	0.0129** (0.0572)
Zone C	0.1907** (0.2772)	0.0209* (0.0292)		0.0075** (0.0322)
Zone D	0.4826** (0.6447)	0.3944** (0.4769)	0.2539** (0.3292)	
		Inversions		

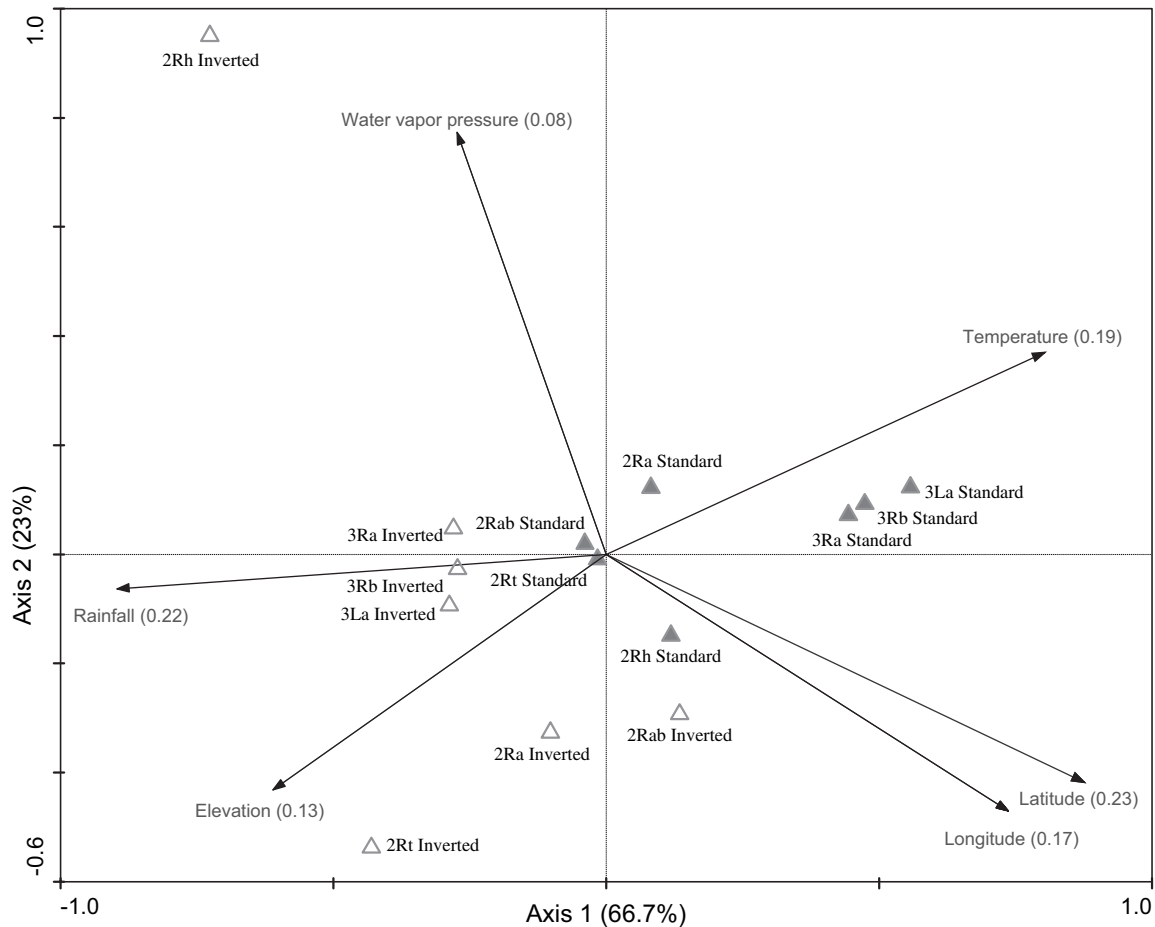
Derived from microsatellite (above diagonal) and chromosomal inversions (below diagonal) frequency data among *A. funestus* populations from Cameroon. *n*: sample size (number of specimens); in brackets, *H<sub>e</sub>*-corrected *F<sub>ST</sub>* (see text). \**P* < 0.05, \*\**P* < 0.01. In bold italics: significant (*P* < 0.05) after accounting for multiple testing (Holm 1979).



**FIG. 2.** Neutrality tests. single-locus  $F_{ST}$  estimates among geographic populations of *A. funestus* in Cameroon, plotted against expected heterozygosity (Nei 1987). (A) Data from 16 microsatellites markers located inside (circles) and outside (triangles) common polymorphic chromosomal inversions. (B) Data from seven chromosomal inversion systems. In both panels, outlier loci are pointed by arrows and referred to by their names. Dashed lines and solid lines represent the 95% and 99% confidence limits under the hypothesis of neutrality of the loci, respectively.

part of the diagram shown in figure 3. Inverted arrangements are associated with more humid and cooler climates and map in the left-hand part of the diagram.

The second axis of the CCA, which explains 23% of total variance, is more difficult to interpret; it represents an environmental gradient mostly affected by the “highlands”



**FIG. 3.** Eco-geographical chromosome inversion distributions. CCA diagram showing the ordination of chromosomal inversions for each homozygote (standard or inverted) along the first two canonical axes (explaining 89.7% of variance). Open triangles represent STANDARD homokaryotes. Full triangles represent INVERTED homokaryotes. EGVs are passively plotted on the graph. Alt, altitude; temp, temperature; rain, rainfall; lat, latitude; long, longitude; WVP: water vapor pressure.

bioclimatic domain correlated with water vapor pressure, altitude, and temperature. Thus, we interpret it as a gradient separating cooler highlands from warmer and humid lowlands and valleys. Inversions 2Rt and 2Rh show the highest level of divergence along this axis: Inversion 2Rt is associated with higher elevation, whereas inversion 2Rh is associated with humid lowlands (fig. 3).

## Discussion

The comparative analysis of chromosomal inversions and microsatellite markers polymorphisms in *A. funestus* populations from Cameroon revealed a contrasting pattern of population genetic subdivision. Presumably, neutral microsatellite markers depicted a weak and shallow population structure, with very low levels of genetic differentiation among populations, whereas chromosomal inversions revealed a very strong population structure. Strong deviation from the expected migration–drift equilibrium based on microsatellite markers suggested that gene flow is limited for chromosomal inversions, and/or that chromosomal inversions are targets for local adaptation. The latter hypothesis was further confirmed by the nonrandom distribution of alternative chromosomal arrangements across different environmental conditions.

Previous genetic structure analysis based on microsatellite markers revealed important gene flow among *A. funestus* populations with a weak, if any, role of geographic distance in shaping natural population structure (Kamau et al. 2003; Cohuet et al. 2005; Michel, Guelbeogo, et al. 2005; Michel, Ingrassi, et al. 2005), as we observed in the present study. As reported in the widely codistributed *A. gambiae*, retention of ancestral polymorphism together with evidence for homoplasy at microsatellite loci and, more importantly, molecular signatures of recent population expansion and gene flow result in a shallow population structure (Donnelly et al. 2001; Krzywinski and Besansky 2003; Lehmann et al. 2003; Michel, Ingrassi, et al. 2005). However, significant levels of genetic differentiation have been reported between sympatric populations characterized by different levels of chromosomal polymorphism (i.e., the “Folonzo” and “Kiribina” chromosomal forms described by Costantini et al. 1999 and Guelbeogo et al. 2005) in Burkina Faso (Michel, Guelbeogo, et al. 2005) or between allopatric populations of *A. funestus* across the inhospitable Great Rift Valley complex (Braginets et al. 2003) with the same or a subset of the microsatellite markers used in this study. Our results indeed suggest that the amount of gene flow between geographical populations of *A. funestus* in Cameroon is high enough to prevent population splitting and demographic independence at this geographical scale. However, the unique experimental design of our study, in which every specimen was both karyotyped on all arms of their chromosomal complement and genotyped at microsatellite loci, allows us to investigate the role of chromosomal inversions in local adaptation and speciation in *A. funestus*. Interestingly, no significant LD was detected between microsatellite markers and inversions. Although

chromosomal inversions reduce recombination, particularly in the vicinity of breakpoints, that is not completely stopped, and the interaction between gene flux and selection is expected to produce a mosaic of more- and less-differentiated genomic regions within the inversion and not necessarily close to the breakpoints (Navarro et al. 1997; White et al. 2007, 2010). Several factors, such as the number of genes under selection or the age of the inversion, can affect the recombination rates within a rearrangement. Recent studies in *A. gambiae* have confirmed that variation in LD patterns within the inversion is not homogeneous (White et al. 2007, 2009). They estimated that, without selection and according to the estimated level of recombination that occurs within the 2La inversion of *A. gambiae* (see above), LD should drop to undetectable levels within less than 4,600 generation (i.e., 400 years) (White et al. 2007, 2009). These general trends are likely to apply to *A. funestus* as well, although no estimate of inversion age or recombination rates in this malaria vector is available yet. However, taking into account the large and standing distribution of these inversions throughout the species range in Africa, we may consider that the age of inversions is an important factor that would explain low LD between inversions and microsatellite markers.

In contrast to microsatellite markers, chromosomal inversions polymorphism depicted a situation whereby alternative chromosomal arrangements (i.e., standard vs. inverted) segregated nonrandomly in different environments across diverse ecological zones in Cameroon. Again, this situation is reminiscent of what has been observed in *A. gambiae* (Coluzzi et al. 1979, 1985; Powell et al. 1999; Toure et al. 1998) and provides strong support for a role of chromosomal inversions or the genes they contain in local adaptation and ecological specialization in both vector species. As such, inversions might contribute to local adaptation by maintaining tight linkage between locally advantageous alleles in the face of gene flow from other populations carrying different sets of alleles. We have shown that estimates of population differentiation based on chromosomal inversions were at odds with the expected migration–drift equilibrium based on presumably neutral markers (fig. 2). Furthermore, alternative chromosomal arrangements segregated differently in the ordination space defined by a number of EGVs (fig. 3). This suggests that the population structure depicted by these chromosomal arrangements is likely the result of local adaptation, which could lead to propitious conditions for incipient speciation, increasing genetic drift between geographically isolated populations (Manoukis et al. 2008). Strong LD between inversions in the face of linkage equilibrium between microsatellite loci and inversions further strengthens this view, suggesting that epistasis between locally adapted alleles captured by different inversions and/or parallel selection–migration balance resulting from local adaptation in multiple traits generated the observed pattern (Li and Nei 1974; Kirkpatrick and Barton 2006; White et al. 2007). In this sense, two recent studies have revealed that both molecular forms in

*A. gambiae* shared the same chromosomal arrangements in areas where they were found in sympatry (Costantini et al. 2009; Simard et al. 2009). The most likely explanation for the observed pattern of chromosomal polymorphism is the action of directional and/or balancing selection acting within the M and S forms on alternative arrangements according to the prevailing eco-geographical conditions. Thus, inversions almost certainly predated the speciation process in *A. gambiae*.

Chromosomal inversions are typically regarded as a genetic mechanism protecting sets of (locally adapted) genes by reducing recombination in heterozygotes (Hoffmann and Rieseberg 2008). The very low genetic differentiation measured at microsatellite markers between populations characterized by alternative arrangements (i.e., standard vs. inverted) was consistent across loci, independently of their physical location within or outside polymorphic chromosomal inversions. This suggests that recombination suppression is not complete within the inversion and that residual gene flow (through double crossing-over or gene conversion) between alternative arrangements is sufficient to homogenize genetic variation in genomic regions not directly targeted by selection. Recent genomic investigations in *A. gambiae* indeed provided evidence that, although recombination is drastically reduced within inversions, the amount of residual gene flow between alternative arrangements is sufficient to allow rapid homogenization (Stump et al. 2007; White et al. 2007, 2009). These empirical findings have been recently supported by theoretical approaches (Feder and Nosil 2009). In *A. funestus*, genetic independence between karyotypes might be prevented by hybridization (see table 3) and migration across the country (estimate of  $D\sigma^2$  based on microsatellite data of 46.35 individuals), homogenizing variation at neutral microsatellite loci, as it has been previously shown (Cohuet et al. 2005). On the other hand, inversions exposed to environmental selection (fig. 3) and possible epistatic effects between inversions (high LD between inversions) might lead to a strong population structure, as observed across heterogeneous landscapes in Cameroon.

Although neither reduced recombination nor selection alone is expected to maintain inversion polymorphism and divergence, migration–selection balance between multiple loci within an inversion and/or selection against deleterious recessives are the most likely to maintain inversion polymorphism in natural populations and to generate high  $F_{ST}$  between ecotypes (Navarro et al. 1997; Kirkpatrick and Barton 2006; White et al. 2007). This was experimentally demonstrated in *Drosophila* (Schaeffer and Anderson 2005) and our data suggest that this general principle may also apply to chromosomal polymorphism and divergence in *A. funestus*. Furthermore, selection against heterozygotes (e.g., underdominance) has been pointed out as a mechanism to establish inversions (King 1993). This is a controversial mechanism because an underdominant inversion is established only by drift in a population and it requires special conditions such as small population size, weak selection against heterokaryotypes, or meiotic drive (Kirkpatrick and

Barton 2006). However, if underdominance is genic (interactions between alleles within an inversion, resulting in maladaptive heterozygotes) and not structural (problems during meiosis) as it seems occur for paracentric inversions in dipteran species (Hoffmann and Rieseberg 2008), local adaptation theory could explain how underdominant inversions are established (Kirkpatrick and Barton 2006). Although several inversions do not show any deficit in heterozygotes across the country, others, such as 3Ra, are dependent upon the eco-geographical zone, suggesting a fitness cost at the heterozygous state. Finally, other mechanisms involving prezygotic barriers and assortative mating could participate to the inversion distribution patterns as it has been described in *Rhagoletis* (Feder et al. 2003). Unfortunately, conventional formal genetics approaches based on experimental crossings are yet impossible to conduct in *A. funestus* because of the scarcity of laboratory colonies available. Furthermore, high-resolution genomics approaches are still awaiting complete genome sequencing. Once available, these tools will provide considerable opportunities for comparative genomics between *A. gambiae* and *A. funestus* and will help elucidate the exact role of common polymorphic chromosomal inversions in ecological adaptation of such devastating pests and pinpoint the genes involved in these processes. Further theoretical and empirical studies aiming at measuring migration rates, fitness, or selection on inversions and on the genes they contain should further help deciphering the biological and ecological mechanisms underlying such adaptive potential.

The ability of both *A. gambiae* and *A. funestus* to colonize a wide range of habitats and to cope with dramatic seasonal fluctuations in environmental conditions have direct consequences on malaria transmission dynamics: It contributes to increase the intensity of parasites transmission locally during the most favorable season and to extend malaria transmission in both space and time (Fontenille and Simard 2004). In these two major African malaria vectors, ecological plasticity is mirrored at the genetic level by high amounts of chromosomal polymorphisms of proven adaptive value (Coluzzi et al. 2002; Pombi et al. 2008). Moreover, behavioral variations with direct relevance for vector efficiency, such as the ability to bite humans and rest indoors, have been shown to be associated with chromosomal polymorphisms in *A. gambiae* (Bryan et al. 1982). Inversions polymorphism within vector populations have also been invoked to explain vector control failures in West Africa (Coluzzi 1982; Molineaux and Gramiccia 1980). Dissecting the molecular make up of chromosomal inversions and unraveling the genetic processes by which they are produced, raise in frequency and are maintained within natural vector populations, is of paramount importance for vector control strategies to be efficient and sustainable. This is particularly true because of the considerable reservoir of standing and possibly cryptic variation embodied in chromosomal inversions (Pombi et al. 2008). A more detailed understanding of the population genetic structure of major African malaria vectors and, more importantly, of the interplay between genetic polymorphisms and ecological or behavioral variations

is required for a number of reasons. First, it will provide a better assessment of the feasibility of innovative vector control strategies based on the use of genetically modified organisms with altered vector competence. Second, it will contribute to devise implementation guidelines and protocols and to insure comprehensive monitoring of the intervention. Finally, such studies are also warranted to assess the impact of anthropogenic modifications of the environment (e.g., urbanization, deforestation, and agriculture) and global climate change on vector distributions and fitness and their consequences on the evolution of malaria transmission risks in Africa (Tanser et al. 2003; Guerra et al. 2006; Pinto et al. 2007; Boete 2009).

## Supplementary Material

Supplementary figures S1 and S2 are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

## Acknowledgments

We thank the Laboratoire de Recherche sur le Paludisme at OCEAC, Yaoundé for their excellent field assistance. F.S. and D.A. gratefully acknowledge Joe Bajo and Kass Telbiehr for continuously prompting useful discussions throughout the study. Authors' contribution: D.A., A.C., D.F., and F.S. conceived and designed the experiments. D.A. and A.C. performed the experiments. D.A., M.C.F., A.C., R.V., and F.S. analyzed the data. D.A., D.F., and F.S. coordinated the fieldwork. D.A., R.V., and F.S. wrote the manuscript with editorial suggestions from the other coauthors. Financial support was provided by the Institut de Recherche pour le Développement. Fieldwork was supported in part by the National Institutes of Health (R01-AI063508 to Nora J. Besansky). D.A. was supported by a student fellowship grant from Fundación CAJA MADRID (Madrid, Spain).

## References

Balanya J, Serra L, Gilchrist GW, Huey RB, Pascual M, Mestres F, Sole E. 2003. Evolutionary pace of chromosomal polymorphism in colonizing populations of *Drosophila subobscura*: an evolutionary time series. *Evolution* 57:1837–1845.

Beaumont MA. 2005. Adaptation and speciation: what can F<sub>st</sub> tell us? *Trends Ecol Evol.* 20:435–440.

Beaumont MA, Balding DJ. 2004. Identifying adaptive genetic divergence among populations from genome scans. *Mol Ecol.* 13:969–980.

Beaumont MA, Nichols R. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proc R Soc Lond Ser B Biol Sci.* 263:1619–1626.

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

Boete C. 2009. *Anopheles* mosquitoes: not just flying malaria vectors especially in the field. *Trends Parasitol.* 25:53–55.

Bonin A, Taberlet P, Miaud C, Pompanon F. 2006. Explorative genome scan to detect candidate loci for adaptation along a gradient of altitude in the common frog (*Rana temporaria*). *Mol Biol Evol.* 23:773–783.

Braginets OP, Minakawa N, Mbogo CM, Yan G. 2003. Population genetic structure of the African malaria mosquito *Anopheles funestus* in Kenya. *Am J Trop Med Hyg.* 69:303–308.

Bryan JH, Deco MA, Petrarca V, Coluzzi M. 1982. Inversion polymorphism and incipient speciation in *Anopheles gambiae* s.str. in The Gambia, West Africa. *Genetica* 59:167–176.

Butlin R, Roper C. 2005. Evolutionary genetics: microarrays and species origins. *Nature* 437:199–201.

Chapuis MP, Estoup A. 2007. Microsatellite null alleles and estimation of population differentiation. *Mol Biol Evol.* 24: 621–631.

Coetzee M, Fontenille D. 2004. Advances in the study of *Anopheles funestus*, a major vector of malaria in Africa. *Insect Biochem Mol Biol.* 34:599–605.

Cohuet A, Dia I, Simard F, Raymond M, Rousset F, Antonio-Nkondjio C, Awono-Ambene PH, Wondji CS, 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.

Cohuet A, Simard F, Berthomieu A, Raymond M, Fontenille D, Weill M. 2002. Isolation and characterization of microsatellite DNA markers in the malaria vector *Anopheles funestus*. *Mol Ecol Notes.* 2:498–500.

Cohuet A, Simard F, Toto JC, Kengne P, Coetzee M, 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.

Coluzzi M. 1982. Spatial distribution of chromosomal inversions and speciation in Anopheline mosquitoes. In: Barigozzi C, editor. Mechanisms of speciation. New York: Alan R Liss. p. 113–115.

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

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.

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

Costantini C, Ayala D, Guelbeogo W, et al. (12 co-authors). 2009. Living at the edge: biogeographic patterns of habitat segregation conform to speciation by niche expansion in *Anopheles gambiae*. *BMC Ecol.* 9:16.

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

della Torre A, Costantini C, Besansky NJ, Caccone A, Petrarca V, Powell JR, Coluzzi M. 2002. Speciation within *Anopheles gambiae*—the glass is half full. *Science* 298:115–117.

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*. *Mol Biol Evol.* 18:1353–1364.

el Mousadik A, Petit RJ. 1996. Chloroplast DNA phylogeography of the argan tree of Morocco. *Mol Ecol.* 5:547–555.

Falush D, Stephens M, Pritchard JK. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567–1587.

Feder JL, Nosil P. 2009. Chromosomal inversions and species differences: when are genes affecting adaptive divergence and reproductive isolation expected to reside within inversions? *Evolution* 63:3061–3075.

Feder JL, Roethele FB, Filchak K, Niedbalski J, Romero-Severson J. 2003. Evidence for inversion polymorphism related to sympatric

- host race formation in the apple maggot fly, *Rhagoletis pomonella*. *Genetics* 163:939.
- Fontaine MC, Baird SJE, Piry S, et al. (19 co-authors). 2007. Rise of oceanographic barriers in continuous populations of a cetacean: the genetic structure of harbour porpoises in Old World waters. *BMC Biol.* 5:30
- Fontenille D, Simard F. 2004. Unravelling complexities in human malaria transmission dynamics in Africa through a comprehensive knowledge of vector populations. *Comp Immunol Microbiol Infect Dis.* 27:357–375.
- Gillies MT, de Meillon B. 1968. The anophelinae of Africa, south of the Sahara. Johannesburg (South Africa): The South African Institute for Medical Research.
- 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/software/fstat.html>
- Guelbeogo WM, Grushko O, Boccolini D, Ouedraogo PA, Besansky NJ, Sagnon NF, Costantini C. 2005. Chromosomal evidence of incipient speciation in the Afrotropical malaria mosquito *Anopheles funestus*. *Med Vet Entomol.* 19:458–469.
- Guerra CA, Snow RW, Hay SI. 2006. A global assessment of closed forests, deforestation and malaria risk. *Ann Trop Med Parasitol.* 100:189–204.
- Hedrick PW. 2005. A standardized genetic differentiation measure. *Evolution* 59:1633–1638.
- Hoffmann AA, Rieseberg LH. 2008. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? *Annu Rev Ecol Syst.* 39:21–42.
- Hoffmann AA, Sgro CM, Weeks AR. 2004. Chromosomal inversion polymorphisms and adaptation. *Trends Ecol Evol.* 19:482–488.
- Hoffmann AA, Willi Y. 2008. Detecting genetic responses to environmental change. *Nat Rev Genet.* 9:421–432.
- Holm S. 1979. A simple sequentially rejective multiple test procedure. *Scand J Stat.* 6:65–70.
- Hunt RH. 1973. A cytological technique for the study of *Anopheles gambiae* complex. *Parassitologia* 15:137–139.
- Kamau L, Munyekenye GO, Koekemoer LL, Hunt RH, Coetzee M. 2003. A survey of the *Anopheles funestus* (Diptera: Culicidae) group of mosquitoes from 10 sites in Kenya with special emphasis on population genetic structure based on chromosomal inversion karyotypes. *J Med Entomol.* 40:664–671.
- King M. 1993. Species evolution: the role of chromosome change. Cambridge (United Kingdom): Cambridge University Press.
- Kirkpatrick M, Barton N. 2006. Chromosome inversions, local adaptation and speciation. *Genetics* 173:419.
- Koekemoer LL, Kamau L, Hunt RH, Coetzee M. 2002. A cocktail polymerase chain reaction assay to identify members of the *Anopheles funestus* (Diptera: Culicidae) group. *Am J Trop Med Hyg.* 66:804–811.
- Krzywinski J, Besansky NJ. 2003. Molecular systematics of *Anopheles*: from subgenera to subpopulations. *Annu Rev Entomol.* 48:111–139.
- Kulathinal RJ, Steverson LS, Noor MAF. 2009. The genomics of speciation in *Drosophila*: diversity, divergence, and introgression estimated using low-coverage genome sequencing. *PLoS Genet.* 5:e1000550.
- Leblois R, Estoup A, Rousset F. 2003. Influence of mutational and sampling factors on the estimation of demographic parameters in a “continuous” population under isolation by distance. *Mol Biol Evol.* 20:491–502.
- Lehmann T, Licht M, Elissa N, Maega BT, Chimumbwa JM, Watsenga FT, Wondji CS, Simard F, Hawley WA. 2003. Population structure of *Anopheles gambiae* in Africa. *J Hered.* 94:133–147.
- Li WH, Nei M. 1974. Stable linkage disequilibrium without epistasis in subdivided populations. *Theor Popul Biol.* 6:173–183.
- Manoukakis NC, Powell JR, Toure MB, Sacko A, Edillo FE, Coulibaly MB, Traore SF, Taylor CE, Besansky NJ. 2008. A test of the chromosomal theory of ecotypic speciation in *Anopheles gambiae*. *Proc Natl Acad Sci U S A.* 105:2940–2945.
- Meirmans PG. 2006. Using the AMOVA framework to estimate a standardized genetic differentiation measure. *Evolution* 60:2399–2402.
- Michel AP, Guelbeogo WM, Grushko O, Schemerhorn BJ, Kern M. 2005. Molecular differentiation between chromosomally defined incipient species of *Anopheles funestus*. *Insect Mol Biol.* 14:375.
- Michel AP, Ingrasci MJ, Schemerhorn BJ, et al. (13 co-authors). 2005. Rangewide population genetic structure of the African malaria vector *Anopheles funestus*. *Mol Ecol.* 14:4235–4248.
- Molineaux L, Gramiccia G. 1980. The Garki Project. Research on Epidemiology and Control of Malaria in the Sudan Savannah of West Africa. *Bull World Health Organ.*
- Morlais I, Poncon N, Simard F, Cohuet A, Fontenille D. 2004. Intraspecific nucleotide variation in *Anopheles gambiae*: new insights into the biology of malaria vectors. *Am J Trop Med Hyg.* 71:795–802.
- Navarro A, Barton NH. 2003. Chromosomal speciation and molecular divergence—accelerated evolution in rearranged chromosomes. *Science* 300:321–324.
- Navarro A, Betran E, Barbadilla A, Ruiz A. 1997. Recombination and gene flux caused by gene conversion and crossing over in inversion heterokaryotypes. *Genetics* 146:695.
- Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* 89: 583–590.
- Nei M. 1987. Molecular evolutionary genetics. New York: Columbia University Press.
- Noor MA, Cunningham AL, Larkin JC. 2001. Consequences of recombination rate variation on quantitative trait locus mapping studies. Simulations based on the *Drosophila melanogaster* genome. *Genetics.* 159:581–588.
- Olivry JC. 1986. Fleuves et Rivières du Cameroun. Paris (France): ORSTOM.
- Ortiz-Barrientos D, Reiland J, Hey J, Noor MAF. 2002. Recombination and the divergence of hybridizing species. *Genetica* 116:167.
- Pinto J, Lynd A, Vicente JL, et al. (13 co-authors). 2007. Multiple origins of knockdown resistance mutations in the Afrotropical mosquito vector *Anopheles gambiae*. *PLoS One.* 2:e1243.
- Pombi M, Caputo B, Simard F, Di Deco MA, Coluzzi M, Della Torre A, Costantini C, Besansky NJ, Petrarca V. 2008. Chromosomal plasticity and evolutionary potential in the malaria vector *Anopheles gambiae* sensu stricto: insights from three decades of rare paracentric inversions. *BMC Evol Biol.* 8:309.
- Powell JR, Petrarca V, della Torre A, Caccone A, Coluzzi M. 1999. Population structure, speciation, and introgression in the *Anopheles gambiae* complex. *Parassitologia* 41:101–113.
- Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- Rieseberg LH. 2001. Chromosomal rearrangements and speciation. *Trends Ecol Evol.* 16:351–358.
- Rodriguez-Trelles F, Alvarez G, Zapata C. 1996. Time-series analysis of seasonal changes of the O inversion polymorphism of *Drosophila subobscura*. *Genetics* 142:179–187.
- Rousset F. 1997. Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145:1219–1228.
- Schaeffer SW, Anderson WW. 2005. Mechanisms of genetic exchange within the chromosomal inversions of *Drosophila pseudoobscura*. *Genetics* 171:1729–1739.
- Scott DW. 1992. Multivariate density estimation: theory, practice, and visualization. New York: John Wiley.

- Service M. 1993. Mosquito ecology field sampling methods. New York: Elsevier Applied Science.
- Sharakhov I, Braginetts O, Grushko O, et al. (12 co-authors). 2004. A microsatellite map of the African human malaria vector *Anopheles funestus*. *J Hered.* 95:29–34.
- Simard F, Ayala D, Kamdem G, et al. 2009. Ecological niche partitioning between *Anopheles gambiae* molecular forms in Cameroon: the ecological side of speciation. *BMC Ecol.* 9:17.
- Sinkins SP, Hackett BJ, Costantini C, Vulule J, Ling YY, Collins FH, Besansky NJ. 2000. Isolation of polymorphic microsatellite loci from the malaria vector *Anopheles funestus*. *Mol Ecol.* 9:490–492.
- Stump AD, Pombi M, Goeddel L, Ribeiro JM, Wilder JA, della Torre A, Besansky NJ. 2007. Genetic exchange in 2La inversion heterokaryotypes of *Anopheles gambiae*. *Insect Mol Biol.* 16:703–709.
- Tanser FC, Sharp B, le Sueur D. 2003. Potential effect of climate change on malaria transmission in Africa. *Lancet.* 362: 1792–1798.
- Ter Braak C. 1987. CANOCO—Fortran program for canonical community ordination. Ithaca (NY): Microcomputer Power.
- Toure YT, Petrarca V, Traore SF, Coulibaly A, Maiga HM, Sankare O, Sow M, Di Deco MA, 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.
- van Doorn GS, Kirkpatrick M. 2007. Turnover of sex chromosomes induced by sexual conflict. *Nature* 449:909.
- Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P. 2004. Micro-checker: software for identifying and correcting genotyping errors in microsatellite data. *Mol Ecol Notes.* 4:535–538.
- Vitalis R, Dawson K, Boursot P. 2001. Interpretation of variation across marker loci as evidence of selection. *Genetics* 158:1811–1823.
- Watts PC, Rousset F, Saccheri IJ, Leblois R, Kemp SJ, Thompson DJ. 2007. Compatible genetic and ecological estimates of dispersal rates in insect (*Coenagrion mercuriale*: Odonata: Zygoptera) populations: analysis of ‘neighbourhood size’ using a more precise estimator. *Mol Ecol.* 16:737–751.
- Weir B, Cockerham C. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370.
- White BJ, Cheng C, Sangare D, Lobo NF, Collins FH, Besansky NJ. 2009. The population genomics of trans-specific inversion polymorphisms in *Anopheles gambiae*. *Genetics* 183:275–288.
- White BJ, Cheng C, Simard F, Costantini C, Besansky NJ. 2010. Genetic association of physically unlinked islands of genomic divergence in incipient species of *Anopheles gambiae*. *Mol Ecol.* 19:925–939.
- White BJ, Hahn MW, Pombi M, Cassone BJ, Lobo NF, Simard F, Besansky NJ. 2007. Localization of candidate regions maintaining a common polymorphic inversion (2La) in *Anopheles gambiae*. *PLoS Genet.* 3:e217.
- Wright S. 1951. The genetical structure of populations. *Ann Eugen.* 15:323–354.
- Wright S. 1973. The origin of the F-statistics for describing the genetic aspects of population structure. Honolulu (HI): University of Hawaii Press.