Chromosomal inversions drive chromosome arm specific patterns of polymorphism in the major African malaria vector, *Anopheles funestus*

Colince Kamdem[1]*, Caroline Fouet[1], Bradley J. White[1,2]*

[1]Department of Entomology, University of California, Riverside, CA 92521

[2]Center for Disease Vector Research, Institute for Integrative Genome Biology, University of California, Riverside, CA 92521

*Corresponding authors: colincek@ucr.edu; bradley.white@ucr.edu

Abstract

Chromosomal inversions facilitate local adaptation of beneficial mutations and modulate genetic polymorphism, but the extent of their effects within the genome is not fully resolved. The genome of *Anopheles funestus*, a malaria mosquito endemic to sub-Saharan Africa, contains an impressive number of paracentric polymorphic inversions, which are unevenly distributed among chromosomes and provide an excellent framework for investigating the genomic impacts of chromosomal rearrangements. Here we present results of a fine-scale analysis of genetic variation within the genome of two *Anopheles funestus* ecotypes present in Cameroon. Specifically, we used F_{ST} outlier and population genomic analyses to highlight the strong relationship that exists between the presence of inversions and the genomic patterns of divergence and polymorphism in this mosquito. We found evidence that local adaptation in this geographic area is driven by divergent selection within three polymorphic chromosomal inversions, which segregate between savannah and forest ecosystems. Importantly, in contrast to the 2L chromosome arm, which is collinear, nucleotide diversity is significantly reduced along the entire length of three autosome arms bearing multiple overlapping chromosomal rearrangements. These findings support the idea that interactions between reduced recombination and selection within chromosomal rearrangements contribute to sculpt nucleotide polymorphism across entire chromosomes in *An. funestus*.

Key words: Chromosomal inversion, *Anopheles funestus*, genetic differentiation, nucleotide diversity, local adaptation.

Introduction

Although much progress has been made in analyzing genetic variation among populations, less is known about the complex interactions between the different driving forces — including mutation, selection, recombination, gene flow, demography or incomplete lineage sorting — that shape nucleotide polymorphism across genomes (Begun & Aquadro 1992; Betancourt & Presgraves 2002; Nordborg & Tavar 2002; Comeron *et al.* 2008; Cutter & Payseur 2013; Gosset & Bierne 2013; Huber *et al.* 2014; Cruickshank & Hahn 2014). Most genetic variability is neutral or nearly so (Kimura 1983), but patterns of polymorphism are also greatly influenced by several types of natural selection whose signatures are heterogeneous in nature and within the genome (Hill & Robertson 1966; Maynard Smith & Haigh 1974; Hudson & Kaplan 1988; Charlesworth *et al.* 1993; Gillespie 1997). One fundamental, unresolved question is how the adaptation of different gene pools to different fitness conditions imposed by spatially varying natural selection in heterogeneous environments (local adaptation) affects their genomic variation (Williams 1966; Kawecki & Dieter 2004; Savolainen *et al.* 2013).

Population genetics theory and empirical models suggest that genomic regions exhibiting high divergence among populations, skewed polymorphism or an

extended correlation among alleles from different loci (linkage disequilibrium) are candidate regions that harbor genes underlying responses to spatially heterogeneous selection pressures (Lewontin & Krakauer 1973; Schlötterer 2002; Beaumont 2005; Nielsen 2005; Storz 2005; Nosil *et al.* 2009). Because migration, gene flow and recombination among diverging populations can break down locally adapted allelic combinations, genetic targets of local adaptation often coincide with genomic regions of reduced recombination – including chromosomal inversions (Noor *et al.* 2001; Rieseberg 2001; Ortiz-Barrientos *et al.* 2002; Butlin 2005; Kirkpatrick & Barton 2006; Joron *et al.* 2011; Roesti *et al.* 2012, 2013; Yeaman 2013; Nishikawa *et al.* 2015).

Structural rearrangements known as chromosomal inversions, which occur when a piece of DNA within a single chromosome breaks and rotates 180° before being reinserted in the reversed orientation (Sturtevant 1921), are widespread in plants and animals (Dobzhansky 1970). A large body of literature supports a major implication of paracentric inversions (which do not encompass the centromere) in evolutionary adaptation in a wide variety of species (see Krimbas & Powell 1992; Powell 1997; Hoffmann *et al.* 2004; Hoffmann & Rieseberg 2008; Kirkpatrick 2010 for a review). This includes many examples of polymorphic inversions whose frequencies are correlated with environmental variables and temporal changes consistent with strong natural selection in dipteran species (Dobzhansky 1943; Mettler *et al.* 1977; Coluzzi *et al.* 1979; Knibb 1982; Krimbas & Powell 1992; De Jong & Bochdanovits 2003; Schaeffer *et al.* 2003; Anderson *et al.* 2005; Schaeffer 2008). Experimental evolution and phenotypic studies also show a remarkable association

between inversions and several fitness-related traits (Wright & Dobzhansky 1946; Dobzhansky 1948; Rako *et al.* 2006; Hoffmann & Weeks 2007; Kennington *et al.* 2007; Lowry & Willis 2010; Lee *et al.* 2011; Fouet *et al.* 2012; Kapun *et al.* 2014, 2016b). More recently, in-depth examinations of DNA sequence polymorphisms have clearly demonstrated the existence of footprints of positive selection within inversions in several species, providing another compelling evidence for the crucial role of chromosomal rearrangements in local adaptation (e.g. Stefansson *et al.* 2005; Corbett-Detig & Hartl 2012; Pool *et al.* 2012; Cheng *et al.* 2012; Marsden *et al.* 2014; O'Loughlin *et al.* 2014; Rane *et al.* 2015; Kapun *et al.* 2016a; Kamdem *et al.* 2017).

Another important property of inversions concerns the significant part they play in genome evolution as a whole. Recombination is strongly reduced between two paired chromosomes that differ by an inversion through several mechanisms, which impede crossing over (reviewed in Roberts 1976). Local recombination rates are expected to affect a myriad of processes including selection, gene conversion, diversity and divergence (Begun & Aquadro 1992; Kliman & Hey 1993; Andolfatto *et al.* 2001; Nordborg *et al.* 2005; Haddrill *et al.* 2007; Kulathinal *et al.* 2008; Comeron *et al.* 2012; Nachman & Payseur 2012; Roesti *et al.* 2013, 2012; Campos *et al.* 2014). Since the pioneering work of Begun and Aquadro (1992), which first demonstrated a strong positive correlation between local rates of recombination and nucleotide diversity within the genome of *Drosophila melanogaster*, multiple studies have provided empirical evidence that reduced crossing over alters genetic variation along chromosomes in a wide variety of organisms (Hellmann *et al.* 2003, 2005; Tenaillon *et al.* 2004; Begun *et al.* 2007; Kulathinal *et al.* 2008, 2009; Pegueroles *et*

103 *al.* 2010; Corbett-Detig & Hartl 2012; McGaugh *et al.* 2012; Pool *et al.* 2012). As
104 coldspots of recombination, inversions thus have the potential to modulate patterns
105 of genetic variation in more or less extended regions of the genome. More
106 specifically, the presence of one or several rearrangements may impede crossing
107 over and reduce diversity along an entire chromosome or chromosome arm, forcing
108 it to evolve differently from the rest of the genome. This scenario is best illustrated
109 by the nonrecombining part of the Y chromosome in mammals, which is maintained
110 by multiple inversions that have suppressed recombination over the entire length of
111 the chromosome (Lahn & Page 1999). The potential of inversions to reduce
112 recombination and diversity across large regions of the genome has long been
113 exploited, especially by *Drosophila* geneticists, to engineer "balancer chromosomes"
114 (Muller 1918; Sturtevant 1921; Hentges & Justice 2004). This consists in
115 chromosomes or chromosomal arms bearing multiple inversions that block
116 recombination, thus providing the possibility to maintain mutations in a
117 heterozygous state. Despite these clear examples, our understanding of how
118 inversions affect polymorphism over the entire length of a chromosome remains
119 limited in part due to a lack of empirical support across taxa. Targeted studies based
120 on fine-scale examinations of genomic variation in species with well-characterized
121 patterns of naturally occurring inversions are particularly important for assessing
122 both the genomic extent and the ubiquity of diversity reduction driven by
123 chromosomal rearrangements.

124 An excellent opportunity to evaluate the role of inversions in genome
125 evolution exists in *An. funestus*, the second most important vector of *Plasmodium*

126 parasites in the Afrotropical region, just behind the best known species, *An. gambiae*
127 (Gillies & De Meillon 1968; Gillies & Coetzee 1987; Sinka *et al.* 2010; Coetzee &
128 Koekemoer 2013). The two taxa share several characteristics, which reflect their
129 increased capacity to track their hosts and to vector malaria parasites, including a
130 near continental distribution, a marked preference for human hosts and
131 intradomicillary host-seeking and resting behavior. The widespread distribution of
132 *An. funestus* is largely due to ongoing adaptive divergence enabling its populations
133 to exploit a significant range of environmental conditions. This mosquito represents
134 at least two relatively differentiated ecotypes whose ecology, phenotypic
135 divergence, distribution range and role in malaria transmission remain obscure
136 (Green & Hunt 1980; Costantini *et al.* 1999; Dia *et al.* 2000; Kamau *et al.* 2002;
137 Boccolini *et al.* 2005; Cohuet *et al.* 2005; Michel *et al.* 2005; Ayala *et al.* 2011; Barnes
138 *et al.* 2017). Cytogenetic studies have identified an impressive number of
139 paracentric polymorphic chromosomal inversions in polytene chromosomes of *An.*
140 *funestus* (Green & Hunt 1980; Sharakhov *et al.* 2004). The existence of stable
141 inversion clines spanning large geographic areas as well as the significant deficit of
142 heterokaryotypes in different parts of the continent clearly indicate the adaptive
143 role of some of these rearrangements (Costantini *et al.* 1999; Dia *et al.* 2000; Kamau
144 *et al.* 2002; Boccolini *et al.* 2005; Ayala *et al.* 2011). Importantly, the non-uniform
145 distribution of chromosomal inversions within the genome of *An. funestus* provides
146 an ideal framework for cross-chromosomal comparison. Indeed, all the autosomal
147 arms are characterized by the presence of numerous large rearrangements except
148 2L, which is perfectly collinear (Sharakhov *et al.* 2004).

149 In this study, we have used Restriction site Associated DNA Sequencing
150 (RAD-Seq) (Baird *et al.* 2008) to examine the effects of inversions on the genetic
151 architecture of selection and diversity across the genome of *An. funestus* ecotypes
152 present in Cameroon. We detected signatures of spatially varying selection affecting
153 large structural variants on chromosomes 2R and 3R. Ecotypic differentiation in this
154 geographic area is driven primarily by genetic divergence within the 3R*a*, 3R*b* and
155 2R*h* rearrangements. Strikingly, genome scans reveal that, in contrast to the
156 collinear 2L arm, the amount of polymorphic sites is drastically reduced along the
157 entire length of three autosomal arms bearing multiple polymorphic inversions.
158 These findings suggest that the combined effects of suppressed recombination and
159 selection within paracentric polymorphic inversions drive diversity reduction at the
160 chromosomal level in *An. funestus*.
161
162 **Materials and methods**
163 **Mosquito samples**
164 Two ecotypes of *An. funestus* are present in Cameroon: one in the northern
165 savannah area and the other across the southern rainforest biome (Cohuet *et al.*
166 2005; Ayala *et al.* 2011). We sequenced 132 mosquitoes collected from two
167 locations representative of these two ecogeographic domains and separated by
168 ~500km (Fig. 1A and Table S1). In Mfou, a small city of the southern forest region,
169 *An. funestus* breeds in an artificial lake and maintains abundant populations
170 throughout the year. In Tibati, which lies within the forest-savannah transition zone,
171 several artificial lakes also provide abundant breeding sites for dense and perenial

172 populations of *An. funestus*. From August to November 2013, we used several
173 sampling methods (Service 1993) to collect *An. funestus* mosquitoes at distinct
174 stages of their life cycle, from different microhabitats and from different time points
175 during the nocturnal biting cycle (Fig. 1A and Table S1). This approach aimed at
176 maximizing the chances that our sample represents a good approximation of the
177 genetic diversity within populations of *An. funestus*. This species is a member of a
178 large taxonomic group comprising at least seven related taxa that are
179 distinguishable by slight morphological variations and a diagnostic PCR, which
180 targets mutations of the ribosomal DNA (Gillies & De Meillon 1968; Gillies & Coetzee
181 1987; Cohuet *et al.* 2003). All samples included in this study were identified as *An.*
182 *funestus*.

183 **Library preparation, sequencing and SNP identification**

184 We extracted genomic DNA of larvae and adult specimens with the DNeasy
185 Blood and Tissue kit (Qiagen) and the Zymo Research MinPrep kit, respectively.
186 Double-digest Restriction-site Associated DNA (ddRAD) libraries were prepared as
187 described in Kamdem *et al.* (2017) following a modified protocol of Peterson *et al.*
188 (2012) and single-end sequenced to 100 base reads on Illumina HiSeq2000.

189 Illumina sequences were sorted according to barcode tag and filtered using
190 the *process_radtags* program of the Stacks v 1.35 software (Catchen *et al.* 2013).
191 Reads with ambiguous barcode, inappropriate restriction site or low sequencing
192 quality score were removed. We checked the depth of sequencing coverage of the
193 final data set after all filtering steps using VCFtools (Danecek *et al.* 2011). To call and
194 genotype SNPs, we aligned the remaining high-quality reads to the *An. funestus*

reference sequence with GSNAP (Wu & Nacu 2010) and used Stacks to identify SNPs within each RAD locus. Following assembly and genotyping, the polymorphism data was further filtered to maximize data quality. We retained only one randomly selected SNP per RAD locus scored in every population and in at least 60% of individuals within population for further analyses. SNP files in different formats used for downstream analyses were created with the *populations* program in Stacks, PLINK v 1.09 and PGDSpider v 2.0.8.2 (Purcell *et al.* 2007; Lischer & Excoffier 2012; Catchen *et al.* 2013).

Population structure and genetic divergence

We examined the genetic relatedness among individuals with a Principal Component Analysis (PCA), a neighbor-joining (NJ) tree and the STUCTURE v 2.3.4 software using SNPs identified in Stacks (Pritchard *et al.* 2000). We utilized the R package *adegenet* to implement the PCA and *ape* to construct individual-based NJ networks using allele frequencies of SNPs and the Euclidian distance (Paradis *et al.* 2004; Jombart 2008; R Development Core Team 2016). In STRUCTURE, we ran five replicates of 1 to 10 assumed clusters (k). Each run consisted of 200000 iterations, and the first 50000 iterations were discarded as burn-in. CLUMPP v1.1.2 (Jakobsson & Rosenberg 2007) was used to aggregate results across multiple STRUCTURE runs and the clustering results were visualized graphically using DISTRUCT v1.1 (Rosenberg 2004). To find the number of genetically distinct clusters, we used both the Discriminant Analysis of Principal Component (DAPC) implemented in *adegenet* and the ad hoc statistic DeltaK of Evanno *et al.* (2005) (Evanno *et al.* 2005; Earl & VonHoldt 2012).

To assess the level of genetic differentiation between ecotypes, we estimated the overall F_{ST} (Weir & Cockerham 1984) in Genodive v1.06 (Meirmans & Van Tienderen 2004) using a subset of 2000 randomly selected SNPs. Additionally, to quantify the contribution of regional and micro-spatial distributions of samples on the genetic variance, we conducted a hierarchical analysis of molecular variance (AMOVA) (Excoffier *et al.* 1992) in GenoDive. We used 10000 permutations to assess the statistical significance of F_{ST} and AMOVA.

Genomic targets of selection

The current *An. funestus* draft genome assembly consists of 12243 scaffolds ranging from 2334 to 3334433bp in length (Giraldo-Calderon *et al.* 2015). Therefore, to perform genome scans and to identify footprints of selection throughout the genome, we ordered and concatenated 104 long scaffolds that had been physically mapped to chromosomes (Neafsey *et al.* 2015) to create "pseudo-chromosomes" corresponding to the five *An. funestus* chromosome arms. Mapped scaffolds accounted for 33% in length of the approximate size of the *An. funestus* genome (Fig. S1 and Table S2).

To detect signatures of selection within the genome of *An. funestus*, we examined the distribution of genetic diversity, allele frequency spectrum and genetic differentiation on RAD loci located on mapped scaffolds. We calculated pairwise nucleotide diversity ($\theta\pi$), Watterson's estimator of segregating sites (θw) and Tajima's *D* in ANGSD v 0.612 (Korneliussen *et al.* 2014) using aligned reads in BAM format as input files (Li *et al.* 2009). ANGSD derives diversity and allele frequency spectrum statistics using genotype likelihoods without SNP calling,

241 thereby alleviating the uncertainty and biases associated with SNP calling in low
242 coverage Next Generation Sequencing (Korneliussen *et al.* 2013). To gain genome-
243 wide view and identify genomic regions of exceptional diversity and/or skewed
244 allele frequency spectra, average values of $\theta\pi$, θw and Tajima's *D* were plotted in
245 non-overlapping 90-kb windows along pseudo-chromosomes.

246 To identify highly differentiated genomic regions that are candidate regions
247 under divergent selection among ecotypes, locus-specific F_{ST} values were estimated
248 with the *populations* program in *Stacks.* The distribution of pairwise genetic
249 differentiation throughout the genome was visualized using SNPs located on
250 mapped scaffolds. SNPs with F_{ST} values above the top 1% of the empirical
251 distribution were considered as outliers of genetic differentiation. Loci with
252 unusually low or high F_{ST} values relative to neutral expectations were also detected
253 using the coalescence-based method FDIST2 (Beaumont & Nichols 1996) as
254 implemented in LOSITAN (Antao *et al.* 2008). The mean neutral F_{ST} across all SNPs
255 was approximated by choosing the "neutral mean F_{ST} option" with 99% confidence
256 interval in LOSITAN. We ran LOSITAN with 100000 simulations and assumed a false
257 discovery rate of 0.1 to minimize the number of false positives. F_{ST} values are
258 dependent on within-population genetic diversity, which may bias estimates of the
259 level of divergence among populations (Noor & Bennett 2009; Cruickshank & Hahn
260 2014). We used two additional statistics to further assess the degree of genetic
261 divergence across the genome — the absolute sequence divergence (d_{xy}) and the
262 proportion of fixed differences between populations (d_f). Both statistics were
263 estimated in ngsTools using genotype likelihood without SNP calling (Fumagalli *et*

al. 2014), and kernel smoothed values were visualized in non-overlapping 90-kb windows along pseudo-chromosomes in R.

Demographic history of *An. funestus*

Estimates of genetic diversity within mosquito populations can be used as a proxy for their adaptive potential and their resilience against vector control measures. To this end, we used the diffusion approximation method implemented in the software package ∂a∂i v 1.6.3 (Gutenkunst *et al.* 2009) to model the genetic variation and infer the demographic history of *An. funestus*. Using 9412 filtered SNPs, we fitted one-population models describing simple demographic scenarios without migration or recombination events to the folded allele frequency spectrum of our samples. The best model was selected using the highest composite log-likelihood, the lowest Akaike Information Criterion (AIC) and the lowest uncorrelated residuals. We used the built-in bootstrap function implemented in ∂a∂i to derive 95% bootstrap confidence intervals for the different demographic parameters.

Results

Genetic differentiation among *An. funestus* ecotypes

Using alignments to the *An. funestus* draft reference genome, we assembled 490871 unique RAD loci. We identified 10687 high-quality biallelic markers across loci that were present in all populations and in at least 60% of individuals within every population while randomly selecting one SNP per locus. A NJ tree and the first three PCA axes based on these variants clearly distinguished two genetic clusters corresponding presumably to the two ecotypes previously described in Cameroon

and across the continent with a diversity of genetic markers (Costantini *et al.* 1999; Dia *et al.* 2000; Kamau *et al.* 2002; Boccolini *et al.* 2005; Cohuet *et al.* 2005; Michel *et al.* 2005; Ayala *et al.* 2011; Barnes *et al.* 2017) (Fig. 1B and 1C). Both the method of Evanno *et al.* (2005) and DAPC confirmed the existence of only two to three distinct gene pools reflecting the ecological divergence known between forest and savannah populations in Cameroon (Fig. 1D and 1E). However, despite this clear geographic segregation, the overall genetic differentiation is low between the two putative subgroups (F_{ST} = 0.033, p < 0.005). In support of weak genetic divergence, STRUCTURE analyses reveal a single admixed population rather than two distinct genetic clusters at k = 2 (Fig. 1F). The moderate genetic differentiation is also illustrated by the results of an AMOVA, which show that the most significant proportion of the genetic variance (98.2%, p < 0.005) among our samples is explained by within-individual variations. The geographic origin accounts for less than 2% of the genetic variance, confirming that extensive gene flow among populations likely overwhelms ongoing local adaptation.

Strong divergent selection within chromosomal inversions

First, we observed a non-random distribution of highly differentiated loci across 10687 SNPs, characterized by an aggregation of the highest F_{ST} values in genomic regions bearing known polymorphic chromosomal inversions (Fig. 2 and Fig. S1). We identified a total of 107 outliers that fell above the 99th percentile of the empirical distribution of F_{ST} throughout the genome. Among these outliers, 31 SNP loci were successfully mapped to concatenated scaffolds and were confirmed as statistical outliers in LOSITAN. Importantly, highly differentiated loci were located

310 exclusively on two chromosomes (3R and 2R), which both harbor at least 20
311 polymorphic inversions (Fig. 2 and Fig. S1) (Sharakhov *et al.* 2004). Precisely, in
312 contrast to chromosomes 2L, 3L and X, which have no discriminatory power, SNPs
313 of 3R and to a lesser extent 2R reproduce the segregation observed between
314 ecotypes at the genome level consistent with the hypothesis that genomic regions
315 that have the greatest impact on local adaptation are located on these two
316 chromosome arms (Fig. 3). The number of SNPs with F_{ST} values above the 1%
317 threshold also varies between the 2R and 3R (9 against 22) implying that mutations
318 or structural variants of the 3R arm are more strongly correlated with genetic
319 differentiation in this geographic region.

320 We looked further into the role of inversions located respectively on 2R and
321 3R in adaptive divergence by examining the population structure separately for four
322 large chromosomal rearrangements present on these two chromosomes. Figure 3
323 shows the genetic relatedness among samples inferred from SNPs located on
324 scaffolds that mapped to the three inversions present on 3R (3R*a*, 3R*b* and 3R*d*) and
325 the 2R*h* that occurs on 2R. Although several inversions coexist and overlap
326 extensively along 2R, six out of nine F_{ST} outliers identified on this chromosome
327 mapped to scaffolds specific to portions of the 2R*h* inversion, making it an
328 interesting candidate locus under divergent selection. SNPs located within the 3R*a*
329 inversion assign more that 90% of individuals to their respective ecotypes, and the
330 3R*b* also separates a significant proportion of individuals between both sampling
331 sites (Fig. 3). These results corroborate the genomic pattern of differentiation,
332 which indicates that 83% of F_{ST} outliers map within these two inversions (Fig. 2). In

addition, consistent with the F_{ST} values, the genome-wide distribution of d_{xy} and fixed differences shows clear peaks within these two large inversions, which provides further support for their significant contribution to ecological divergence (Fig. 4). The population structure of 2R*h* is comparable with that of 3R*b*, indicating that this inversion also contributes to the savannah-forest segregation of mosquitoes. Conversely, SNPs located outside the 3R*a* and 3R*b*, or within the 3R*d* have no discriminatory power (Fig. 3). In summary, the population structure of inversions located on 2R and 3R largely reflects the genomic distribution of differentiated loci, which suggests that signatures of divergent selection between *An. funestus* ecotypes in Cameroon are particularly strong within two polymorphic inversions (3R*a* and 3R*b*) and to a lesser extent the 2R*h* rearrangement. This strong divergent selection translates into F_{ST} values, which increase dramatically, from 0.033 at the genome level to 0.053, 0.08 and 0.22 within the 2R*h,* 3R*b* and 3R*a* inversions, respectively. Similarly, the proportion of the genetic variance explained by the geographic origin of samples rises from 1.8% for the genome-wide SNPs to 3.2%, 4.7% and 12.9% when only variants present within the 2R*h*, 3R*b* and 3R*a* rearrangements, respectively, are included.

Chromosome arm-specific diversity associated with inversions

We further explored patterns of polymorphism across the genome using genome scans based on estimates of nucleotide diversity ($\theta\pi$ and θw) (Fig. 5). Surprisingly, nucleotide diversity is not particularly depressed within the 3R*a*, 3R*b* or 2R*h* inversions relative to the rest of the genome as expected due to the strong divergent selection observed across these loci. Instead, we noted that the amount of

356 polymorphic sites (θw) varies substantially between chromosome arms (Fig. 5 and
357 Fig. 6). The highest values of θw are found on the single chromosome arm that
358 harbours no known inversion (2L). In addition, the average value of θw differs
359 significantly between 2L and each of the three other autosomes (Wilcoxon rank sum
360 test, $p < 0.001$) (Fig. 6). More precisely, θw is reduced by 34.1% on 3R, 24.0% on 2R
361 and 13.2% on 3L relative to the 2L arm, regardless of the sampling location. The
362 amount of polymorphic sites is also significantly different along each of the five
363 chromosome arms (Wilcoxon rank sum test, $p < 0.001$) consistent with a different
364 demographic history between the two ecotypes. Forest populations from Mfou are
365 fixed for the major polymorphic chromosomal inversions (notably 3R*a* and 3R*b*)
366 whose frequencies fluctuate in Tibati (savannah) (Cohuet *et al.* 2005; Ayala *et al.*
367 2011). This may at least in part contribute to the stronger reduction of nucleotide
368 diversity on inverted autosomes observed in Mfou. The drop of the average θw on
369 inverted autosomes relative to 2L is 34.7% on 3R, 29.2% on 2R and 13.2% on 3L in
370 Mfou compared with 33.5% on 3R, 18.8% on 2R and 13.2% on 3L in Tibati (Fig. 5
371 and Table 1). The X chromosome harbours no known chromosomal rearrangement,
372 but presumably because of its particular divergence and its distorted effective
373 population size, the average θw diversity is also weak compared with 2L (21.1%
374 reduction relative to 2L) regardless of the sampling site. A significant reduction in
375 pairwise nucleotide diversity ($\theta\ \pi$) relative to the 2L is also detected on
376 chromosomes 3R, X and 2R to a lesser extent (Fig. 5, Fig. 6 and Table 1). However,
377 the specificity of the collinear chromosome arm compared with the rest of the

genome is less evident, and at the exception of 2R, there is no statistically significant difference in the distribution of $\theta\pi$ at the level of chromosomes between ecotypes.

Local recombination rates are known to be strongly correlated with genetic diversity and the effective population size (Begun & Aquadro 1992). Therefore, it is possible and parsimonious that the chromosome-bias polymorphism observed among autosomes in *An. funestus* is due to the presence of multiple overlapping inversions that are under selection in different geographic contexts throughout the species' range. These inversions block meiotic crossover and depress genetic variation along the 2R, 3R and 3L autosomes, which contain multiple rearrangements. On the 3R arm for instance, three inversions (3R*a*, 3R*b* and 3R*d*) account for at least 75% of the length of this chromosomal arm (Fig. S1). Similarly, at least 80% of the length of 2R and 3L are affected by chromosomal rearrangements, likely leading to minimal crossover rates over the entire chromosome arm. The hypothesis that the combined effects of suppressed recombination and selection within multiple inversions drive polymorphism reduction throughout the three autosomal arms is clearly supported by the diversity pattern exhibited by the collinear chromosome arm 2L, which acts as a negative control. Conceivably, polymorphism is highest on this chromosome arm due to the absence of inversions, which preserves substantial levels of recombination.

Demographic history

The overall diversity falls in the range described in other highly polymorphic *Anopheles* populations with RADseq markers ($\theta w = 0.0083$ and $\theta\pi = 0.0042$ in Mfou; $\theta w = 0.0097$ and $\theta\pi = 0.0043$ in Tibati) (O'Loughlin *et al.* 2014; Fouet *et al.*

2017; Kamdem *et al.* 2017). To infer the demographic history of our populations, we first examined the genome-wide distribution of Allele Frequency Spectra (AFS), summarized as Tajima's *D* (Tajima 1989). The range of values of Tajima's *D* (from -2.47 to -1.58 among Tibati samples and from -2.20 to -1.19 in Mfou) was shifted towards negative values suggestive of recent population expansion leading to an accumulation of low-frequency variants (Fig. 5) (Donnelly *et al.* 2001). Unsurprisingly, Tajima's *D* is more negative on the collinear 2L chromosome, which exhibits the greatest difference between the amount of polymorphic sites and the pairwise nucleotide diversity in both ecotypes. We next modeled the AFS to infer the demographic history of populations in ∂a∂i. We used 9412 genome-wide SNPs while excluding variants located on the X chromosome to prevent the confounding effects due to the smaller population effective size of this chromosome. Also, to avoid strong disparities in sample sizes that may affect our results, we selected 35 individuals from each sampling site, prioritizing, whenever possible, individuals with high sequencing depths. As suggested by the genome-wide distribution of Tajima's *D*, the autosomal polymorphism in both Tibati and Mfou samples reflects population expansion models. Precisely, the best-fit model shows exponential growth in Tibati and two-epoch growth in Mfou (Table S3). Consistent with the greater genome-wide diversity observed in savannah populations, model parameters also reveal more substantial population growth in Tibati (Table S3).

Discussion

Recombination, inversions and genetic divergence

Our analysis based on 10687 genome-wide SNPs confirms the weak genetic differentiation previously described between *An. funestus* ecotypes that are adapted respectively to moist habitats typical of the southern forest region and to more xeric areas in the North of Cameroon (Cohuet *et al.* 2005; Ayala *et al.* 2011). As also suggested by previous studies from different parts of the continent, divergence between ecotypes is limited to a few loci bearing structural rearrangements, which segregate along latitudinal clines in certain countries (Green & Hunt 1980; Costantini *et al.* 1999; Dia *et al.* 2000; Cohuet *et al.* 2005; Ayala *et al.* 2011). Our conclusions come with caveats due to the potential limitations of the sampling scheme, the reference genome and the RAD sequencing approach we used. For instance, it is likely that some genomic signatures of selection remain undetected because of the limited genomic coverage of RAD tags and pseudo-chromosomes (Arnold *et al.* 2013; Tiffin & Ross-Ibarra 2014). Bearing in mind these caveats, the clustering of genetic differentiation within the 3R*a*, 3R*b* and 2R*h* inversions appears to be consistent with the fact that these rearrangements are the major genetic targets of local adaptation in Cameroon. The 3R*a* and 3R*b* inversions in particular have long been suspected to be the most important drivers of divergent selection due to the clinal distribution of alternative karyotypes and the significant deficit in heterokaryotypes observed in several countries (Costantini *et al.* 1999; Dia *et al.* 2000; Ayala *et al.* 2011).

445 The phenotypic, behavioral or ecological variations associated with
446 inversions under divergent selection remain unclear. In principle, the segregation
447 between the arid zone and the humid forest may potentially involve fitness traits
448 and phenotypes contributing to thermal tolerance. Indeed, examples of inversions
449 whose alternative karyotypes provide selective advantages in different moisture
450 conditions are well known in dipteran species (e.g. the 2L*a* inversion and to a lesser
451 extent the 2R*b* in *An. gambiae*) (Coluzzi *et al.* 1979; Gray *et al.* 2009; Lee *et al.* 2009;
452 Fouet *et al.* 2012; Cheng *et al.* 2012). Coincidentally, comparative cytogenetic
453 studies also showed a high proportion of conserved genes between the 3R*b* in *An.*
454 *funestus* and the 2L*a* in *An. gambiae* and between the 2R*h* in *An. funestus* and the
455 2R*b* in *An. gambiae* (Sharakhova *et al.* 2011). Although this hypothesis awaits
456 thorough investigation, the two pairs of inversions may have evolved similar
457 phenotypic patterns through shared ancestry or convergent evolution between *An.*
458 *funestus* and *An. gambiae*.

459 Theoretical models and empirical data support the idea that, at early stages
460 of divergence as in the case of *An. funestus* ecotypes, genetic differentiation is
461 restricted to a few genomic regions because the effects of natural selection at this
462 step are very localized (Feder *et al.* 2012; Nosil & Feder 2012; Andrew & Rieseberg
463 2013; Seehausen *et al.* 2014). Also, selection at a locus affects the level of genetic
464 differentiation among populations and reciprocally the degree of subdivision within
465 a population impacts patterns of variation at selected loci (Lewontin & Krakauer
466 1973; Charlesworth *et al.* 1997; Slatkin & Wiehe 1998; Majewski & Cohan 1999; Kim
467 & Maruki 2011; Schneider & Kim 2013). In fact, it has been proposed that selection

468 modulates levels of divergence locally and globally across the genome through two
469 types of genetic hitchhiking — divergence hitchhiking and genomic hitchhiking
470 (Feder *et al.* 2012). Divergent hitchhiking (the spread of a mutation and linked
471 neutral sequences within a population) (Maynard Smith & Haigh 1974) causes the
472 diffusion of favorable alleles at loci important for local adaptation and ecological
473 differentiation. Limited gene flow that occurs between divergent lineages can
474 naturally give rise to genomic hitchhiking, which reflects the global accumulation of
475 differences at the genome level through different selective or demographic
476 mechanisms (Feder *et al.* 2012). In *An. funestus*, it is plausible that divergent
477 hitchhiking within the three important inversions, 3R*a*, 3R*b* and 2R*h*, is so far
478 strongly counterbalanced by extensive gene flow and the weak overall number of
479 selected loci across the genome, leading to minimal genomic divergence among
480 ecotypes.

481 Although some controversy persists (see McGaugh *et al.* 2012), the idea that
482 strongly differentiated genomic regions between species or between populations
483 within the same species accumulate disproportionately in regions of low genetic
484 recombination including chromosome centers and chromosomal rearrangements
485 has been strongly supported in many species notably humans, *Drosophila* flies,
486 stickleback fish and maize (Hellmann *et al.* 2003, 2005; Tenaillon *et al.* 2004;
487 Kulathinal *et al.* 2008; Cai *et al.* 2009; Keinan & Reich 2010; Nachman & Payseur
488 2012; McGaugh & Noor 2012; Roesti *et al.* 2012, 2013). This coincidence can be
489 explained by a mechanical effect of recombination, which reduces genetic
490 polymorphism within populations and thereby generates sequence divergence

between populations as a byproduct of diversity reduction (Begun & Aquadro 1992; Roesti *et al.* 2012). Concordance between limited rate of crossing over and increased divergence may also be directly correlated with more pronounced effects of divergent selection in genomic regions in which recombination is less frequent (Charlesworth *et al.* 1997; Charlesworth 1998; Nachman 2002; Cutter & Payseur 2013; Cruickshank & Hahn 2014). Within inversions in particular, reduced recombination enhances divergent selection acting on locally adapted gene complexes, which co-segregate with or without epistatic interactions (Dobzhansky 1950; Kirkpatrick & Barton 2006).

Recombination, inversions and genetic polymorphism

We have observed chromosome arm specific patterns of polymorphism in *An. funestus* characterized by a significant difference between collinear and inverted chromosomes. We first conducted several tests to insure that uncertainties associated with our sequencing and analytical approach cannot account for the observed disparities in genetic polymorphism between chromosomes. We compared the distribution and the length of mapped scaffolds and confirmed that each chromosome arm is represented by a substantial number of long scaffolds (Table S2 and Fig. S1). The length of pseudo-chromosomes is also proportional to the length of chromosome arms, ranging from 6.8 Mb on X (the smallest) to 27.4 Mb on 2R (the largest). Therefore, we can reasonably rule out the possibility that the chromosome-bias diversity is due to a systematic error associated with our reference sequence. Additionally, mapped scaffolds are evenly distributed along the length of chromosomes, suggesting that such important variations in the amount of

514 polymorphism between chromosome arms are not due to centromere- and/or
515 telomere-proximal effects (Fig. S1) (Aguade *et al.* 1989; Stephan & Langley 1989).
516 Finally, we noted that the mean depth of sequencing coverage per mapped scaffolds
517 was consistent across the five chromosome arms (Fig. S2), implying that
518 chromosome-specific diversity is not simply a covariate of sequencing biases.

519 As estimates of the number of segregating sites (θw) also known as the
520 population-scaled mutation rate (Watterson 1975) are the most drastically affected
521 by between-chromosome variations observed in *An. funestus*, the stark contrast in
522 the amount of polymorphism among inverted and collinear autosomes may in
523 theory be due to lower mutation rates on inverted chromosomes. However, as
524 shown in other insect species, such a variability in mutation rates between
525 chromosomes or between large segments of the genome is unlikely (Begun *et al.*
526 2007; Keightley *et al.* 2015). Instead, the occurrence of chromosome-specific
527 patterns of diversity in *An. funestus* is consistent with the presence of weakly
528 recombining autosomes bearing multiple overlapping chromosomal
529 rearrangements. These same causes produce similar effects along the non-
530 recombining portion of the Y chromosome in mammals (Lahn & Page 1999) or along
531 balancer chromosomes described in *Drosophila* and rodent species (Muller 1918;
532 Hentges & Justice 2004). Indeed, population genetic analyses across genomes of
533 diverse taxa have found a positive correlation between the rate of recombination
534 and genetic variation (Aguade *et al.* 1989; Stephan & Langley 1989; Begun &
535 Aquadro 1992; Hellmann *et al.* 2003, 2005; Tenaillon *et al.* 2004; Begun *et al.* 2007;
536 Kulathinal *et al.* 2008, 2009). This correlation translates into specific patterns of

537 diversity that can be observed at the species, genomic or chromosomal levels. At the
538 species level, plants and some animal species, which reproduce by self-fertilization
539 have reduced overall genomic diversity due to low genetic recombination
540 (Nordborg *et al.* 1996; Akhunov *et al.* 2010; Cutter & Choi 2010; Andersen *et al.*
541 2012; Thomas *et al.* 2015). The local effects of recombination on genetic
542 polymorphism are also particularly evident across centromere and telomeres of
543 chromosomes and within genomic regions bearing structural rearrangements
544 (Aguade *et al.* 1989; Stephan & Langley 1989; Andolfatto *et al.* 2001; Corbett-Detig &
545 Hartl 2012; Pool *et al.* 2012).

546 The relationship between diversity and crossing over has been ascribed to
547 two processes: selection and mutagenesis. Recombination may influence
548 polymorphism because of new mutations created during crossing over, which
549 increase diversity (the mutagenic effect of recombination) (Magni & Von Borstel
550 1962). However, multiple studies that have analyzed fine-scale recombination in
551 humans, yeast, *Arabidopsis*, *Anopheles*, *Drosophila*, and *Caenorhabditis* have
552 undermined the notion that recombination, or some correlate, generally exerts
553 mutagenic effects on genomes (Betancourt & Presgraves 2002; Stump *et al.* 2005;
554 Spencer *et al.* 2006; Wright *et al.* 2006; Begun *et al.* 2007; Noor 2008; Denver *et al.*
555 2009; Cutter & Choi 2010; Mcgaugh *et al.* 2012). Alternatively, recombination may
556 affect diversity indirectly by modulating the effects of positive or negative selection
557 across the genome. Indeed, the hitchhiking of favorable alleles (selective sweep) or
558 the removal of recurrent deleterious mutations (background selection) affect linked
559 neutral sequences more strongly in low-recombination genomic regions, which

gives rise to a positive correlation between diversity and recombination rate (Maynard Smith & Haigh 1974; Kaplan *et al.* 1989; Begun & Aquadro 1992; Charlesworth *et al.* 1993, 1997; Nordborg *et al.* 1996; Andolfatto 2001; Nachman 2002). Another indirect effect of recombination on genetic diversity may be driven by the Hill-Roberston interference, which occurs when two sites under weak selection are in physical linkage and as a result, selection at one site interferes with selection at another site (Hill & Robertson 1966; Felsenstein 1974). Hill-Roberston interference can also be thought of as a reduction in the effective population size (N_e) caused by selection at linked loci (Comeron *et al.* 2008; Cutter & Payseur 2013; Castellano *et al.* 2016). Recombination, by alleviating interference between linked sites, alleviates this reduction in N_e leading to a positive correlation between recombination rate and levels of neutral polymorphism. Because of the relationship between recombination and polymorphism, measures of the skew in the allele frequency spectrum (AFS), such as Tajima's *D* values (Tajima 1989), are expected to positively correlate with the rate of recombination (Braverman *et al.* 1995). The *An. funestus* genome shows a strong relationship between Tajima's *D* values and recombination, and as expected, the chromosome 2L, which recombines freely, exhibits the sharpest skew in the AFS.

Conclusions and perspectives

Our results reveal that *An. funestus* has a large effective population size and consequently a significant adaptive potential necessary to buffer the stochastic effects of population declines often associated with large-scale vector controls such as the programs that are currently underway in most Sub-Saharan Africa (WHO

583 2016). We found evidence that three chromosomal rearrangements among the
584 many identified in *An. funestus* are the primary targets of local adaptation among
585 ecotypes present in Cameroon. The other rearrangements are likely either
586 cosmopolitan or endemic inversions under selection at various geographic extents
587 that have yet to be resolved. Our data support the idea that interactions between
588 recombination and selection — which amplify the effects of selective sweeps or
589 background selection within genomic regions where recombination is blocked by
590 multiple chromosomal inversions — account for the strong disparity in nucleotide
591 diversity observed between autosomes in this mosquito.

592 To deepen our understanding of the adaptive role of inversions and their
593 contribution to chromosome-scale diversity patterns, a complete reference genome
594 assembly and extensive sampling across the species range are needed to conduct
595 more sensitive tests including the functional characterization of footprints of
596 selection and their detailed signatures among individuals and populations. The
597 presence of hallmarks of both reduced recombination and linked selection across
598 large genomic sequences in *An. funestus* highlights the important contribution of
599 multiple aspects of linkage in genome evolution. These aspects have significant
600 implications for the detection of genomic signatures of natural selection in species
601 whose genome contains multiple polymorphic chromosomal inversions.

602

603

Acknowledgements

Funding for this project was provided by the University of California Riverside and NIH grants 1R01AI113248 and 1R21AI115271 to BJW.

Author contributions

CK, CF and BJW conceived, designed and performed the experiments, analysed the data and wrote the paper.

Data Archiving Statement

Data for this study are available at: to be completed after manuscript is accepted for publication

References

Aguade M, Miyashita N, Langley CH (1989) Reduced variation in the yellow-achaete-scute region in natural populations of Drosophila melanogaster. *Genetics*, **122**, 607–15.

Akhunov ED, Akhunova AR, Anderson OD *et al.* (2010) Nucleotide diversity maps reveal variation in diversity among wheat genomes and chromosomes. *BMC genomics*, **11**, 702.

Andersen EC, Gerke JP, Shapiro J a *et al.* (2012) Chromosome-scale selective sweeps shape Caenorhabditis elegans genomic diversity. *Nature genetics*, **44**, 285–90.

Anderson AR, Hoffmann AA, McKechnie SW, Umina PA, Weeks AR (2005) The latitudinal cline in the In (3R) Payne inversion polymorphism has shifted in the last 20 years in Australian Drosophila melanogaster populations. *Molecular Biology and Evolution*, **14**, 851–858.

Andolfatto P (2001) Adaptive hitchhiking effects on genome variability. *Current Opinion in Genetics and Development*, **11**, 635–641.

Andolfatto P, Depaulis F, Navarro A (2001) Inversion polymorphisms and nucleotide variability in Drosophila. *Genetical research*, **77**, 1–8.

Andrew RL, Rieseberg LH (2013) Divergence is focused on few genomic regions early in speciation: incipient speciation of sunflower ecotypes. *Evolution; international journal of organic evolution*, **67**, 2468–82.

Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G (2008) LOSITAN: a workbench to detect molecular adaptation based on a Fst-outlier method. *BMC bioinformatics*, **9**, 323.

640 Arnold B, Corbett-Detig RB, Hartl D, Bomblies K (2013) RADseq underestimates
641 diversity and introduces genealogical biases due to nonrandom haplotype
642 sampling. *Molecular Ecology*, **22**, 3179–3190.
643 Ayala D, Fontaine MC, Cohuet A *et al.* (2011) Chromosomal inversions, natural
644 selection and adaptation in the malaria vector Anopheles funestus. *Molecular*
645 *biology and evolution*, **28**, 745–758.
646 Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP Discovery and Genetic
647 Mapping Using Sequenced RAD Markers. *PLoS ONE*, **3**, e3376.
648 Barnes KG, Weedall GD, Ndula M *et al.* (2017) Genomic Footprints of Selective
649 Sweeps from Metabolic Resistance to Pyrethroids in African Malaria Vectors
650 Are Driven by Scale up of Insecticide-Based Vector Control. *PLoS Genetics*,
651 **13(2): e10**, 1–22.
652 Beaumont MA (2005) Adaptation and speciation : what can F st tell us ? *Trends in*
653 *Ecology & Evolution*, **20**, 435–440.
654 Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of
655 population structure. *Proceedings of the Royal Society B: Biological Sciences*,
656 **263**, 1619–1626.
657 Begun D, Aquadro C (1992) Levels of naturally occurring DNA polymorphism
658 correlate with recombination rates in D. melanogaster. *Nature*, **356**, 519–520.
659 Begun DJ, Holloway AK, Stevens K *et al.* (2007) Population genomics: whole-genome
660 analysis of polymorphism and divergence in Drosophila simulans. *PLoS biology*,
661 **5**, e310.
662 Betancourt AJ, Presgraves DC (2002) Linkage limits the power of natural selection in

663 Drosophila. *Proceedings of the National Academy of Sciences of the United States*
664 *of America*, **99**, 13616–20.
665 Boccolini D, Carrara GC, Dia I *et al.* (2005) Chromosomal differentiation of
666 Anopheles funestus from Luanda and Huambo Provinces, western and central
667 Angola. *American Journal of Tropical Medicine and Hygiene*, **73**, 1071–1076.
668 Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W (1995) The
669 hitchhiking effect on the site frequency spectrum of DNA polymorphisms.
670 *Genetics*, **140**, 783–796.
671 Butlin RK (2005) Recombination and speciation. *Molecular Ecology*, **14**, 2621–2635.
672 Cai JJ, Macpherson JM, Sella G, Petrov DA (2009) Pervasive hitchhiking at coding and
673 regulatory sites in humans. *PLoS Genetics*, **5**.
674 Campos JL, Halligan DL, Haddrill PR, Charlesworth B (2014) The relation between
675 recombination rate and patterns of molecular evolution and variation in
676 drosophila melanogaster. *Molecular Biology and Evolution*, **31**, 1010–1028.
677 Castellano D, Coronado-Zamora M, Campos JL, Barbadilla A, Eyre-Walker A (2016)
678 Adaptive evolution is substantially impeded by hill-Robertson interference in
679 drosophila. *Molecular Biology and Evolution*, **33**, 442–455.
680 Catchen J, Hohenlohe P a., Bassham S, Amores A, Cresko W. (2013) Stacks: An
681 analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.
682 Charlesworth B (1998) Measures of divergence between populations and the effect
683 of forces that reduce variability. *Molecular Biology and Evolution*, **15**, 538–543.
684 Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious
685 mutations on neutral molecular variation. *Genetics*, **134**, 1289–1303.

686 Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection,
687 balanced polymorphism and background selection on equilibrium patterns of
688 genetic diversity in subdivided populations. *Genet. Res., Camb*, **70**, 155–174.
689 Cheng C, White BJ, Kamdem C *et al.* (2012) Ecological genomics of Anopheles
690 gambiae along a latitudinal cline: a population-resequencing approach.
691 *Genetics*, **190**, 1417–32.
692 Coetzee M, Koekemoer LL (2013) Molecular systematics and insecticide resistance
693 in the major African malaria vector Anopheles funestus. *Annual Review of*
694 *Entomology*, **58**, 393–412.
695 Cohuet A, Dia I, Simard F *et al.* (2005) Gene flow between chromosomal forms of the
696 malaria vector Anopheles funestus in Cameroon, Central Africa, and its
697 relevance in malaria fighting. *Genetics*, **169**, 301–311.
698 Cohuet A, Simard F, Toto J-C *et al.* (2003) Species identification within the Anopheles
699 funestus group of malaria vectors in Cameroon and evidence for a new species.
700 *The American journal of tropical medicine and hygiene*, **69**, 200–205.
701 Coluzzi M, Sabatini A, Petrarca V, Di Deco M (1979) Chromosomal differentiation
702 and adaptation to human environments in the Anopheles gambiae complex.
703 *Trans R Soc Trop Med Hyg*, **73**, 483–97.
704 Comeron JM, Ratnappan R, Bailin S (2012) The Many Landscapes of Recombination
705 in Drosophila melanogaster. *PLoS Genetics*, **8**, 33–35.
706 Comeron JM, Williford a, Kliman RM (2008) The Hill-Robertson effect: evolutionary
707 consequences of weak selection and linkage in finite populations. *Heredity*, **100**,
708 19–31.

709 Corbett-Detig RB, Hartl DL (2012) Population Genomics of Inversion
710 Polymorphisms in Drosophila melanogaster. *PLoS Genetics*, **8**.
711 Costantini C, Sagnon N, Ilboudo-Sanogo E, Coluzzi M, Boccolini D (1999)
712 Chromosomal and bionomic heterogeneities suggest incipient speciation in
713 Anopheles funestus from Burkina Faso. *Parassitologia*, **41**, 595–611.
714 Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of
715 speciation are due to reduced diversity, not reduced gene flow. *Molecular*
716 *ecology*, **23**, 3133–57.
717 Cutter AD, Choi JY (2010) Natural selection shapes nucleotide polymorphism across
718 the genome of the nematode Caenorhabditis briggsae Natural selection shapes
719 nucleotide polymorphism across the genome of the nematode Caenorhabditis
720 briggsae. *Genome Research*, 1103–1111.
721 Cutter AD, Payseur BA (2013) Genomic signatures of selection at linked sites:
722 unifying the disparity among species. *Nature reviews. Genetics*, **14**, 262–74.
723 Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools.
724 *Bioinformatics*, **27**, 2156–2158.
725 Denver DR, Dolan PC, Wilhelm LJ *et al.* (2009) A genome-wide view of
726 Caenorhabditis elegans base-substitution mutation processes. *Proceedings of*
727 *the National Academy of Sciences of the United States of America*, **106**, 16310–4.
728 Dia I, Lochouarn L, Boccolini D, Costantini C, Fontenille D (2000) Spatial and
729 temporal variations of the chromosomal inversion polymorphism of Anopheles
730 funestus in Senegal. *Parasite*, **7**, 179–184.
731 Dobzhansky T (1943) Genetics of natural populations. IX. Temporal changes in the

composition of populations of Drosophila pseudoobscura. *Genetics*, **28**, 162–186.

Dobzhansky T (1948) Genetics of natural populations .XVIII. Experiments on chromosomes of Drosophila pseudoobscura from different geographic regions. *Genetics*, **33**, 588–602.

Dobzhansky T (1950) Genetics of natural populations. XIX. Origin of heterosis through natural selection in populations of Drosophila pseudoobscura. *Genetics*, **35**, 288–302.

Dobzhansky T (1970) *Genetics of the evolutionary process*. New York.

Donnelly MJ, Licht MC, Lehmann T (2001) Evidence for recent population expansion in the evolutionary history of the malaria vectors Anopheles arabiensis and Anopheles gambiae. *Molecular biology and evolution*, **18**, 1353–1364.

Earl DA, VonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**.

Evanno G, Goudet J, Regnaut S (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.

Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends in Genetics*, **28**, 342–350.

755 Felsenstein J (1974) The evolutionary advantage of recombination. II. Individual
756 selection for recombination. *Genetics*, **78**, 737–756.
757 Fouet C, Gray E, Besansky NJ, Costantini C (2012) Adaptation to aridity in the
758 malaria mosquito Anopheles gambiae: chromosomal inversion polymorphism
759 and body size influence resistance to desiccation. *PloS one*, **7**, e34841.
760 Fouet C, Kamdem C, Gamez S, White BJ (2017) Extensive genetic diversity among
761 populations of the malaria mosquito Anopheles moucheti revealed by
762 population genomics. *Infection, Genetics and Evolution*, **48**, 27–33.
763 Fumagalli M, Vieira FG, Linderoth T (2014) ngsTools : methods for population
764 genetics analyses from Next-Generation Sequencing data. *Bioinformatics*.
765 Gillespie JH (1997) Junk ain't what junk does: Neutral alleles in a selected context.
766 *Gene*, **205**, 291–299.
767 Gillies MT, Coetzee M (1987) *A supplement to the Anophelinae of Africa south of the*
768 *Sahara*. The South African Institute for Medical Research, Johannesburg.
769 Gillies MT, De Meillon B (1968) *The Anophelinae of Africa South of the Sahara*.
770 Publications of the South African Institute for Medical Research, Johannesburg.
771 Giraldo-Calderon GI, Emrich SJ, MacCallum RM *et al.* (2015) VectorBase: an updated
772 bioinformatics resource for invertebrate vectors and other organisms related
773 with human diseases. *Nucleic Acids Research*, **43**, D707–D713.
774 Gosset CC, Bierne N (2013) Differential introgression from a sister species explains
775 high FST outlier loci within a mussel species. *Journal of Evolutionary Biology*,
776 **26**, 14–26.
777 Gray EM, Rocca K a C, Costantini C, Besansky NJ (2009) Inversion 2La is associated

778 with enhanced desiccation resistance in Anopheles gambiae. *Malaria journal*, **8**,
779 215.
780 Green CA, Hunt RH (1980) Interpretation of variation in ovarian polytene
781 chromosomes of Anopheles funestus Giles, A. parensis Gillies, and A. aruni ?
782 *Genetica*, **51**, 187–195.
783 Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the
784 joint demographic history of multiple populations from multidimensional SNP
785 frequency data. *PLoS genetics*, **5**, e1000695.
786 Haddrill PR, Halligan DL, Tomaras D, Charlesworth B (2007) Reduced efficacy of
787 selection in regions of the Drosophila genome that lack crossing over. *Genome*
788 *biology*, **8**, R18.
789 Hellmann I, Ebersberger I, Ptak SE, Pääbo S, Przeworski M (2003) A Neutral
790 Explanation for the Correlation of Diversity with Recombination Rates in
791 Humans. *The American Journal of Human Genetics*, **72**, 1527–1535.
792 Hellmann I, Prüfer K, Ji H *et al.* (2005) Why do human diversity levels vary at a
793 megabase scale? *Genome Research*, **15**, 1222–1231.
794 Hentges KE, Justice MJ (2004) Checks and balancers : balancer chromosomes to
795 facilitate genome annotation. *Trends in Genetics*, **20**.
796 Hill WG, Robertson A (1966) The effect of linkage on the limits of artificial selection.
797 *Genetical Research*, **8**, 269–294.
798 Hoffmann A a, Rieseberg LH (2008) Revisiting the Impact of Inversions in Evolution:
799 From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation?
800 *Annual review of ecology, evolution, and systematics*, **39**, 21–42.

801 Hoffmann AA, Sgro CM, Weeks AR (2004) Chromosomal inversion polymorphisms
802 and adaptation. *Trends in Ecology & Evolution*, **19**, 482–488.
803 Hoffmann AA, Weeks AR (2007) Climatic selection on genes and traits after a 100
804 year-old invasion : a critical look at the temperate-tropical clines in Drosophila
805 melanogaster from eastern Australia. *Genetica*, **129**, 133–147.
806 Huber CD, Nordborg M, Hermisson J, Hellmann I (2014) Keeping It Local: Evidence
807 for Positive Selection in Swedish Arabidopsis thaliana. *Molecular Biology and*
808 *Evolution*, **31**, 3026–3039.
809 Hudson RR, Kaplan NL (1988) The coalescent process in models with selection and
810 recombination. *Genetics*, **120**, 819–829.
811 Jakobsson M, Rosenberg N (2007) CLUMPP: a cluster matching and permutation
812 program for dealing with multimodality in analysis of population structure.
813 *Bioinformatics*, **23**, 1801– 1806.
814 Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic
815 markers. *Bioinformatics*, **24**, 1403–1405.
816 De Jong G, Bochdanovits Z (2003) Latitudinal clines in Drosophila melanogaster :
817 body size , allozyme frequencies , inversion frequencies , and the insulin-
818 signalling pathway. *Journal of Genetics*, **82**, 207–223.
819 Joron M, Frezal L, Jones RT *et al.* (2011) Chromosomal rearrangements maintain a
820 polymorphic supergene controlling butterfly mimicry. *Nature*, **477**, 203–206.
821 Kamau L, Hunt RH, Coetzee M (2002) Analysis of the population structure of
822 Anopheles funestus (Diptera: Culicidae) from western and coastal Kenya using
823 paracentric chromosomal inversion frequencies. *Journal of Medical Entomology*,

824 **39**, 78–83.

825 Kamdem C, Fouet C, Gamez S, White BJ (2017) Pollutants and insecticides drive local
826 adaptation in African malaria mosquitoes. *Molecular Biology and Evolution,* **34**,
827 1261–1275.

828 Kaplan NL, Hudson RR, Langley CH (1989) The "hitchhiking effect" revisited.
829 *Genetics*, **123**, 887–899.

830 Kapun M, Fabian DK, Goudet J, Flatt T (2016a) Genomic Evidence for Adaptive
831 Inversion Clines in Drosophila melanogaster. *Molecular Biology and Evolution*,
832 **33**, 1317–1336.

833 Kapun M, Van Schalkwyk H, McAllister B, Flatt T, Schlötterer C (2014) Inference of
834 chromosomal inversion dynamics from Pool-Seq data in natural and laboratory
835 populations of Drosophila melanogaster. *Molecular Ecology*, **23**, 1813–1827.

836 Kapun M, Schmidt C, Durmaz E, Schmidt PS, Flatt T (2016b) Parallel effects of the
837 inversion In (3R) Payne on body size across the North American and
838 Australian clines in Drosophila melanogaster. *Journal of Evolutionary Biology*,
839 **29**, 1059–1072.

840 Kawecki TJ, Dieter E (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**,
841 1225–1241.

842 Keightley PD, Pinharanda A, Ness RW *et al.* (2015) Estimation of the Spontaneous
843 Mutation Rate in Heliconius melpomene. *Molecular Biology and Evolution*, **32**,
844 239–243.

845 Keinan A, Reich D (2010) Human population differentiation is strongly correlated
846 with local recombination rate. *PLoS Genetics*, **6**.

847 Kennington WJ, Hoffmann AA, Partridge L (2007) Mapping Regions Within
848 Cosmopolitan Inversion In(3R)Payne Associated With Natural Variation in
849 Body Size in Drosophila melanogaster. *Genetics*, **177**, 549–556.
850 Kim Y, Maruki T (2011) Hitchhiking effect of a beneficial mutation spreading in a
851 subdivided population. *Genetics*, **189**, 213–226.
852 Kimura MK (1983) *The Neutral Theory of Molecular Evolution*. Cambridge, UK.
853 Kirkpatrick M (2010) How and why chromosome inversions evolve. *PLoS Biology*, **8**,
854 e1000501.
855 Kirkpatrick M, Barton N (2006) Chromosome inversions, local adaptation and
856 speciation. *Genetics*, **173**, 419–434.
857 Kliman R, Hey J (1993) Reduced natural selection associated with low
858 recombination in Drosophila melanogaster. *Molecular biology and evolution*, **10**,
859 1239–1258.
860 Knibb WR (1982) Chromosome inversion polymorphisms in Drosophila
861 melanogaster II. Geographic clines and climatic associations in Australasia ,
862 North America and Asia. *Genetica*, **58**, 213–221.
863 Korneliussen T, Albrechtsen A, Nielsen R (2014) ANGSD: Analysis of Next
864 Generation Sequencing Data. *BMC Bioinformatics*, **15**, 356.
865 Korneliussen TS, Moltke I, Albrechtsen A, Nielsen R (2013) Calculation of Tajima's D
866 and other neutrality test statistics from low depth next-generation sequencing
867 data. *BMC bioinformatics*, **14**, 289.
868 Krimbas C, Powell JR (1992) Introduction. In: *Drosophila inversion polymorphism*
869 (eds Krimbas C, Powell JR), p. . London.

870 Kulathinal RJ, Bennett SM, Fitzpatrick CL, Noor MAF (2008) Fine-scale mapping of
871 recombination rate in Drosophila refines its correlation to diversity and
872 divergence. *Proceedings of the National Academy of Sciences*, **105**, 10051–
873 10056.
874 Kulathinal RJ, Stevison LS, Noor MAF (2009) The genomics of speciation in
875 Drosophila: Diversity, divergence, and introgression estimated using low-
876 coverage genome sequencing. *PLoS Genetics*, **5**, e1000550.
877 Lahn BT, Page DC (1999) Chromosome Four Evolutionary Strata on the Human X
878 Chromosome. *Science*, **286**, 964–967.
879 Lee SF, Chen Y, Varan AK *et al.* (2011) Molecular Basis of Adaptive Shift in Body Size
880 in Drosophila melanogaster : Functional and Sequence Analyses of the Dca Gene
881 Research article. *Molecular Biology and Evolution*, **28**, 2393–2402.
882 Lee Y, Meneses CR, Fofana A, Lanzaro GC (2009) Desiccation Resistance Among
883 Subpopulations of Anopheles gambiae s.s. From Selinkenyi, Mali. *Journal of*
884 *medical entomology*, **46**, 316–320.
885 Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the
886 theory of the selective neutrality of polymorphisms. *Genetics*, **74**, 175–195.
887 Li H, Handsaker B, Wysoker A *et al.* (2009) The Sequence Alignment/Map format
888 and SAMtools. *Bioinformatics*, **25**, 2078–2079.
889 Lischer HEL, Excoffier L (2012) PGDSpider: An automated data conversion tool for
890 connecting population genetics and genomics programs. *Bioinformatics*, **28**,
891 298–299.
892 Lowry DB, Willis JH (2010) A Widespread Chromosomal Inversion Polymorphism

893 Contributes to a Major Life-History Transition , Local Adaptation , and
894 Reproductive Isolation. *PLoS Biology*, **8**, e1000500.
895 Magni GE, Von Borstel RC (1962) Different Rates of Spontaneous Mutation during
896 Mitosis and Meiosis in Yeast. *Genetics*, **47**, 1097–1108.
897 Majewski J, Cohan FM (1999) Adapt globally, act locally: The effect of selective
898 sweeps on bacterial sequence diversity. *Genetics*, **152**, 1459–1474.
899 Marsden CD, Lee Y, Kreppel K *et al.* (2014) Diversity, Differentiation, and Linkage
900 Disequilibrium: Prospects for Association Mapping in the Malaria Vector
901 Anopheles arabiensis. *G3 (Bethesda, Md.)*, **4**, 121–31.
902 Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene.
903 *Genetical research*, **23**, 23–35.
904 Mcgaugh SE, Heil CSS, Manzano-winkler B *et al.* (2012) Recombination Modulates
905 How Selection Affects Linked Sites in Drosophila. *PLoS Biology*, **10**, e1001422.
906 McGaugh SE, Noor M a F (2012) Genomic impacts of chromosomal inversions in
907 parapatric Drosophila species. *Philosophical transactions of the Royal Society of*
908 *London. Series B, Biological sciences*, **367**, 422–9.
909 Meirmans P, Van Tienderen P (2004) GENOTYPE and GENODIVE: two programs for
910 the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*,
911 **4**, 792–794.
912 Mettler LE, Voelker RA, Mukai T (1977) Inversion clines in populations of
913 Drosophila melanogaster. *Genetics*, **87**, 169–176.
914 Michel a. P, Ingrasci MJ, Schemerhorn BJ *et al.* (2005) Rangewide population genetic
915 structure of the African malaria vector Anopheles funestus. *Molecular Ecology*,

916 **14**, 4235–4248.

917 Muller HJ (1918) Genetic variability, twin hybrids, and constant hybrids, in a case of
918 balanced lethal factors. *Genetics*, **3**, 422–499.

919 Nachman MW (2002) Variation in recombination rate across the genome: Evidence
920 and implications. *Current Opinion in Genetics and Development*, **12**, 657–663.

921 Nachman MW, Payseur BA (2012) Recombination rate variation and speciation :
922 theoretical predictions and empirical results from rabbits and mice.
923 *Philosophical Transactions of the Royal Society B: Biological Sciences*, **367**, 409–
924 421.

925 Neafsey DE, Waterhouse RM, Abai MR *et al.* (2015) Highly evolvable malaria vectors:
926 The genomes of 16 Anopheles mosquitoes. *Science*, **347**, 1258522–1258522.

927 Nielsen R (2005) Molecular Signatures of Natural Selection. *Annual review of*
928 *genetics*, **39**, 197–218.

929 Nishikawa H, Iijima T, Kajitani R *et al.* (2015) A genetic mechanism for female-
930 limited Batesian mimicry in Papilio butterfly. *Nature genetics*, **47**, 405–9.

931 Noor MAF (2008) Mutagenesis from meiotic recombination is not a primary driver
932 of sequence divergence between Saccharomyces species. *Molecular Biology and*
933 *Evolution*, **25**, 2439–2444.

934 Noor M a F, Bennett SM (2009) Islands of speciation or mirages in the desert?
935 Examining the role of restricted recombination in maintaining species. *Heredity*,
936 **103**, 439–444.

937 Noor MAF, Grams KL, Bertucci LA, Reiland J (2001) Chromosomal inversions and the
938 reproductive isolation of species. *Proceedings of the National Academy of*

939 *Sciences*, **98**, 12084–12088.

940 Nordborg M, Charlesworth B, Charlesworth B (1996) Increased levels of

941 polymorphism surrounding selectively maintained sites in highly selfing

942 species. *Proceedings of the Royal Society B: Biological Sciences*, **263**, 1033–1039.

943 Nordborg M, Hu TT, Ishino Y *et al.* (2005) The pattern of polymorphism in

944 Arabidopsis thaliana. *PLoS Biology*, **3**, 1289–1299.

945 Nordborg M, Tavar S (2002) Linkage disequilibrium : what history has to tell us.

946 *Trends in Genetics*, **18**, 83–90.

947 Nosil P, Feder JL (2012) Widespread yet heterogeneous genomic divergence.

948 *Molecular Ecology*, **21**, 2829–2832.

949 Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous

950 genomic divergence. *Molecular Ecology*, **18**, 375–402.

951 O'Loughlin SM, Magesa S, Mbogo C *et al.* (2014) Genomic Analyses of Three Malaria

952 Vectors Reveals Extensive Shared Polymorphism but Contrasting Population

953 Histories. *Molecular biology and evolution*, **31**, 889-902.

954 Ortiz-Barrientos D, Reiland J, Hey J, Noor MAF (2002) Recombination and the

955 divergence of hybridizing species. *Genetica*, **116**, 167–178.

956 Paradis E, Claude J, Strimmer K (2004) Analyses of Phylogenetics and Evolution in R

957 language. *Bioinformatics*, **20**, 289–290.

958 Pegueroles C, Ordóñez V, Mestres F, Pascual M (2010) Recombination and selection

959 in the maintenance of the adaptive value of inversions. *Journal of Evolutionary*

960 *Biology*, **23**, 2709–2717.

961 Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest

962 RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in
963 Model and Non-Model Species. *PLoS ONE*, **7**, e37135.
964 Pool JE, Corbett-Detig RB, Sugino RP *et al.* (2012) Population Genomics of Sub-
965 Saharan Drosophila melanogaster: African Diversity and Non-African
966 Admixture. *PLoS Genetics*, **8**, e1003080.
967 Powell JR (1997) *Progress and prospects in evolutionary biology: the Drosophila*
968 *model*. New York.
969 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using
970 multilocus genotype data. *Genetics*, **155**, 945–959.
971 Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: a toolset for whole-genome
972 association and population-based linkage analysis. *American Journal of Human*
973 *Genetics*, **81**, 559-75.
974 R Development Core Team (2016) *R: A language and environment for statistical*
975 *computing*. R Foundation for Statistical Computing, Vienna, Austria.
976 Rako L, Anderson AR, Sgro CM, Stocker AJ, Hoffmann AA (2006) The association
977 between inversion In (3R) Payne and clinally varying traits in Drosophila
978 melanogaster. *Genetica*, **128**, 373–384.
979 Rane AR V, Rako L, Lee SF, Hoffmann A a (2015) Genomic evidence for role of
980 inversion 3RP of Drosophila melanogaster in facilitating climate change
981 adaptation Abstract : *Molecular Ecology*, **24**, 2423–2432.
982 Rieseberg LH (2001) Chromosomal rearrangements and speciation. *Trends in*
983 *Ecology and Evolution*, **16**, 351–358.
984 Roberts P (1976) The genetics of chromosomal aberration. In: *The genetics and*

985 *biology of Drosophila* (eds Ashburner M, Novitski E), pp. 67–184. London.
986 Roesti M, Hendry AP, Salzburger W, Berner D (2012) Genome divergence during
987 evolutionary diversification as revealed in replicate lake-stream stickleback
988 population pairs. *Molecular ecology*, **21**, 2852–62.
989 Roesti M, Moser D, Berner D (2013) Recombination in the threespine stickleback
990 genome - Patterns and consequences. *Molecular Ecology*, **22**, 3014–3027.
991 Rosenberg N (2004) DISTRUCT: a program for the graphical display of population
992 structure. *Molecular Ecology Resources*, **4**, 137–138.
993 Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation.
994 *Nature Reviews Genetics,* **14**, 807–820.
995 Schaeffer SW (2008) Selection in heterogeneous environments maintains the gene
996 arrangement polymorphism of drosophila pseudoobscura. *Evolution*, **62**, 3082–
997 3099.
998 Schaeffer SW, Goetting-minesky MP, Kovacevic M *et al.* (2003) Evolutionary
999 genomics of inversions in Drosophila pseudoobscura : Evidence for epistasis.
1000 *Proceedings of the National Academy of Sciences*, **100**, 8319–8324.
1001 Schlötterer C (2002) Towards a molecular characterization of adaptation in local
1002 populations. *Current Opinion in Genetics and Development*, **12**, 683–687.
1003 Schneider KA, Kim Y (2013) Genetic Hitchhiking under Heterogeneous Spatial
1004 Selection Pressures. *PLoS ONE*, **8**, e61742.
1005 Seehausen O, Butlin RK, Keller I *et al.* (2014) Genomics and the origin of species.
1006 *Nature reviews. Genetics*, **15**, 176–92.
1007 Service MW (1993) *Mosquito ecology: field sampling methods* (UK Elsevier Applied

1008 Science, London, Ed,).

1009 Sharakhova M V, Xia A, Leman SC, Sharakhov I V (2011) Arm-specific dynamics of

1010 chromosome evolution in malaria mosquitoes. *BMC evolutionary biology*, **11**,

1011 91.

1012 Sharakhov I, Braginets O, Grushko O *et al.* (2004) A Microsatellite Map of the African

1013 Human Malaria Vector Anopheles funestus. *Journal of Heredity*, **95**, 29–34.

1014 Sinka ME, Bangs MJ, Manguin S *et al.* (2010) The dominant Anopheles vectors of

1015 human malaria in Africa, Europe and the Middle East: occurrence data,

1016 distribution maps and bionomic précis. *Parasites & vectors*, **3**, 117.

1017 Slatkin M, Wiehe T (1998) Genetic hitch-hiking in a subdivided population. *Genetical*

1018 *Research*, **71**, 155–160.

1019 Spencer CCA, Deloukas P, Hunt S *et al.* (2006) The influence of recombination on

1020 human genetic diversity. *PLoS Genetics*, **2**, 1375–1385.

1021 Stefansson H, Helgason A, Thorleifsson G *et al.* (2005) A common inversion under

1022 selection in Europeans. *Nature genetics*, **37**, 129–137.

1023 Stephan W, Langley CH (1989) Molecular genetic variation in the centromeric region

1024 of the X chromosome in three Drosophila ananassae populations. I. Contrasts

1025 between the vermilion and forked loci. *Genetics*, **121**, 89–99.

1026 Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive

1027 population divergence. *Molecular Ecology*, **14**, 671–688.

1028 Stump AD, Fitzpatrick MC, Lobo NF *et al.* (2005) Centromere-proximal

1029 differentiation and speciation in Anopheles gambiae. *Proceedings of the*

1030 *National Academy of Sciences of the United States of America*, **102**, 15930–5.

1031 Sturtevant A (1921) A case of rearrangement of genes in Drosophila. *Proceedings of*
1032 *the National Academy of Sciences*, **7**, 235–237.
1033 Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by
1034 DNA polymorphism. *Genetics*, **123**, 585–595.
1035 Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS (2004) Selection versus demography: A
1036 multilocus investigation of the domestication process in maize. *Molecular*
1037 *Biology and Evolution*, **21**, 1214–1225.
1038 Thomas CG, Wang W, Jovelin R *et al.* (2015) Full-genome evolutionary histories of
1039 selfing , splitting , and selection in Caenorhabditis. *Genome Research*, **25**, 667–
1040 678.
1041 Tiffin P, Ross-Ibarra J (2014) Advances and limits of using population genetics to
1042 understand local adaptation. *Trends in Ecology & Evolution*, **29**, 673–680.
1043 Watterson GA (1975) On the number of segregating sites in genetical models
1044 without recombination. *Theoretical Population Biology*, **7**, 256–276.
1045 Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population
1046 structure. *Evolution*, **38**, 1358–1370.
1047 WHO (2016) *World malaria report 2016*. pp 186 (ISBN: 978 92 4 151171 1)
1048 Williams GC (1966) *Adaptation and Natural Selection: A Critique of Some Current*
1049 *Evolutionary Thought*. Princeton, NJ.
1050 Wright S, Dobzhansky T (1946) Genetics of natural populations. XII. Experimental
1051 reproduction of some of the changes caused by natural selection in certain
1052 populations of Drosophila pseudoobscura. *Genetics*, **31**, 125–156.
1053 Wright SI, Foxe JP, DeRose-Wilson L *et al.* (2006) Testing for effects of

recombination rate on nucleotide diversity in natural populations of Arabidopsis lyrata. *Genetics*, **174**, 1421–1430.

Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics*, **26**, 873–881.

Yeaman S (2013) Genomic rearrangements and the evolution of clusters of locally adaptive loci. *Proceedings of the National Academy of Sciences*, **110**, E1743-51.

Figure legends

Figure 1: Geographic origin and genetic relatedness of *An. funestus* ecotypes in Cameroon. (A) Map showing Mfou and Tibati, the two locations where *An. funestus* mosquitoes were collected. A delimitation of the approximate distribution ranges of the two ecotypes found in Cameroon is shown (continuous line in the middle of the map). (B) and (C) Population genetic structure as revealed respectively by a neighbor-joining tree and a PCA. The percentage of variance explained is indicated on each PCA axis. (D) and (E) Confirmation of the presence of two *An. funestus* ecotypes with DAPC and the delta k method of Evanno *et al.* (2005). The lowest Bayesian Information Criterion (BIC) and cross-validation error and the highest delta k indicate the most probable number of clusters (2-3). (F) Bayesian clustering in STRUCTURE.

Figure 2: Estimates of pairwise population differentiation (F_{ST}) based on SNPs ordered by position along pseudo-chromosomes representing the five chromosome arms of *An. funestus*. F_{ST} values on top of the red line are above the 99th percent of the empirical distribution. Arrows indicate the genomic coordinates of the 3R*a* and 3R*b* inversions.

Figure 3: Population genetic structure revealed by each chromosome arm and by four polymorphic chromosomal inversions (2R*h*, 3R*a*, 3R*b* and 3R*d*). The number of SNPs is indicated in parenthesis.

Figure 4: Genome-wide distribution of d_{xy} and fixed differences (d_f) across 90-kb non-overlapping windows along the five pseudo-chromosomes among Mfou and

Tibati populations. Strong divergent selection within the 3R*a* and 3R*b* inversions is characterized by the presence of peaks of d_{xy} and d_f at these loci.

Figure 5: Estimates of nucleotide diversity ($\theta\pi$ and θw) and allele frequency spectrum (Tajima's *D*) across 90-kb non-overlapping windows illustrating the uneven distribution of genetic diversity along the five chromosome arms. The collinear autosome (2L) is the most polymorphic in both Mfou and Tibati populations. In contrast, autosomes bearing multiple inversions exhibit a significant reduction in the amount of polymorphic sites.

Figure 6: Box plot depicting the distribution of $\theta\pi$ and θw among chromosomes in Mfou and Tibati.

Supplemental Figure Legends

Figure S1: The positions of the 104 scaffolds that were placed on the *An. funestus* physical map are indicated by red rectangles (modified from Sharakhov *et al*. 2004).

Figure S2: Distribution of the average coverage depth along mapped scaffolds.

Tables

Table 1: Reduction in nucleotide diversity relative to the 2L chromosome arm (%).

		2R	3L	3R	X
Mfou	θw	29.22 *	13.24 *	34.69 *	23.35 *
	θπ	7.92 *	-3.23	19.23 *	13.92 *
Tibati	θw	18.79 *	13.24 *	33.54 *	18.94 *
	θπ	3.23	-4.84	15.61 *	11.86 *

*Statistically significant (Wilcoxon rank sum test, $p < 0.001$)

Supplemental Tables

Table S1: Description of *An. funestus* samples included in this study.

Ecogeographic region	Sampling location	Geographic coordinates	Sampling method				Total	Gender		
			HLC-OUT	HLC-IN	LC	SPRAY		Female	Male	NA
Savannah	Tibati	6°28'08"N, 12°37'44"E	8	7	0	32	47	43	4	0
Forest	Mfou	3°58'00"N, 11°56'00"E	2	8	22	53	85	55	8	22
	Total						132	98	12	22

HLC-OUT, human landing catches performed outdoor; HLC-IN, human landing catches performed indoor; LC, larval collection; SPRAY, spray catches; NA, not available

Table S2: Characteristics of the five pseudo-chromosomes of *An. funestus* used for genome scans.

Chromosome	Number of mapped scaffolds	Total length (bp)	Number of SNPs
2R	30	27 415 570	1 587
2L	18	19 110 819	987
3R	24	15 197 945	745
3L	22	12 545 136	652
X	10	6 812 527	296

Table S3: Parameters of the best-fit demographic models in ∂a∂i.

Population	Best model	Log Likelihood	Final population size[b] (95% CI)		Time[c] (95% CI)	
Tibati	*Growth*	-88.8	77.6	(51.8 - 138.1)	3.8	(1.8 - 8.0)
Mfou	*Two-epoch*	-84.7	13.6	(12.2 - 15.7)	1.4	(1.2 - 1.7)

[b]Relative to ancestral population size.

[c]Expressed in units 2Ne generations from start of growth to present

Figures

Figure1

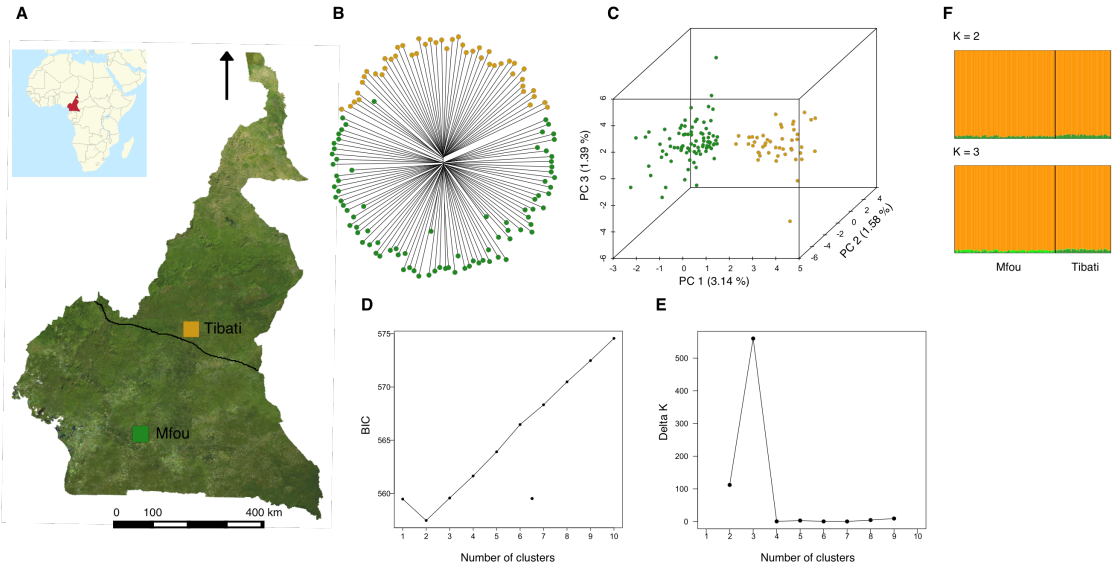

Figure 2

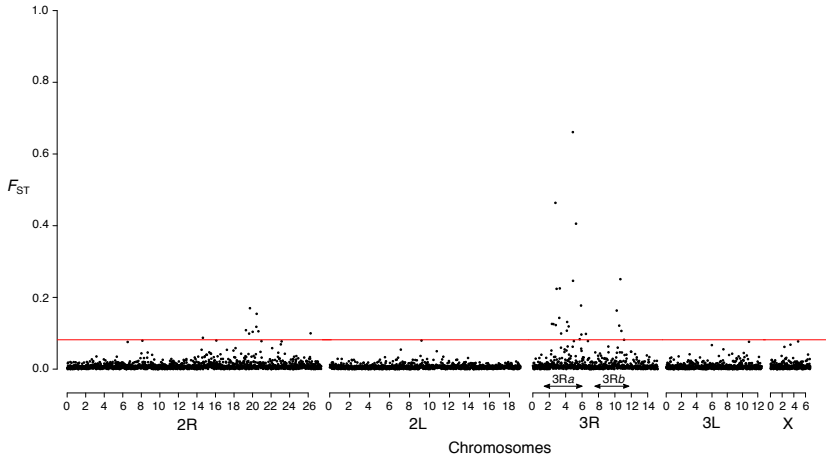

Figure 3

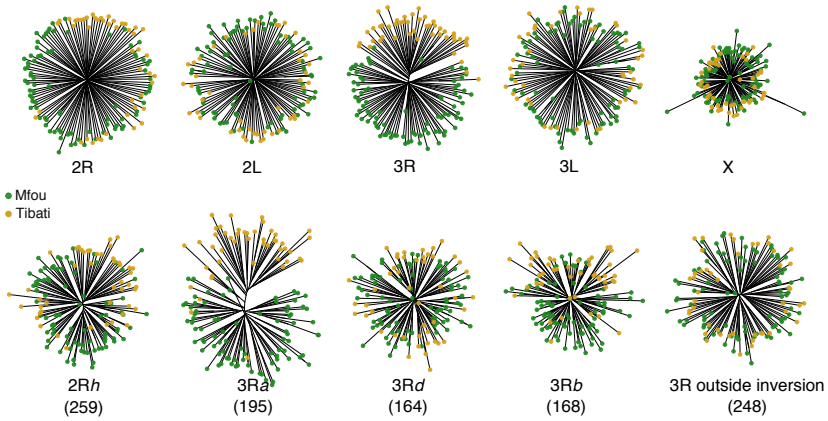

Figure 4

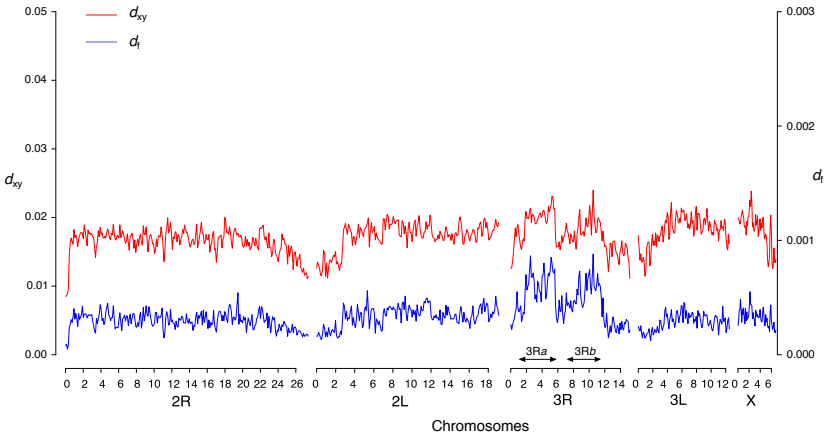

Figure 5

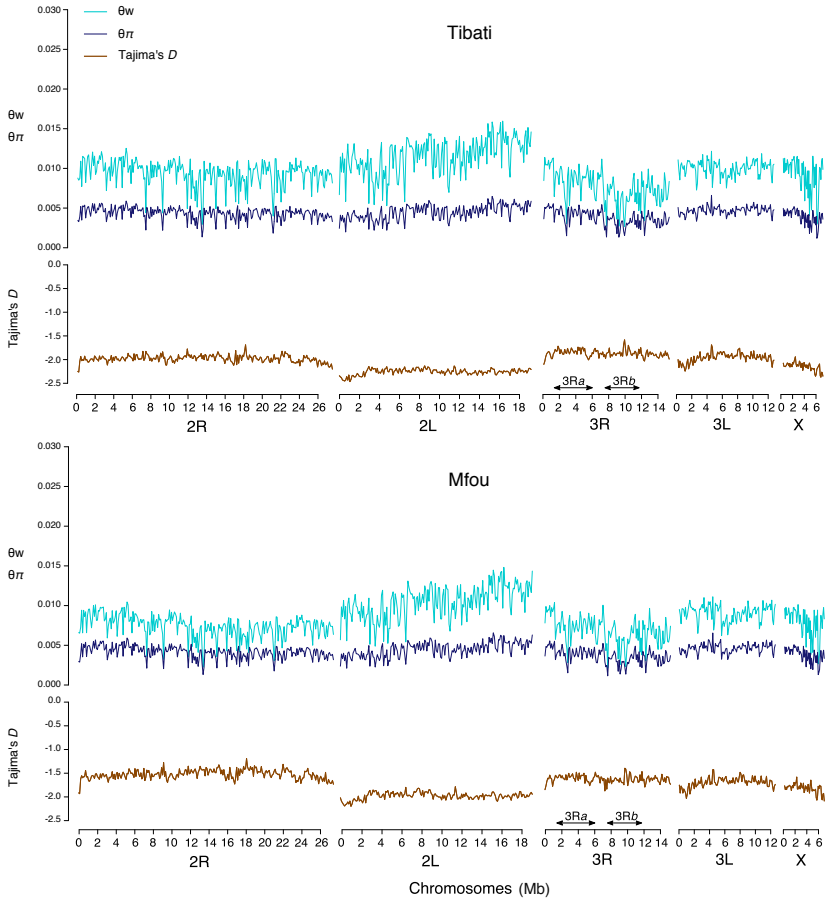

Figure 6

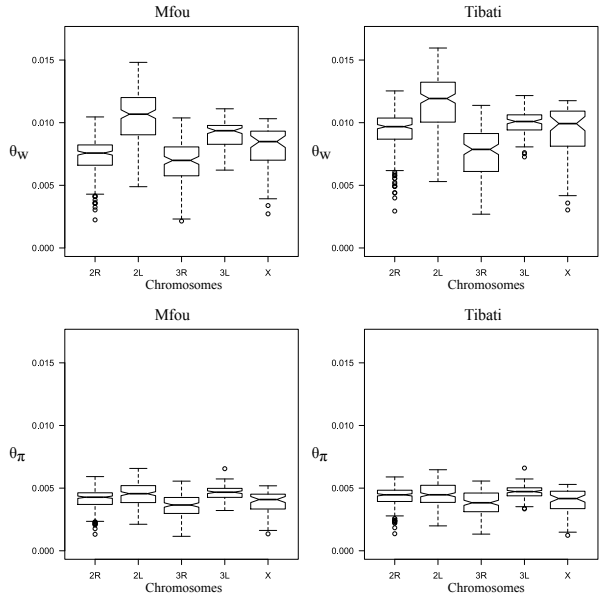

Figure S1

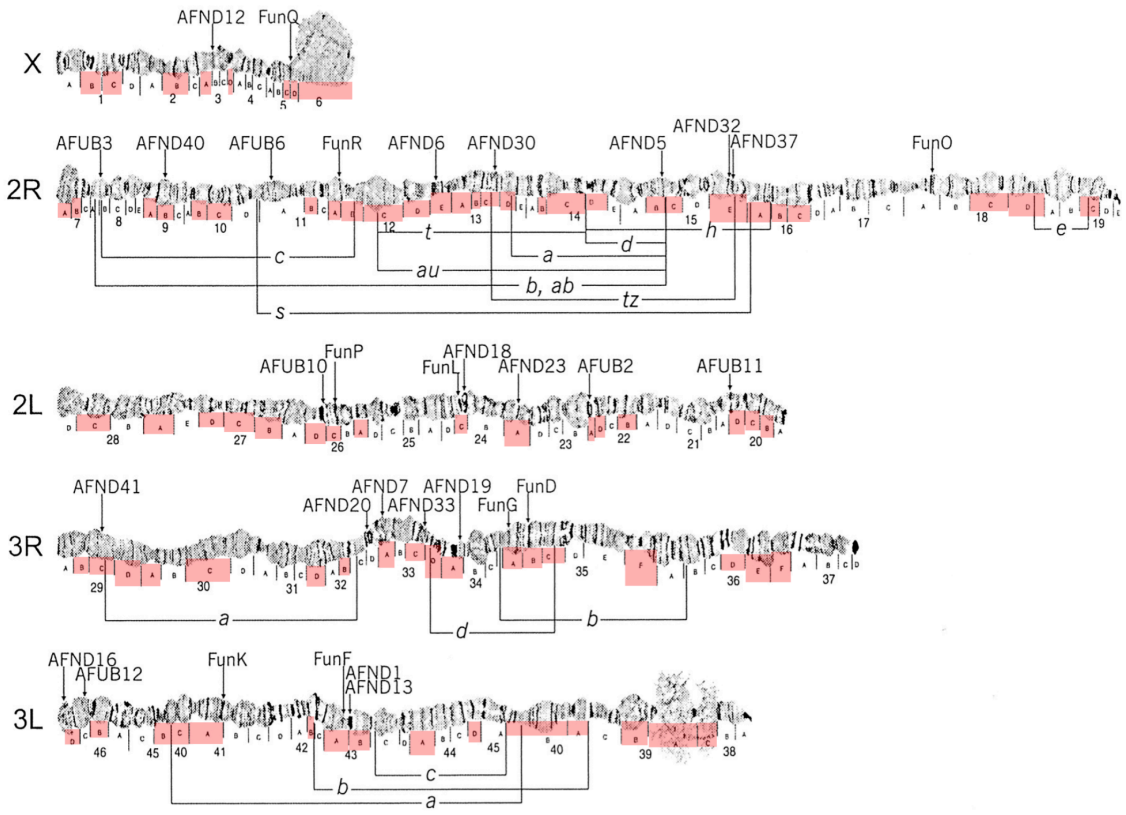

Figure S2