Chromosome arm specific patterns of polymorphism associated with chromosomal inversions in the major African malaria vector, Anopheles funestus

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

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

*Corresponding author: colincek@ucr.edu

[#] Present address: Verily Life Sciences, South San Francisco, CA 94080, USA

Abstract

Chromosomal inversions facilitate local adaptation of beneficial mutations and modulate genetic polymorphism, but the extent of their effects within the genome is still insufficiently understood. 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 weakly differentiated populations of *Anopheles funestus* inhabiting contrasting moisture conditions in Cameroon. Using population genomic analyses, we found that genetic divergence between the two populations is centered on regions of the genome corresponding to three inversions, which are characterized by high values of F_{ST}, absolute sequence divergence and fixed differences. 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 natural selection within inversions contribute to sculpt nucleotide polymorphism across chromosomes in *An. funestus*.

Key words: *Anopheles funestus*, chromosomal inversion, recombination, genetic divergence, nucleotide diversity.

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 — which 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 within species are also influenced by several types of natural selection whose signatures are heterogeneous in nature and across 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 candidate genomic regions, which contain genes underlying responses to spatially heterogeneous selection pressures exhibit high divergence among populations,

55 skewed polymorphism or an extended correlation among alleles from different loci
56 (linkage disequilibrium) (Lewontin & Krakauer 1973; Schlötterer 2002; Beaumont
57 2005; Nielsen 2005; Storz 2005; Nosil *et al.* 2009). Because migration, gene flow and
58 recombination among diverging populations can break down locally adapted allelic
59 combinations, genetic targets of local adaptation are often protected within genomic
60 regions of reduced recombination – including chromosomal inversions (Noor *et al.*
61 2001; Rieseberg 2001; Ortiz-Barrientos *et al.* 2002; Butlin 2005; Kirkpatrick &
62 Barton 2006; Joron *et al.* 2011; Roesti *et al.* 2012, 2013; Yeaman 2013; Nishikawa *et*
63 *al.* 2015).

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

78 inversion and several fitness-related traits (Wright & Dobzhansky 1946;
79 Dobzhansky 1948; Rako *et al.* 2006; Hoffmann & Weeks 2007; Kennington *et al.*
80 2007; Lowry & Willis 2010; Lee *et al.* 2011; Fouet *et al.* 2012; Kapun *et al.* 2014,
81 2016).

82 Another important property of inversions concerns the significant role they
83 play in genome evolution as a whole. Recombination is strongly reduced between
84 two paired chromosomes that differ by an inversion through several mechanisms,
85 which impede crossing over (reviewed in Roberts 1976). Local recombination rates
86 affect a myriad of processes including selection, gene conversion, diversity and
87 divergence throughout the genome (Begun & Aquadro 1992; Kliman & Hey 1993;
88 Andolfatto *et al.* 2001; Nordborg *et al.* 2005; Haddrill *et al.* 2007; Kulathinal *et al.*
89 2008; Comeron *et al.* 2012; Nachman & Payseur 2012; Roesti *et al.* 2012, 2013;
90 Campos *et al.* 2014; Ortiz-barrientos *et al.* 2016). Since the pioneering work of
91 Begun and Aquadro (1992), which demonstrated a positive correlation between
92 recombination rate and nucleotide diversity within the genome of *Drosophila*
93 *melanogaster*, multiple studies have provided empirical evidence that reduced
94 crossing over alters genetic variation along chromosomes in a wide variety of
95 organisms (Hellmann *et al.* 2003, 2005; Tenaillon *et al.* 2004; Begun *et al.* 2007;
96 Kulathinal *et al.* 2008, 2009; Pegueroles *et al.* 2010; Corbett-Detig & Hartl 2012;
97 McGaugh *et al.* 2012; Pool *et al.* 2012). Consequently, inversions, which are
98 coldspots of recombination, have the potential to modulate patterns of genetic
99 variation across relatively significant regions of the genome. More specifically, the
100 presence of one or several rearrangements may reduce crossing over and diversity

101 along an entire chromosome or chromosome arm, which in turn evolves differently
102 from the rest of the genome. This scenario is best illustrated by the nonrecombining
103 part of the Y chromosome in mammals, which is maintained by multiple inversions
104 that have suppressed recombination over the entire length of the chromosome
105 (Lahn & Page 1999). The potential of inversions to reduce recombination and
106 diversity in large regions of the genome has long been exploited to engineer
107 chromosomes or chromosomal arms with multiple inversions that block
108 recombination, allowing lethal or sterile mutations to be stably maintained in a
109 heterozygous state (balancer chromosomes) (Muller 1918; Sturtevant 1921;
110 Hentges & Justice 2004; Edgley *et al.* 2006). Despite these clear examples, our
111 understanding of how inversions affect polymorphism over the entire length of a
112 chromosome remains limited in part due to the lack of empirical support across
113 taxa. Targeted studies based on fine-scale examinations of genomic variation in
114 species with well-characterized patterns of naturally occurring inversions are
115 important for assessing both the genomic extent and the ubiquity of diversity
116 reduction associated with chromosomal rearrangements.

117 An excellent opportunity to investigate the role of inversions in genome
118 evolution exists in *An. funestus*, the second most important vector of *Plasmodium*
119 parasites in the Afrotropical region behind the best known species, *An. gambiae*
120 (Gillies & De Meillon 1968; Gillies & Coetzee 1987; Sinka *et al.* 2010; Coetzee &
121 Koekemoer 2013). The two taxa share several characteristics including a near
122 continental distribution, a marked preference for human hosts and intradomicillary
123 host-seeking and resting behavior, which reflect their efficiency as vector of malaria

124 parasites (Dia *et al.* 2013; Lanzaro & Lee 2013). Numerous studies have suggested
125 that ongoing adaptive divergence contributes in part to the significant
126 environmental flexibility, which underlies the widespread distribution of *An.*
127 *funestus* populations. This species has been described as an amalgam of at least two
128 relatively differentiated ecotypes whose ecology, phenotypic divergence,
129 distribution range and role in malaria transmission remain obscure (Green & Hunt
130 1980; Costantini *et al.* 1999; Dia *et al.* 2000; Kamau *et al.* 2002; Boccolini *et al.* 2005;
131 Cohuet *et al.* 2005; Michel *et al.* 2005; Ayala *et al.* 2011; Barnes *et al.* 2017). A very
132 high number of paracentric polymorphic chromosomal inversions have been
133 identified in polytene chromosomes of *An. funestus* via cytogenetic studies (Green &
134 Hunt 1980; Sharakhov *et al.* 2004). Some of these rearrangements are spread along
135 stable geographic clines and exhibit a significant deficit in heterokaryotypes,
136 suggesting that they may play a role in environmental adaptation (Costantini *et al.*
137 1999; Dia *et al.* 2000; Kamau *et al.* 2002; Boccolini *et al.* 2005; Ayala *et al.* 2011).
138 Importantly, dozens of large rearrangements have been detected on all autosomes
139 except for 2L (Sharakhov *et al.* 2004). This uneven distribution of inversions
140 between chromosome arms provides an ideal framework for testing hypotheses
141 about the impacts of rearrangements at the chromosome level via cross-
142 chromosomal comparisons.

143 In this study, we have used Restriction site Associated DNA Sequencing
144 (RAD-Seq) (Baird *et al.* 2008) to address the effects of inversions on the genetic
145 architecture of selection and diversity across the genome of *An. funestus* populations
146 collected from Cameroon. Although the genome-wide level of differentiation is

147 weak, signals of divergence are apparent within three large chromosomal
148 rearrangements whose frequencies vary along moisture gradients in this region
149 (3R*a*, 3R*b* and 2R*h*). Genome scans show that, in contrast to the collinear 2L arm, the
150 proportion of polymorphic sites is drastically reduced along the entire length of
151 three autosomal arms bearing multiple polymorphic inversions. These findings
152 suggest that the evolution of chromosomes is impacted by the combined effects of
153 suppressed recombination and selection within paracentric polymorphic inversions
154 in *An. funestus*.
155

156 **Materials and methods**

157 **Mosquito samples**

158 Populations of *An. funestus* that occur in Cameroon have been assimilated to
159 two weakly differentiated ecotypes distributed along a moisture gradient spanning
160 the whole country (Costantini *et al.* 1999; Cohuet *et al.* 2005; Ayala *et al.* 2011). We
161 sequenced 132 mosquitoes collected from two locations separated by ~500km,
162 representative of the savannah and the forest ecogeographic domains (Fig. 1A and
163 Table S1). In Mfou, a small city of the forest region, *An. funestus* mosquitoes breed in
164 an artificial lake and maintain abundant populations throughout the year. In Tibati,
165 which lies within the forest-savannah transition zone, several artificial lakes also
166 provide abundant breeding sites for dense and perennial populations of *An. funestus*.
167 Between August and November 2013, we used several sampling methods (Service
168 1993) to collect *An. funestus* larvae in breeding sites and adults seeking human hosts
169 in and around human dwellings at night, or resting indoors during daytime. (Fig. 1A

170 and Table S1). With this diversified sampling, we aimed to maximize the chances
171 that our sample represents a good approximation of the genetic diversity of *An.*
172 *funestus* populations, as genetic differentiation within malaria vectors species
173 sometimes overlap with subtle microgeographic and temporal segregations (e.g.
174 Riehle *et al.* 2011). *An. funestus* belong to a large taxonomic unit comprising at least
175 seven taxa that are identified by slight morphological variations and a diagnostic
176 PCR based on mutations of the ribosomal DNA (Gillies & De Meillon 1968; Gillies &
177 Coetzee 1987; Cohuet *et al.* 2003). We verified, using morphology and PCR, that all
178 samples included in this study belonged to the nominal species and only important
179 malaria vector of the group: *An. funestus*.

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

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

186 Illumina sequences were sorted according to barcode tag and filtered using
187 the *process_radtags* program of the Stacks v 1.35 software (Catchen *et al.* 2013).
188 Reads with ambiguous barcode, inappropriate restriction site or low sequencing
189 quality score were removed. We checked the depth of sequencing coverage of the
190 final data set after all filtering steps using VCFtools (Danecek *et al.* 2011). To call and
191 genotype SNPs, we first aligned the remaining high-quality reads to the *An. funestus*
192 reference sequence with GSNAP (Wu & Nacu 2010), with a maximum of five

193 mismatches allowed and terminal alignments prevented. We then used Stacks to
194 build consensus RAD loci and to identify SNPs within each locus. We set the
195 minimum number of reads required to form a "stack" to three and allowed two
196 mismatches during catalogue creation. Following assembly and genotyping, the
197 polymorphism data was further filtered to maximize data quality. To do this, we
198 retained only RAD loci scored in every population and in at least 60% of individuals
199 within each population and we randomly selected one SNP per locus for further
200 analyses. SNP files in different formats used for downstream analyses were created
201 with the *populations* program in Stacks, PLINK v 1.09 and PGDSpider v 2.0.8.2
202 (Purcell *et al.* 2007; Lischer & Excoffier 2012; Catchen *et al.* 2013).

203 **Population structure and genetic divergence**

204 We examined the genetic relatedness among individuals with a Principal
205 Component Analysis (PCA), a neighbor-joining (NJ) tree and the STUCTURE v 2.3.4
206 software (Pritchard *et al.* 2000) using SNPs identified in Stacks. We utilized the R
207 package *adegenet* to implement the PCA and *ape* to compute a genotype-based
208 Euclidian distance matrix between individuals and to infer individual-based NJ
209 networks (Paradis *et al.* 2004; Jombart 2008; R Development Core Team 2016). In
210 STRUCTURE, we ran five replicates of 1 to 10 assumed clusters (k). Each run
211 consisted of 200000 iterations, and the first 50000 iterations were discarded as
212 burn-in. CLUMPP v1.1.2 (Jakobsson & Rosenberg 2007) was used to aggregate
213 results across multiple STRUCTURE runs and the clustering results were visualized
214 graphically using DISTRUCT v1.1 (Rosenberg 2004). To find the optimal number of
215 genetically distinct clusters, we used both the Discriminant Analysis of Principal

216 Component (DAPC) implemented in *adegenet* and the ad hoc statistic DeltaK of
217 Evanno *et al.* (2005) (Evanno *et al.* 2005; Earl & VonHoldt 2012).

218 To assess the level of genetic differentiation between savannah and forest
219 populations, we estimated the overall F_{ST} (Weir & Cockerham 1984) in Genodive
220 v1.06 (Meirmans & Van Tienderen 2004) using a subset of 2000 randomly selected
221 SNPs. To examine to what extent the geographic region contributes to the genetic
222 variance among samples, we conducted a hierarchical analysis of molecular variance
223 (AMOVA) (Excoffier *et al.* 1992) in GenoDive. We used 10000 permutations to
224 assess the statistical significance of F_{ST} and AMOVA.

225 **Genomic targets of selection**

226 The current *An. funestus* draft genome assembly consists of 12243 scaffolds
227 ranging from 2334 to 3334433bp in length (Giraldo-Calderon *et al.* 2015; Neafsey *et*
228 *al.* 2015). Therefore, to perform genome scans and inspect footprints of selection
229 throughout the genome, we ordered and concatenated 104 long scaffolds whose
230 positions on the physical map have been inferred via alignment and orthology
231 (Neafsey *et al.* 2015) and we created "pseudo-chromosomes" corresponding to the
232 five chromosome arms of *An. funestus*. These mapped scaffolds accounted for 33%
233 in length of the approximate size of the *An. funestus* genome (Fig. S1 and Table S2).

234 To delineate genomic signatures of selection, we performed an outlier
235 analysis in order to detect genomic regions exhibiting exceptional differentiation or
236 diversity that are putative targets of selection (Lewontin & Krakauer 1973; Storz
237 2005). Locus-specific F_{ST} values between savannah and forest populations were
238 estimated with the *populations* program in *Stacks* and visualized graphically using

239 the subset of SNPs located on mapped scaffolds. SNPs with F_{ST} values above the top
240 1% of the empirical distribution were considered as outliers of genetic
241 differentiation. Loci with unusually low or high F_{ST} values relative to neutral
242 expectations were also detected using the coalescence-based method FDIST2
243 (Beaumont & Nichols 1996) as implemented in LOSITAN (Antao *et al.* 2008). The
244 mean neutral F_{ST} across all SNPs was approximated by choosing the "neutral mean
245 F_{ST} option" with 99% confidence interval in LOSITAN. We ran LOSITAN with 100000
246 simulations and assumed a false discovery rate of 0.1 to minimize the number of
247 false positives. F_{ST} values are dependent on within-population genetic diversity,
248 which may bias estimates of the level of divergence among populations (Noor &
249 Bennett 2009; Cruickshank & Hahn 2014). To alleviate this effect in our analyses, we
250 used two complementary statistics to assess the degree of genetic divergence among
251 populations across the genome — the absolute sequence divergence (d_{xy}) and the
252 proportion of fixed differences between populations (d_f). Both statistics were
253 estimated in ngsTools using genotype likelihood without SNP calling (Fumagalli *et*
254 *al.* 2014), and kernel smoothed values were visualized in non-overlapping 90-kb
255 windows along pseudo-chromosomes in R.

256 To inspect genomic patterns of genetic diversity and allele frequency spectra,
257 we calculated pairwise nucleotide diversity (θ_π), Watterson's estimator of
258 segregating sites (θ_w) and Tajima's *D* across RAD loci located on mapped scaffolds in
259 ANGSD v 0.612 (Korneliussen *et al.* 2014). This program derives diversity and allele
260 frequency spectrum statistics using genotype likelihoods without SNP calling,
261 thereby alleviating some of the uncertainties and biases associated with SNP calling

262 in low coverage Next Generation Sequencing (Korneliussen *et al.* 2013). To gain a
263 genome-wide view and identify genomic regions of exceptional diversity and/or
264 skewed allele frequency spectra, average values of θ_{π}, θ_{w} and Tajima's *D* were
265 determined in non-overlapping 90-kb windows and plotted along pseudo-
266 chromosomes.
267
268 **Results**
269 **Genetic differentiation within *An. funestus***
270 Using alignments to the draft reference genome, we assembled 490871
271 unique RAD loci. We identified a total of 10687 high-quality biallelic markers by
272 randomly choosing one SNP across loci that were present in all populations and in at
273 least 60% of individuals within every population. A NJ tree and the first three PCA
274 axes based on these variants clearly distinguished two genetic clusters likely
275 corresponding to the two ecotypes previously described in Cameroon and in several
276 other countries with a diversity of genetic markers (Costantini *et al.* 1999; Dia *et al.*
277 2000; Kamau *et al.* 2002; Boccolini *et al.* 2005; Cohuet *et al.* 2005; Michel *et al.* 2005;
278 Ayala *et al.* 2011; Barnes *et al.* 2017) (Fig. 1B and 1C). The method of Evanno *et al.*
279 (2005) and DAPC confirmed the occurrence of two to three distinct gene pools
280 reflecting the ecological divergence known between forest and savannah
281 populations in Cameroon (Fig. 1D and 1E). However, despite this apparent
282 geographic segregation, the overall genetic differentiation is low between the two
283 putative subgroups (F_{ST} = 0.033, p < 0.005). Consistent with weak genetic
284 divergence, STRUCTURE analyses revealed a single cluster of individuals with

285 admixed ancestry at k = 2 (Fig. 1F). The moderate genetic differentiation between
286 savannah and forest populations is also illustrated by the results of an AMOVA,
287 which show that the greatest proportion of the genetic variance (98.2%, $p < 0.005$)
288 among our samples is explained by within-individual variations. The geographic
289 origin of individuals accounts for less than 2% of the genetic variance. Overall, the
290 genetic differentiation of *An. funestus* in Cameroon suggests a low level of genomic
291 divergence, consistent with extensive gene flow and/or recent split between the two
292 putative ecotypes.

293 **Divergence within chromosomal inversions**

294 We scanned the genome of savannah and forest populations to identify the
295 few regions of the genome that diverge within a largely homogeneous background.
296 We found that, in our panel of 10687 SNPs, highly differentiated loci are non-
297 randomly distributed, the highest F_{ST} values being clustered in genomic regions
298 bearing known polymorphic chromosomal inversions that have been previously
299 verified in specimen collected from this area (Cohuet *et al.* 2005; Ayala *et al.* 2011)
300 (Fig. 2 and Fig. S1). We identified a total of 107 outliers that fell above the 99th
301 percentile of the empirical distribution of F_{ST}, including 31 SNPs that were
302 successfully mapped to concatenated scaffolds and were confirmed as statistical
303 outliers in LOSITAN. F_{ST} outliers were located exclusively on two chromosome arms
304 containing nearly 20 polymorphic inversions (3R and 2R) (Fig. 2 and Fig. S1)
305 (Sharakhov *et al.* 2004). Moreover, in contrast to 2L, 3L and X, which have no
306 discriminatory power, SNPs that mapped to 3R or 2R reproduce the segregation
307 observed between the savannah and the forest at the genome level (Fig. 3). The

308 number of SNPs with F_{ST} values above the 1% threshold also varies between the 2R
309 and 3R (9 against 22) implying that mutations or structural variants of the 3R arm
310 are more strongly correlated with the genetic differentiation between the two
311 populations. Overall, the genomic distribution of F_{ST} outliers suggests that divergent
312 genomic regions between ecotypes are located within polymorphic inversion on two
313 chromosome arms.

314 To further explore the role of inversions in genetic divergence, we examined
315 the population structure separately for three large chromosomal rearrangements
316 present on 3R (3R*a*, 3R*b* and 3R*d*) and for the 2R*h* that occurs on 2R. Although
317 several inversions coexist and overlap along 2R, six out of nine F_{ST} outliers identified
318 on this chromosome mapped to scaffolds specific to portions of the 2R*h* inversion,
319 making it an interesting candidate locus. SNPs located within the 3R*a* assign more
320 that 90% of individuals to their respective ecotypes, and the 3R*b* also separates a
321 significant proportion of individuals between both sampling sites (Fig. 3). In
322 addition, consistent with the genomic pattern of differentiation, which indicates that
323 83% of F_{ST} outliers map to these two inversions (Fig. 2), the genome-wide
324 distribution of d_{xy} and fixed differences shows clear peaks within these two large
325 inversions (Fig. 4). The population structure of 2R*h* is comparable with that of 3R*b*,
326 suggesting that this inversion also contributes to the savannah-forest segregation of
327 *An. funestus* populations in Cameroon. Conversely, SNPs identified outside the 3R*a*
328 and 3R*b*, or within the 3R*d* have no discriminatory power (Fig. 3). In summary, the
329 population structure of inversions on 2R and 3R reflects the genomic distribution of
330 differentiated loci, which suggests that divergence centered on three chromosomal

331 rearrangements (3R*a*, 3R*b* and 2R*h*) maintains some level of genomic integrity
332 despite low genome-wide differentiation between ecotypes of *An. funestus* in this
333 geographic area. This divergence translates into F_{ST} values, which increase from
334 0.033 at the genome level to 0.053, 0.08 and 0.22 in 2R*h,* 3R*b* and 3R*a*, respectively.
335 Similarly, the proportion of the genetic variance explained by the geographic origin
336 of samples rises from 1.8% for the genome-wide SNPs to 3.2%, 4.7% and 12.9%
337 when only variants present within the 2R*h*, 3R*b* and 3R*a* rearrangements,
338 respectively, are included.

339 **Chromosome arm-specific diversity associated with inversions**

340 We investigated patterns of polymorphism across the genome using scans
341 based on estimates of nucleotide diversity (θ_π and θ_w) and Tajima's *D* (Fig. 5). The
342 proportion of polymorphic sites (θ_w) varies substantially between chromosome
343 arms, the highest values being found on the only collinear chromosome arm (2L)
344 (Fig. 5 and Fig. 6). Precisely, the average value of θ_w differs significantly between 2L
345 and each of the three other autosomes (Wilcoxon rank sum test, $p < 0.001$) (Fig. 6).
346 θ_w is reduced by 34.1% on 3R, 24.0% on 2R and 13.2% on 3L relative to the 2L arm,
347 regardless of the sampling location. Estimates of the number of polymorphic sites
348 along each of the five chromosome arms are also significantly different between
349 Mfou (forest) and Tibati (savannah) populations (Wilcoxon rank sum test, $p <$
350 0.001). Additionally, the drop of the average θ_w on inverted autosomes relative to 2L
351 is 34.7% on 3R, 29.2% on 2R and 13.2% on 3L in Mfou compared with 33.5% on 3R,
352 18.8% on 2R and 13.2% on 3L in Tibati (Fig. 5 and Table 1). A different demographic
353 history between ecotypes or diverse other factors including the variation in sample

354 size between the two locations can account for this difference in the level of
355 polymorphism between populations. Nevertheless, since most of the polymorphic
356 chromosomal inversions (notably 3R*a* and 3R*b*) are fixed in forest populations and
357 fluctuate in Tibati (Cohuet *et al.* 2005; Ayala *et al.* 2011), it is likely that variations in
358 karyotype frequencies of inversions contribute at least in part to the stronger
359 reduction of nucleotide diversity on inverted autosomes observed in Mfou
360 compared with Tibati. The X chromosome bears no known chromosomal
361 rearrangement, but likely due to its particular divergence and its distorted effective
362 population size, its average θ_w is also weak compared with 2L (21.1% reduction
363 relative to 2L) regardless of the sampling site. A significant reduction in pairwise
364 nucleotide diversity (θ_π) relative to the 2L is found on chromosomes 3R, X and 2R to
365 a lesser extent, but there are no clear patterns among chromosomes and sampling
366 locations as is the case with θ_w (Fig. 5, Fig. 6 and Table 1).

367 Local recombination rates are strongly correlated with genetic diversity and
368 the effective population size (Begun & Aquadro 1992). Therefore, it is possible and
369 parsimonious that the strong variation in polymorphism observed among
370 autosomes in *An. funestus* is due to the presence of multiple overlapping inversions
371 that are under selection in different geographic contexts throughout the species'
372 range. These inversions reduce meiotic crossover and genetic variation along the
373 2R, 3R and 3L autosomes, which contain multiple rearrangements. For instance,
374 three inversions (3R*a*, 3R*b* and 3R*d*) account for at least 75% of the length of the 3R
375 arm (Fig. S1). Similarly, nearly 80% of the length of 2R and 3L is affected by
376 chromosomal rearrangements. The idea that the combined effects of low

377 recombination and selection within multiple inversions contributes to reduce
378 polymorphism throughout the three inverted autosomal arms is supported by the
379 pattern of diversity observed on the collinear chromosome arm 2L, which acts as a
380 negative control. Conceivably, polymorphism is highest on this chromosome arm
381 due to the absence of inversions, which preserves more substantial levels of
382 recombination.

383 The genome-wide diversity falls in the range described in other highly
384 polymorphic *Anopheles* populations with RADseq markers (θ_w = 0.0083 and θ_π =
385 0.0042 in Mfou; θ_w = 0.0097 and θ_π = 0.0043 in Tibati) (O'Loughlin *et al.* 2014;
386 Fouet *et al.* 2017; Kamdem *et al.* 2017). The range of values of Tajima's *D* (from -
387 2.47 to -1.58 among Tibati samples and from -2.20 to -1.19 in Mfou) was shifted
388 towards negative values suggestive of recent population expansion leading to an
389 accumulation of low-frequency variants, although the effects of population
390 structure, selective sweeps, background selection or changes in sample size can
391 contribute to skewed site frequency spectra across the genome (Donnelly *et al.*
392 2001; Thornton 2005; Gattepaille *et al.* 2013) (Fig. 5). Unsurprisingly, Tajima's *D* is
393 more negative on the 2L chromosome, which exhibits the greatest difference
394 between the number of polymorphic sites and the pairwise nucleotide diversity in
395 both ecotypes.

396 **Discussion**

397 **Recombination, inversions and genetic divergence**

398 Our analysis based on more than 10000 genome-wide SNPs revealed that *An.*
399 *funestus* populations collected from two different moisture conditions in Cameroon
400 are very weakly differentiated. These results corroborate two previous studies
401 based on microsatellites and polymorphic chromosomal inversions, which have
402 found similar levels of divergence between *An. funestus* populations in this region
403 (Cohuet *et al.* 2005; Ayala *et al.* 2011). The population genetic structure of this
404 mosquito throughout Africa remains largely unresolved, but based on many
405 previous studies in different parts of the continent, it has been proposed that at least
406 two ecotypes segregate within this species (Green & Hunt 1980; Costantini *et al.*
407 1999; Dia *et al.* 2000; Kamau *et al.* 2002; Boccolini *et al.* 2005; Cohuet *et al.* 2005;
408 Michel *et al.* 2005; Ayala *et al.* 2011; Barnes *et al.* 2017). These studies have also
409 indicated that polymorphic chromosomal inversions play a key role in ecological
410 divergence in *An. funestus* (Green & Hunt 1980; Costantini *et al.* 1999; Dia *et al.*
411 2000; Cohuet *et al.* 2005; Ayala *et al.* 2011). The 3R*a* and 3R*b* inversions in
412 particular have long been suspected to be strongly associated with ecological
413 divergence due to the clinal distribution of alternative karyotypes and the
414 significant deficit in heterokaryotypes observed in several countries (Costantini *et*
415 *al.* 1999; Dia *et al.* 2000; Ayala *et al.* 2011). Here we provide the first genomic
416 evidence that genetic divergence within this species is limited to a few loci bearing
417 large chromosomal rearrangements. Our conclusions come with some caveats due
418 to the potential limitations of the sampling scheme, the reference genome and the

419 RAD sequencing approach we used. For instance, it is likely that some genomic
420 signatures of selection are not detected because of the limited genomic coverage of
421 RAD tags and pseudo-chromosomes (Arnold *et al.* 2013; Tiffin & Ross-Ibarra 2014).
422 Bearing in mind these caveats, the clustering of genetic differentiation within the
423 3R*a*, 3R*b* and 2R*h* inversions appears to be consistent with the hypothesis that these
424 rearrangements are the major genetic targets of divergent selection in Cameroon.

425 The phenotypic, behavioral or ecological variations associated with
426 inversions in *An. funestus* remain unknown. In principle, the segregation between
427 the arid zone and the humid forest may potentially involve fitness traits and
428 phenotypes contributing to thermal tolerance. Indeed, examples of inversions
429 whose alternative karyotypes provide selective advantages in different moisture
430 conditions are well known in dipteran species (e.g. the 2L*a* inversion and to a lesser
431 extent the 2R*b* in *An. gambiae*) (Coluzzi *et al.* 1979; Gray *et al.* 2009; Lee *et al.* 2009;
432 Fouet *et al.* 2012; Cheng *et al.* 2012). Comparative cytogenetic studies also showed a
433 high proportion of conserved genes between the 3R*b* in *An. funestus* and the 2L*a* in
434 *An. gambiae* and between the 2R*h* in *An. funestus* and the 2R*b* in *An. gambiae*
435 (Sharakhova *et al.* 2011). Although this hypothesis awaits thorough investigation,
436 the two pairs of inversions may have inherited similar phenotypes via a common
437 ancestor or via convergent evolution between *An. funestus* and *An. gambiae*.

438 Theoretical models and empirical data support the idea that, at early stages
439 of divergence as in the case of *An. funestus* ecotypes, genetic differentiation is
440 restricted to a few genomic regions because the effects of natural selection at this
441 step are very localized (Feder *et al.* 2012; Nosil & Feder 2012; Andrew & Rieseberg

442 2013; Seehausen *et al.* 2014). Also, selection at a locus affects the level of genetic
443 differentiation among populations and reciprocally the degree of subdivision within
444 a population impacts patterns of variation at selected loci (Lewontin & Krakauer
445 1973; Charlesworth *et al.* 1997; Slatkin & Wiehe 1998; Majewski & Cohan 1999; Kim
446 & Maruki 2011; Schneider & Kim 2013). In fact, it has been proposed that selection
447 modulates levels of divergence locally and globally across the genome through two
448 types of genetic hitchhiking — divergence hitchhiking and genomic hitchhiking
449 (Feder *et al.* 2012). Divergent hitchhiking (the spread of a mutation and linked
450 neutral sequences within a population) (Maynard Smith & Haigh 1974) causes the
451 diffusion of favorable alleles at loci important for local adaptation and ecological
452 differentiation. Limited gene flow that occurs between divergent lineages can
453 naturally lead to genomic hitchhiking, which reflects the global accumulation of
454 differences at the genome level through different selective or demographic
455 mechanisms (Feder *et al.* 2012). In *An. funestus*, it is plausible that divergent
456 hitchhiking within the three important inversions, 3R*a*, 3R*b* and 2R*h*, is so far
457 strongly counterbalanced by extensive gene flow and the weak overall number of
458 selected loci across the genome, leading to minimal genomic divergence among
459 ecotypes.

460 Although some controversy persists (see McGaugh *et al.* 2012), the idea that
461 strongly differentiated genomic regions between species or between populations
462 within the same species accumulate disproportionately in regions of low genetic
463 recombination including chromosome centers and chromosomal rearrangements
464 has been strongly supported in many species notably humans, *Drosophila* flies,

465 stickleback fish and maize (Hellmann *et al.* 2003, 2005; Tenaillon *et al.* 2004;
466 Kulathinal *et al.* 2008; Cai *et al.* 2009; Keinan & Reich 2010; Nachman & Payseur
467 2012; McGaugh & Noor 2012; Roesti *et al.* 2012, 2013). This coincidence can be
468 explained by a mechanical effect of recombination, which reduces genetic
469 polymorphism within populations and thereby generates sequence divergence
470 between populations as a simple byproduct of diversity reduction (Begun &
471 Aquadro 1992; Roesti *et al.* 2012). Concordance between limited rate of crossing
472 over and increased divergence may also be directly correlated with more
473 pronounced effects of divergent selection in genomic regions in which
474 recombination is less frequent (Charlesworth *et al.* 1997; Charlesworth 1998;
475 Nachman 2002; Cutter & Payseur 2013; Cruickshank & Hahn 2014). Within
476 inversions in particular, reduced recombination enhances divergent selection acting
477 on locally adapted gene complexes, which co-segregate with or without epistatic
478 interactions (Dobzhansky 1950; Kirkpatrick & Barton 2006).

479 **Recombination, inversions and genetic polymorphism**

480 We have found chromosome arm specific patterns of polymorphism in *An.*
481 *funestus* characterized by a significant difference between collinear and inverted
482 chromosomes. We first conducted several tests to insure that uncertainties
483 associated with our sequencing and analytical approach cannot account for the
484 observed disparities in genetic polymorphism between chromosomes. We
485 compared the distribution and the length of mapped scaffolds and confirmed that
486 each chromosome arm is represented by a substantial number of long scaffolds
487 (Table S2 and Fig. S1). The length of pseudo-chromosomes is also proportional to

488 the length of chromosome arms, ranging from 6.8 Mb on X (the smallest) to 27.4 Mb
489 on 2R (the largest). Therefore, we can reasonably rule out the possibility that the
490 chromosome-bias diversity is due to a systematic error associated with our
491 reference sequence. Additionally, mapped scaffolds are evenly distributed along the
492 length of chromosomes, suggesting that such important variations in the amount of
493 polymorphism between chromosome arms are not due to centromere- and/or
494 telomere-proximal effects (Fig. S1) (Aguade *et al.* 1989; Stephan & Langley 1989).
495 Finally, we noted that the mean depth of sequencing coverage per mapped scaffolds
496 was consistent across the five chromosome arms (Fig. S2), implying that
497 chromosome-specific diversity is not simply a covariate of sequencing biases.

498 As estimates of the number of segregating sites (θ_w) also known as the
499 population-scaled mutation rate (Watterson 1975) are the most drastically affected
500 by between-chromosome variations observed in *An. funestus*, the stark contrast in
501 the amount of polymorphism among inverted and collinear autosomes may in
502 theory be due to lower mutation rates on inverted chromosomes. However, as
503 shown in other insect species, such a variability in mutation rates between
504 chromosomes or between large segments of the genome is unlikely (Begun *et al.*
505 2007; Keightley *et al.* 2015). Instead, the occurrence of chromosome-specific
506 patterns of diversity in *An. funestus* is consistent with the presence of weakly
507 recombining autosomes bearing multiple overlapping chromosomal
508 rearrangements. These same causes produce similar effects along the non-
509 recombining portion of the Y chromosome in mammals (Lahn & Page 1999) or along
510 balancer chromosomes described in *Drosophila*, *Caenorhabditis* and rodent species

511 (Muller 1918; Hentges & Justice 2004; Edgley *et al.* 2006). Indeed, population
512 genetic analyses across genomes of diverse taxa have found a positive correlation
513 between the rate of recombination and genetic variation (Aguade *et al.* 1989;
514 Stephan & Langley 1989; Begun & Aquadro 1992; Hellmann *et al.* 2003, 2005;
515 Tenaillon *et al.* 2004; Begun *et al.* 2007; Kulathinal *et al.* 2008, 2009). This
516 correlation translates into specific patterns of diversity that can be observed at the
517 species, genomic or chromosomal levels. At the species level, plants and some
518 animal species, which reproduce by self-fertilization have reduced overall genomic
519 diversity due to low genetic recombination (Nordborg *et al.* 1996; Akhunov *et al.*
520 2010; Cutter & Choi 2010; Andersen *et al.* 2012; Thomas *et al.* 2015). The local
521 effects of recombination on genetic polymorphism are also particularly evident
522 across centromere and telomeres of chromosomes and within genomic regions
523 bearing structural rearrangements (Aguade *et al.* 1989; Stephan & Langley 1989;
524 Andolfatto *et al.* 2001; Corbett-Detig & Hartl 2012; Pool *et al.* 2012).

525 The relationship between diversity and crossing over has been ascribed to
526 two processes: selection and mutagenesis. Recombination may influence
527 polymorphism because of new mutations created during crossing over, which
528 increase diversity (the mutagenic effect of recombination) (Magni & Von Borstel
529 1962). However, multiple studies that have analyzed fine-scale recombination in
530 humans, yeast, *Arabidopsis*, *Anopheles*, *Drosophila*, and *Caenorhabditis* have
531 undermined the notion that recombination, or some correlate, generally exerts
532 mutagenic effects on genomes (Betancourt & Presgraves 2002; Stump *et al.* 2005;
533 Spencer *et al.* 2006; Wright *et al.* 2006; Begun *et al.* 2007; Noor 2008; Denver *et al.*

534 2009; Cutter & Choi 2010; Mcgaugh *et al.* 2012). Alternatively, recombination may
535 affect diversity indirectly by modulating the effects of positive or negative selection
536 across the genome. Indeed, the hitchhiking of favorable alleles (selective sweep) or
537 the removal of recurrent deleterious mutations (background selection) affect linked
538 neutral sequences more strongly in low-recombination genomic regions, which
539 gives rise to a positive correlation between diversity and recombination rate
540 (Maynard Smith & Haigh 1974; Kaplan *et al.* 1989; Begun & Aquadro 1992;
541 Charlesworth *et al.* 1993, 1997; Nordborg *et al.* 1996; Andolfatto 2001; Nachman
542 2002). Another indirect effect of recombination on genetic diversity may be driven
543 by the Hill-Roberston interference, which occurs when two sites under weak
544 selection are in physical linkage and as a result, selection at one site interferes with
545 selection at another site (Hill & Robertson 1966; Felsenstein 1974). Hill-Roberston
546 interference can also be thought of as a reduction in the effective population size
547 (N_e) caused by selection at linked loci (Comeron *et al.* 2008; Cutter & Payseur 2013;
548 Castellano *et al.* 2016). Recombination, by alleviating interference between linked
549 sites, alleviates this reduction in N_e leading to a positive correlation between
550 recombination rate and levels of neutral polymorphism. Because of the relationship
551 between recombination and polymorphism, measures of the skew in the allele
552 frequency spectrum, such as Tajima's *D* values (Tajima 1989), are expected to
553 positively correlate with the rate of recombination (Braverman *et al.* 1995). The *An.*
554 *funestus* genome shows a strong relationship between Tajima's *D* values and
555 recombination, and as expected, the chromosome 2L, whose recombination is not
556 limited by inversions, exhibits the sharpest skew in the AFS.

557 **Conclusions and perspectives**

558 We found evidence that, among the many chromosomal rearrangements
559 identified in *An. funestus*, genomic footprints of divergence are centered on three
560 inversions that are potential targets of selection among ecotypes in Cameroon. The
561 other rearrangements are likely either cosmopolitan or endemic inversions under
562 selection at various geographic extents that have yet to be resolved. Our data
563 support the idea that interactions between recombination and selection — which
564 amplify the effects of selective sweeps or background selection within genomic
565 regions where recombination is blocked by multiple chromosomal inversions —
566 account for the strong disparity in nucleotide diversity observed between
567 autosomes in this mosquito.

568 To deepen our understanding of the adaptive role of inversions and their
569 contribution to chromosome specific patterns of diversity, a complete reference
570 genome assembly and extensive sampling across the species range are needed.
571 These resources will make it possible to design more sensitive tests including the
572 functional characterization of footprints of selection and their detailed signatures
573 among individuals and populations. The presence of hallmarks of both reduced
574 recombination and linked selection across large genomic sequences in *An. funestus*
575 highlights the important contribution of multiple aspects of linkage in genome
576 evolution. These aspects have significant implications for the detection of genomic
577 signatures of adaptation in species whose genome contains multiple polymorphic
578 chromosomal inversions.

579

Acknowledgements

We thank the associate editor and six reviewers for their constructive comments. This work was supported by NIH grant 1R01AI113248 to BJW.

Author contributions

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

Data Archiving Statement

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

594 **References**

595 Aguade M, Miyashita N, Langley CH (1989) Reduced variation in the yellow-achaete-
596 scute region in natural populations of Drosophila melanogaster. *Genetics*, **122**,
597 607–15.

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

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

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

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

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

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

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

617 Arnold B, Corbett-Detig RB, Hartl D, Bomblies K (2013) RADseq underestimates
618 diversity and introduces genealogical biases due to nonrandom haplotype
619 sampling. *Molecular Ecology*, **22**, 3179–3190.
620 Ayala D, Fontaine MC, Cohuet A *et al.* (2011) Chromosomal inversions, natural
621 selection and adaptation in the malaria vector Anopheles funestus. *Molecular*
622 *biology and evolution*, **28**, 745–758.
623 Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP Discovery and Genetic
624 Mapping Using Sequenced RAD Markers. *PLoS ONE*, **3**, e3376.
625 Barnes KG, Weedall GD, Ndula M *et al.* (2017) Genomic Footprints of Selective
626 Sweeps from Metabolic Resistance to Pyrethroids in African Malaria Vectors
627 Are Driven by Scale up of Insecticide-Based Vector Control. *PLoS Genetics*,
628 **13(2): e10**, 1–22.
629 Beaumont MA (2005) Adaptation and speciation : what can F st tell us ? *Trends in*
630 *Ecology & Evolution*, **20**, 435–440.
631 Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of
632 population structure. *Proceedings of the Royal Society B: Biological Sciences*,
633 **263**, 1619–1626.
634 Begun D, Aquadro C (1992) Levels of naturally occurring DNA polymorphism
635 correlate with recombination rates in D. melanogaster. *Nature*, **356**, 519–520.
636 Begun DJ, Holloway AK, Stevens K *et al.* (2007) Population genomics: whole-genome
637 analysis of polymorphism and divergence in Drosophila simulans. *PLoS biology*,
638 **5**, e310.
639 Betancourt AJ, Presgraves DC (2002) Linkage limits the power of natural selection in

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

663 Charlesworth B, Nordborg M, Charlesworth D (1997) The effects of local selection,
664 balanced polymorphism and background selection on equilibrium patterns of
665 genetic diversity in subdivided populations. *Genet. Res., Camb*, **70**, 155–174.
666 Cheng C, White BJ, Kamdem C *et al.* (2012) Ecological genomics of Anopheles
667 gambiae along a latitudinal cline: a population-resequencing approach.
668 *Genetics*, **190**, 1417–32.
669 Coetzee M, Koekemoer LL (2013) Molecular systematics and insecticide resistance
670 in the major African malaria vector Anopheles funestus. *Annual Review of*
671 *Entomology*, **58**, 393–412.
672 Cohuet A, Dia I, Simard F *et al.* (2005) Gene flow between chromosomal forms of the
673 malaria vector Anopheles funestus in Cameroon, Central Africa, and its
674 relevance in malaria fighting. *Genetics*, **169**, 301–311.
675 Cohuet A, Simard F, Toto J-C *et al.* (2003) Species identification within the Anopheles
676 funestus group of malaria vectors in Cameroon and evidence for a new species.
677 *The American journal of tropical medicine and hygiene*, **69**, 200–205.
678 Coluzzi M, Sabatini A, Petrarca V, Di Deco M (1979) Chromosomal differentiation
679 and adaptation to human environments in the Anopheles gambiae complex.
680 *Trans R Soc Trop Med Hyg*, **73**, 483–97.
681 Comeron JM, Ratnappan R, Bailin S (2012) The Many Landscapes of Recombination
682 in Drosophila melanogaster. *PLoS Genetics*, **8**, 33–35.
683 Comeron JM, Williford a, Kliman RM (2008) The Hill-Robertson effect: evolutionary
684 consequences of weak selection and linkage in finite populations. *Heredity*, **100**,
685 19–31.

686 Corbett-Detig RB, Hartl DL (2012) Population Genomics of Inversion
687 Polymorphisms in Drosophila melanogaster. *PLoS Genetics*, **8**.
688 Costantini C, Sagnon N, Ilboudo-Sanogo E, Coluzzi M, Boccolini D (1999)
689 Chromosomal and bionomic heterogeneities suggest incipient speciation in
690 Anopheles funestus from Burkina Faso. *Parassitologia*, **41**, 595–611.
691 Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of
692 speciation are due to reduced diversity, not reduced gene flow. *Molecular*
693 *ecology*, **23**, 3133–57.
694 Cutter AD, Choi JY (2010) Natural selection shapes nucleotide polymorphism across
695 the genome of the nematode Caenorhabditis briggsae Natural selection shapes
696 nucleotide polymorphism across the genome of the nematode Caenorhabditis
697 briggsae. *Genome Research*, 1103–1111.
698 Cutter AD, Payseur BA (2013) Genomic signatures of selection at linked sites:
699 unifying the disparity among species. *Nature reviews. Genetics*, **14**, 262–74.
700 Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools.
701 *Bioinformatics*, **27**, 2156–2158.
702 Denver DR, Dolan PC, Wilhelm LJ *et al.* (2009) A genome-wide view of
703 Caenorhabditis elegans base-substitution mutation processes. *Proceedings of*
704 *the National Academy of Sciences of the United States of America*, **106**, 16310–4.
705 Dia I, Guelbeogo M, Ayala D (2013) Advances and Perspectives in the Study of the
706 Malaria Mosquito Anopheles funestus. In: *Anopheles mosquitoes - New insights*
707 *into malaria vectors* (ed Manguin S)
708 Dia I, Lochouarn L, Boccolini D, Costantini C, Fontenille D (2000) Spatial and

709 temporal variations of the chromosomal inversion polymorphism of Anopheles
710 funestus in Senegal. *Parasite*, **7**, 179–184.
711 Dobzhansky T (1943) Genetics of natural populations. IX. Temporal changes in the
712 composition of populations of Drosophila pseudoobscura. *Genetics*, **28**, 162–
713 186.
714 Dobzhansky T (1948) Genetics of natural populations .XVIII. Experiments on
715 chromosomes of Drosophila pseudoobscura from different geographic regions.
716 *Genetics*, **33**, 588–602.
717 Dobzhansky T (1950) Genetics of natural populations. XIX. Origin of heterosis
718 through natural selection in populations of Drosophila pseudoobscura. *Genetics*,
719 **35**, 288–302.
720 Dobzhansky T (1970) *Genetics of the evolutionary process*. New York.
721 Donnelly MJ, Licht MC, Lehmann T (2001) Evidence for recent population expansion
722 in the evolutionary history of the malaria vectors Anopheles arabiensis and
723 Anopheles gambiae. *Molecular biology and evolution*, **18**, 1353–1364.
724 Earl DA, VonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for
725 visualizing STRUCTURE output and implementing the Evanno method.
726 *Conservation Genetics Resources*, **4**.
727 Edgley ML, Baillie DL, Riddle DL, Rose AM (2006) *Genetic balancers* (Wormbook,
728 Ed,).
729 Evanno G, Goudet J, Regnaut S (2005) Detecting the number of clusters of
730 individuals using the software structure: a simulation study. *Molecular Ecology*,
731 **14**, 2611–2620.

732 Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred
733 from metric distances among DNA haplotypes: Application to human
734 mitochondrial DNA restriction data. *Genetics*, **131**, 479–491.
735 Feder JL, Egan SP, Nosil P (2012) The genomics of speciation-with-gene-flow. *Trends*
736 *in Genetics*, **28**, 342–350.
737 Felsenstein J (1974) The evolutionary advantage of recombination. II. Individual
738 selection for recombination. *Genetics*, **78**, 737–756.
739 Fouet C, Gray E, Besansky NJ, Costantini C (2012) Adaptation to aridity in the
740 malaria mosquito Anopheles gambiae: chromosomal inversion polymorphism
741 and body size influence resistance to desiccation. *PloS one*, **7**, e34841.
742 Fouet C, Kamdem C, Gamez S, White BJ (2017) Extensive genetic diversity among
743 populations of the malaria mosquito Anopheles moucheti revealed by
744 population genomics. *Infection, Genetics and Evolution*, **48**, 27–33.
745 Fumagalli M, Vieira FG, Linderoth T (2014) ngsTools : methods for population
746 genetics analyses from Next-Generation Sequencing data. *Bioinformatics*.
747 Gattepaille LM, Jakobsson M, Blum MGB (2013) Inferring population size changes
748 with sequence and SNP data: lessons from human bottlenecks. *Heredity*, **110**,
749 409–19.
750 Gillespie JH (1997) Junk ain't what junk does: Neutral alleles in a selected context.
751 *Gene*, **205**, 291–299.
752 Gillies MT, Coetzee M (1987) *A supplement to the Anophelinae of Africa south of the*
753 *Sahara*. The South African Institute for Medical Research, Johannesburg.
754 Gillies MT, De Meillon B (1968) *The Anophelinae of Africa South of the Sahara*.

755 Publications of the South African Institute for Medical Research, Johannesburg.
756 Giraldo-Calderon GI, Emrich SJ, MacCallum RM *et al.* (2015) VectorBase: an updated
757 bioinformatics resource for invertebrate vectors and other organisms related
758 with human diseases. *Nucleic Acids Research*, **43**, D707–D713.
759 Gosset CC, Bierne N (2013) Differential introgression from a sister species explains
760 high FST outlier loci within a mussel species. *Journal of Evolutionary Biology*,
761 **26**, 14–26.
762 Gray EM, Rocca K a C, Costantini C, Besansky NJ (2009) Inversion 2La is associated
763 with enhanced desiccation resistance in Anopheles gambiae. *Malaria journal*, **8**,
764 215.
765 Green CA, Hunt RH (1980) Interpretation of variation in ovarian polytene
766 chromosomes of Anopheles funestus Giles, A. parensis Gillies, and A. aruni ?
767 *Genetica*, **51**, 187–195.
768 Haddrill PR, Halligan DL, Tomaras D, Charlesworth B (2007) Reduced efficacy of
769 selection in regions of the Drosophila genome that lack crossing over. *Genome*
770 *biology*, **8**, R18.
771 Hellmann I, Ebersberger I, Ptak SE, Pääbo S, Przeworski M (2003) A Neutral
772 Explanation for the Correlation of Diversity with Recombination Rates in
773 Humans. *The American Journal of Human Genetics*, **72**, 1527–1535.
774 Hellmann I, Prüfer K, Ji H *et al.* (2005) Why do human diversity levels vary at a
775 megabase scale? *Genome Research*, **15**, 1222–1231.
776 Hentges KE, Justice MJ (2004) Checks and balancers : balancer chromosomes to
777 facilitate genome annotation. *Trends in Genetics*, **20**.

Hill WG, Robertson A (1966) The effect of linkage on the limits of artificial selection. *Genetical Research*, **8**, 269–294.

Hoffmann A a, Rieseberg LH (2008) Revisiting the Impact of Inversions in Evolution: From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation? *Annual review of ecology, evolution, and systematics*, **39**, 21–42.

Hoffmann AA, Sgro CM, Weeks AR (2004) Chromosomal inversion polymorphisms and adaptation. *Trends in Ecology & Evolution*, **19**, 482–488.

Hoffmann AA, Weeks AR (2007) Climatic selection on genes and traits after a 100 year-old invasion : a critical look at the temperate-tropical clines in Drosophila melanogaster from eastern Australia. *Genetica*, **129**, 133–147.

Huber CD, Nordborg M, Hermisson J, Hellmann I (2014) Keeping It Local: Evidence for Positive Selection in Swedish Arabidopsis thaliana. *Molecular Biology and Evolution*, **31**, 3026–3039.

Hudson RR, Kaplan NL (1988) The coalescent process in models with selection and recombination. *Genetics*, **120**, 819–829.

Jakobsson M, Rosenberg N (2007) CLUMPP: a cluster matching and permutation program for dealing with multimodality in analysis of population structure. *Bioinformatics*, **23**, 1801– 1806.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

De Jong G, Bochdanovits Z (2003) Latitudinal clines in Drosophila melanogaster : body size , allozyme frequencies , inversion frequencies , and the insulin-signalling pathway. *Journal of Genetics*, **82**, 207–223.

801 Joron M, Frezal L, Jones RT *et al.* (2011) Chromosomal rearrangements maintain a
802 polymorphic supergene controlling butterfly mimicry. *Nature*, **477**, 203–206.
803 Kamau L, Hunt RH, Coetzee M (2002) Analysis of the population structure of
804 Anopheles funestus (Diptera: Culicidae) from western and coastal Kenya using
805 paracentric chromosomal inversion frequencies. *Journal of Medical Entomology*,
806 **39**, 78–83.
807 Kamdem C, Fouet C, Gamez S, White BJ (2017) Pollutants and insecticides drive local
808 adaptation in African malaria mosquitoes. *Molecular Biology and Evolution*, **34**,
809 1261–1275.
810 Kaplan NL, Hudson RR, Langley CH (1989) The "hitchhiking effect" revisited.
811 *Genetics*, **123**, 887–899.
812 Kapun M, Van Schalkwyk H, McAllister B, Flatt T, Schlötterer C (2014) Inference of
813 chromosomal inversion dynamics from Pool-Seq data in natural and laboratory
814 populations of Drosophila melanogaster. *Molecular Ecology*, **23**, 1813–1827.
815 Kapun M, Schmidt C, Durmaz E, Schmidt PS, Flatt T (2016) Parallel effects of the
816 inversion In (3R) Payne on body size across the North American and
817 Australian clines in Drosophila melanogaster. *Journal of Evolutionary Biology*,
818 **29**, 1059–1072.
819 Kawecki TJ, Dieter E (2004) Conceptual issues in local adaptation. *Ecology Letters*, **7**,
820 1225–1241.
821 Keightley PD, Pinharanda A, Ness RW *et al.* (2015) Estimation of the Spontaneous
822 Mutation Rate in Heliconius melpomene. *Molecular Biology and Evolution*, **32**,
823 239–243.

824 Keinan A, Reich D (2010) Human population differentiation is strongly correlated
825 with local recombination rate. *PLoS Genetics*, **6**.
826 Kennington WJ, Hoffmann AA, Partridge L (2007) Mapping Regions Within
827 Cosmopolitan Inversion In(3R)Payne Associated With Natural Variation in
828 Body Size in Drosophila melanogaster. *Genetics*, **177**, 549–556.
829 Kim Y, Maruki T (2011) Hitchhiking effect of a beneficial mutation spreading in a
830 subdivided population. *Genetics*, **189**, 213–226.
831 Kimura MK (1983) *The Neutral Theory of Molecular Evolution*. Cambridge, UK.
832 Kirkpatrick M (2010) How and why chromosome inversions evolve. *PLoS Biology*, **8**.
833 Kirkpatrick M, Barton N (2006) Chromosome inversions, local adaptation and
834 speciation. *Genetics*, **173**, 419–434.
835 Kliman R, Hey J (1993) Reduced natural selection associated with low
836 recombination in Drosophila melanogaster. *Molecular biology and evolution*, **10**,
837 1239–1258.
838 Knibb WR (1982) Chromosome inversion polymorphisms in Drosophila
839 melanogaster II. Geographic clines and climatic associations in Australasia ,
840 North America and Asia. *Genetica*, **58**, 213–221.
841 Korneliussen T, Albrechtsen A, Nielsen R (2014) ANGSD: Analysis of Next
842 Generation Sequencing Data. *BMC Bioinformatics*, **15**, 356.
843 Korneliussen TS, Moltke I, Albrechtsen A, Nielsen R (2013) Calculation of Tajima's D
844 and other neutrality test statistics from low depth next-generation sequencing
845 data. *BMC bioinformatics*, **14**, 289.
846 Krimbas C, Powell JR (1992) Introduction. In: *Drosophila inversion polymorphism*

847 (eds Krimbas C, Powell JR), p. . London.

848 Kulathinal RJ, Bennett SM, Fitzpatrick CL, Noor MAF (2008) Fine-scale mapping of
849 recombination rate in Drosophila refines its correlation to diversity and
850 divergence. *Proceedings of the National Academy of Sciences*, **105**, 10051–
851 10056.

852 Kulathinal RJ, Stevison LS, Noor MAF (2009) The genomics of speciation in
853 Drosophila: Diversity, divergence, and introgression estimated using low-
854 coverage genome sequencing. *PLoS Genetics*, **5**.

855 Lahn BT, Page DC (1999) Chromosome Four Evolutionary Strata on the Human X
856 Chromosome. *Science*, **286**, 964–967.

857 Lanzaro GC, Lee Y (2013) Speciation in Anopheles gambiae — The Distribution of
858 Genetic Polymorphism and Patterns of Reproductive Isolation Among Natural
859 Populations. In: *Anopheles mosquitoes - New insights into malaria vectors* (ed
860 Manguin S)

861 Lee SF, Chen Y, Varan AK *et al.* (2011) Molecular Basis of Adaptive Shift in Body Size
862 in Drosophila melanogaster : Functional and Sequence Analyses of the Dca Gene
863 Research article. *Molecular Biology and Evolution*, **28**, 2393–2402.

864 Lee Y, Meneses CR, Fofana A, Lanzaro GC (2009) Desiccation Resistance Among
865 Subpopulations of Anopheles gambiae s.s. From Selinkenyi, Mali. *Journal of*
866 *medical entomology*, **46**, 316–320.

867 Lewontin RC, Krakauer J (1973) DISTRIBUTION OF GENE FREQUENCY AS A TEST
868 OF THE THEORY O F THE SELECTIVE NEUTRALITY OF and LEWONTIN.
869 *Genetics*, **74**, 175–195.

870 Lischer HEL, Excoffier L (2012) PGDSpider: An automated data conversion tool for
871 connecting population genetics and genomics programs. *Bioinformatics*, **28**,
872 298–299.
873 Lowry DB, Willis JH (2010) A Widespread Chromosomal Inversion Polymorphism
874 Contributes to a Major Life-History Transition , Local Adaptation , and
875 Reproductive Isolation. *PLoS Biology*, **8**.
876 Magni GE, Von Borstel RC (1962) Different Rates of Spontaneous Mutation during
877 Mitosis and Meiosis in Yeast. *Genetics*, **47**, 1097–1108.
878 Majewski J, Cohan FM (1999) Adapt globally, act locally: The effect of selective
879 sweeps on bacterial sequence diversity. *Genetics*, **152**, 1459–1474.
880 Maynard Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene.
881 *Genetical research*, **23**, 23–35.
882 Mcgaugh SE, Heil CSS, Manzano-winkler B *et al.* (2012) Recombination Modulates
883 How Selection Affects Linked Sites in Drosophila. *PLoS Biology*, **10**.
884 McGaugh SE, Noor M a F (2012) Genomic impacts of chromosomal inversions in
885 parapatric Drosophila species. *Philosophical transactions of the Royal Society of*
886 *London. Series B, Biological sciences*, **367**, 422–9.
887 Meirmans P, Van Tienderen P (2004) GENOTYPE and GENODIVE: two programs for
888 the analysis of genetic diversity of asexual organisms. *Molecular Ecology Notes*,
889 **4**, 792–794.
890 Mettler LE, Voelker RA, Mukai T (1977) Inversion clines in populations of
891 Drosophila melanogaster. *Genetics*, **87**, 169–176.
892 Michel a. P, Ingrasci MJ, Schemerhorn BJ *et al.* (2005) Rangewide population genetic

structure of the African malaria vector Anopheles funestus. *Molecular Ecology*, **14**, 4235–4248.

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

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

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

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

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

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

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

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

Noor MAF, Grams KL, Bertucci LA, Reiland J (2001) Chromosomal inversions and the

916 reproductive isolation of species. *Proceedings of the National Academy of*
917 *Sciences*, **98**, 12084–12088.
918 Nordborg M, Charlesworth B, Charlesworth B (1996) Increased levels of
919 polymorphism surrounding selectively maintained sites in highly selfing
920 species. *Proceedings of the Royal Society B: Biological Sciences*, **263**, 1033–1039.
921 Nordborg M, Hu TT, Ishino Y *et al.* (2005) The pattern of polymorphism in
922 Arabidopsis thaliana. *PLoS Biology*, **3**, 1289–1299.
923 Nordborg M, Tavar S (2002) Linkage disequilibrium : what history has to tell us.
924 *Trends in Genetics*, **18**, 83–90.
925 Nosil P, Feder JL (2012) Widespread yet heterogeneous genomic divergence.
926 *Molecular Ecology*, **21**, 2829–2832.
927 Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous
928 genomic divergence. *Molecular Ecology*, **18**, 375–402.
929 O'Loughlin SM, Magesa S, Mbogo C *et al.* (2014) Genomic Analyses of Three Malaria
930 Vectors Reveals Extensive Shared Polymorphism but Contrasting Population
931 Histories. *Molecular biology and evolution*, 1–14.
932 Ortiz-barrientos D, Engelstädter J, Rieseberg LH (2016) Recombination Rate
933 Evolution and the Origin of Species. *Trends in Ecology & Evolution*, **31**, 226–236.
934 Ortiz-Barrientos D, Reiland J, Hey J, Noor MAF (2002) Recombination and the
935 divergence of hybridizing species. *Genetica*, **116**, 167–178.
936 Paradis E, Claude J, Strimmer K (2004) Analyses of Phylogenetics and Evolution in R
937 language. *Bioinformatics*, **20**, 289–290.
938 Pegueroles C, Ordóñez V, Mestres F, Pascual M (2010) Recombination and selection

939 in the maintenance of the adaptive value of inversions. *Journal of Evolutionary*
940 *Biology*, **23**, 2709–2717.
941 Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest
942 RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in
943 Model and Non-Model Species. *PLoS ONE*, **7**, e37135.
944 Pool JE, Corbett-Detig RB, Sugino RP *et al.* (2012) Population Genomics of Sub-
945 Saharan Drosophila melanogaster: African Diversity and Non-African
946 Admixture. *PLoS Genetics*, **8**.
947 Powell JR (1997) *Progress and prospects in evolutionary biology: the Drosophila*
948 *model*. New York.
949 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using
950 multilocus genotype data. *Genetics*, **155**, 945–959.
951 Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: a toolset for whole-genome
952 association and population-based linkage analysis. *American Journal of Human*
953 *Genetics*, **81**.
954 R Development Core Team (2014) *R: A language and environment for statistical*
955 *computing*. R Foundation for Statistical Computing, Vienna, Austria.
956 Rako L, Anderson AR, Sgro CM, Stocker AJ, Hoffmann AA (2006) The association
957 between inversion In (3R) Payne and clinally varying traits in Drosophila
958 melanogaster. *Genetica*, **128**, 373–384.
959 Riehle MM, Guelbeogo WM, Gneme A *et al.* (2011) A cryptic subgroup of Anopheles
960 gambiae is highly susceptible to human malaria parasites. *Science (New York,*
961 *N.Y.)*, **331**, 596–8.

962 Rieseberg LH (2001) Chromosomal rearrangements and speciation. *Trends in*
963 *Ecology and Evolution*, **16**, 351–358.
964 Roberts P (1976) The genetics of chromosomal aberration. In: *The genetics and*
965 *biology of Drosophila* (eds Ashburner M, Novitski E), pp. 67–184. London.
966 Roesti M, Hendry AP, Salzburger W, Berner D (2012) Genome divergence during
967 evolutionary diversification as revealed in replicate lake-stream stickleback
968 population pairs. *Molecular ecology*, **21**, 2852–62.
969 Roesti M, Moser D, Berner D (2013) Recombination in the threespine stickleback
970 genome - Patterns and consequences. *Molecular Ecology*, **22**, 3014–3027.
971 Rosenberg N (2004) DISTRUCT: a program for the graphical display of population
972 structure. *Molecular Ecology Resources*, **4**, 137–138.
973 Savolainen O, Lascoux M, Merilä J (2013) Ecological genomics of local adaptation. ,
974 **14**.
975 Schaeffer SW (2008) Selection in heterogeneous environments maintains the gene
976 arrangement polymorphism of drosophila pseudoobscura. *Evolution*, **62**, 3082–
977 3099.
978 Schaeffer SW, Goetting-minesky MP, Kovacevic M *et al.* (2003) Evolutionary
979 genomics of inversions in Drosophila pseudoobscura : Evidence for epistasis.
980 *Proceedings of the National Academy of Sciences*, **100**, 8319–8324.
981 Schlötterer C (2002) Towards a molecular characterization of adaptation in local
982 populations. *Current Opinion in Genetics and Development*, **12**, 683–687.
983 Schneider KA, Kim Y (2013) Genetic Hitchhiking under Heterogeneous Spatial
984 Selection Pressures. *PLoS ONE*, **8**.

985 Seehausen O, Butlin RK, Keller I *et al.* (2014) Genomics and the origin of species.
986 *Nature reviews. Genetics*, **15**, 176–92.
987 Service MW (1993) *Mosquito ecology: field sampling methods* (UK Elsevier Applied
988 Science, London, Ed,).
989 Sharakhova M V, Xia A, Leman SC, Sharakhov I V (2011) Arm-specific dynamics of
990 chromosome evolution in malaria mosquitoes. *BMC evolutionary biology*, **11**,
991 91.
992 Sharakhov I, Braginets O, Grushko O *et al.* (2004) A Microsatellite Map of the African
993 Human Malaria Vector Anopheles funestus. *Journal of Heredity*, **95**, 29–34.
994 Sinka ME, Bangs MJ, Manguin S *et al.* (2010) The dominant Anopheles vectors of
995 human malaria in Africa, Europe and the Middle East: occurrence data,
996 distribution maps and bionomic précis. *Parasites & vectors*, **3**, 117.
997 Slatkin M, Wiehe T (1998) Genetic hitch-hiking in a subdivided population. *Genetical*
998 *Research*, **71**, 155–160.
999 Spencer CCA, Deloukas P, Hunt S *et al.* (2006) The influence of recombination on
1000 human genetic diversity. *PLoS Genetics*, **2**, 1375–1385.
1001 Stephan W, Langley CH (1989) Molecular genetic variation in the centromeric region
1002 of the X chromosome in three Drosophila ananassae populations. I. Contrasts
1003 between the vermilion and forked loci. *Genetics*, **121**, 89–99.
1004 Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive
1005 population divergence. *Molecular Ecology*, **14**, 671–688.
1006 Stump AD, Fitzpatrick MC, Lobo NF *et al.* (2005) Centromere-proximal
1007 differentiation and speciation in Anopheles gambiae. *Proceedings of the*

1008 *National Academy of Sciences of the United States of America*, **102**, 15930–5.
1009 Sturtevant A (1921) A case of rearrangement of genes in Drosophila. *Proceedings of*
1010 *the National Academy of Sciences*, **7**, 235–237.
1011 Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by
1012 DNA polymorphism. *Genetics*, **123**, 585–595.
1013 Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS (2004) Selection versus demography: A
1014 multilocus investigation of the domestication process in maize. *Molecular*
1015 *Biology and Evolution*, **21**, 1214–1225.
1016 Thomas CG, Wang W, Jovelin R *et al.* (2015) Full-genome evolutionary histories of
1017 selfing , splitting , and selection in Caenorhabditis. , 667–678.
1018 Thornton K (2005) Recombination and the properties of Tajima's D in the context of
1019 approximate-likelihood calculation. *Genetics*, **171**, 2143–8.
1020 Tiffin P, Ross-Ibarra J (2014) Advances and limits of using population genetics to
1021 understand local adaptation. *Trends in Ecology & Evolution*, **29**, 673–680.
1022 Watterson GA (1975) On the number of segregating sites in genetical models
1023 without recombination. *Theoretical Population Biology*, **7**, 256–276.
1024 Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population
1025 structure. *Evolution*, **38**, 1358–1370.
1026 Williams GC (1966) *Adaptation and Natural Selection: A Critique of Some Current*
1027 *Evolutionary Thought*. Princeton, NJ.
1028 Wright S, Dobzhansky T (1946) Genetics of natural populations. XII. Experimental
1029 reproduction of some of the changes caused by natural selection in certain
1030 populations of Drosophila pseudoobscura. *Genetics*, **31**, 125–156.

1031 Wright SI, Foxe JP, DeRose-Wilson L *et al.* (2006) Testing for effects of
1032 recombination rate on nucleotide diversity in natural populations of
1033 Arabidopsis lyrata. *Genetics*, **174**, 1421–1430.
1034 Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and
1035 splicing in short reads. *Bioinformatics*, **26**, 873–881.
1036 Yeaman S (2013) Genomic rearrangements and the evolution of clusters of locally
1037 adaptive loci. *Proceedings of the National Academy of Sciences*, **110**, E1743-51.
1038

Figure legends

Figure 1: Geographic origin and genetic relatedness of *An. funestus* populations in Cameroon. (A) Map showing Mfou and Tibati, the two locations where mosquitoes were collected. A delimitation of the approximate distribution ranges of the two ecotypes that occur 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* clusters 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 in Mfou and Tibati

populations. Strong sequence divergence 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 (θ_π and θ_w) 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 θ_π and θ_w 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. Inversions are delimitated by boxes and indicated by alphabet letters (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	θ_w	29.22 *	13.24 *	34.69 *	23.35 *
	θ_π	7.92 *	-3.23	19.23 *	13.92 *
Tibati	θ_w	18.79 *	13.24 *	33.54 *	18.94 *
	θ_π	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

Figures

Figure1

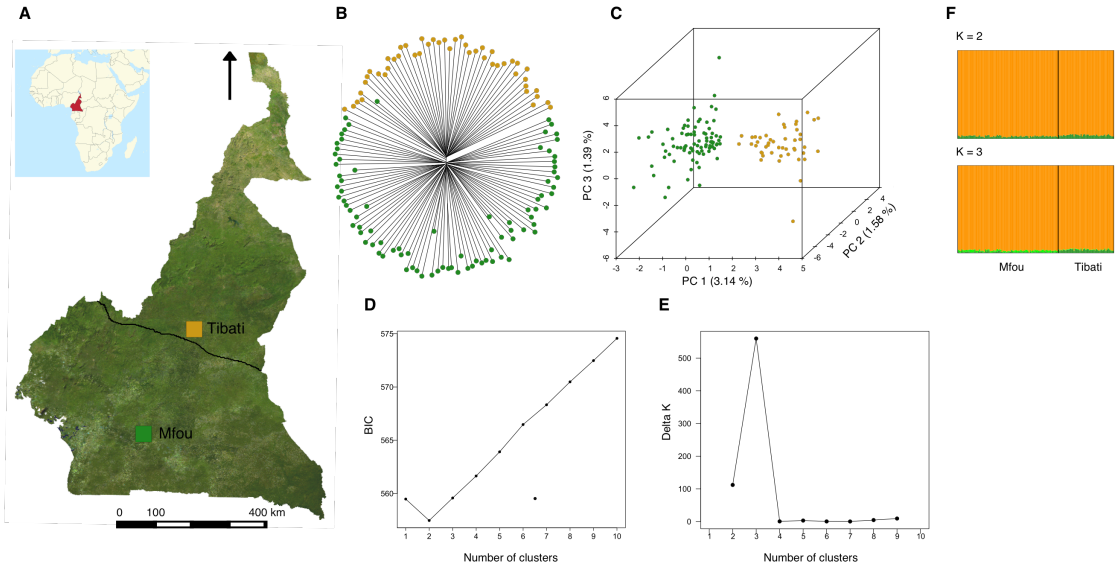


Figure 2

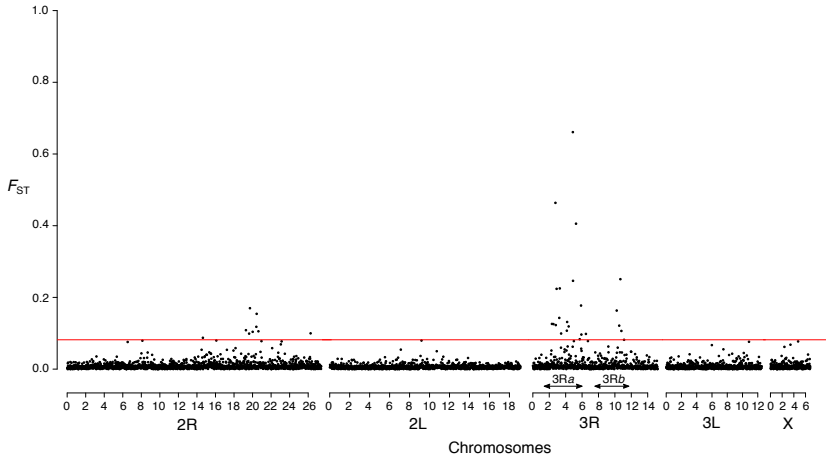

1.0
0.8
0.6
F_{ST}
0.4
0.2
0.0
3Ra
3Rb
0 2 4 6 8 10 12 14 16 18 20 22 24 26
2R
0 2 4 6 8 10 12 14 16 18
2L
0 2 4 6 8 10 12 14
3R
0 2 4 6 8 10 12
3L
0 2 4 6
X
Chromosomes

Figure 3

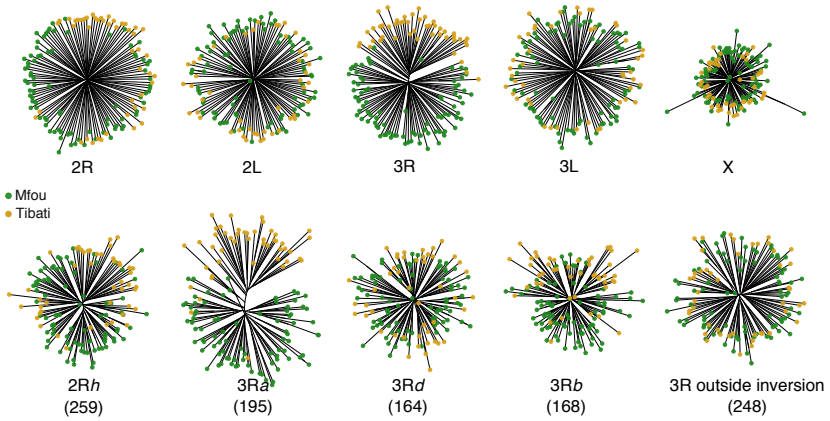

2R
2L
3R
3L
X
Mfou
Tibati
2R*h*
(259)
3R*a*
(195)
3R*d*
(164)
3R*b*
(168)
3R outside inversion
(248)

Figure 4

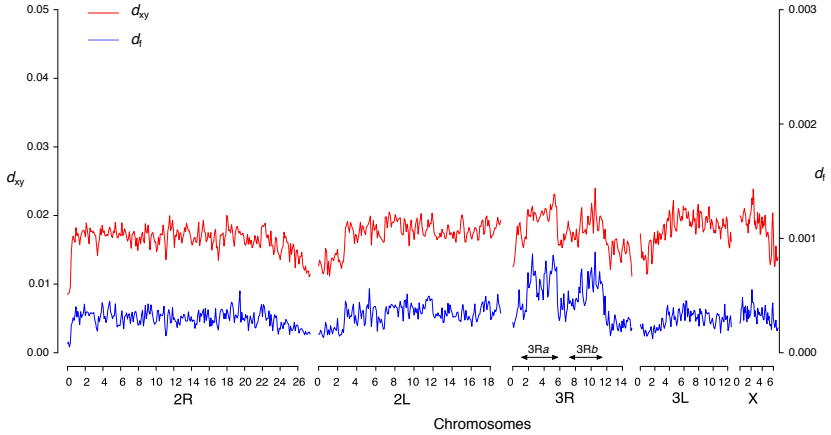

d_{xy}
d_f
0.05
0.04
0.03
0.02
0.01
0.00
0.003
0.002
0.001
0.000
3Ra
3Rb
0 2 4 6 8 10 12 14 16 18 20 22 24 26
2R
0 2 4 6 8 10 12 14 16 18
2L
0 2 4 6 8 10 12 14
3R
0 2 4 6 8 10 12
3L
0 2 4 6
X
Chromosomes

Figure 5

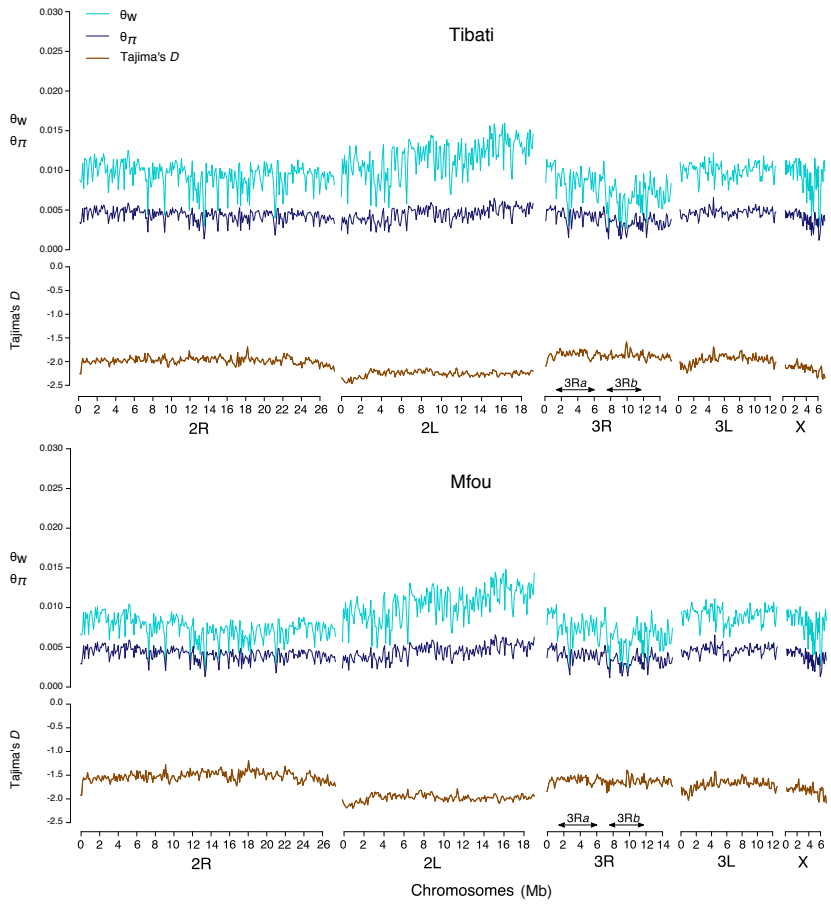

θW
θπ
Tajima's D
Tibati
Mfou
3Ra
3Rb
2R
2L
3R
3L
X
Chromosomes (Mb)

Figure 6

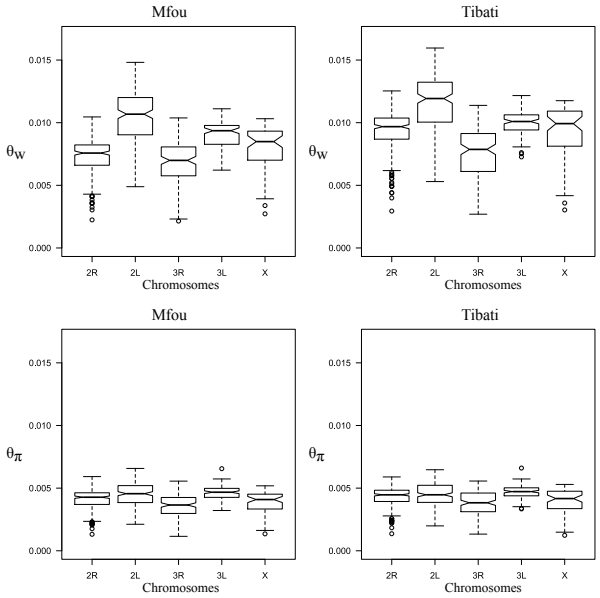

Mfou
Tibati
Mfou
Tibati
θW
θπ
0.000
0.005
0.010
0.015
2R
2L
3R
3L
X
Chromosomes

Figure S1

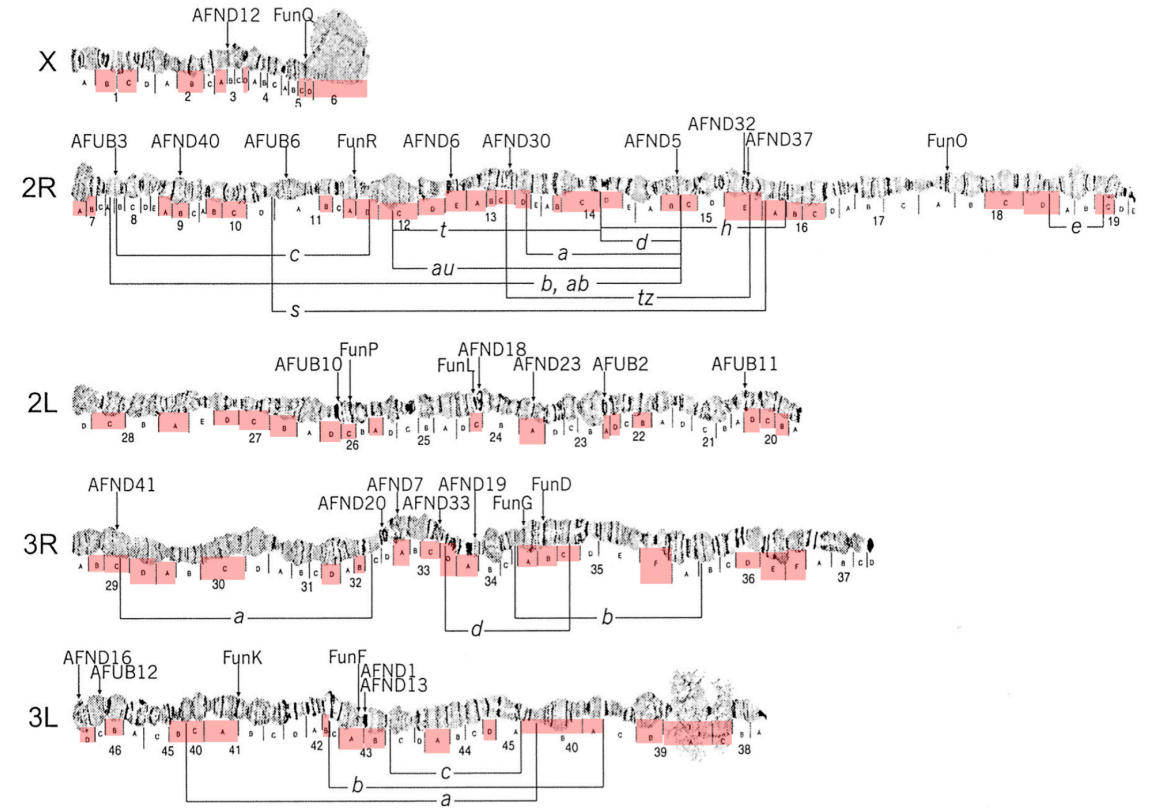
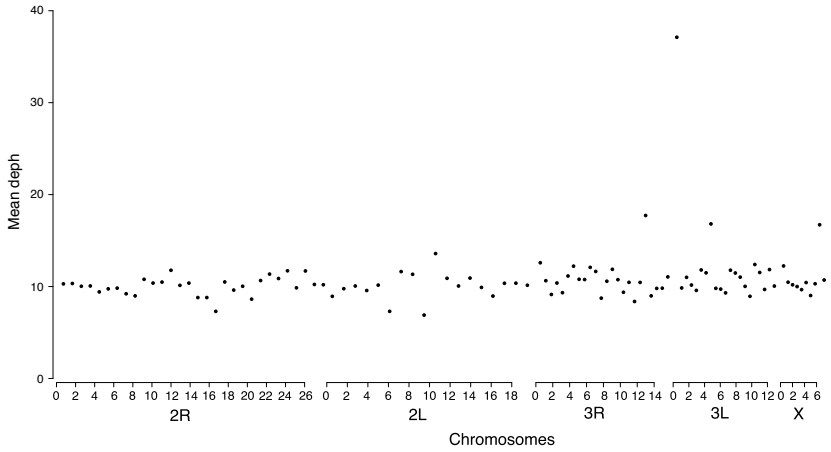

X
AFND12
FunQ
2R
AFUB3
AFND40
AFUB6
FunR
AFND6
AFND30
AFND5
AFND32
AFND37
FunO
t
c
au
a
d
h
e
b, ab
tz
s
2L
AFUB10
FunP
FunL
AFND18
AFND23
AFUB2
AFUB11
3R
AFND41
AFND20
AFND7
AFND33
AFND19
FunG
FunD
a
d
b
3L
AFND16
AFUB12
FunK
FunF
AFND1
AFND13
c
b
a

Figure S2


Mean deph
40
30
20
10
0
0 2 4 6 8 10 12 14 16 18 20 22 24 26
2R
0 2 4 6 8 10 12 14 16 18
2L
0 2 4 6 8 10 12 14
3R
0 2 4 6 8 10 12
3L
0 2 4 6
X
Chromosomes