Chromosomal inversions facilitate chromosome-scale evolution in *Anopheles funestus*

Caroline Fouet[1]*, Colince Kamdem[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: caroline.fouet@ucr.edu; bradley.white@ucr.edu

10 **Abstract**

11 Understanding the neutral and selective forces that shape the genetic diversity
12 along chromosomes is one of the central goals of evolutionary genomics. Contrary to
13 expectations, combined actions of recombination and selection can result in reduced
14 diversity along an entire chromosome, but this hypothesis still lacks strong
15 empirical supports. Here we show that nucleotide diversity is specific to
16 chromosome arms in the African malaria mosquito *Anopheles funestus*. We analysed
17 the population history and genetic diversity in 132 wild-caught individuals
18 representing two diverging ecotypes using RADseq markers. Genome scans revealed
19 that genetic differentiation between ecotypes is accumulated within segregating
20 polymorphic chromosomal inversions. Importantly, the presence of multiple
21 overlapping inversions is associated with strong reductions in diversity throughout
22 entire chromosome arms. The depression in genetic diversity is correlated to the
23 frequency of inversions, being stronger in forest populations that are fixed for
24 several adaptive inversions. Our results emphasize the role of polymorphic
25 inversions in the genome architecture of adaptive speciation in *An. funestus* and
26 provide a case of chromosome-scale evolution associated with the interplay
27 between population demography, selection and recombination.

28

29 **Key words:** Divergent selection, chromosomal inversion, recombination, nucleotide
30 diversity

31

Introduction

Investigating the population history and genetic diversity in emerging species should provide important insights into the genomic architecture of speciation driven by natural selection (Nosil 2012). Advances in genomics now allow robust descriptions of genetic diversity along chromosomes in almost any living organism (Davey et al. 2011; Ellegren 2014). However, while obtaining detailed estimates of genomic diversity has become very common, the evolutionary signals that contribute to shape this diversity remain difficult to understand. In principle, patterns of diversity that are consistent throughout the genome are generally thought to result from neutral demographic processes. In contrast, signatures of recombination and those of positive or negative selection are expected to be restricted to more or less small portions of chromosomes (O'Reilly et al. 2008). A vast majority of the massive genome scans that are accumulating owing to increased availability of high-throughput sequencing largely support this view (Hohenlohe et al. 2010; Lawniczak et al. 2010; Nielsen et al. 2011; Via 2012; Gagnaire et al. 2013; Kamdem et al. 2016). The known exceptions include rearranged chromosomal regions of chromosomes and collinear regions around the centromere and the telomere where disproportionately depressed levels of polymorphism due to low recombination can extend over several megabases (Neafsey et al. 2010; Ellegren et al. 2012; Marsden et al. 2014). However, in reality, the nature of the genetic diversity across genomes can be more complicated to explain than this, mainly because of the potentially confounding signals produced by several neutral and selective forces. For example, analyses of genetic variation in several *Drosophila*

55 species have revealed widespread footprints of purifying selection across the
56 genome (Sella et al. 2009). Likewise, recurrent bottlenecks can obscure signals of
57 positive selection (Chang and Hartl 2014). Further, the power to detect molecular
58 signatures of natural selection varies depending on whether the demographic
59 history is explicitly taken into account or not (Huber et al. 2014). Overall, the
60 dichotomy between localized effects of selection/recombination and genome-wide
61 effects of demography is sufficient to explain common patterns of genetic diversity
62 along chromosomes, but it is much less straightforward to judge when the
63 evolutionary history of the species is too complex or when diversity is altered over
64 extremely large portions of the genome.
65
66 One example of these puzzling scenarios involves patterns of nucleotide diversity
67 that are specific to whole chromosomes or chromosome arms. Reduced diversity on
68 sex chromosomes relative to the rest of the genome is common and well
69 documented in the heterogametic sex (Ellegren 2009). This discrepancy is largely
70 due to neutral effects since the effective population size of sex chromosomes is
71 lower in the heterogametic sex, but can also in some cases be correlated to
72 widespread selection along the chromosome (e.g., the Y chromosome of humans)
73 (Wilson Sayres et al. 2014). In contrast, autosome-specific diversity is very unusual,
74 more difficult to explain, and certainly involves a complex interplay between drift,
75 selection and recombination that is yet to be elucidated. The strongest evidence of
76 autosome-specific diversity so far comes from *Drosophila* (Kulathinal et al. 2009;
77 Corbett-Detig and Hartl 2012; Pool et al. 2012). Genome scans of worldwide

78 populations of *Drosophila melanogaster* revealed that some populations have either
79 reduced or elevated nucleotide diversity specific to autosomal arms (Corbett-Detig
80 and Hartl 2012; Pool et al. 2012). Interestingly, both the elevation and the reduction
81 of diversity along a chromosome relative to the rest of the genome are due to the
82 presence of inversions (Pool et al. 2012). Chromosomal inversions have been the
83 focus of intense research for several decades and their role in local adaptation and
84 speciation is widely recognized (Hoffmann and Rieseberg 2008; Kirkpatrick 2010).
85 Notably, frequencies of certain inversions match latitudinal clines in flies and
86 mosquitoes providing strong evidence of divergent selection on alternative
87 karyotypes (Coluzzi et al. 2002; Hoffmann and Rieseberg 2008; Ayala et al. 2011).
88 The explicit selective mechanisms at play within chromosomal inversions remain
89 poorly elucidated. It has been suggested that inversions reduce local recombination
90 and bring adapted alleles into closer physical association, which causes their spread
91 in favourable environments (Kirkpatrick 2010; Lohse et al. 2015). Despite their
92 importance as a major mechanism for reorganizing the genome, little progress has
93 been made towards understanding how and why inversions evolve. Even though the
94 generalization across taxa remains insufficient, the recent realization that inversions
95 affect sequence polymorphism not only within their physical limits but also
96 chromosome-wide reminds us that their impacts in genome evolution are perhaps
97 far more complex than we dare suppose. The fine-scale examination of genetic
98 diversity along chromosomes of species bearing numerous well-characterized
99 polymorphic chromosomal inversions will provide crucial insights into the physical
100 extent of selective processes in and around inversions.

Mosquitoes of the genus *Anopheles* provide an excellent opportunity for studying the dynamics of genome evolution during speciation. The short generation time allow the accumulation of recombination events that can be easily observed. Recent adaptive processes can also be investigated within clades with large effective population sizes and rapid divergence (Donnelly et al. 2001; Coluzzi et al. 2002; Sinka et al. 2010; Kamdem et al. 2012; Kamdem et al. 2016). Further, the evolutionary proximity of anopheline mosquitoes with *Drosophila* flies allows comparisons based on data from some of the best-studied model species. Importantly, *Anopheles* species appear particularly attractive to understand the role of inversions on the evolution of chromosomes because of the presence of hundreds of inversion polymorphisms and strong karyotypic variability among and within species (Sharakhova et al. 2011).

In this paper, we focused on *Anopheles funestus*, the second most important vector of human malaria parasites in Sub-Saharan Africa. This highly anthropophilic mosquito is splitting into at least two ecotypes widespread across the continent (Michel et al. 2005; Ayala et al. 2011; Coetzee and Koekemoer 2013; Dia et al. 2013). Microsatellites and cytogenetic analyses suggest that chromosomal inversions have played a crucial role in the adaptive evolution of *An. funestus* (Ayala et al. 2011). Indeed, not only an impressive number of polymorphic chromosomal inversions have been identified in the polytene chromosome, but gradients of inversion frequencies are well characterized in natural populations of *An. funestus* (Sharakhov

et al. 2004; Ayala et al. 2011; Dia et al. 2013). The main goal of this study was to use Next Generation Sequencing to characterize targets of selection in the *An. funestus* genome. We investigated the genomic architecture of divergence and selection in the two ecotypes using ~10,000 SNP markers identified with RAD sequencing. We conducted genome scans in 132 individuals from Cameroon and we showed that strong signatures of divergent selection are clustered within two major polymorphic inversions. We found that strong selection and low recombination in inversions result in reduced diversity throughout entire chromosomal arms supporting the idea of a chromosome-scale evolution facilitated by the presence of rearrangements.

Results

SNP genotyping

We sequenced 132 *An. funestus* mosquitoes collected from two locations separated by ~500km (Fig. 1A, Table S1). Mfou is a small city belonging to the southern forest region. *An. funestus* breeds in an artificial lake and maintain abundant populations throughout the year across the city. Tibati lies within the forest-savannah transition zone. Likewise, several artificial lakes provide abundant breeding sites for dense and perenial populations of *An. funestus* in this village. Using alignments to the draft reference genome, we identified 490,871 unique 96-bp RAD loci. We removed loci that were no present in all populations and in at least 60% of individuals within every population and we randomly selected one SNP identified within these loci resulting in 10687 high-quality biallelic markers.

Genetic and geographic differentiation

Based on 10687 genome-wide SNPs, the NJ tree and the first three PCA axes clearly distinguished two genetic clusters corresponding presumably to the two subgroups previously described in Cameroon and across the continent with microsatellites and inversion polymorphism markers (Cohuet et al. 2005; Michel et al. 2005; Ayala et al. 2011) (Fig. 1B). In Cameroon, in particular, populations of *An. funestus* are spread along a latitudinal cline of about 1500km and bear tens of polymorphic inversions, but remain weakly differentiated (Cohuet et al. 2005; Ayala et al. 2011). Both the method of Evanno et al and the DAPC confirmed the existence of only two to three demes corresponding to the forest and the savannah ecotypes among our samples (Fig. 1D,E). Despite this subdivision, the overall genetic differentiation is low (F_{ST} =

156 0.033, p < 0.005) between the two subgroups and a STRUCTURE analysis with k = 2
157 simply depicts a single population with admixed ancestry rather than two distinct
158 subunits. The moderate geographic differentiation is also reflected by the results of
159 an AMOVA, which shows that the most significant proportion of the genetic variance
160 (98.2%) among our samples is explained by within-individual variations. The
161 geographic origin explained less than 2% of the genetic variance, confirming that
162 extensive gene flow among populations overwhelms ongoing local adaptations. The
163 low number of sampling locations can be a handicap for the interpretation of
164 patterns of genetic structure we observed. However, our sampling includes the two
165 main subdivisions described in *An. funestus* with a variety of genetic markers and
166 thereby provided an unbiased picture of the genetic differentiation of this species in
167 West Africa.

168 **Evidence for strong divergent selection in chromosomal inversions**

169 We estimated and plotted the locus-specific F_{ST} along the pseudochromosomes to
170 assess the genomic differentiation between ecotypes and identify outlier loci. First,
171 using 10687 SNPs, we observed a non-random distribution of highly differentiated
172 sites, characterized by an aggregation of the highest F_{ST} values in genomic regions
173 bearing known polymorphic chromosomal inversions (Fig. 2). Using the 99% cutoff,
174 we identify 876 outliers of F_{ST} throughout the genome, 31 of which were mapped to
175 concatenated scaffolds and were confirmed as statistical outliers in LOSITAN. All
176 these outliers resided on chromosomes 3R and 2R, which also concentrate the
177 highest number of polymorphic inversions identified in the *An. funestus* (Fig. 2, S1).
178 We also noted an imbalance in the number of outliers between the 2R and 3R (9

179 against 22, respectively) implying that inversions located on the 3R are the most
180 significant in local adaptation. The strong divergence of inversions and inverted
181 chromosomes is particularly evident in the genomic distribution of d_{xy} and fixed
182 differences. In keeping with F_{ST} values, the highest values of d_{xy} are clustered within
183 the 3R*a* and 3R*b* inversions on the 3R chromosome arm (Fig. 3), which further
184 emphasizes the role of these two large inversions in ecological divergence.
185 Unsurprisingly, the 3R arm has by far the highest proportion of fixed differences
186 among autosomes. Finally, two distinct peaks of both d_{xy} and d_f can be clearly
187 identified in regions corresponding to the 3R*a* and 3R*b* inversions (Fig. 3).
188 To further assess the magnitude of the role of inversions in adaptive divergence, we
189 performed NJ analyses separately for each chromosome arm as well as four
190 different inversions (2R*h*, 3R*a*, 3R*b* and 3R*d*). As expected, the inverted
191 chromosome 3R and to a lesser extent the 2R reproduced the segregation between
192 ecotypes observed with genome-wide SNPs (Fig. 4). Most notably, variations within
193 the 3Ra inversion assigned 100% of individuals to their respective ecotypes, and the
194 3*Rb* also separated a significant proportion of individuals whereas the two
195 populations were not discriminated by SNPs from regions outside the two
196 inversions (Fig. 4). In contrast, the two inversions, 2R*h* and 3R*d*, had no
197 discriminatory power, which confirmed that the 3R*a* and 3R*b*, and especially the
198 3Ra, are the two main targets of divergent selection. This strong selection translates
199 into F_{ST} values, which increase from 0.033 genome-wide to 0.08 within the 2R*b*
200 inversion and reach 0.22 within the 3Ra inversion. Finally, the crucial role of the two
201 inversions in the geographic structuring of the genetic diversity in *An. funestus* is

202 also reflected by the drastic increase in the proportion of the genetic variance
203 explained by the geographic origin of samples within inversions. This proportion
204 rises from 1.8% for the genome-wide SNPs to 4.7% and 12.9% when only variants
205 included in the 3*Rb* and the 3Ra inversions, respectively, are used. Overall, our
206 results provide multiple lines of evidence that strong signatures of spatially varying
207 selection are clustered within polymorphic chromosomal inversions in a moderately
208 differentiated genomic background in *An. funestus*.

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

210 To further characterize the signatures of selection, we conducted genome scans
211 using estimates of nucleotide diversity ($\theta\pi$ and θw) and Tajima's *D* (Fig. 5).
212 Surprisingly, there was no obvious depression in diversity and Tajima's *D* within the
213 3R*a* and the 3R*b* inversions relative to the rest of the genome as could be expected
214 due to strong divergent selection observed within these inversions. Instead, we
215 noted that, in contrast to $\theta\pi$ that is fairly constant across the entire length of all
216 chromosomes, θw varies significantly between chromosome arms (Fig. 5, 6). The
217 highest values of θw are concentrated on the uninverted chromosome 2L, and there
218 are statistically significant (Wilcoxon rank sum test, *p-value* < 0.001) differences in
219 nucleotide diversity between the 2L and each of the 3 other autosomes (Fig. 6).
220 Precisely, inverted autosomes showed reduced average θw of 34.1% for the 3R,
221 24.0% for the 2R and 13.2% for the 3L relative to the 2L arm. The X chromosome
222 bears no known inversion, but presumably because of its particular divergence and
223 distorted effective size, the average θw diversity was also weaker relative to the 2L
224 arm (21.1% reduction relative to the 2L). Before examining the biological meaning

225 of these particular patterns of genetic diversity, we conducted several tests to insure
226 that uncertainties associated with our sequencing and analytical approach cannot
227 account for the observed chromosomal-scale disparities. First, we compared the
228 distribution and lengths of mapped scaffolds and we confirmed that every
229 chromosome arm was represented by a substantial number of scaffolds (Table S2
230 and Fig. S2). The length of pseudochromosomes was also proportional to the actual
231 length of chromosomes, ranging from 6.8 Mb on the X chromosome (the smallest) to
232 27.4 Mb on the 2R (the largest). Therefore, we excluded the possibility that the
233 chromosome-bias diversity was due to a systematic error associated with the
234 reference sequence. Further, mapped scaffolds were evenly distributed along all
235 chromosomes, suggesting that inconsistencies in genome-wide diversity were not
236 linked to centromere or telomere-proximal effects (Fig. S1). Second, although our
237 estimates of genetic diversity were derived using a "low coverage approach"
238 designed to alleviate the uncertainty associated with low sequencing depth
239 (Korneliussen et al. 2013; Korneliussen et al. 2014), we examined the distribution of
240 the average depth of coverage along chromosome arms to insure that the
241 chromosome-specific diversity was not correlated to inconsistent sequencing across
242 the genome. We ruled out this possibility because the mean depth of sequencing
243 coverage per mapped scaffolds was consistent along all the five chromosome arms
244 (Fig. S2). As a result, the intriguing pattern of chromosome arm-specific diversity
245 observed in *An. funestus* genome clearly has a biological meaning. Below, we are
246 using several analyses to better understand this process.

First, θw, which varies between chromosomes, is an estimate of the number of segregating sites also known as the population scaled mutation rate and represents the "raw" amount of polymorphisms in one population. In theory, the drastic variability in average polymorphism between inverted and uninverted autosomes can be due to a correspondingly low mutation rate on inverted chromosomes. However, as shown in other insect species (Keightley et al. 2015), such an important variability in mutation rates *per se* between chromosomes is very unlikely. Also, neutral processes such as the variation of effective size among chromosomes often observed between autosomes and the sex chromosome cannot solely explain our findings since all individuals possess all autosomal arms. Our results are most likely due to a combination of recombination and selective events, and probably other more stochastic forces that differ significantly between inverted and uninverted chromosome arms. The distribution of the neutral recombination rate remains unknown in the genome of *An. funestus*, but it is widely acknowledged that recombination is rare within inversions compared with genome-wide backgrounds and that this low crossover rate can also extends beyond the limits of inversions (Hoffmann and Rieseberg 2008). Therefore, the autosomal arms bearing polymorphic inversions presumably have average recombination rates that are substantially lower than that of the 2L. The strong correlation between reduced recombination and depressed diversity is both conceptually obvious and largely documented by empirical evidence.

We have also provided compelling evidence of strong signatures of divergent selection on adaptive inversions in *An. funestus*, notably the peaks of d_{xy} and fixed

differences observed in the loci of the 3R*a* and 3R*b* inversions. As we noted previously, it is hard to tell from our genome scans if there is a specific reduction of polymorphism within the 3R*a* and the 3R*b* inversion loci, but the low recombination and strong selection should clearly lead to lower nucleotide diversity. Although we can explain the depression in polymorphism expected within inversions by straightforward recombination and selective mechanisms, we however lack a comprehensive understanding of the processes that extend this diversity reduction beyond the limits of the inversions. This intriguing pattern of chromosome-scale reduction in diversity has been described only in several species of *Drosophila* and is hard to explain through common paradigms like the selective sweeps and cold spots of recombination. Further studies will be needed to clarify different hypotheses in *An. funestus* in particular. In the first place, a direct comparison of diversity and long-range linkage disequilibrium between inversion-bearing haplotypes and standard uninverted haplotypes from a sympatric area like Tibati using a more comprehensive genome sequencing method will be crucial to more clearly demonstrate the association between the presence of inversions and chromosome-scale diversity (Corbett-Detig and Hartl 2012; Pool et al. 2012). On the other hand, more than 100 inversions have been identified in the polytene chromosome map of *An. funestus* (Sharakhov et al. 2004), and how exactly the number and the natural frequencies of inversions bore by an autosomal arm are important in its overall diversity remains to be clarified. Each inverted chromosome of *An. funestus* contains multiple overlapping inversions of large size, the greatest number of common inversions being observed on the 2R arm (Fig. S1). It is therefore tempting to

293 imagine that perhaps selection on multiple overlapping inversions can degrade
294 diversity across the entire chromosome arm, but this hypothesis holds only if all
295 inversions are simultaneously under significant selection. Nevertheless, as
296 illustrated with the example of the non-recombining Y chromosome of humans,
297 selection may be particularly potent when there is no recombination, so that
298 deleterious mutations in one area of the chromosome could reduce levels of genetic
299 diversity across the entire chromosome (Wilson Sayres et al. 2014). Likewise, if
300 recombination is rare along entire chromosomal arms in *An. funestus* because of the
301 presence of multiple overlapping inversions, a strong selection within a single
302 inversion is sufficient to reduce genetic polymorphism across the chromosome arm.
303 Our findings have revealed obvious footprints of spatially varying selection only
304 within the 3Ra and 3*Rb*, which represent ~50% of the length of the 3R arm, but we
305 can not exclude that other inversions may also be subjected to strong divergent
306 selection at other spatial scales throughout the species' range. For instance, the 2*Ra*
307 and 3*La* inversions are widespread in continental populations and are suspected to
308 provide selective advantage in certain environments (Dia et al. 2013). Rare and
309 endemic inversions like the 2R*t*, 2R*h* and 2R*d* may also be strong targets of local
310 selection in some settings (Ayala et al. 2011). Further studies will focus on using
311 more extensive sampling to examine the adaptive status as well as the footprints
312 and strengths of positive or negative selection in other common and rare inversions
313 in *An. funestus*.
314 Finally, we noted variations in patterns of chromosome-scale diversity between
315 Mfou and Tibati samples consistent with perceptible differences in the population

demographic history despite high overall gene flow between ecotypes, which raise the question of the contribution of stochastic processes in the creation of chromosome-specific diversity. Indeed, the average genome-wide θw is 1.3 greater in Tibati than in Mfou populations. The reduction of nucleotide diversity on inverted autosomes relative to the 2L is also stronger in Mfou compare with Tibati samples. The drop of the average θw of inverted chromosomes relative to the 2L is 34.7% on the 3R, 29.2% on the 2R and 13.2% on the 3L in Mfou and 33.5% on the 3R, 18.8% on the 2R and 13.2 on the 3L in Tibati. This difference likely reflects the spatial variation in the intensity of selection within inversion since forest populations (Mfou) are fixed for the major polymorphic chromosomal inversions (notably the 3R*a* and 3R*b*) whose frequencies fluctuate in Tibati (savannah) (Cohuet et al. 2005; Ayala et al. 2011). The effective size should also be reduced on inverted related to uninverted chromosome because selection on inverted chromosome lowers the number of individuals carrying certain polymorphisms. We address this particular aspect later in this section by fitting different growth models on Site Frequency Spectra (SFS) of the different autosomal arms. Beside obvious indirect effects of the specific demographic history of chromosome arms on their diversity, it remains to be determined if the population-level demography have any impact on chromosome-scale evolution. Both the overall diversity (θw = 0.0083 and $\theta\pi$ = 0.0042 in Mfou; θw = 0.0097 and $\theta\pi$ = 0.0043 in Tibati) and the percentage of sites that are polymorphic within populations (64.0 % in Mfou and 65.2% in Tibati) suggest that *An. funestus* is a highly polymorphic species. As a result, it is also crucial to understand how important is a population expansion that maintains high

339 baseline genome-wide diversity to chromosome-specific diversity. To infer the
340 demographic history of our populations, we first examined the genome-wide
341 distribution of allele frequency spectra, summarized as Tajima's *D* (TD). Values of
342 TD ranged from -2.47 to -1.58 among Tibati samples and from -2.20 to -1.19 in Mfou
343 and were shifted towards negative values as expected in case of recent population
344 expansion leading to an accumulation of low-frequency variants. Unsurprisingly, TD
345 is more negative on the 2L chromosome in both ecotypes since this chromosome
346 freely accumulates mutations in contrast to inverted chromosomes that are
347 constrained by strong selection. We next modeled the SFS to infer the demographic
348 history of populations in ∂a∂i. We used 9412 genome-wide SNPs while excluding
349 variants located on the X chromosome to prevent the confounding effects associated
350 with the lesser population effective size of this chromosome. We used a dataset
351 comprising 35 samples from Tibati and 35 samples from Mfou selected based on the
352 best sequencing coverage criterion. We found that a population expansion model
353 matched the average observed levels of autosomal polymorphism in both Tibati and
354 Mfou populations. The best-fit model depicts an exponential growth in Tibati and a
355 two-epoch growth in Mfou. Model parameters depict a more important population
356 growth in Tibati consistent with the greater genome-wide diversity observed in
357 savannah populations (Table 1). As we mentioned previously, we also modeled the
358 population growth on each chromosome using arm-specific SFS in ∂a∂i. Although
359 the true history of chromosome arms is likely explained by more complex models
360 including drift, selection and local recombination rates, simple growth models fitted
361 the SFS of all autosomes in both Tibati and Mfou populations. As expected, the

362 growth seems relatively more important on the 2L relative to inverted autosomes
363 (Table 1). Overall, the simple growth models fitted to chromosomes do not
364 integrally reproduce the ratio of diversity described among chromosomal arms
365 because chromosome-specific diversity is due to compound factors as we specified
366 before. However, these models well illustrate the fact that the 3R and to a lesser
367 extent the 2R arm have gone through recent drastic reduction in population
368 effective size associated with strong selection within polymorphic chromosomal
369 inversions.
370

Discussion

Population history of *An. funestus*

We conducted the first genome-wide analysis of natural polymorphism in *An. funestus* populations, which revealed that genetic divergence between the two weakly differentiated ecotypes of this mosquito species is almost exclusively located in a few inverted segments of the genome. Moreover, strong selection and reduced recombination within these polymorphic chromosomal inversions are also associated with an unusual pattern of chromosome-specific diversity. We analyzed the genetic structure of populations from Cameroon and found that genome-wide SNPs corroborated previous results obtained with microsatellites and/or polymorphic chromosomal inversions markers. Low-resolution markers have indicated that populations of *An. funestus* globally belong to two clusters that have been sometimes called “Kiribina” and “Folonzo”: the names of two locations in Burkina Faso where the subgroups were described for the first time (Cohuet et al. 2005; Michel et al. 2005; Ayala et al. 2011; Dia et al. 2013). In Cameroon, an intensive sampling including 305 sites along a north-to-south transect and subsequent examination of the genetic structure with microsatellites and several polymorphic inversion loci showed that *An. funestus* is separated into two weakly differentiated subgroups corresponding probably to Kiribina and Folonzo (Ayala et al. 2011). Our detailed genome-wide analysis with more than 10000 SNPs is in perfect agreement with this study. PCA, Neighbor-joining trees and Bayesian clustering confirmed the segregation based on the geographic origin (the savannah against the forest). A massive ongoing gene flow and a weak genetic differentiation

394 between the two ecotypes translate into values of F_{ST} near zero. As a result, we
395 emphasize that the putative subgroups of *An. funestus* are far from incipient species,
396 but represent at best two ecotypes at early stages of genetic divergence.
397 Additionally, as we document further in the next paragraphs, populations of *An.*
398 *funestus* are globally undifferentiated even at the genomic level except within loci of
399 a few polymorphic chromosomal inversions. The consistent subdivision observed
400 across the species range is primarily a translation of karyotypic variability along
401 geographic clines of these inversions. *An. funestus* larvae breed almost exclusively in
402 lakes that are most of the time artificial lakes created by human activity in
403 geographically distant locations. In Cameroon, historic populations of *An. funestus*
404 are generally observed around a network of ~5 relatively isolated artificial lakes
405 (Fig 1A). Our findings and previous studies (Cohuet et al. 2005; Ayala et al. 2011)
406 show that this apparent lack of connectivity among populations does not prevent a
407 massive gene flow.
408 The distribution and genetic structure of *Anopheles* populations across
409 microhabitats and time scales is markedly heterogeneous, especially among leading
410 African malaria vectors. This heterogeneity is being exacerbated by the recent scale-
411 up of malaria control measures including the widespread use of Insecticide-Treated
412 Nets (ITNs) and the multiplication of indoor insecticide spraying campaigns to
413 which mosquito populations respond by shifting their spatial and temporal
414 distributions (Derua et al. 2012; Mwangangi et al. 2013). With regard to *An. funestus*
415 precisely, a detectable change in the nocturnal biting cycle has been noticed in West
416 and East Africa after the scale-up of mosquito nets ownership and use (Moiroux et

al. 2012). These shifts may lead to the emergence of cryptic subgroups if specialized populations become genetically differentiated (Riehle et al. 2011). Fortunately, our results do not indicate any genetic structure beyond the major geographic subdivisions expected within *An. funestus* populations. In Cameroon, as in many other Sub-Saharan African countries, massive ITNs distribution campaigns have been implemented recently and studies are ongoing to evaluate the short or long-term effects of these measures on vector populations (Bowen 2013). We surveyed 28 locations, including historical foci of *An. funestus* population Cameroon, and we recorded the presence of the species in only two villages. This rarity may suggest a recent contraction in the distribution range of *An. funestus*, which still need to be confirmed by a careful examination of the temporal dynamics. Importantly, a high-confidence genome-scale analysis allows us to clearly emphasize that *An. funestus* has undergone no shift in its population genetic structure so far.

Spatially varying selection in chromosomal inversions

We estimated locus-specific population differentiation using the Weir and Cockerham's unbiased estimator of the F_{ST}, the absolute sequence divergence (d_{xy}) and the proportion of fixed differences to characterize the level of divergence at each variant site sampled across the genome. We found a strong heterogeneity in genetic divergence across the genome. Notably, despite the lack of differentiation between ecotypes (F_{ST} = 0.033), more than 5% of SNPs have disproportionately elevated F_{ST} (F_{ST} outliers). Importantly, these F_{ST} outliers are exclusively found on two of the five chromosome arms (2R and 3R), especially within two segregating inversions (3Ra and 2*Rb*). The genome-wide organization of genetic divergence

during speciation has been increasingly examined in diverse taxa, but the conclusions remain contentious. Two concurrent models have been proposed to explain the architecture of genetic divergence end reproductive isolation along genomes of emerging species. The genetic divergence is thought to be either "heterogeneous", meaning that genomic regions of high differentiation are scattered throughout the genome, or "restricted" when only a few regions that have been sometimes called "speciation island" bear signatures of strong divergence (Nosil and Feder 2012; Nosil and Feder 2013). Empirical data support the existence of both models at unknown frequencies in nature (Lawniczak et al. 2010; Nosil and Feder 2012; Roesti et al. 2012; Andrew and Rieseberg 2013; Nosil and Feder 2013; Renaut et al. 2013). However, divergence is clearly focused on few genomic regions early in speciation due to localized effects of natural selection (Andrew and Rieseberg 2013). During the speciation process, when the ecological and genetic segregation between splitting lineages becomes significant, the combination of selective and demographic processes expands the genetic divergence throughout the genome. Moreover, it is very common to find high levels of genetic divergence that are accumulated in regions of low recombination especially at early stages of speciation (Andrew and Rieseberg 2013). Based on our findings, *An. funestus* appears as excellent model system to understand the architecture of genomic divergence at early stages of adaptive divergence. Unsurprisingly, the genomic divergence between the two undifferentiated ecotypes is restricted to a few regions of low recombination, which correspond to polymorphic chromosomal inversions in this species. Clear signals of divergent selection within two inversions translate into

463 peaks of F_{ST}, d_{xy} and fixed differences relative to the genome-wide average. The
464 prime role of the two inversions in ecological divergence is also illustrated by the
465 fact that except these inversions, no other locus or chromosome is able to reproduce
466 the structuring of *An. funestus* into two populations as expected. Cytogenetic
467 analyses and microsatellite loci have also suggested that the 3R*a* inversion may have
468 a central role in local adaptation in Cameroonian populations (Ayala et al. 2011).
469 Chromosomal inversions have been found in virtually all species that have been well
470 scrutinized so far (Kirkpatrick 2010). Nearly a century of research on these
471 chromosomal rearrangements has provided multiple lines of evidence suggesting
472 that strong selection governs the distribution of inversions in insects, plants and
473 animals (Hoffmann and Rieseberg 2008). Notably, chromosomal inversions play a
474 crucial role in adaptive speciation as shown in several insect species: the most
475 prominent examples being *Drosophila*, *Rhagoletis* and *Anopheles* (Feder et al. 2003;
476 Ayala and Coluzzi 2005; Rane et al. 2015). In another African malaria mosquito, *An.*
477 *gambiae*, known for its reticulate speciation leading to numerous cryptic species and
478 emerging subgroups, high levels of genetic differentiation are also accumulated in
479 selective sweeps found within segregating polymorphic chromosomal (Kamdem et
480 al. 2016), especially the 2L*a* inversion involved in resistance to aridity (Cheng et al.
481 2012). Likewise, inversions are responsible for the majority of the genetic structure
482 observed in cosmopolitan populations of *Drosophila melanogaster* (Corbett-Detig
483 and Hartl 2012; Pool et al. 2012). Chromosomal rearrangements are also involved in
484 adaptation to new host plants and sympatric speciation in *Rhagoletis pomonella*
485 (Feder et al. 2003). Inversions are thought to contribute to the spread of

combinations of favourable alleles because recombination is suppressed between the inverted sequence and standard karyotype, leading to divergence between inversions and homo-sequential regions and the potential for evolution of coadapted gene complexes (Kirkpatrick and Barton 2006; Hoffmann and Rieseberg 2008). One of the direct effects of low recombination within inversion is also to minimize the competition that can occur between loci under selection to reduce diversity across their nearby linked loci (Hill–Robertson effect) (Comeron et al. 2008). These combined actions of recombination and selection then result in a population effective size that is considerably reduced within inversions relative to genomic regions with higher recombination. In principle, selection in inversions can be either strong and infrequent or weak and common, but empirical supports about the type and magnitude of selection within inversions remain rare. However, impressive clines of inversions frequencies observed in natural populations, especially in *Drosophila* and mosquitoes illustrate the powerful effects of recombination and selection within inversion in genomic divergence.

Chromosome-scale evolution

The genetic hitchhiking model (Via 2012) well explain the reduced diversity that can be encountered within loci of segregating inversion subjected to divergent natural selection. Beyond the accumulation of outliers of F_{ST} and other divergence metrics within the 3R*a* and 3R*b* inversions, we haven't observed other characteristic signals of divergent selection such as the depression of nucleotide diversity and skewed allele frequency spectrum (Fig. 5). A more sophisticated version of the reference genome assembly and annotation of *An. funestus* will allow us to conduct a more

509 sensitive sequential and functional characterization of footprints of selection
510 observed within these key inversions. Presumably, as the two inversions segregate
511 along a latitudinal gradient, genes involved in resistance to aridity or desiccation can
512 be suspected to be among the functional classes potentially accumulated within the
513 two inversions. In support of this hypothesis, comparative cytogenetic have
514 revealed that the 3R chromosomal arm of *An. funestus* and the 2L arm of *An.*
515 *gambiae* share a very close evolutionary history (Sharakhova et al. 2011). It is still
516 unclear whether this chromosomal similarity is the result of a convergent evolution
517 or something that the two species share since they diverged from their common
518 ancestor. Importantly, the comparative cytogenetic study also stipulates that the
519 3*Rb* inversion of *An. funestus* probably correspond to the 2*La* inversion known to be
520 crucial for desiccation tolerance in *An. gambiae* populations (Fouet et al. 2012).
521 Instead of reduced diversity or skewed AFS in loci of inversions subjected to strong
522 selection, we have found that the genetic diversity was rather specific to whole
523 chromosomal arms. We conducted several analyses to first rule out the possibility
524 that chromosome-bias diversity was due to errors and limits of our experimental
525 approach. There is no simple explanation that can translate the biological meaning
526 of chromosome-specific diversity, but our work and previous studies on *Drosophila*
527 (Corbett-Detig and Hartl 2012; Pool et al. 2012) pinpoint the critical role of spatially
528 varying selection and reduced recombination within chromosomal inversions as the
529 primary catalyzers. In *An. funestus* chromosome-scale evolution of diversity is likely
530 due to the presence of multiple overlapping inversions subjected to spatially varying
531 divergent selection all across the geographic range of the species. The models of

532 population demography of our samples suggest an exponential population growth,
533 but we do not know if high levels of genome-wide polymorphism are key to the
534 onset of chromosome-specific diversity or not.
535 Chromosome-scale evolution is an emerging field or research and it is premature to
536 provide general rules or a definitive explanation of forces that affect the genetic
537 variation at large scales across genomes. So far, patterns of chromosome-specific
538 diversity have been described mainly along sex chromosomes (Ellegren 2009).
539 Particularly, it has recently been hypothesized that the reduced diversity along the Y
540 chromosome relative to autosomes in humans is not only due to predictable effects
541 of a lower population effective size of the sex chromosome but also to pervasive
542 effects of selection (Wilson Sayres et al. 2014). Precisely, the absence of
543 recombination along the Y chromosome renders background selection particularly
544 efficient, so that deleterious mutations in one area of the chromosome can
545 potentially reduce levels of genetic diversity across the entire chromosome if the
546 selection is sufficiently strong to overcome the effects of interference with
547 neighboring loci (Hill-Robertson effect) (Wilson Sayres et al. 2014). A similar
548 process can be possible on inverted chromosomal arms in *An. funestus* where
549 recombination is suppressed because of the presence of multiple inversions.
550 The reduction of diversity due to low recombination across large portions of the
551 genome is sufficiently documented, especially around centromeric and telomeric
552 regions of chromosomes (Neafsey et al. 2010; Marsden et al. 2014). In plants,
553 polyploidy and loss of recombination result in substantial reduction of diversity
554 along large chromosomal regions (Akhunov et al. 2010). Likewise, species with low

555 levels of genome-wide recombination, such as highly self-fertilizing species, show
556 reduced genetic diversity compared with their outcrossing relatives. Indeed,
557 extensive linkage disequilibrium and low recombination due to self-fertilization are
558 thought to interact with selection to shape chromosome-scale patterns of genetic
559 diversity in *Caenorhabditis* (Cutter and Choi 2010; Andersen et al. 2012; Thomas et
560 al. 2015). The really intriguing cases of chromosome-scale diversity in outcrossing
561 species are observed so far in *Drosophila* and *An. funestus*. In *Drosophila*
562 *melanogaster*, *Drosophila pseudoobscura* and related species, comparison of
563 inversion-bearing haplotypes and individuals with standard arrangements lead to
564 the conclusion that, in contrast to *An. funestus* where the action of overlapping
565 inversions can be suggested, the presence of only one polymorphic inversion can
566 drive diversity reduction or increase throughout the entire chromosome arm
567 (Corbett-Detig and Hartl 2012; McGaugh and Noor 2012; Pool et al. 2012).
568

Conclusions and implications

We described chromosome-wide diversity supporting recent evidence from *Drosophila* that genome evolution and patterns of variation at genome or chromosome-scales can be facilitated by selection and recombination within polymorphic chromosomal inversions. Our results should motivate other studies in *Anopheles* and other insect species in order to provide additional empirical support for this process, which has strong implications in our understanding of the evolutionary forces that shape genetic diversity along chromosomes. In *An. funestus*, further directions could be to improved genomic resources of this nonmodel organism, notably a recombination map and an improved version of the reference genome. This will allow a more precise understanding of the genetic variation and functional characterization of local adaptation in this mosquito. Notably, a high-quality reference genome combine with a more intensive sampling, a comprehensive genomic sequencing and additional cytogenetic studies will be crucial for mapping chromosomal inversions and karyotyping individuals across the species range. As such, we will be able to conduct direct tests for the effects of inclusion or exclusion of inversions on the genetic diversity similar to what has been done in *Drosophila melanogaster* (Pool et al. 2012). Finally, *An. funestus* is the second most important vector of human malaria parasites in Africa and investigating the genomic bases of its evolutionary diversification will be useful to assess the adaptive response of this vector to ongoing expansive insecticide-based malaria control campaigns.

Materials and methods

Sample collection

An. funestus is endemic to Africa and is structured in two genetically differentiated ecotypes distributed along a latitudinal cline with a wide contact zone (Cohuet et al. 2003; Cohuet et al. 2005; Ayala et al. 2011). We surveyed 28 locations across a north-to-south transect in Cameroon, from August to November 2013, to collect *An. funestus* specimens (Fig. 1A, Table S1). Most of the locations we visited were selected because previous surveys indicated the presence of *An. funestus* populations (Cohuet et al. 2003; Ayala et al. 2011) or because of the presence of lakes, which are typical larval breeding sites of *An. funestus*. Others were based on opportunistic sampling. We used several sampling methods to collect mosquitoes at distinct stages of the life cycle, from different microhabitats and from different time points during the nocturnal biting cycle (Service 1993). This approach maximizes the chances that our sampling represents a best approximation of the species genetic diversity. We identified *An. funestus* individuals using morphological characters and PCR (Gillies and De Meillon 1968; Gillies and Coetzee 1987; Cohuet et al. 2003).

Library preparation and sequencing

We extracted genomic DNA of larvae and adult specimens with the DNeasy Blood and Tissue kit (Qiagen) and the Zymo Research MinPrep kit. Double-digest Restriction-site Associated DNA (ddRAD) libraries were prepared following a modified protocol of Peterson *et al.* 2012. Each library contained DNA fragments resulting from simultaneous digestion with *MluC1* and *NlaIII* restriction enzymes

615 and annealed to unique short nucleotide sequences (barcode adaptors) that enabled
616 the identification of the individual associated with each read. RAD tags of
617 individually-barcoded mosquitoes were pooled and size-selected to target 400-bp
618 sequences that were enriched via PCR. One extremity of the resulting 400-bp RAD
619 tags was sequenced on Illumina HiSeq2000 to generate single-end reads of 101 bp.

620 **Alignment and SNP calling**

621 Raw reads were demultiplexed and filtered using the *process_radtags* program of
622 the Stacks v 1. 35 software (Catchen et al. 2013). Reads with ambiguous barcodes or
623 restriction sites and those with low sequencing quality score (average Phred score <
624 33) were removed. To call and identify SNP among our sampling, we aligned the
625 remaining high-quality reads along the *An. funestus* reference sequence using GSNAP
626 (Wu and Nacu 2010). A maximum of five nucleotide mismatches were allowed and
627 terminal alignments prevented. We used Stacks v 1. 35 to identify polymorphisms
628 and to call SNPs within each RAD locus. Following assembly and genotyping, the
629 polymorphism data was further filtered to maximize data quality. We retained only
630 one randomly selected SNP per RAD locus and we kept only loci scored in every
631 population and in at least 60 % of individuals within population for further analyses.
632 We created SNP files in different formats using the *populations* program in Stacks v
633 1. 35, PLINK and PGDSPIDER (Purcell et al. 2007; Lischer and Excoffier 2012;
634 Catchen et al. 2013). We checked the depth of sequencing coverage of the final data
635 set after all filtering steps using VCFtools (Danecek et al. 2011). The *An. funestus*
636 draft genome assembly we used consisted of 12,243 scaffolds ranging from 2,334 to
637 3,334,433 bp in length. Therefore, to perform genome scans and to effectively

identify footprints of selection throughout the genome, we placed SNPs on the five *An. funestus* chromosome arms using 104 scaffolds that had been physically mapped to chromosome (Neafsey et al. 2015). Mapped scaffolds corresponded to 7% in number and 33% in length of all *An. funestus* scaffolds (Table S2).

Population structure and genetic divergence

We analyzed the genetic structure of populations with a Principal Component Analysis (PCA), neighbor-joining (NJ) trees and the STUCTURE v 2.3.4 software (Pritchard et al. 2000). PCA and NJ analysis were performed respectively with the packages *adegenet* and *ape* in R (Paradis et al. 2004; Jombart 2008; R Development Core Team 2014). In STRUCTURE, we ran five replicates of 1 to 10 assumed clusters (k). Each run consisted of 200,000 iterations and the first 50000 iterations were discarded as burn-in. CLUMPP v1.1.2 (Jakobsson and 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 and VonHoldt 2012).

To quantify the extent of the genetic differentiation between pairs of populations, we calculated the overall F_{ST} (Weir and Cockerham 1984) in Genodive v1.06 (Meirmans and Van Tienderen 2004) using a subset of 2000 randomly selected SNPs. Additionally, to assess the contribution of regional and micro-spatial distributions of samples on the genetic variance, we conducted a hierarchical

660 analysis of molecular variance (AMOVA) (Excoffier et al. 1992) in Genodive. We used
661 10000 permutations to assess the statistical significance of F_{ST} and AMOVA.

662 **Genomic targets of selection**

663 To search for footprints of selection in the genome of *An. funestus*, we examined the
664 distribution of genetic diversity, allele frequency spectrum and genetic
665 differentiation on RAD loci located on mapped scaffolds. We computed the pairwise
666 nucleotide diversity ($\theta\pi$), the Watterson's estimator of segregating sites (θw) and
667 Tajima's *D* with ANGSD v 0.612 (Korneliussen et al. 2014). This program estimates
668 the genetic diversity and Site Frequency Spectrum (SFS) statistics using genotype
669 likelihoods without SNP calling thereby alleviating the uncertainty and biases
670 associated with SNP calling in low coverage Next Generation Sequencing
671 (Korneliussen et al. 2013). To gain genome-wide view and identify outliers regions
672 of diversity and Tajima's D, average values of these metrics were plotted in non-
673 overlapping 90-kb windows along mapped scaffolds that we previously ordered and
674 concatenated into "pseudo-chromosomes" corresponding to different *An. funestus*
675 chromosome arms. F_{ST} values per SNP locus were also calculated with the
676 *populations* program in *Stacks* and plotted across the pseudo-chromosomes. To
677 identify F_{ST} outliers that are likely genomic regions under strong selection, we used
678 both empirical and model-based approaches. First, loci above the top 1% of the
679 empirical distribution were identified as putative outliers of F_{ST}. Second, we used the
680 coalescence-based method FDIST2 (Beaumont and Nichols 1996) as implemented in
681 LOSITAN (Antao et al. 2008) to identify loci with unusually low or high F_{ST} values
682 compared to neutral expectations (outlier of F_{ST}). The mean neutral F_{ST} across all

683 SNPs in the data set was approximated by choosing the "neutral mean F_{ST} option"
684 with 99% confidence interval in LOSITAN. We ran LOSITAN with 100000
685 simulations and assumed a False Discovery Rate (FDR) of 0.1 to minimize the
686 number of false positives. Estimates of F_{ST} are affected by within-population
687 diversity, which can confound assessments of population divergence (Cruickshank
688 and Hahn 2014). It has been suggested that the use of several divergence metrics
689 including the d_{xy} provides a more reliable approach to examine the genetic
690 divergence among populations (Noor and Bennett 2009; Cruickshank and Hahn
691 2014). We estimated the absolute sequence divergence (d_{xy}) and the proportion of
692 fixed differences between populations (d_f) using ngsTools (Fumagalli et al. 2014)
693 and we derived and plotted average values across non-overlapping 10-kb windows
694 along pseudochromosomes in R.

695 **Demographic history of An. funestus**

696 We used the diffusion approximation method implemented in the software package
697 ∂a∂i v 1.6.3 (Gutenkunst et al. 2009) to infer the demographic history of *An. funestus*
698 from Site Frequency Spectrum (SFS). Different demographic histories leave distinct
699 signatures in SNPs across the genome. ∂a∂i applies composite likelihoods to
700 evaluate how well the spectrum of allele frequencies fit different simple or
701 compound demographic scenarios. Using 9412 filtered SNPs, we fitted four
702 alternative one-population models (*neutral, growth, two-epoch, bottle-growth*),
703 describing simple demographic scenarios without migration or recombination
704 events, to the folded allele frequency spectrum of our samples. We chose the best
705 model as the one with the highest composite log likelihood, the lowest Akaike

706 Information Criterion (AIC) and the lowest uncorrelated residuals. We used the
707 built-in bootstrap function implemented in ∂a∂i to derive 95% bootstrap confidence
708 intervals for the different demographic parameters.

Acknowledgements

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

Author contributions

Conceived and designed the experiments: CF CK BJW. Performed the experiments: CF CK BJW. Analyzed the data: CF CK BJW. Wrote the paper: CF CK BJW.

715 **References**

716 Akhunov ED, Akhunova AR, Anderson OD, Anderson J a, Blake N, Clegg MT, Coleman-
717 Derr D, Conley EJ, Crossman CC, Deal KR, et al. 2010. Nucleotide diversity maps
718 reveal variation in diversity among wheat genomes and chromosomes. BMC
719 Genomics 11:702.

720 Andersen EC, Gerke JP, Shapiro J a, Crissman JR, Ghosh R, Bloom JS, Félix M-A,
721 Kruglyak L. 2012. Chromosome-scale selective sweeps shape Caenorhabditis
722 elegans genomic diversity. Nat. Genet. 44:285–290.

723 Andrew RL, Rieseberg LH. 2013. Divergence is focused on few genomic regions early
724 in speciation: incipient speciation of sunflower ecotypes. Evolution 67:2468–
725 2482.

726 Antao T, Lopes A, Lopes RJ, Beja-Pereira A, Luikart G. 2008. LOSITAN: a workbench
727 to detect molecular adaptation based on a Fst-outlier method. BMC
728 Bioinformatics 9:323.

729 Ayala D, Fontaine MC, Cohuet A, Fontenille D, Vitalis R, Simard F. 2011.
730 Chromosomal inversions, natural selection and adaptation in the malaria vector
731 Anopheles funestus. Mol. Biol. Evol. 28:745–758.

732 Ayala FJ, Coluzzi M. 2005. Chromosome speciation: humans, Drosophila, and
733 mosquitoes. Proc. Natl. Acad. Sci. U. S. A. 102 Suppl :6535–6542.

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

736 Bowen HL. 2013. Impact of a mass media campaign on bed net use in Cameroon.
737 Malar. J. 12:36.

738 Catchen J, Hohenlohe P a., Bassham S, Amores A, Cresko W. 2013. Stacks: An analysis
739 tool set for population genomics. Mol. Ecol. 22:3124–3140.
740 Chang H-H, Hartl DL. 2014. Recurrent bottlenecks in the malaria life cycle obscure
741 signals of positive selection. Parasitology:1–10.
742 Cheng C, White BJ, Kamdem C, Mockaitis K, Costantini C, Hahn MW, Besansky NJ.
743 2012. Ecological genomics of Anopheles gambiae along a latitudinal cline: a
744 population-resequencing approach. Genetics 190:1417–1432.
745 Coetzee M, Koekemoer L. 2013. Molecular Systematics and Insecticide Resistance in
746 the Major African Malaria Vector Anopheles funestus. Annu. Rev. Entomol.
747 Cohuet A, Dia I, Simard F, Raymond M, Rousset F, Antonio-Nkondjio C, Awono-
748 Ambene PH, Wondji CS, Fontenile D. 2005. Gene flow between chromosomal
749 forms of the malaria vector Anopheles funestus in Cameroon, Central Africa,
750 and its relevance in malaria fighting. Genetics 169:301–311.
751 Cohuet A, Simard F, Toto J-C, Kengne P, Coetzee M, Fontenille D. 2003. Species
752 identification within the Anopheles funestus group of malaria vectors in
753 Cameroon and evidence for a new species. Am. J. Trop. Med. Hyg. 69:200–205.
754 Coluzzi M, Sabatini A, Torre A, Angela M, Deco D, Petrarca V. 2002. A Polytene
755 Chromosome Analysis of the Anopheles gambiae Species Complex. Science (80-.
756). 298:1415–1419.
757 Comeron JM, Williford a, Kliman RM. 2008. The Hill-Robertson effect: evolutionary
758 consequences of weak selection and linkage in finite populations. Heredity
759 (Edinb). 100:19–31.
760 Corbett-Detig RB, Hartl DL. 2012. Population Genomics of Inversion Polymorphisms

761 in Drosophila melanogaster. PLoS Genet. 8.

762 Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of
763 speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol.
764 23:3133–3157.

765 Cutter AD, Choi JY. 2010. Natural selection shapes nucleotide polymorphism across
766 the genome of the nematode Caenorhabditis briggsae Natural selection shapes
767 nucleotide polymorphism across the genome of the nematode Caenorhabditis
768 briggsae. Genome Res.:1103–1111.

769 Danecek P, Auton A, Abecasis G, Albers C a., Banks E, DePristo M a., Handsaker RE,
770 Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools.
771 Bioinformatics 27:2156–2158.

772 Davey JW, Hohenlohe P a, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011.
773 Genome-wide genetic marker discovery and genotyping using next-generation
774 sequencing. Nat. Rev. Genet. 12:499–510.

775 Derua Y a, Alifrangis M, Hosea KM, Meyrowitsch DW, Magesa SM, Pedersen EM,
776 Simonsen PE. 2012. Change in composition of the Anopheles gambiae complex
777 and its possible implications for the transmission of malaria and lymphatic
778 filariasis in north-eastern Tanzania. Malar. J. 11:188.

779 Dia I, Guelbeogo M, Ayala D. 2013. Advances and Perspectives in the Study of the
780 Malaria Mosquito Anopheles funestus.

781 Donnelly MJ, Licht MC, Lehmann T. 2001. Evidence for recent population expansion
782 in the evolutionary history of the malaria vectors Anopheles arabiensis and
783 Anopheles gambiae. Mol. Biol. Evol. 18:1353–1364.

784 Earl DA, VonHoldt BM. 2012. STRUCTURE HARVESTER: a website and program for
785 visualizing STRUCTURE output and implementing the Evanno method. Conserv.
786 Genet. Resour. 4.
787 Ellegren H, Smeds L, Burri R, Olason PI, Backström N, Kawakami T, Künstner A,
788 Mäkinen H, Nadachowska-Brzyska K, Qvarnström A, et al. 2012. The genomic
789 landscape of species divergence in Ficedula flycatchers. Nature 491:756–760.
790 Ellegren H. 2009. The different levels of genetic diversity in sex chromosomes and
791 autosomes. Trends Genet. 25:278–284.
792 Ellegren H. 2014. Genome sequencing and population genomics in non-model
793 organisms. Trends Ecol. Evol. 29:51–63.
794 Evanno G, Goudet J, Regnaut S. 2005. Detecting the number of clusters of individuals
795 using the software structure: a simulation study. Mol. Ecol. 14:2611–2620.
796 Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred
797 from metric distances among DNA haplotypes: Application to human
798 mitochondrial DNA restriction data. Genetics 131:479–491.
799 Feder JL, Roethele JB, Filchak K, Niedbalski J, Romero-Severson J. 2003. Evidence for
800 inversion polymorphism related to sympatric host race formation in the apple
801 maggot fly, Rhagoletis pomonella. Genetics 163:939–953.
802 Fouet C, Gray E, Besansky NJ, Costantini C. 2012. Adaptation to aridity in the malaria
803 mosquito Anopheles gambiae: chromosomal inversion polymorphism and body
804 size influence resistance to desiccation. PLoS One 7:e34841.
805 Fumagalli M, Vieira FG, Linderoth T. 2014. ngsTools : methods for population
806 genetics analyses from Next-Generation Sequencing data. Bioinformatics.

807 Gagnaire P-A, Pavey S a, Normandeau E, Bernatchez L. 2013. The genetic
808 architecture of reproductive isolation during speciation-with-gene-flow in lake
809 whitefish species pairs assessed by RAD sequencing. Evolution 67:2483–2497.
810 Gillies MT, Coetzee M. 1987. A supplement to the Anophelinae of Africa south of the
811 Sahara. Johannesburg: The South African Institute for Medical Research
812 Gillies MT, De Meillon B. 1968. The Anophelinae of Africa South of the Sahara.
813 Second Edi. Johannesburg: Publications of the South African Institute for
814 Medical Research
815 Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the
816 joint demographic history of multiple populations from multidimensional SNP
817 frequency data. PLoS Genet. 5:e1000695.
818 Hoffmann A a, Rieseberg LH. 2008. Revisiting the Impact of Inversions in Evolution:
819 From Population Genetic Markers to Drivers of Adaptive Shifts and Speciation?
820 Annu. Rev. Ecol. Evol. Syst. 39:21–42.
821 Hohenlohe P a, Bassham S, Etter PD, Stiffler N, Johnson E a, Cresko W a. 2010.
822 Population genomics of parallel adaptation in threespine stickleback using
823 sequenced RAD tags. PLoS Genet. 6:e1000862.
824 Huber CD, Nordborg M, Hermisson J, Hellmann I. 2014. Keeping It Local: Evidence
825 for Positive Selection in Swedish Arabidopsis thaliana. Mol. Biol. Evol. 31:3026–
826 3039.
827 Jakobsson M, Rosenberg N. 2007. CLUMPP: a cluster matching and permutation
828 program for dealing with multimodality in analysis of population structure.
829 Bioinformatics 23:1801–1806.

830 Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic
831 markers. Bioinformatics 24:1403–1405.
832 Kamdem C, Fouet C, Gamez S, White BJ. 2016. Pollutants and insecticides drive local
833 adaptation in African malaria mosquitoes. bioRxiv.
834 Kamdem C, Tene Fossog B, Simard F, Etouna J, Ndo C, Kengne P, Boussès P, Etoa F-X,
835 Awono-Ambene P, Fontenille D, et al. 2012. Anthropogenic Habitat Disturbance
836 and Ecological Divergence between Incipient Species of the Malaria Mosquito
837 Anopheles gambiae. PLoS One 7:e39453.
838 Keightley PD, Pinharanda A, Ness RW, Simpson F, Dasmahapatra KK, Mallet J, Davey
839 JW, Jiggins CD. 2015. Estimation of the Spontaneous Mutation Rate in
840 Heliconius melpomene. Mol. Biol. Evol. 32:239–243.
841 Kirkpatrick M, Barton N. 2006. Chromosome inversions, local adaptation and
842 speciation. Genetics 173:419–434.
843 Kirkpatrick M. 2010. How and why chromosome inversions evolve. PLoS Biol. 8.
844 Korneliussen T, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next Generation
845 Sequencing Data. BMC Bioinformatics 15:356.
846 Korneliussen TS, Moltke I, Albrechtsen A, Nielsen R. 2013. Calculation of Tajima's D
847 and other neutrality test statistics from low depth next-generation sequencing
848 data. BMC Bioinformatics 14:289.
849 Kulathinal RJ, Stevison LS, Noor MAF. 2009. The genomics of speciation in
850 Drosophila: Diversity, divergence, and introgression estimated using low-
851 coverage genome sequencing. PLoS Genet. 5.
852 Lawniczak MKN, Emrich SJ, Holloway a K, Regier a P, Olson M, White B, Redmond S,

853 Fulton L, Appelbaum E, Godfrey J, et al. 2010. Widespread divergence between
854 incipient Anopheles gambiae species revealed by whole genome sequences.
855 Science 330:512–514.
856 Lischer HEL, Excoffier L. 2012. PGDSpider: An automated data conversion tool for
857 connecting population genetics and genomics programs. Bioinformatics
858 28:298–299.
859 Lohse K, Clarke M, Ritchie MG, Etges WJ. 2015. Genome-wide tests for introgression
860 between cactophilic Drosophila implicate a role of inversions during speciation.
861 Evolution (N. Y). 69:1178–1190.
862 Marsden CD, Lee Y, Kreppel K, Weakley A, Cornel A, Ferguson HM, Eskin E, Lanzaro
863 GC. 2014. Diversity, Differentiation, and Linkage Disequilibrium: Prospects for
864 Association Mapping in the Malaria Vector Anopheles arabiensis. G3
865 (Bethesda). 4:121–131.
866 McGaugh SE, Noor M a F. 2012. Genomic impacts of chromosomal inversions in
867 parapatric Drosophila species. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 367:422–
868 429.
869 Meirmans P, Van Tienderen P. 2004. GENOTYPE and GENODIVE: two programs for
870 the analysis of genetic diversity of asexual organisms. Mol. Ecol. Notes 4:792–
871 794.
872 Michel a. P, Ingrasci MJ, Schemerhorn BJ, Kern M, Le Goff G, Coetzee M, Elissa N,
873 Fontenille D, Vulule J, Lehmann T, et al. 2005. Rangewide population genetic
874 structure of the African malaria vector Anopheles funestus. Mol. Ecol. 14:4235–
875 4248.

876 Moiroux N, Gomez MB, Pennetier C, Elanga E, Djènontin A, Chandre F, Djègbé I, Guis
877 H, Corbel V. 2012. Changes in anopheles funestus biting behavior following
878 universal coverage of long-lasting insecticidal nets in benin. J. Infect. Dis.
879 206:1622–1629.
880 Mwangangi JM, Mbogo CM, Orindi BO, Muturi EJ, Midega JT, Nzovu J, Gatakaa H,
881 Githure J, Borgemeister C, Keating J, et al. 2013. Shifts in malaria vector species
882 composition and transmission dynamics along the Kenyan coast over the past
883 20 years. Malar. J. 12:13.
884 Neafsey DE, Lawniczak MKN, Park DJ, Redmond SN, Coulibaly MB, Traoré SF, Sagnon
885 N, Costantini C, Johnson C, Wiegand RC, et al. 2010. SNP genotyping defines
886 complex gene-flow boundaries among African malaria vector mosquitoes.
887 Science 330:514–517.
888 Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, Amon J,
889 Arca B, Arensburger P, Artemov G, et al. 2015. Highly evolvable malaria vectors:
890 The genomes of 16 Anopheles mosquitoes. Science (80-.). 347:1258522–
891 1258522.
892 Nielsen R, Paul JS, Albrechtsen A, Song YS. 2011. Genotype and SNP calling from
893 next-generation sequencing data. Nat. Rev. Genet. 12:443–451.
894 Noor M a F, Bennett SM. 2009. Islands of speciation or mirages in the desert?
895 Examining the role of restricted recombination in maintaining species. Heredity
896 (Edinb). 103:439–444.
897 Nosil P, Feder JL. 2012. Widespread yet heterogeneous genomic divergence. Mol.
898 Ecol. 21:2829–2832.

899 Nosil P, Feder JL. 2013. Genome evolution and speciation: toward quantitative
900 descriptions of pattern and process. Evolution 67:2461–2467.
901 Nosil P. 2012. Ecological Speciation. Oxford Ser.
902 O'Reilly PF, Birney E, Balding DJ. 2008. Confounding between recombination and
903 selection, and the Ped/Pop method for detecting selection. Genome Res.
904 18:1304–1313.
905 Paradis E, Claude J, Strimmer K. 2004. Analyses of Phylogenetics and Evolution in R
906 language. Bioinformatics 20:289–290.
907 Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double Digest
908 RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in
909 Model and Non-Model Species. PLoS One 7:e37135.
910 Pool JE, Corbett-Detig RB, Sugino RP, Stevens K a., Cardeno CM, Crepeau MW,
911 Duchen P, Emerson JJ, Saelao P, Begun DJ, et al. 2012. Population Genomics of
912 Sub-Saharan Drosophila melanogaster: African Diversity and Non-African
913 Admixture. PLoS Genet. 8.
914 Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using
915 multilocus genotype data. Genetics 155:945–959.
916 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, Maller J, Sklar P,
917 de Bakker P, Daly M, et al. 2007. PLINK: a toolset for whole-genome association
918 and population-based linkage analysis. Am. J. Hum. Genet. 81.
919 R Development Core Team. 2014. R: A language and environment for statistical
920 computing. R Foundation for Statistical Computing, Vienna, Austria
921 Rane AR V, Rako L, Lee SF, Hoffmann A a. 2015. Genomic evidence for role of

922 inversion 3RP of Drosophila melanogaster in facilitating climate change
923 adaptation Abstract : :1–24.
924 Renaut S, Grassa CJ, Yeaman S, Moyers BT, Lai Z, Kane NC, Bowers JE, Burke JM,
925 Rieseberg LH. 2013. Genomic islands of divergence are not affected by
926 geography of speciation in sunflowers. Nat. Commun. 4:1827.
927 Riehle MM, Guelbeogo WM, Gneme A, Eiglmeier K, Holm I, Bischoff E, Garnier T,
928 Snyder GM, Li X, Markianos K, et al. 2011. A cryptic subgroup of Anopheles
929 gambiae is highly susceptible to human malaria parasites. Science 331:596–
930 598.
931 Roesti M, Hendry AP, Salzburger W, Berner D. 2012. Genome divergence during
932 evolutionary diversification as revealed in replicate lake-stream stickleback
933 population pairs. Mol. Ecol. 21:2852–2862.
934 Rosenberg N. 2004. DISTRUCT: a program for the graphical display of population
935 structure. Mol. Ecol. Resour. 4:137–138.
936 Sella G, Petrov D a, Przeworski M, Andolfatto P. 2009. Pervasive natural selection in
937 the Drosophila genome? PLoS Genet. 5:e1000495.
938 Service MW. 1993. Mosquito ecology: field sampling methods. (Elsevier Applied
939 Science, London UK, editor.).
940 Sharakhov I, Braginets O, Grushko O, Cohuet A, Guelbeogo WM, Boccolini D, Weill M,
941 Costantini C, Sagnon N, Fontenille D, et al. 2004. A Microsatellite Map of the
942 African Human Malaria Vector Anopheles funestus. J. Hered. 95:29–34.
943 Sharakhova M V, Xia A, Leman SC, Sharakhov I V. 2011. Arm-specific dynamics of
944 chromosome evolution in malaria mosquitoes. BMC Evol. Biol. 11:91.

945 Sinka ME, Bangs MJ, Manguin S, Coetzee M, Mbogo CM, Hemingway J, Patil AP,
946 Temperley WH, Gething PW, Kabaria CW, et al. 2010. The dominant Anopheles
947 vectors of human malaria in Africa, Europe and the Middle East: occurrence
948 data, distribution maps and bionomic précis. Parasit. Vectors 3:117.
949 Thomas CG, Wang W, Jovelin R, Ghosh R, Lomasko T, Trinh Q, Kruglyak L, Stein LD,
950 Cutter AD. 2015. Full-genome evolutionary histories of selfing , splitting , and
951 selection in Caenorhabditis. :667–678.
952 Via S. 2012. Divergence hitchhiking and the spread of genomic isolation during
953 ecological speciation-with-gene-flow. Philos. Trans. R. Soc. B Biol. Sci. 367:451–
954 460.
955 Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population
956 structure. Evolution (N. Y). 38:1358–1370.
957 Wilson Sayres MA, Lohmueller KE, Nielsen R. 2014. Natural Selection Reduced
958 Diversity on Human Y Chromosomes. PLoS Genet. 10.
959 Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and
960 splicing in short reads. Bioinformatics 26:873–881.
961
962
963

964 **Figure legends**

965 **Figure 1**: Geographic origin and genetic relatedness of *An. funestus* samples in
966 Cameroon. (A) Map of sampling locations showing the 28 sites surveyed (small
967 circles) and the two villages (large orange and green squares) where *An. funestus*
968 samples were collected. The line in the middle of the map indicates an arbitrary
969 delimitation of the distribution ranges of the two ecotypes found in Cameroon
970 (Cohuet et al. 2005; Ayala et al. 2011). (B), (C) and (D) Population genetic structure
971 as revealed by neighbor-joining tree, PCA and Bayesian clustering (STRUCTURE).
972 The percentage of variance explained is indicated on each PCA axis. (E) and (F)
973 Confirmation of the presence of two ecotypes with DAPC and the delta k method of
974 Evanno et al. The decrease of the Bayesian Information Criterion (BIC) (DAPC) and
975 the increase of delta k as a function of the number of k indicate a subdivision of *An.*
976 *funestus* into two 2-3 genetic clusters.

977 **Figure 2**: Estimates of pairwise population differentiation (F_{ST}) based on 10687
978 SNPs ordered by position along "pseudochromosomes" representing the five
979 chromosomal arms of *An. funestus*. F_{ST} values above the red line are above the 99th
980 percent of the empirical distribution. Arrows indicate the genomic coordinates of
981 the 3Ra and 3R*b* inversions.

982 **Figure 3**: Genome-wide distribution of d_{xy} and fixed differences (d_f) in *An. funestus*
983 among Mfou and Tibati populations. Estimates are smoothed with kernel smoothing
984 and plotted along the five chromosomal arms.

985 **Figure 4**: Population genetic structure revealed by each chromosomal arm and by
986 four polymorphic chromosomal inversions (2R*h*, 3R*a*, 3R*b* and 3R*d*).

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 chromosomal arms. The only uninverted autosome (2L) is the most polymorphic both in Mfou and Tibati.

Figure 6: Box plot depicting the variation of θw among chromosomes in Mfou and Tibati.

Supplemental Figure Legends

Figure S1: Positions of the 104 scaffolds placed on the physical map of *An. funestus* indicated by grey rectangles below the chromosomes (modified from Sharakhov *et al*. 2004).

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

Tables

Table 1: Parameters of the best-fitted demographic model in ∂a∂i for the genome-wide and chromosome-specific diversity in Mfou and Tibati populations.

Population	Chromosome	Best model	Log Likelihood	Final population size[b] (95% CI)		Time[c] (95% CI)	
Tibati	Autosomes	*Growth*	-88.8	77.6	(51.8 - 138.1)	3.8	(1.8 - 8.0)
	2R	*Growth*	-52.93	92.5	(54.3 - 187.2)	5.3	(2.4 - 11.9)
	2L	*Growth*	-45.86	141.4	(69.2 - 199.3)	4.5	(2.0 - 10.1)
	3R	*Growth*	-47.35	38.8	(24.8 - 87.9)	3.8	(0.6 - 3.9)
	3L	*Growth*	-42.52	83	(47.4 - 185.1)	4.6	(1.3 - 12.3)
Mfou	Autosomes	*Two-epoch*	-84.7	13.6	(12.2 - 15.7)	1.4	(1.2 - 1.7)
	2R	*Two-epoch*	-56.45	11.4	(8.4 - 17.1)	1.5	(1.0 - 2.8)
	2L	*Two-epoch*	-44.27	26.9	(16.4 - 52.5)	2.2	(1.1 - 4.8)
	3R	*Two-epoch*	-44.65	15.1	(10.7 - 23.4)	1.2	(0.8 - 2.2)
	3L	*Two-epoch*	-41.65	19	(5.1 - 36.0)	2.2	(0.2 - 4.4)

[b]Relative to ancestral population size.

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

Supplemental Tables

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

Ecogeographic regions	Sampling locations	Geographic coordinates	Sampling methods				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

Table S2: Characteristics of the five "pseudochromosomes" of *An. funestus* used for genome scans.

Chromosome	Number of 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

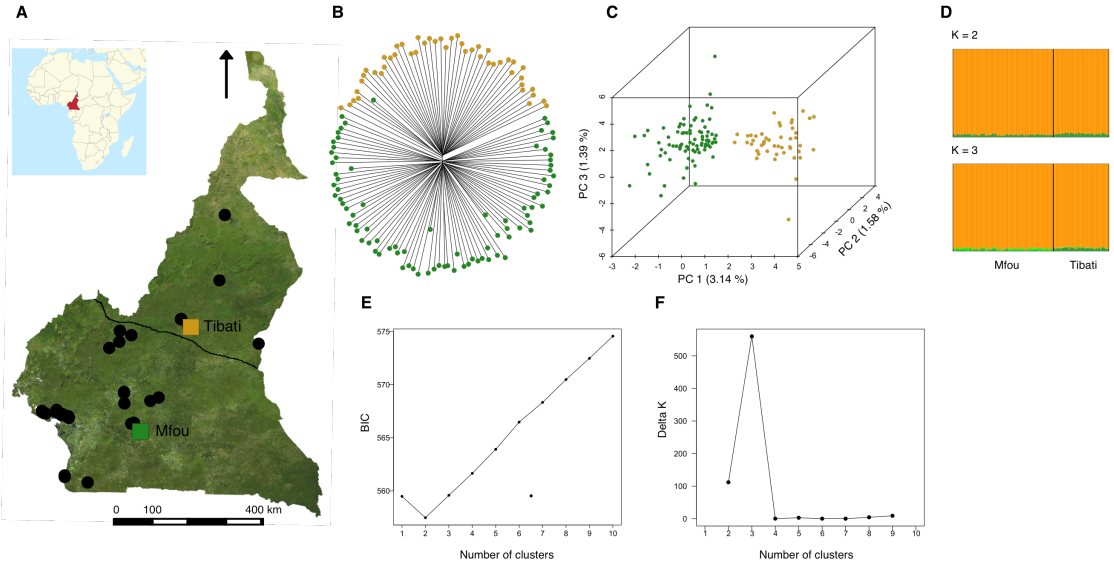

Figure 2

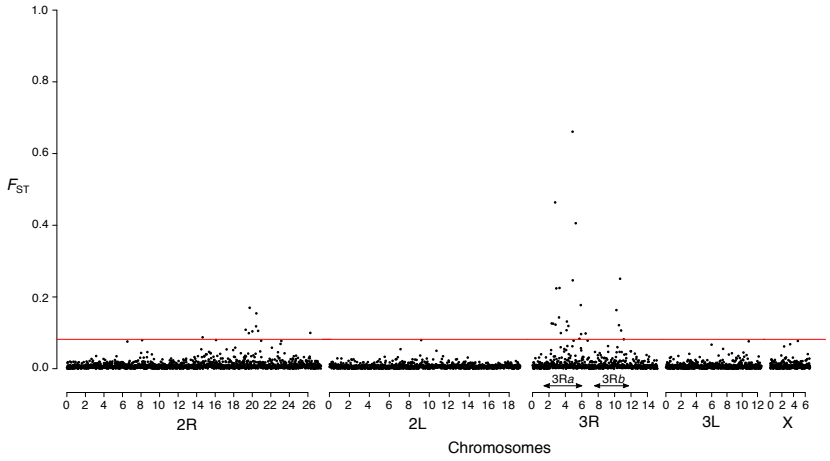

Figure 3

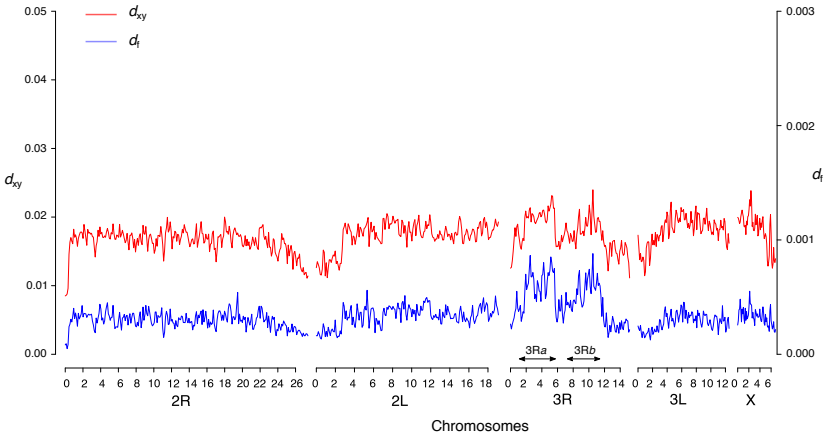

Figure 4

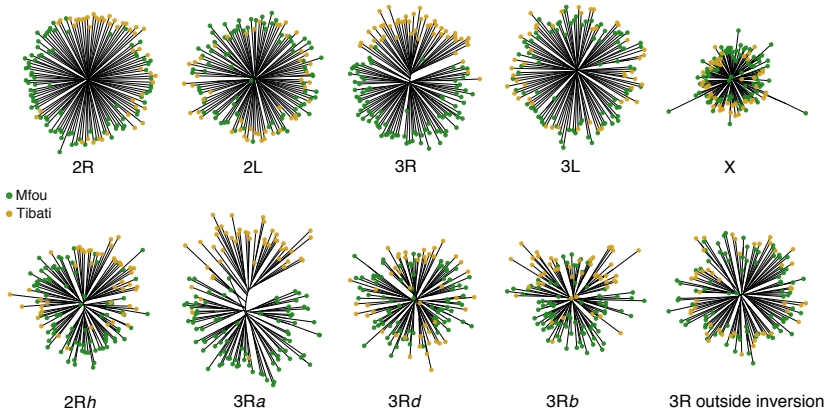

Figure 5

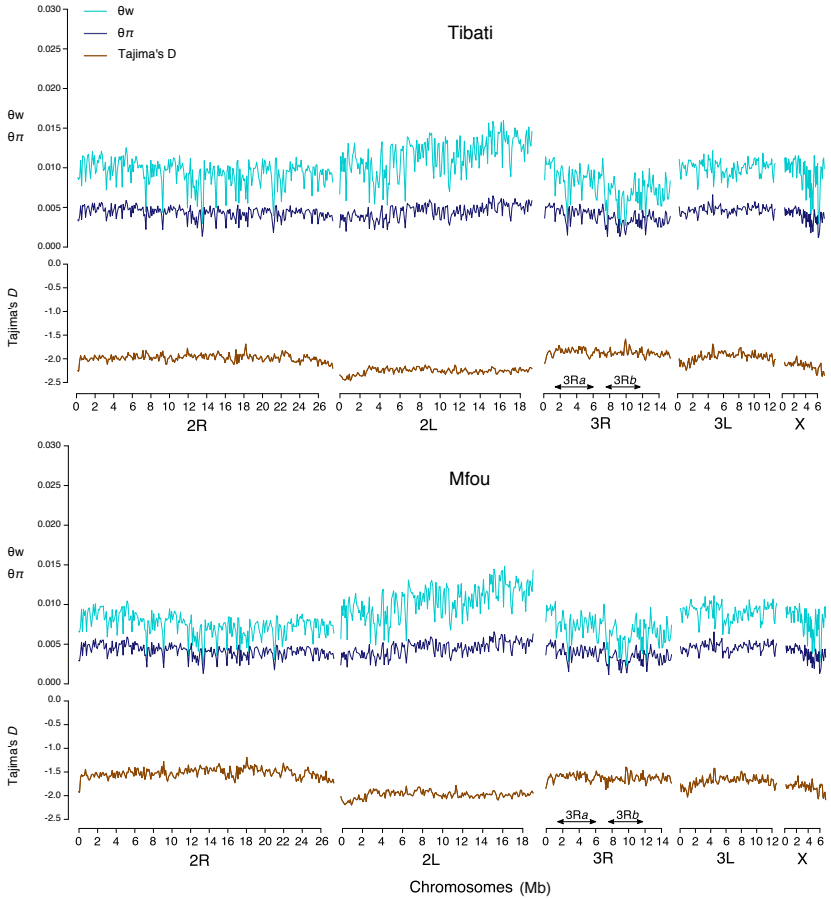

Figure 6

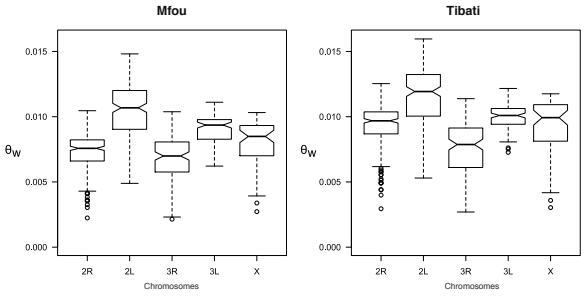

Figure S1

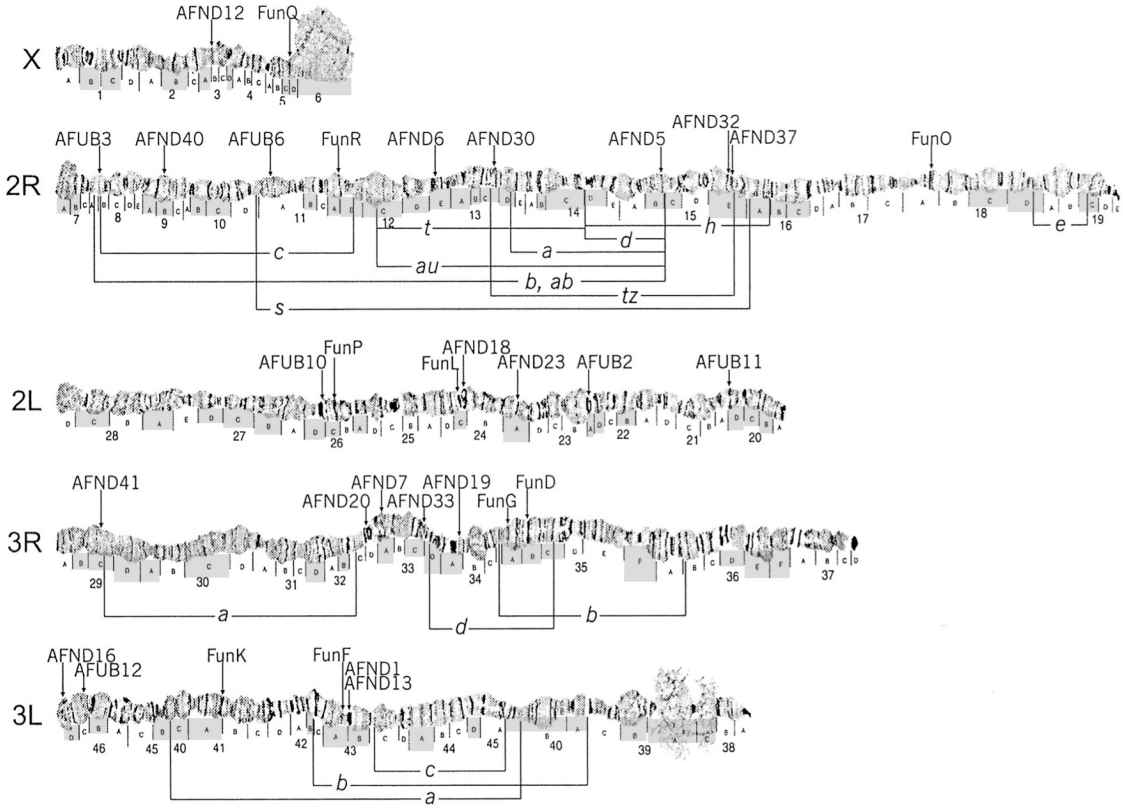

Figure S2