



HAL
open science

Evolutionary implications of recombination differences across diverging populations of *Anopheles*

Joel T Nelson, Omar E Cornejo, The *Anopheles Gambiae* 1000 Genomes
Consortium (including Mc Fontaine)

► **To cite this version:**

Joel T Nelson, Omar E Cornejo, The *Anopheles Gambiae* 1000 Genomes Consortium (including Mc Fontaine). Evolutionary implications of recombination differences across diverging populations of *Anopheles*. 2021. hal-03415333

HAL Id: hal-03415333

<https://hal.science/hal-03415333>

Preprint submitted on 4 Nov 2021

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

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

1 Evolutionary implications of recombination differences across diverging
2 populations of *Anopheles*

3 **Authors: Joel T. Nelson¹, Omar E. Cornejo^{1*} and Ag1000G Consortium²**

4

5 ¹ **School of Biological Sciences, Washington State University, Pullman WA. 99164**

6 ² **Wellcome Sanger Institute, Hinxton, Cambridgeshire, CB10 1SD, United Kingdom**

7 (<https://www.malariagen.net/projects/ag1000g#people>)

8

9 **Corresponding: omar.cornejo@wsu.edu**

10

11

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32 **Abstract**

33 Recombination is one of the main evolutionary mechanisms responsible for changing the
34 genomic architecture of populations; and in essence, it is the main mechanism by which novel
35 combinations of alleles, haplotypes, are formed. A clear picture that has emerged across study
36 systems is that recombination is highly variable, even among closely related species. However, it
37 is only until very recently that we have started to understand how recombination variation
38 between populations of the same species impact genetic diversity and divergence. Here, we used
39 whole-genome sequence data to build fine-scale recombination maps for nine populations within
40 two species of *Anopheles*, *Anopheles gambiae* and *Anopheles coluzzii*. The genome-wide
41 recombination averages were on the same order of magnitude for all populations except one. Yet,
42 we identified significant differences in fine-scale recombination rates among all population
43 comparisons. We report that effective population sizes, and presence of a chromosomal inversion
44 has major contribution to recombination rate variation along the genome and across populations.
45 We identified over 400 highly variable recombination hotspots across all populations, where
46 only 9.6% are shared between two or more populations. Additionally, our results are consistent
47 with recombination hotspots contributing to both genetic diversity and absolute divergence (dxy)
48 between populations and species of *Anopheles*. However, we also show that recombination has a
49 small impact on population genetic differentiation as estimated with F_{ST} . The minimal impact
50 that recombination has on genetic differentiation across populations represents the first empirical
51 evidence against recent theoretical work suggesting that variation in recombination along the
52 genome can mask or impair our ability to detect signatures of selection. Our findings add new
53 understanding to how recombination rates vary within species, and how this major evolutionary
54 mechanism can maintain and contribute to genetic variation and divergence within a prominent
55 malaria vector.

56

57

58

59

60

61

62

63 **Introduction**

64 Meiotic recombination is the main evolutionary mechanism that shapes haplotypic
65 variation in sexually reproducing species (Felsenstein 1974; San Filippo et al. 2008; Beeson et al.
66 2019). A main result of recombination is the movement of alleles onto different genetic
67 backgrounds, which has the potential to create novel combinations of beneficial haplotypes and
68 novel haplotypes (Posada et al. 2002; Smagulova et al. 2016; Korunes and Noor 2017).
69 Recombination can reduce interference between linked loci, thereof reducing mutational load,
70 and it allows for beneficial mutations to sweep through a population more rapidly (Gabriel et al.
71 1993; McVean, G. A. and Charlesworth 2000; Otto and Barton 2001; Hartfield and Keightley
72 2012; Ritz et al. 2017). It has been suggested that elevated recombination rates can contribute to
73 increased patterns of genetic diversity and divergence, though this relationship is not well
74 understood (Smukowski and Noor 2011). Across most species, recombination is highly variable
75 within and across chromosomes but tends to cluster in local regions identified as recombination
76 hotspots (Hellsten et al. 2013; Beeson et al. 2019). Both fine-scale recombination rates and
77 hotspots are poorly conserved across mammals and plants with significant differences even
78 among closely related species (Spencer et al. 2006; Choi et al. 2016; Beeson et al. 2019; Dreissig
79 et al. 2019). However, less is known about the onset of variation in recombination rates across
80 diverging populations of the same species (Schwarzkopf et al. 2020) and how it impacts patterns
81 of genetic variation. Comparing recombination maps across multiple populations of humans has
82 recently been investigated; however, this question has not been applied to other model systems
83 and was aimed at identifying the impact of demographic changes on the recombination landscape
84 and not how elevated recombination impacts patterns of diversity and divergence (Spencer et al.
85 2006). Thus, comparing recombination maps across closely related populations and quantifying
86 its impact on genetic variation is imperative for elucidating how this mechanism shapes patterns
87 of evolution.

88 Comparing recombination maps across closely related populations could help predict
89 shifts in the recombination landscape that are responsible for species-level differences in
90 recombination (Smukowski and Noor 2011). For example, comparing the recombination
91 landscape across genomes with varying selective sweeps, effective populations sizes, and
92 chromosomal inversions. These aspects of genomic evolution are known to impact
93 recombination events, however the contribution they have towards recombination rate

94 differentiation across populations in generally unknown (Keightley and Otto 2006; Feder and
95 Nosil 2009; Yang et al. 2018). Furthermore, comparing recombination landscapes among
96 populations can also have significant contributions towards understanding the impact that
97 elevated recombination rates have on patterns of genetic diversity and divergence (Brown and
98 Jiricny 1987; Brown et al. 1989; Papavasiliou and Schatz 2000; Smukowski and Noor 2011). For
99 instance, because of the hitchhiking effects of linked selection, local regions of the genome with
100 significantly larger recombination rates (recombination hotspots) may act as reservoirs for
101 genetic diversity. Previous studies have identified an increase in nucleotide diversity within
102 regions of elevated recombination; however, there is a discord among studies for the number of
103 differences between species within the same regions (Smukowski and Noor 2011; Roesti et al.
104 2013). In concordance to genetic diversity and divergence, little is known about the impact that
105 elevated recombination rates have on localized patterns of genetic differentiation (F_{ST}); and if
106 lower levels of differentiation are expected in regions of the genome with elevated
107 recombination rates due to reduced linkage. Thus, understanding these patterns is an important
108 step in elucidating the early onset of recombination rate evolution among closely related
109 populations and the impact that elevated rates have on levels of genetic diversity and divergence.
110 A perhaps more subtle, but equally important relevance of recombination in modern genomic
111 studies is the potential impact that differences in recombination have on our ability to identify
112 signatures of hard or soft sweeps in the genome. Using simulations, it has been shown that
113 recombination acting on regions close to a hard sweep can generate patterns of polymorphism
114 that resemble those found in soft sweeps, rendering them indistinguishable (Pennings and
115 Hermisson 2006; Schrider et al. 2015).

116 To address recombination rate variation among closely related populations we used
117 sequence data from two different species of *Anopheles*, *A. gambiae* and *A. coluzzii*; which are the
118 major contributors to the spread of malaria (Holt et al. 2002; Anopheles gambiae 1000 Genomes
119 Consortium 2017). There are several known populations of *A. gambiae* and *A. coluzzii* that have
120 a large distribution throughout Africa, spanning across prominent ecological gradients
121 (Anopheles gambiae 1000 Genomes Consortium 2017). Previous studies have identified several
122 populations that have undergone expansions and bottlenecks, as well as containing the presence
123 of selective sweeps and chromosomal inversions, characteristics that contribute to differences in
124 the recombination architecture of populations (Lehmann et al. 1999; Anopheles gambiae 1000

125 Genomes Consortium 2017). Populations within this system have low background levels of
126 genetic differentiation (White et al. 2010; Anopheles gambiae 1000 Genomes Consortium 2017).
127 The combination of multiple population pairwise comparisons and low levels of genetic
128 differentiation allows for robust detection of increased genetic differentiation and divergence in
129 localized regions of the genome where recombination is elevated (e.g. recombination hotspots).
130 There is a large body of evidence showing a correlation between recombination and nucleotide
131 diversity in *Drosophila melanogaster* (Begun and Aquadro 1992; Andolfatto and Przeworski
132 2001), with no consequence on divergence, which has led to suggest that natural selection is an
133 important force structuring patterns of polymorphism along the genome (Wright et al. 2006). The
134 pattern described for *Drosophila* has not always been replicated in other organisms, where
135 divergence is at times correlated with recombination, suggesting a lack of generality on this
136 observation (Nachman et al. 1998; Nachman 2001; Lercher and Hurst 2002; Cutter and Payseur
137 2003; Hellmann, Ines et al. 2003; Hellmann, I. et al. 2005). We believe that a more extensive
138 analysis of the architecture of recombination and genome-wide polymorphism is necessary to
139 shed light into this phenomenon.

140

141 Here, we use whole genome data to infer fine-scale recombination maps across nine
142 diverging populations of *A. gambiae* and *A. coluzzii* (Figure 1). Specifically, we 1) quantified the
143 recombination landscape across populations and identified the impact that different
144 evolutionary/genomic features have on recombination rates, 2) identified regions of the genome
145 where recombination rates were significantly greater than average background levels (hotspots),
146 and 3) quantified the impact of elevated recombination on population diversity and divergence.
147 We expected to see large variations in recombination rates, even among closely related
148 populations (fine-scale), yet smaller difference when comparing genome-wide rates (broad-
149 scale). Based on what has been observed in studies in other organisms, we expected that
150 differences in recombination hotspots will be similar to variation in recombination rates where
151 most hotspots would be unique to a specific population. Simulation studies have shown that
152 recombination differences can reduce our ability to detect regions of the genome under selection,
153 evidenced by excessive differentiation. We expect that our analyses of recombination will allow
154 to investigate the extent of these expectations in natural populations.

155 **Results**

156 *Comparing fine-scale recombination maps across diverging populations*

157 Average population-scaled recombination rate ($\rho = 4Ne r$, where Ne is the effective
158 population size and r is the per site recombination rate; measures one recombination event per
159 kb) across the entire genome was 70.9ρ . Population averages ranged from 0.58ρ (KES) to
160 140.26ρ (CMS) (Table 1 & 2) (see Supplementary Figure S1 – S3 for summarized
161 recombination maps for each species). Genome-wide recombination distributions followed a
162 positively skewed distributions where most recombination events throughout the genome were
163 low with the exceptions to localized regions with elevated recombination (Figure 2). Results
164 from our Wilcoxon Rank test suggests significant difference in recombination rates across all
165 population comparisons ($p < 7.39e-14$) (Supplementary Table S1 & S2). After accounting for
166 effective population size average per-site recombination rates (r) were all on the same order of
167 magnitude expect for KES (Table 2). Average recombination rates (ρ/bp) were roughly two
168 orders of magnitude lower than fine-scale estimates across species of *Drosophila* (Chan et al.
169 2012). Indeed, fine-scale recombination rates within *Anopheles* were in the lower quantile when
170 including recombination estimates from other insect systems (Stapley et al. 2017).

171 *Ne, chromosomal location, inversions influence recombination variation, and sweeps*

172 Populations with a smaller effective population size (AOM, GAS, and KES) exhibited a
173 lower variance across the recombination landscape, while populations with a large effective
174 population size reported nearly 100 times larger variance in the recombination landscape (Figure
175 2). When considering chromosomal location, we identified large reductions in recombination
176 rates within centromeric regions of each chromosome, however, this patten was more prominent
177 along the X chromosome (Figure 2 & 3). As expected, for populations that were polymorphic
178 for chromosomal inversions, all recombination estimates (with the exception to BFS) fell below
179 the genome average within inverted regions and then increased once outside of the inversion
180 (Table 3; Figure 3). This pattern was not observed for populations that were homozygous for the
181 inverted or wild-type genotypes (Figure 3), further confirming the characterization of the
182 polymorphism of the inversion in these populations. Furthermore, Wilcoxon Rank test suggest
183 significant difference among all pairwise comparisons in recombination rates within genomic
184 regions containing the 2La and 2Rb inversion ($p < 2.2e-16$) (Supplementary Table S3 & S4).
185 Lastly, we also showed that recombination rates were largely suppressed within insecticide
186 resistant (IR) genes that have been identified as under strong selection in previous studies

187 (Anopheles gambiae 1000 Genomes Consortium 2017). For example, the decrease in
188 recombination rates in a 1 Mb region within the Cyp6p gene and a 3Mb region within the
189 Cyp9k1 gene (Figure 4; also see Supplementary Figure S4-S9). For IR genes that were putatively
190 evolving under neutral conditions (or weak signals of selection), we did not identify a prominent
191 reduction in recombination rates or change in the topography of the recombination map.

192 *Identifying unique and shared recombination hotspots*

193 We identified a total of 435 robust signatures consistent with a recombination hotspot
194 (Figure 5; Supplementary Table S5). In most populations, the average recombination rate within
195 hotspots was at least twice as large as the adjacent background; however, the average
196 recombination rates within hotspots were only slightly larger than the average genome-wide
197 recombination rate across all populations ($HS = \sim 71.0 \rho$, genome = $\sim 70.0 \rho$). Moreover, for
198 autosomal chromosomes, we identified 383 hotspots with an average of 48 hotspots per
199 population. Of the 383 autosomal hotspots, 344 were unique to a single population while 39 were
200 shared between two or more populations. For the X chromosome, we retained a total of 52
201 recombination hotspots with an average of six hotspots per population. Among the 52 hotspots,
202 51 were unique to a single population while only a single hotspot was shared between two
203 populations (GWA and KES). A total of 26 population pairwise comparisons shared
204 recombination hotspots.

205 Despite finding very little evidence of overlapping (shared) hotspots, empirical overlaps
206 were on average larger than expected when hotspots were randomly shuffled throughout the
207 genome, resulting in a positive net difference (empirical hotspot overlap – shuffled overlap;
208 Supplementary Table S6). Furthermore, pairwise Fishers Exact test showed that a total of 19 (out
209 of 26) comparisons had significantly larger hotspot overlaps than stochastic expectation ($p <$
210 0.05 ; Supplementary Table S6). Of the 19 significant comparisons, most were found between
211 populations of *A. gambiae* (14 comparisons) while the remaining comparisons were between *A.*
212 *gambiae* and *A. coluzzii*; there were no significant hotspot overlaps within *A. coluzzii*.

213 *Consequences of recombination hotspots on population diversity, and divergence*

214 Average nucleotide diversity was larger within recombination hotspots ($0.02578 \pm$
215 0.00073) compared to genomic background estimates (0.02257 ± 0.00002201) (Figure 6). Across
216 the genome, we show that within most populations, nucleotide diversity was significantly larger
217 ($p < 0.05$) within recombination hotspots than compared to the surrounding background levels

218 (Figure 6). Moreover, for autosomal chromosomes, the average nucleotide diversity was
219 significantly larger within recombination hotspots (0.02893 ± 0.00038) than compared to the
220 average background (0.02567 ± 0.00001322). Across the X chromosome, nucleotide diversity
221 followed similar patterns where nucleotide diversity was significantly larger within
222 recombination hotspots (0.02264 ± 0.00109) than compared to the genomic background
223 (0.01947 ± 0.0000308) (Figure 6).

224 Similar to nucleotide diversity, average estimates of dxy proved to be larger within
225 recombination hotspots (0.0244 ± 0.0004) when compared to a genome-wide estimate ($0.0215 \pm$
226 0.0000122). Between autosomal and the X chromosome most population pairwise comparisons
227 support evidence suggesting significant differences ($p < 0.05$) in the number of per site
228 differences within recombination hotspots (Figure 7A). Moreover, across autosomal
229 chromosomes average dxy was larger within recombination hotspots (0.0272 ± 0.000325) than
230 average dxy estimates across the genome (0.0241 ± 0.0000113). For the X chromosome, average
231 dxy estimates were larger within recombination hotspots (0.0215 ± 0.000475) than compared to
232 dxy along the X chromosome (0.0189 ± 0.0000130) (Supplementary Figure S10). Regions
233 within recombination hotspots experience ~3.0% more nucleotide differences per site than the
234 entire X chromosome.

235 When we examine the patterns of F_{ST} we find that of the 72 comparisons, 51 (70.8%)
236 showed an increase in F_{ST} within recombination hotspots, whereas the remaining 21 comparisons
237 (29.2%) showed a decrease in F_{ST} within hotspots (Figure 7B). However, the difference ratio
238 between the genomic background F_{ST} and adjacent hotspot F_{ST} zero, suggesting very little
239 change in relative genetic differentiation between hotspots and the adjacent background for all
240 comparisons. Despite this, we still identified 13 comparisons where F_{ST} was significantly greater
241 than what was predicted from the distribution of F_{ST} for non-hotspot regions. Variance of F_{ST}
242 followed similar patterns where across all hotspots it was higher (0.00742) when compared to the
243 F_{ST} variance of the adjacent genomic backgrounds (0.00702); however, this difference was not
244 significant ($p = 0.0660$).

245

246 Discussion

247 In this study, we report the first recombination landscape for multiple populations of *A.*
248 *gambiae* and *A. coluzzii*. We identified large variations in recombination rates across all

249 populations. We showed that effective population size, chromosomal location (centromeric
250 regions), chromosomal inversions, and selective sweeps all contribute to recombination
251 differentiation among populations. We identified several hotspots on both X and autosomal
252 chromosomes suggesting regions of the genome with significantly higher recombination rates.
253 Lastly, we showed significant associations between recombination hotspots, nucleotide diversity,
254 and relative/absolute divergence, suggesting that recombination hotspots may be playing a role
255 in not only maintaining within population diversity, but also contributing to between population
256 divergence.

257 *Fine-scale recombination rates across populations of Anopheles*

258 Consistent with other study systems, our results suggest differences in recombination
259 rates among populations (Kong et al. 2002; Nachman 2002; Smukowski and Noor 2011) (Figure
260 2; Table 1). Our results suggest that these differences are not just the consequence of overall
261 changes in recombination rate among populations, but the result of differences in fine-scale
262 recombination rates along the genome (Table 2), consistent with recent findings (Spence and
263 Song 2019). Importantly, these patterns were consistent with previous studies showing high
264 levels of recombination divergence at fine scale, whereas more conservation in recombination
265 rates across the entire genome (Smukowski and Noor 2011). It has been proposed that variation
266 in recombination rates is due to evolutionary constraints that allows for a lower and upper bound
267 of recombination (Smukowski and Noor 2011; Ritz et al. 2017). Lower recombination rates may
268 prove to be maladaptive as the genome lacks the ability to create novel haplotypes and reduce
269 interference of deleterious mutations. Conversely, extremely high levels of recombination have
270 the potential to break up beneficial haplotypes and gene complexes, proving to be maladaptive.
271 These bounds may be responsible for the conservation of the average genome-wide
272 recombination rates, however, the localized differences in recombination may represent the
273 variation in recombination within evolutionary constraints., or the presence of selective sweeps.
274 Regions of the genome with recently fixed selective sweeps might display excessive LD and low
275 recombination (Kim and Nielsen 2004; Sibley and Ajioka 2008). We see this pattern reflected in
276 the recombination landscape of *Anopheles* when we focus on genes involved in insecticide
277 resistance, and more thoroughly discuss this results in the following sections.

278

279 *Evolutionary and genomic features shaping the recombination landscape*

280 We showed that the recombination landscape is heavily influenced by evolutionary and
281 genomic features such as effective population size, chromosomal location, inversions, and
282 selective sweeps. For example, populations that are south of the Congo River basin (AOM and
283 GAS) and east of the African Rift valley (KES) have shown genomic evidence of recent
284 bottlenecks (Lehmann et al. 1999; Anopheles gambiae 1000 Genomes Consortium 2017). Indeed
285 AOM, GAS, and KES had the lowest variance, as well as reduced population-scaled
286 recombination (ρ/kb) (Figure 2A). However, when calculating the per site recombination rate to
287 account for effective population size, KES had an average recombination rate that was an order
288 of magnitude larger than the other populations. As demonstrated within previous studies,
289 populations with a small effective population size may exhibit higher recombination rates
290 associated to an increase accumulation of deleterious mutations in homogenized genomes
291 (Keightley and Otto 2006; Kumar et al. 2019; Schwarzkopf et al. 2020). Because KES has had a
292 small historical N_e ($\sim 10,000$), we hypothesize that the deleterious effects of a low N_e have
293 contributed to a increased overall recombination rates within KES; however, this hypothesis
294 needs to be tested.

295 Chromosomal location is also a major influence on the recombination landscape. For
296 example, recombination was reduced near the centromeric regions for all populations (Figure 2
297 & 3). Reduced recombination near centromeres has been previously observed (Nambiar and
298 Smith 2016). Recombination within centromeric regions is often considered maladaptive where
299 it can break up conserved blocks of genes and interfere with proper chromosome segregation
300 during meiosis, which has the potential to create aneuploid progeny (Nambiar and Smith 2016).
301 In the case of *A. gambiae* and *A. coluzzii*, each species is grouped by a molecular form based off
302 diagnostic genetic differences in the ribosomal DNA, near centromeric regions of the X
303 chromosome (Coetzee et al. 2013; Aboagye-Antwi et al. 2015). These fixed differences have
304 been shown to play a role in positive assortative mating, thus reducing gene flow between the
305 two species (Aboagye-Antwi et al. 2015). It has long been speculated that low levels of
306 recombination rates along the X chromosome have contributed to reproductive isolation between
307 *A. gambiae* and *A. coluzzii*. Here, for the first time, we show drastic reduction in recombination
308 rates along the centromeric regions of the X chromosomes across all sampled populations that
309 supports the underlining hypothesis of little recombination near the centromere of the X
310 chromosome.

311 Of the six inversions within *Anopheles*, two of them, 2Rb and 2La, show prominent
312 associations for specific environments (rain-dependent vs. rain-independent) (Coluzzi et al.
313 1979; Lanzaro and Lee 2013; Cheng et al. 2018). For populations that are homozygous for either
314 chromosomal inversion (2La or 2Rb), or the wild type (no inversion), there were no reductions in
315 recombination rates across the 2La and 2Rb chromosomal inversions. Indeed, the inversion loop
316 does not need to form during meiosis (Figure 3; Table 3), and therefore, does not result in
317 maladaptive indels within recombinant chromosomes. However, for populations that were
318 polymorphic for the inversions, recombination rates within the inversion were largely
319 suppressed. This drastic reduction in recombination was also followed by an increase in
320 recombination along regions that flank the inversions (Figure 3; Table 3). Though, for
321 populations that were polymorphic for chromosomal inversions we still detected evidence for
322 successful recombination events within these inversions, and in some cases, recombination
323 hotspots. Successful recombination within inversions could be explained by a double cross-over
324 within inversion loops which, can potentially prevent the major insertions/deletions that occur
325 between inverted and non-inverted homologous chromosomes, resulting in successful
326 recombinant progeny. Previous studies focusing on the clinal variation of chromosomal
327 inversions have suggested that relaxed pressure for the 2La inversion in non-arid regions may
328 lead to an increase in recombination (through double cross-overs) within inverted regions (Cheng
329 et al. 2012).

330

331 *Recombination rates within insecticide resistance genes*

332 Because of long-term exposure to insecticides, recent selection analyses show signatures
333 of molecular adaptation across several genes known to be involved with insecticide resistance
334 (*Anopheles gambiae* 1000 Genomes Consortium 2017). Although this exposure has been
335 relatively long in an ecological timeframe; the selection imposed by insecticides on *Anopheles*
336 populations have only been acting recently in an evolutionary time scale. The implications for
337 the evolution of resistance in these populations is that we expect to see reduced recombination in
338 areas surrounding the genes involved in insecticide resistance. These regions are of great
339 importance because as insecticide resistance increases, *Anopheles* becomes increasingly difficult
340 to control and indirectly making the spread of malaria tough to mitigate. Recombination is
341 especially important because of its ability to generate novel haplotypes which may increase

342 insecticide resistance, especially under models that apply an evolutionary arms race (Clay and
343 Kover 1996). However, here we show that within IR genes, especially within Cyp6p and Cyp9k1
344 (both known for the resistance towards DDT, and pyrethroid-based compounds), recombination
345 is reduced within 1-3Mb of the IR gene (Figure 4; Supplementary Figure S2-S7). These genes
346 were found to be under selection in UGS and BFM, however, they are neutrally (or show a weak
347 signal of selection) evolving in other populations such as GAS and AOM (*Anopheles gambiae*
348 1000 Genomes Consortium 2017). For those populations with neutrally evolving IR genes, we
349 found no change in the recombination rates within the same regions (See Figure 4). For those
350 populations with neutrally evolving IR genes, we found no change in the recombination rates
351 within the same regions. This is consistent with the recent selective sweeps containing little
352 variation and reduced signature of recombination. An alternative explanation, is that
353 recombination within IR genes may break apart favorable haplotypes and otherwise prove to be
354 maladaptive in terms of insecticide resistance, with natural selection preventing the spread of
355 recombinants in the population. Perhaps a comprehensive study involving the differences in
356 fitness across recombinant and parental chromosomes when exposed to varying levels of
357 insecticides could help us shed some light between the role that recombination and selection play
358 in the evolution of this locus.

359

360 *Identifying recombination hotspots*

361 Because of the absence of recombination hotspots in *Drosophila* (Chan et al. 2012), a
362 main question in recombination studies is if this pattern is true across other insect systems,
363 including those that are closely related to *Drosophila*? Implementing LD methods, which have
364 not been applied to *Drosophila* systems in order to identify hotspots, we identified a total of 436
365 genomic regions that were consistent with recombination hotspots within *Anopheles* (Figure 5;
366 Supplementary Table S5). Majority of recombination hotspots were unique to a specific
367 population, however, only 9.6% (40 hotspots) were shared between two or more populations.

368 For all population pairwise comparisons that shared a hotspot, all but one had empirical
369 overlaps that were larger than the overlaps generate from stochastic hotspot shuffling
370 (Supplementary Table S6). Moreover, for the 25 comparisons where the empirical overlap was
371 larger than the simulated overlaps, Fisher's Exact test showed that 19 comparisons were
372 statistically significant (p -value < 0.05), suggesting a sharing in recombination hotspots more so

373 than what is expected under a stochastic assumption (Supplementary Table S6). It should be
374 noted that most of the significant overlaps were within populations of *A. gambiae*, however there
375 were few significant overlaps between *A. gambiae* and *A. coluzzii*, and none within *A. coluzzii*.
376 This is a very interesting finding because it demonstrates the potential conservation of few
377 recombination hotspots within the *Anopheles* genome, which may be serving a functionally
378 important role in the generation of novel variation within the genome.

379

380 *Consequences of recombination hotspots on population diversity, and divergence and*
381 *implications for the identification of adaptive variation*

382 Within autosomal chromosomes, we identified significant differences between nucleotide
383 diversity within recombination hotspots and average genomic levels. Specifically, nucleotide
384 diversity was significantly higher in recombination hotspots for all populations with the
385 exception of KES (Figure 6). Because of the recent, and severe, bottleneck for *Anopheles* within
386 Kenya, it is not surprising that there were no significant differences in nucleotide diversity.
387 Similarly, for the X chromosome, all nucleotide diversity comparisons also proved to be
388 significant (Figure 6). Identifying significant increase in nucleotide diversity within
389 recombination hotspot is an interesting finding because it suggests an evolutionary role for
390 recombination in the maintenance of genetic variation. More specifically, the presence of
391 recombination hotspots could be acting as protected reservoirs for genetic variation by reducing
392 the effects of linked selection.

393 Much like nucleotide diversity, the number of per-site differences (dxy) between
394 populations followed similar patterns along the autosomal chromosomes, where dxy within
395 recombination hotspots were significantly larger than background estimates for 66 of 72 (91.6%)
396 comparisons, suggesting an increase in population divergence within these regions (Figure 7A).
397 The only population comparisons that did not show a significant increase in dxy within hotspots
398 were within AOM and KES. These two populations have smaller effective population sizes and
399 could be a contributing factor to why these populations experienced lower overall changes in
400 dxy. Similarly, along the X chromosome, we identified significant increases in dxy within
401 recombination hotspots for 43 of 46 (93.4%) comparisons, suggesting that population divergence
402 is significantly larger within these regions (Supplementary Figure S10). Much like the
403 autosomes, two comparisons within AOM and one within GAS did not reflect significant

404 differences in d_{xy} . Effective population size for GAS was estimated to be on the same order of
405 magnitude as AOM, further supporting an increased role of genetic drift across the genome
406 resulting in higher background levels of genetic divergence. With the exception to nine
407 comparisons, the remaining comparisons support a mutagenic hypothesis where the mechanistic
408 properties of recombination not only influence variation within a population, but also drives
409 localized regions of the genome to have elevated levels of divergence. The other nine
410 comparisons support a linked selection hypothesis where recombination is maintaining genetic
411 diversity within populations, but elevated levels of divergence are larger where recombination
412 rates are lower. Importantly, these results suggest that recombination hotspots are maintaining,
413 and potentially generating, genetic diversity and contributing to absolute genetic divergence
414 within and between species of *Anopheles*.

415 The relationship between the recombination landscape and F_{ST} is more complex than
416 patterns of diversity and divergence (as measured by d_{xy}). For example, we showed that only
417 29.2% of our pairwise comparisons suggest that F_{ST} was lower within recombination hotspots
418 when compared to the average genome (Figure 7B). Despite the differences in background and
419 hotspot genetic differentiation being close to zero (suggesting little difference between the two
420 comparisons) 13 pairwise comparisons were significantly higher F_{ST} within recombination
421 hotspot compared to background estimates. These empirical data suggest that elevated
422 recombination rates may not play an essential role in reducing levels of genetic differentiation.
423 However, our results provide little insight on the efficacy of elevated F_{ST} within regions of low
424 recombination and warrants further investigation. Patterns of genetic differentiation are a
425 common statistic in population genetics used to infer differences in allele frequencies and
426 sometime selection. Here we provide evidence that suggests that regions of the genome with
427 increased recombination rates can retain genetic variation and suppress patterns of genetic
428 differentiation between populations. However, the data also suggest that elevated recombination
429 rates can still lead to patterns of increased genetic differentiation. Overall, we find that the
430 recombination landscape can have significant impacts on patterns of genetic diversity,
431 divergence, but potentially a smaller impact on patterns of F_{ST} . We suggest that future studies
432 involved with identifying patterns of genetic diversity and divergence across populations also
433 incorporate the recombination landscape, to better infer the patterns of genetic variation.

434 **Conclusion**

435 In this study, we detailed the fine-scale recombination maps and hotspot locations across
436 nine populations containing two species of *Anopheles*; *A. gambiae* and *A. coluzzii*. We identified
437 patterns of recombination rates that are similar to other systems that show large rate variation.
438 We also show that changes in the landscape of recombination maps strongly depends on location
439 along the chromosome, chromosomal inversions, and selective sweeps within IR genes. The
440 changes in recombination rates along the genome are even more prominent near the centromeric
441 regions of the X chromosome where recombination rates are minimal and may be responsible for
442 maintaining species boundaries between *A. gambiae* and *A. coluzzii*. The proportion of unique
443 and shared hotspots support the findings of other studies suggesting that there is a quick
444 evolutionary turnover of hotspots and recombination rates; however, few hotspot locations may
445 be more conserved, potentially serving important biological functions within the genome. Lastly,
446 we show evidence that recombination hotspots have the potential to increase both nucleotide
447 diversity within populations and the number of nucleotide differences between populations. This
448 was also true for patterns of relative genetic differentiation; however, these results were less
449 consistent and show that elevated recombination can also reduce levels of genetic differentiation.
450 Overall, our results provide significant insight into the evolutionary role of recombination across
451 populations of *Anopheles*, which can be implemented for further understanding evolutionary
452 trajectory of a prominent biological vector.

453

454 **Materials and Methods**

455 *Sample and Sequence data*

456 For this study we used full genome data from previously published work that was aimed
457 at characterizing the genomic structure, patterns of selection, migration, and estimates of
458 effective population size (N_e) for *A. gambiae* and *A. coluzzii* (*Anopheles gambiae* 1000 Genomes
459 Consortium 2017). Specifically, sequencing was performed on an Illumina HiSeq 2000 platform
460 at the Wellcome Trust Sanger Institute. All sequence reads were aligned to the AgamP3
461 reference genome (Sharakhova et al. 2007) using bwa (Li and Durbin 2009) and Single
462 Nucleotide Polymorphisms (SNPs) were discovered using GATK under their best practices
463 protocol (DePristo et al. 2011; Van der Auwera, Geraldine A et al. 2013). For more information
464 pertaining to the sequence data used and sample collection see (*Anopheles gambiae* 1000
465 Genomes Consortium 2017). We used a total of 441 individuals (882 genomes) sampled over

466 nine populations, eight countries, and two species of *Anopheles*. Sample sizes ranged from 88 to
467 120 genomes per population (Figure 1; Supplementary Table S7). Prior to estimating
468 recombination rates, we filtered bi-allelic phased SNP data to account for a minor allele
469 frequency of at least three individuals, missing data, and fixed sites.

470 *Estimating fine-scale recombination rates*

471 To estimate fine-scale recombination rates, we used LDhat which is a package of
472 programs that estimates population recombination rates by implementing the composite
473 likelihood method of Hudson et al. (2001) (Hudson 2001; McVean, Gil and Auton 2007).
474 Because obtaining fine-scale estimates of recombination across a genome for a large number of
475 samples is computationally exhaustive, we decomposed all genomic data in two different
476 components to help reduce computational resources. First, LDhat was ran in parallel within each
477 population allowing the split of individuals and reduced memory usage. Second, instead of
478 estimating recombination rates for the entire genome at one time, each chromosome arm was
479 divided into sliding windows, further reducing the amount of memory required for each
480 population/chromosome. Each window contained 2,000 SNPs including a 500 SNP overlap
481 between adjacent windows and ran for 100,000,000 iterations with a burn-in period of
482 50,000,000 iterations. LDhat sampled every 10,000 iterations, providing 10,000 sampled
483 recombination rates per site under the INTERVAL function. After obtaining recombination rates
484 the STATS function was implemented in LDhat which calculates the mean, median, upper 95%,
485 and lower 95% recombination rate for each SNP. Windows were then trimmed of their overlaps
486 and aligned to their respected positions along the chromosome. This pipeline was performed on
487 both autosomal and X chromosomes, however, within the *Anopheles* system males are the
488 heterogametic sex. We, therefore, included only females for building the recombination maps for
489 the X chromosome. In total we used 417 females for estimating recombination rates along the X
490 chromosome (see Supplementary Table S7).

491 *Detecting recombination hotspots*

492 In order to detect recombination hotspots, we used LDhot, which is a program designed
493 to identify regions of the genome where recombination rates are significantly larger than
494 background levels (McVean, G. A. et al. 2004; Myers et al. 2005). Specifically, LDhot defines a
495 Poisson distribution with a given lambda coupled with a sliding window approach to identify
496 regions of the genome that have recombination rates explained by a different lambda and Poisson

497 distribution (McVean, G. A. et al. 2004). If localized patterns of recombination rates are better
498 explained by a different Poisson distribution, then LDhot flags the window as a recombination
499 hotspot. Poisson distributions were defined based off 1,000 simulations. Only sliding windows
500 with a p-value ≤ 0.01 were retained and then merged together using BedTools MERGE (Quinlan
501 and Hall 2010) to identify the size and location of the recombination hotspot.

502 However, because LDhot is known to lose power when population-scaled recombination
503 rates are low ($\rho < 5$) (Johnston and Cutler 2012; Wall and Stevison 2016) we retained hotspots
504 that only showed large deviations in recombination rates from the surrounding background.
505 Specifically, we compared recombination rates within hotspots to the recombination rates for the
506 surrounding 50 kb background (the same size flanks that LDhot uses to detect hotspots). We
507 used the difference in recombination rates between hotspots and the adjacent 50kb background to
508 rank the intensity of recombination hotspots along a given background (supplemental Figures
509 S11-S16). We retained only the top 5% of hotspots with the largest difference in rho when
510 compared to the surrounding background rates for each population. Though this method is
511 conserved, we have a greater confidence in our filtered hotspots as being true hotspots and not an
512 artifact of demographic and computational limitations.

513 After hotspots were identified, we used UpSet in R to view the union and intersection of
514 hotspots for all population pairwise comparisons (Conway et al. 2017). For all pairwise
515 comparisons that shared at least one recombination hotspot, we wanted to test if these hotspots
516 were more likely shared due to stochastic processes (random hotspot location within the genome)
517 or if there was potential evidence of hotspot conservation between populations (shared more than
518 random expectations). Here, the degree of overlap was determined by how many base pairs were
519 shared within a given recombination hotspot. Specifically, to perform stochastic shuffling of
520 recombination hotspots, we used BedTools SHUFFLE (Quinlan and Hall 2010) to randomly
521 shuffle known hotspot locations for each chromosome arm. Recombination hotspots were
522 shuffled for 10,000 iterations where the degree of overlap (in bp) was calculated for each
523 iteration. The differences between average simulated and empirical overlaps were then
524 calculated; a positive net difference suggests that empirical overlaps were larger than the
525 shuffled overlaps. To complement stochastic simulations, we performed a Fishers Exact Test,
526 using BedTools, to identify significant deviations from our null hypothesis of stochastic hotspot
527 sharing across all empirical overlaps. We defined a significant overlap with a p-value < 0.05 .

528 *Consequences of elevated recombination on diversity and divergence*

529 We compared levels of diversity within and between populations to understand potential
530 consequences that recombination has on genetic diversity and divergence. For this, levels of
531 nucleotide diversity (π) were calculated within recombination hotspots and compared to the per
532 site genome-wide estimates of π for each population. Because of prominent reproductive
533 isolation along the X chromosome, this analysis was performed on the autosomal and X
534 chromosome separately. To determine significant differences for π between hotspots and the
535 genomic background a Wilcoxon Rank test was performed; a non-parametric statistical test that
536 does not assume the data are normally distributed, which is the case for estimates of π (p-value \leq
537 0.05). To test for elevated levels of divergence within recombination hotspots we calculated two
538 different metrics of genetic divergence: absolute divergence (dxy) and relative divergence (F_{ST}).
539 Specifically, for dxy, we calculated the number of per site differences for each recombination
540 hotspot for each population pairwise comparison and compared dxy estimates across the average
541 genomic background level.. A Wilcoxon Rank Test was performed for each dxy pairwise
542 comparison was used to identify significant deviations (p-value \leq 0.05).

543 We used a randomization approximation to test the significance of the difference in F_{ST}
544 between regions with HS of recombination and Background. For this, each the F_{ST} values for
545 each region were labeled as HS or Background, and a shuffling procedure of the labels was used
546 to generate 1000 pseudo-replicates. The average difference in F_{ST} between relabeled HS and
547 Background regions was estimated for each pseudo-replicate and a distribution was generated.
548 An empirical p-value (p-value \leq 0.05) for the significance of the observed difference was
549 determined by looking at the proportion of pseudo-replicates with values larger than 95% of the
550 observed difference.

551

552

553

554

555

556

557

References

- 558 Aboagye-Antwi F, Alhafez N, Weedall GD, Brothwood J, Kandola S, Paton D, Fofana A,
559 Olohan L, Betancourth MP, Ekechukwu NE et al. 2015. Experimental swap of *Anopheles*
560 *gambiae*'s assortative mating preferences demonstrates key role of X-chromosome divergence
561 island in incipient sympatric speciation. *PLoS Genet.* 11:e1005141.
- 562 Andolfatto P, Przeworski M. 2001. Regions of lower crossing over harbor more rare variants in
563 African populations of *Drosophila melanogaster*. *Genetics.* 158:657-665.
- 564 *Anopheles gambiae* 1000 Genomes Consortium. 2017. Genetic diversity of the African malaria
565 vector *Anopheles gambiae*. *Nature.* 552:96.
- 566 Beeson SK, Mickelson JR, McCue ME. 2019. Exploration of fine-scale recombination rate
567 variation in the domestic horse. *Genome Res.* 29:1744-1752.
- 568 Begun DJ, Aquadro CF. 1992. Levels of naturally occurring DNA polymorphism correlate with
569 recombination rates in *D. melanogaster*. *Nature.* 356:519.
- 570 Brown TC, Jiricny J. 1987. A specific mismatch repair event protects mammalian cells from loss
571 of 5-methylcytosine. *Cell.* 50:945-950.
- 572 Brown TC, Zbinden I, Cerutti PA, Jiricny J. 1989. Modified SV40 for analysis of mismatch
573 repair in simian and human cells. *Mutation Research/Reviews in Genetic Toxicology.* 220:115-
574 123.
- 575 Chan AH, Jenkins PA, Song YS. 2012. Genome-wide fine-scale recombination rate variation in
576 *Drosophila melanogaster*. *PLoS Genet.* 8:e1003090.
- 577 Cheng C, Tan JC, Hahn MW, Besansky NJ. 2018. Systems genetic analysis of inversion
578 polymorphisms in the malaria mosquito *Anopheles gambiae*. *Proc Natl Acad Sci U S A.*
579 115:7005-7014.
- 580 Cheng C, White BJ, Kamdem C, Mockaitis K, Costantini C, Hahn MW, Besansky NJ. 2012.
581 Ecological genomics of *Anopheles gambiae* along a latitudinal cline: a population-resequencing
582 approach. *Genetics.* 190:1417-1432.
- 583 Choi K, Reinhard C, Serra H, Ziolkowski PA, Underwood CJ, Zhao X, Hardcastle TJ, Yelina
584 NE, Griffin C, Jackson M. 2016. Recombination rate heterogeneity within *Arabidopsis* disease
585 resistance genes. *PLoS Genetics.* 12:e1006179.
- 586 Clay K, Kover PX. 1996. The Red Queen hypothesis and plant/pathogen interactions. *Annu Rev*
587 *Phytopathol.* 34:29-50.

- 588 Coetzee M, Hunt RH, Wilkerson R, Della Torre A, Coulibaly MB, Besansky NJ. 2013.
589 *Anopheles coluzzii* and *Anopheles amharicus*, new members of the *Anopheles gambiae*
590 complex. *Zootaxa*. 3619:246-274.
- 591 Coluzzi M, Sabatini A, Petrarca V, Di Deco M. 1979. Chromosomal differentiation and
592 adaptation to human environments in the *Anopheles gambiae* complex. *Trans R Soc Trop Med*
593 *Hyg.* 73:483-497.
- 594 Conway JR, Lex A, Gehlenborg N. 2017. UpSetR: an R package for the visualization of
595 intersecting sets and their properties. *Bioinformatics*. 33:2938-2940.
- 596 Cutter AD, Payseur BA. 2003. Selection at linked sites in the partial selfer *Caenorhabditis*
597 *elegans*. *Mol Biol Evol.* 20:665-673.
- 598 DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Del
599 Angel G, Rivas MA, Hanna M. 2011. A framework for variation discovery and genotyping using
600 next-generation DNA sequencing data. *Nat Genet.* 43:491.
- 601 Dreissig S, Mascher M, Heckmann S. 2019. Variation in recombination rate is shaped by
602 domestication and environmental conditions in barley. *Mol Biol Evol.* 36:2029-2039.
- 603 Feder JL, Nosil P. 2009. Chromosomal inversions and species differences: when are genes
604 affecting adaptive divergence and reproductive isolation expected to reside within inversions?
605 *Evolution: International Journal of Organic Evolution.* 63:3061-3075.
- 606 Felsenstein J. 1974. The evolutionary advantage of recombination. *Genetics.* 78:737-756.
- 607 Gabriel W, Lynch M, Bürger R. 1993. Muller's ratchet and mutational meltdowns. *Evolution.*
608 47:1744-1757.
- 609 Hartfield M, Keightley PD. 2012. Current hypotheses for the evolution of sex and
610 recombination. *Integrative Zoology.* 7:192-209.
- 611 Hellmann I, Ebersberger I, Ptak SE, Pääbo S, Przeworski M. 2003. A neutral explanation for the
612 correlation of diversity with recombination rates in humans. *The American Journal of Human*
613 *Genetics.* 72:1527-1535.
- 614 Hellmann I, Prüfer K, Ji H, Zody MC, Paabo S, Ptak SE. 2005. Why do human diversity levels
615 vary at a megabase scale? *Genome Res.* 15:1222-1231.
- 616 Hellsten U, Wright KM, Jenkins J, Shu S, Yuan Y, Wessler SR, Schmutz J, Willis JH, Rokhsar
617 DS. 2013. Fine-scale variation in meiotic recombination in *Mimulus* inferred from population
618 shotgun sequencing. *Proc Natl Acad Sci U S A.* 110:19478-19482.

- 619 Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker P, Clark
620 AG, Ribeiro JM, Wides R et al. 2002. The genome sequence of the malaria mosquito *Anopheles*
621 *gambiae*. *Science*. 298:129-149.
- 622 Hudson RR. 2001. Two-locus sampling distributions and their application. *Genetics*. 159:1805-
623 1817.
- 624 Johnston HR, Cutler DJ. 2012. Population demographic history can cause the appearance of
625 recombination hotspots. *The American Journal of Human Genetics*. 90:774-783.
- 626 Keightley PD, Otto SP. 2006. Interference among deleterious mutations favours sex and
627 recombination in finite populations. *Nature*. 443:89-92.
- 628 Kim Y, Nielsen R. 2004. Linkage disequilibrium as a signature of selective sweeps. *Genetics*.
629 167:1513-1524.
- 630 Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B,
631 Sigurdardottir S, Barnard J, Hallbeck B, Masson G. 2002. A high-resolution recombination map
632 of the human genome. *Nat Genet*. 31:241.
- 633 Korunes KL, Noor MA. 2017. Gene conversion and linkage: effects on genome evolution and
634 speciation. *Mol Ecol*. 26:351-364.
- 635 Kumar R, Duhamel M, Coutant E, Ben-Nahia E, Mercier R. 2019. Antagonism between BRCA2
636 and FIGL1 regulates homologous recombination. *Nucleic Acids Res*. 47:5170-5180.
- 637 Lanzaro GC, Lee Y. 2013. Speciation in *Anopheles gambiae*—The distribution of genetic
638 polymorphism and patterns of reproductive isolation among natural populations. In: Anonymous
639 *Anopheles mosquitoes-New insights into malaria vectors*. ed Manguin S (Intech, Rijeka,
640 Croatia).
- 641 Lehmann T, Hawley W, Grebert H, Danga M, Atieli F, Collins F. 1999. The Rift Valley complex
642 as a barrier to gene flow for *Anopheles gambiae* in Kenya. *Journal of Heredity*, 90(6):613-621.
- 643 Lercher MJ, Hurst LD. 2002. Human SNP variability and mutation rate are higher in regions of
644 high recombination. *Trends in Genetics*. 18:337-340.
- 645 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform.
646 *Bioinformatics*. 25:1754-1760.
- 647 McVean G, Auton A. 2007. LDhat 2.1: A package for the population genetic analysis of
648 recombination. Oxford, United Kingdom: Department of Statistics. Accessed
649 at: <http://www.stats.ox.ac.uk/~mcvean/LDhat/manual.pdf>
650 . Department of Statistics, Oxford, OX1 3TG, UK.

- 651 McVean GA, Charlesworth B. 2000. The effects of Hill-Robertson interference between weakly
652 selected mutations on patterns of molecular evolution and variation. *Genetics*. 155:929-944.
- 653 McVean GA, Myers SR, Hunt S, Deloukas P, Bentley DR, Donnelly P. 2004. The fine-scale
654 structure of recombination rate variation in the human genome. *Science*. 304:581-584.
- 655 Myers S, Bottolo L, Freeman C, McVean G, Donnelly P. 2005. A fine-scale map of
656 recombination rates and hotspots across the human genome. *Science*. 310:321-324.
- 657 Nachman MW. 2002. Variation in recombination rate across the genome: evidence and
658 implications. *Curr Opin Genet Dev*. 12:657-663.
- 659 Nachman MW. 2001. Single nucleotide polymorphisms and recombination rate in humans.
660 *TRENDS in Genetics*. 17:481-485.
- 661 Nachman MW, Bauer VL, Crowell SL, Aquadro CF. 1998. DNA variability and recombination
662 rates at X-linked loci in humans. *Genetics*. 150:1133-1141.
- 663 Nambiar M, Smith GR. 2016. Repression of harmful meiotic recombination in centromeric
664 regions. *Genetics*. 54:188-197.
- 665 Otto SP, Barton NH. 2001. Selection for recombination in small populations. *Evolution*.
666 55:1921-1931.
- 667 Papavasiliou FN, Schatz DG. 2000. Cell-cycle-regulated DNA double-strand breaks in somatic
668 hypermutation of immunoglobulin genes. *Nature*. 408:216.
- 669 Pennings PS, Hermisson J. 2006. Soft sweeps III: the signature of positive selection from
670 recurrent mutation. *PLoS Genet*. 2:e186.
- 671 Posada D, Crandall KA, Holmes EC. 2002. Recombination in evolutionary genomics. *Annu Rev*
672 *Genet*. 36:75-97.
- 673 Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic
674 features. *Bioinformatics*. 26:841-842.
- 675 Ritz KR, Noor MA, Singh ND. 2017. Variation in recombination rate: adaptive or not? *Trends in*
676 *Genetics*. 33:364-374.
- 677 Roesti M, Moser D, Berner D. 2013. Recombination in the threespine stickleback genome—
678 patterns and consequences. *Mol Ecol*. 22:3014-3027.
- 679 San Filippo J, Sung P, Klein H. 2008. Mechanism of eukaryotic homologous recombination.
680 *Annu Rev Biochem*. 77:229-257.

- 681 Schridder DR, Mendes FK, Hahn MW, Kern AD. 2015. Soft shoulders ahead: spurious signatures
682 of soft and partial selective sweeps result from linked hard sweeps. *Genetics*. 200:267-284.
- 683 Schwarzkopf EJ, Motamayor JC, Cornejo OE. 2020. Genetic differentiation and intrinsic
684 genomic features explain variation in recombination hotspots among cocoa tree populations.
685 *BMC Genomics*. 21:1-16.
- 686 Sharakhova MV, Hammond MP, Lobo NF, Krzywinski J, Unger MF, Hillenmeyer ME,
687 Bruggner RV, Birney E, Collins FH. 2007. Update of the *Anopheles gambiae* PEST genome
688 assembly. *Genome Biol*. 8:R5.
- 689 Sibley LD, Ajioka JW. 2008. Population structure of *Toxoplasma gondii*: clonal expansion
690 driven by infrequent recombination and selective sweeps. *Annu Rev Microbiol*. 62:329-351.
- 691 Smagulova F, Brick K, Pu Y, Camerini-Otero RD, Petukhova GV. 2016. The evolutionary
692 turnover of recombination hot spots contributes to speciation in mice. *Genes Dev*. 30:266-280.
- 693 Smukowski C, Noor M. 2011. Recombination rate variation in closely related species. *Heredity*.
694 107:496.
- 695 Spence JP, Song YS. 2019. Inference and analysis of population-specific fine-scale
696 recombination maps across 26 diverse human populations. *Science Advances*. 5:eaaw9206.
- 697 Spencer CC, Deloukas P, Hunt S, Mullikin J, Myers S, Silverman B, Donnelly P, Bentley D,
698 McVean G. 2006. The influence of recombination on human genetic diversity. *PLoS Genetics*.
699 2:e148.
- 700 Stapley J, Feulner PG, Johnston SE, Santure AW, Smadja CM. 2017. Variation in recombination
701 frequency and distribution across eukaryotes: patterns and processes. *Philosophical Transactions
702 of the Royal Society B: Biological Sciences*. 372:20160455.
- 703 Van der Auwera, Geraldine A, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy Moonshine
704 A, Jordan T, Shakir K, Roazen D, Thibault J. 2013. From FastQ data to high confidence variant
705 calls: the genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*.
706 43:11.10. 1-11.10. 33.
- 707 Wall JD, Stevison LS. 2016. Detecting Recombination Hotspots from Patterns of Linkage
708 Disequilibrium. *G3 (Bethesda)*. 6:2265-2271.
- 709 White BJ, Cheng C, Simard F, Costantini C, Besansky NJ. 2010. Genetic association of
710 physically unlinked islands of genomic divergence in incipient species of *Anopheles gambiae*.
711 *Mol Ecol*. 19:925-939.
- 712 Wright SI, Foxe JP, DeRose-Wilson L, Kawabe A, Looseley M, Gaut BS, Charlesworth D. 2006.
713 Testing for effects of recombination rate on nucleotide diversity in natural populations of
714 *Arabidopsis lyrata*. *Genetics*. 174:1421-1430.

715 Yang L, Ouyang H, Fang Z, Zhu W, Wu E, Luo G, Shang L, Zhan J. 2018. Evidence for
716 intragenic recombination and selective sweep in an effector gene of *Phytophthora infestans*.
717 *Evolutionary Applications*. 11:1342-1353.

718

719

720

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746 Table 1. The average population-scaled (ρ) recombination rate for the entire genome, each
747 autosomal chromosome arm, and the X chromosome.

Population	ρ (avg.)	2L	2R	3L	3R	X	Variance
AOM	7.8	6.5	7.4	9.9	9.2	6.1	39.7765
BFM	111	107.1	60.7	138	131.9	117.5	6910.2
GWA	64.6	20.4	16.1	95.4	89	102.2	4965.31
BFS	119.8	70.5	107.8	151.2	145.3	124.1	10373.1
CMS	129.3	58.3	105.8	158.9	159.5	164	12530.1
GAS	4.4	4.4	4.5	5.3	5.6	2.1	14.2874
GNS	110.9	54.4	93.8	125.5	124	157	7618.57
KES	0.5	0.3	0.3	0.3	0.2	1.3	1.69677
UGS	89.9	36	85.6	78.6	113.3	135.8	8032.67
Average	70.9	39.8	53.6	84.8	86.5	90	5609.52

748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769

770 Table 2. The average per-site (r) recombination rate for the entire genome, each autosomal
771 chromosome arm, and the X chromosome.

Population	r (avg.)	2L (r)	2R (r)	3L (r)	3R (r)	X (r)	cM/Mb
AOM	1.95E-09	1.62E-09	1.85E-09	2.48E-09	2.29E-09	1.51E-09	0.19509262
BFM	2.78E-09	2.68E-09	1.52E-09	3.45E-09	3.30E-09	2.94E-09	0.27760314
GWA	3.26E-09	5.11E-09	4.02E-09	2.38E-09	2.23E-09	2.55E-09	0.161564
BFS	2.99E-09	1.76E-09	2.70E-09	3.78E-09	3.63E-09	3.10E-09	0.29947138
CMS	3.23E-09	1.46E-09	2.65E-09	3.97E-09	3.99E-09	4.10E-09	0.32330976
GAS	1.10E-09	1.09E-09	1.13E-09	1.33E-09	1.41E-09	5.34E-10	0.1099353
GNS	2.77E-09	1.36E-09	2.35E-09	3.14E-09	3.10E-09	3.92E-09	0.2773442
KES	1.23E-08	7.38E-09	8.49E-09	6.46E-09	5.87E-09	3.33E-08	1.22950744
UGS	3.87E-09	9.00E-09	2.14E-09	1.97E-09	2.83E-09	3.40E-09	0.22465757
Average	3.81E-09	3.50E-09	2.98E-09	3.22E-09	3.18E-09	6.15E-09	0.34427616

772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792

793 Table 3. Averaged population-scaled recombination rate for the seven different inversions across
794 sampled populations of *Anopheles*. The shaded boxes represents populations where there are at
795 least three heterozygous genotypes for the 2Rb or the 2La inversion.

Population	ρ (avg.)	2Rb(avg.)	2La(avg.)
AOM	8.22	11.17	7.06
BFM	104.54	20.41	103.77
GWA	50.61	15.58	8.29
BFS	117.71	146.71	32.79
CMS	118.96	87.55	15.04
GAS	4.95	5.45	4.22
GNS	98.78	73.21	7.84
KES	0.29	0.30	0.27
UGS	80.60	56.32	11.40

796

797

798

799

800

801

802

803

804

805

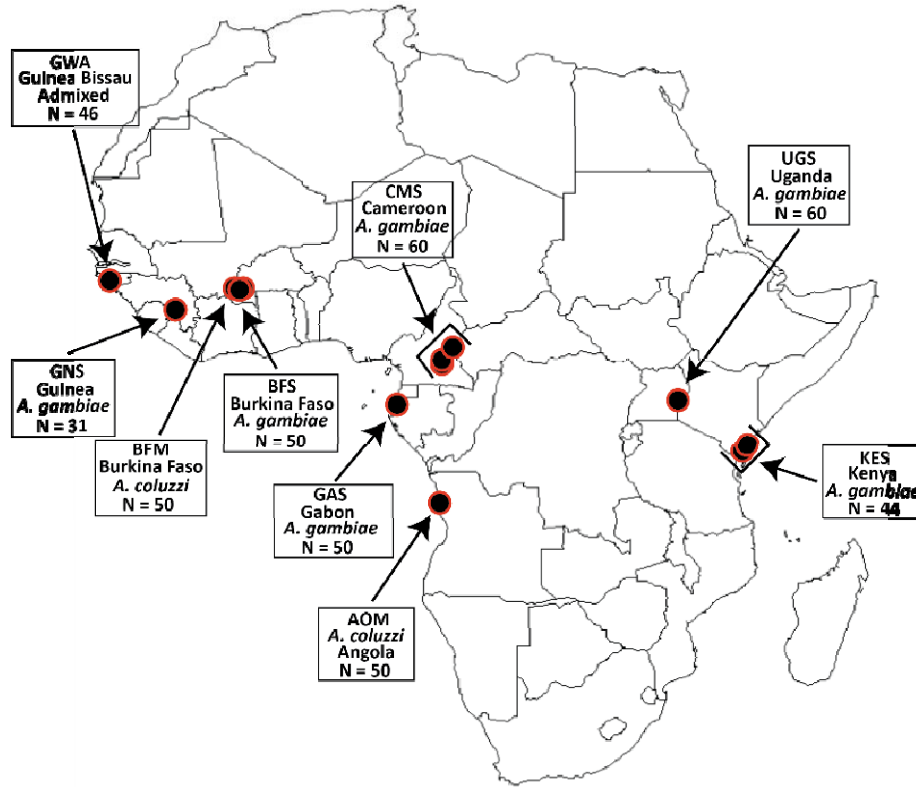
806

807

808

809

810



811

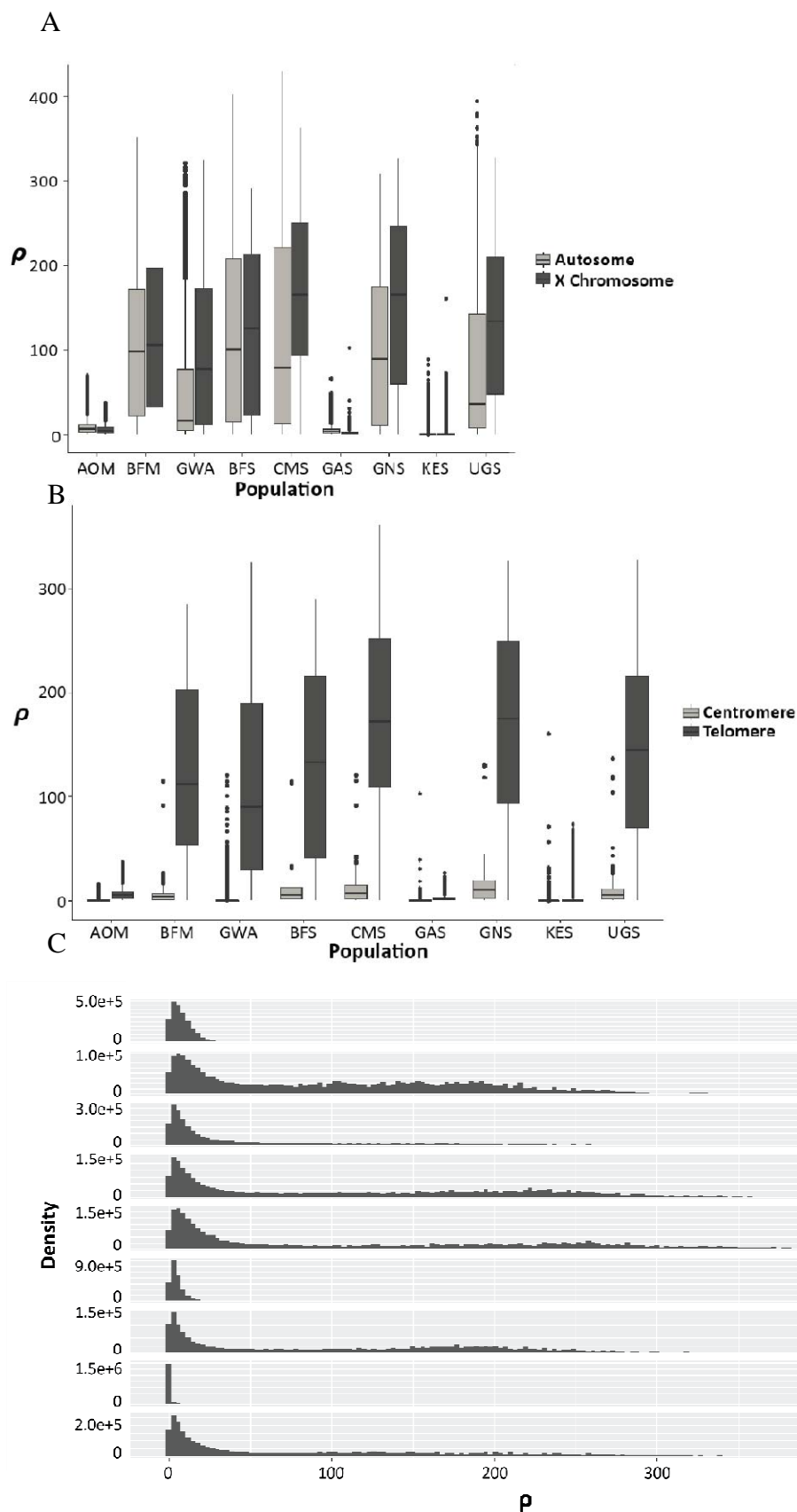
812 Figure 1. Map of Africa depicting the locations and sample sizes of each populations used for
813 estimating recombination rates. Note that for several populations there multiple sampling
814 locations.

815

816

817

818



819

820

821

822

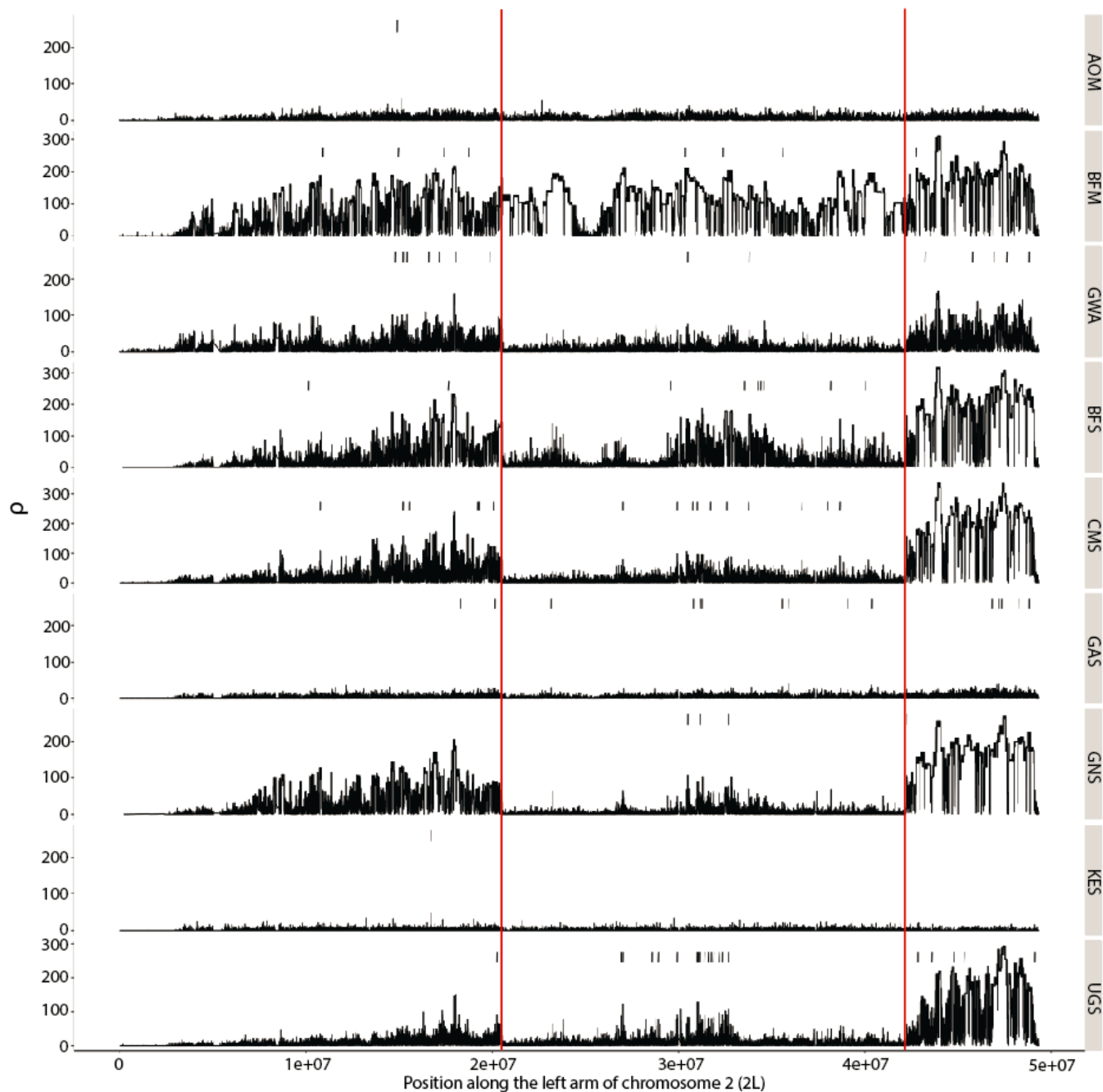
Figure 2. Top panel (A) depicts a boxplot showing the degree of variation in population-scaled recombination rates within and across populations for the autosomal genome and the X

823 chromosome. The middle panel (B) shows the difference in population-scaled recombination
824 rates between the first 18Mb (Telomere) and the last 6Mb (Centromere) of the X chromosome
825 for each population. The bottom panel (C) represents individual histograms of the genome-wide
826 (autosomal and the X chromosome) recombination rates for each population. Here, ρ is the
827 population-scaled recombination rate which is the product of $4N_e r$, where N_e is the effective
828 population size and r is the per-site recombination rate.
829

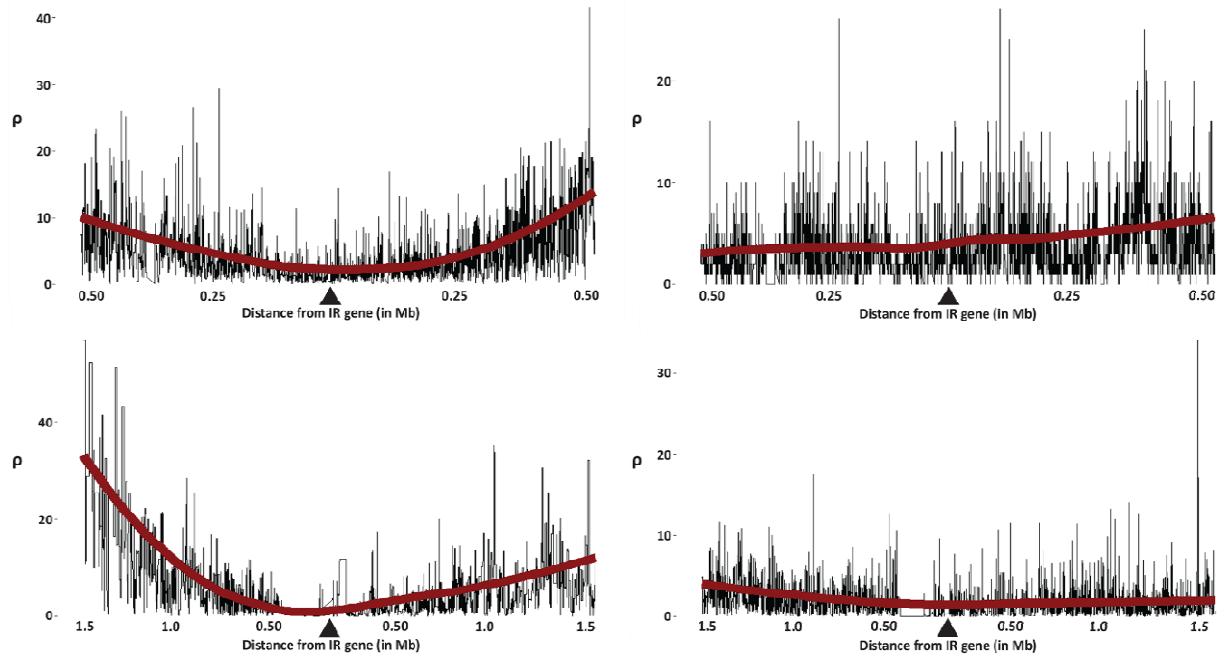
830

831

832

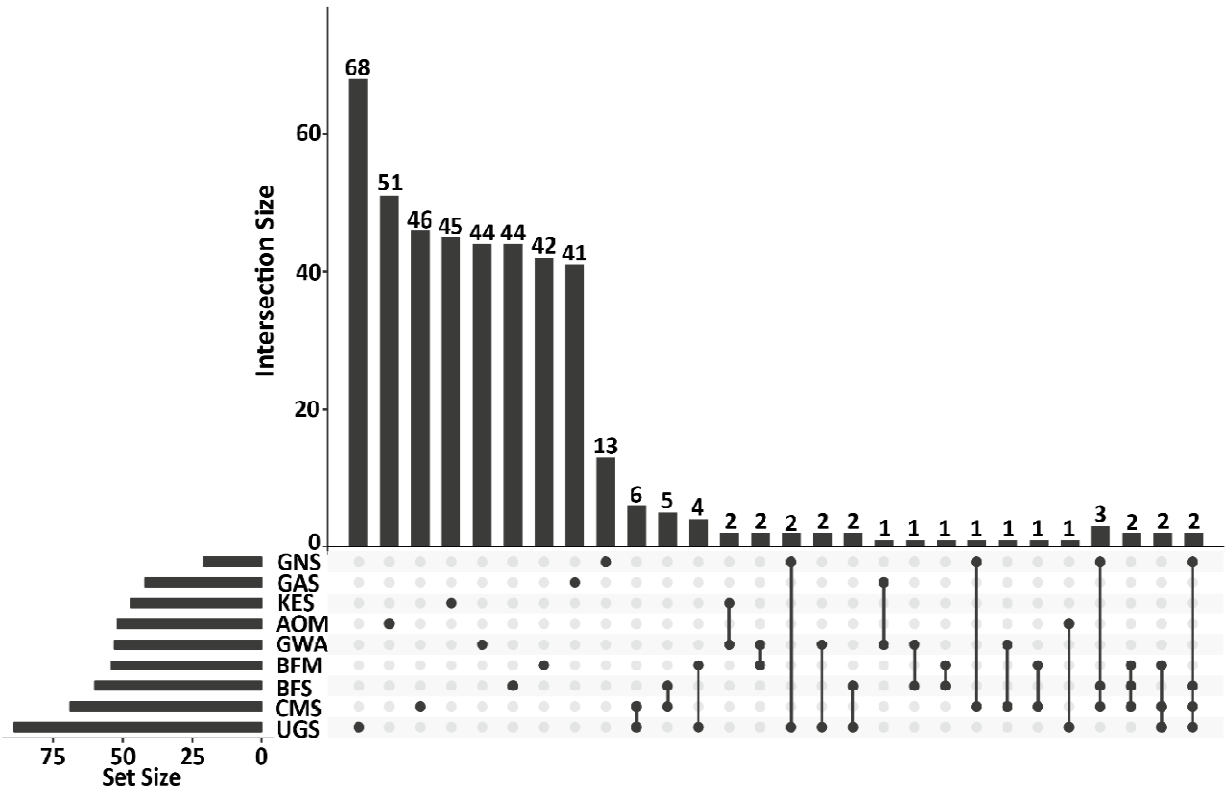


833
834 Figure 3. The estimated population-scaled recombination map of the left arm of the second
835 chromosome for all populations. The small black vertical lines at the top of each map represents
836 the location of a recombination hotspot identified from LDhot. The red lines denote the location
837 of the 2La chromosomal inversions.
838
839



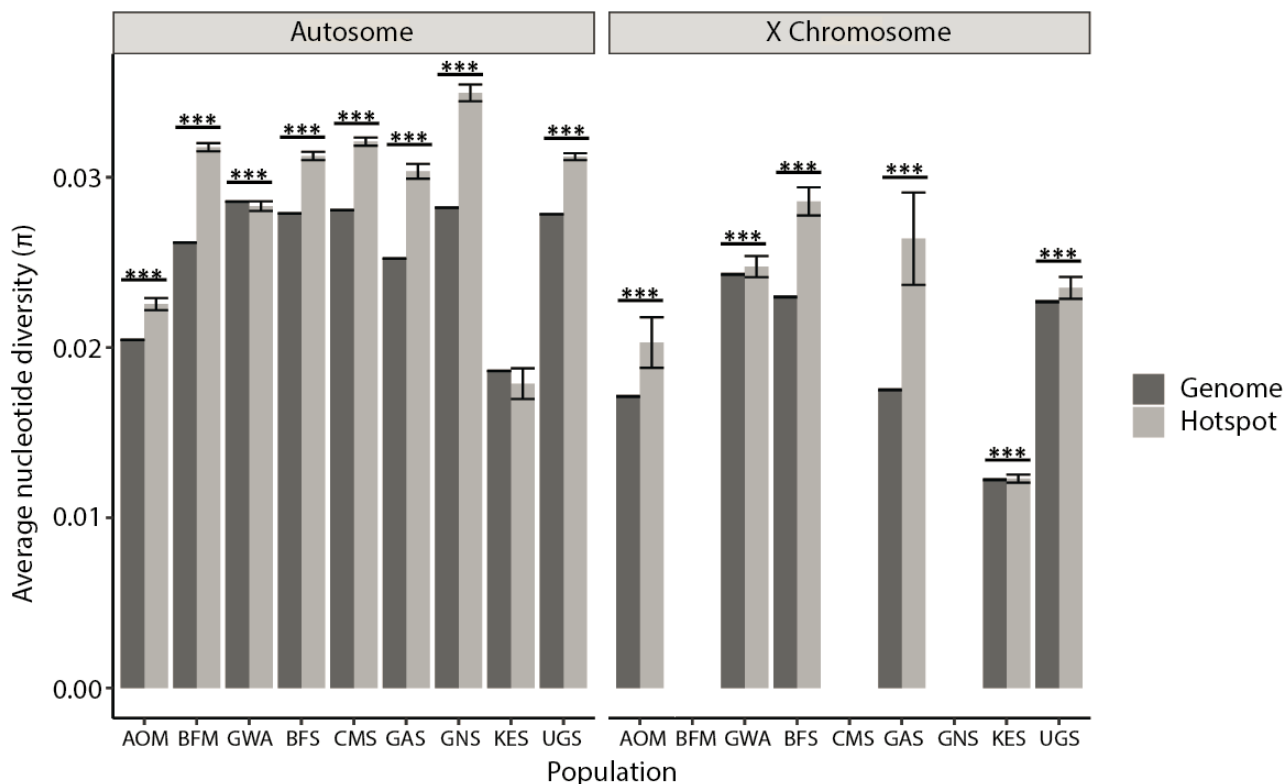
840
841 Figure 4. Population-scaled recombination rate for a genomic window containing two known IR
842 genes that are under selection in some populations but are evolving neutrally in others. The red
843 line denotes a smoothed average of recombination rates for the given region. Top panel shows
844 the changes in recombination rate in a 1 Mb window containing the Cyp6p gene. Note that this
845 gene shows evidence of selection in UGS (top left) but not in GAS (top right). The lower panel
846 shows the changes in recombination rate in a 3 Mb window containing the Cyp9k1 gene.
847 Previous work has found evidence for selection in BFM but weakly selected (or nearly neutral)
848 in AOM. Here, the black triangle denotes the position of the given IR gene.

849
850
851
852
853
854
855
856
857
858
859

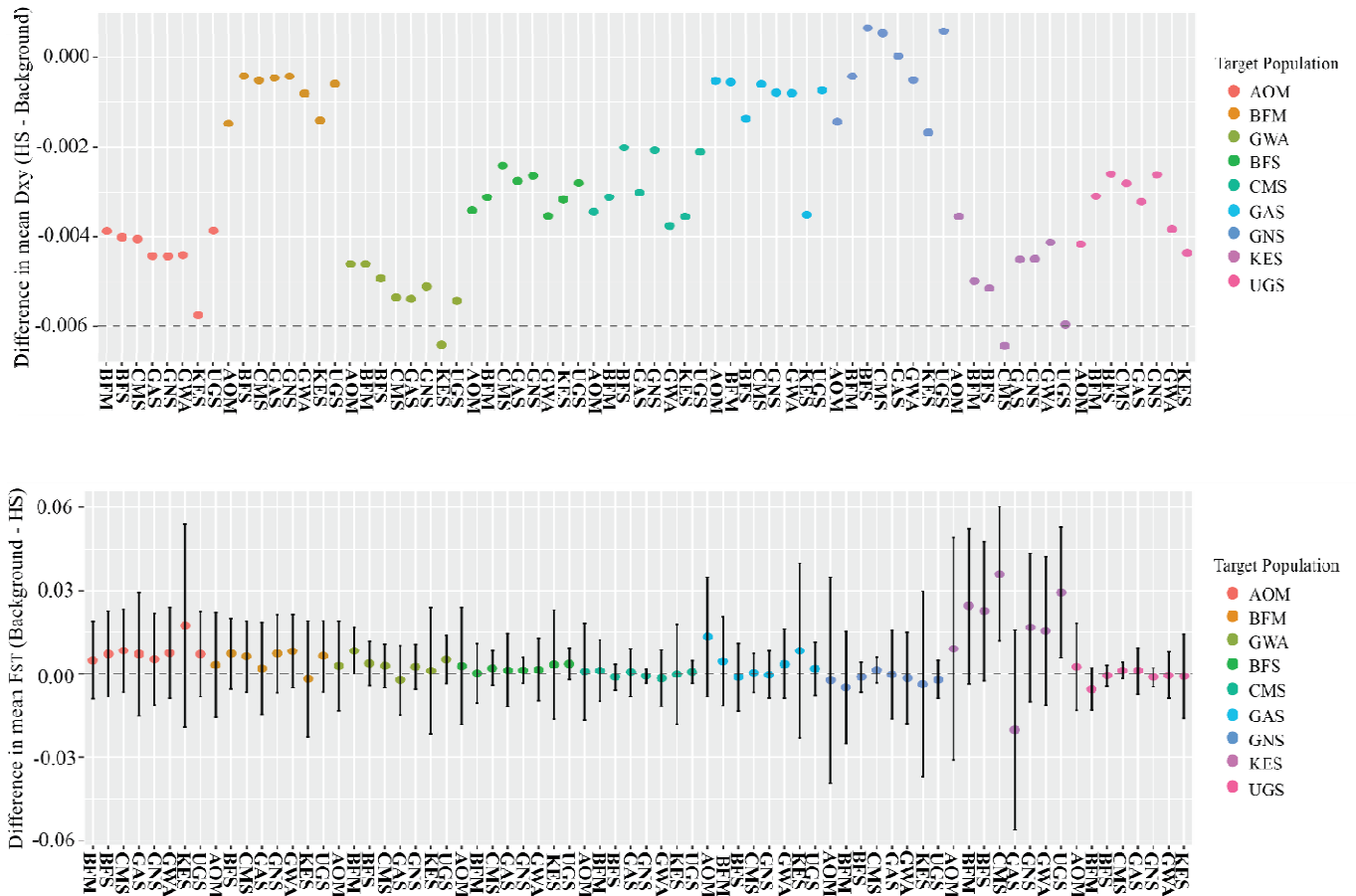


860
 861 Figure 5. An Upset analysis depicting the distribution of unique and shared recombination
 862 hotspots. Within this figure the set size represents the total number of hotspots found within each
 863 population, while the intersection size defines the number of hotspots that were identified in a
 864 single population (single dots) and the number of hotspots shared across multiple populations
 865 (connected dots). The numbers at the top of each bar define the number of hotspot found within
 866 that intersection.

867
 868
 869
 870
 871
 872
 873
 874



875
 876 Figure 6. Barplot comparing nucleotide diversity within recombination hotspots and genome-
 877 wide nucleotide diversity for the autosomes and X chromosomes. Vertical lines at the end of
 878 each bar represents the standard error for each measurement. Solid horizontal bars and asterisks
 879 denote significant differences between recombination hotspots and the overall genome. The
 880 asterisks above the bars denotes the level of significance (* = p-value \leq 0.05, ** = p-value \leq
 881 0.01, and *** = p-value \leq 0.001).
 882
 883
 884
 885
 886
 887
 888
 889
 890
 891
 892
 893
 894
 895
 896
 897
 898
 899



900
901

902
903
904
905
906
907
908
909
910

Figure 7. Barplots that show the ratio between dxy (top panel) and F_{ST} (bottom panel) within recombination hotspots compared to background estimates for autosomal chromosomes. For each plot, the target population is the population in which within recombination hotspots are being compared to the same genomic regions with the populations labeled along the X-axis. The dashed black line signifies no difference in divergence metrics (dxy or F_{ST}) between the genomic background and hotspots. The asterisks above represents significant pairwise comparisons.