Natural diversity of the malaria vector Anopheles gambiae — **Source link** ↗

Alistair Miles, Nicholas J. Harding, Giordano Bottà, Chris S Clarkson ...+57 more authors

Related papers:

- Natural diversity of the malaria vector Anopheles gambiae

- The Anopheles gambiae transcriptome - a turning point for malaria control.

- Anopheles gambiae genome: perspectives for malaria control.

- Malaria entomological profile in Tanzania from 1950 to 2010: a review of mosquito distribution, vectorial capacity and insecticide resistance.

- Rapid ageing and species identification of natural mosquitoes for malaria surveillance

Natural diversity of the malaria vector

Anopheles gambiae

The *Anopheles gambiae* 1000 Genomes Consortium[*]

The sustainability of malaria control in Africa is threatened by rising levels of insecticide resistance, and new tools to prevent malaria transmission are urgently needed. To gain a better understanding of the mosquito populations that transmit malaria, we sequenced the genomes of 765 wild specimens of *Anopheles gambiae* and *Anopheles coluzzii* sampled from 15 locations across Africa. The data reveal high levels of genetic diversity, with over 50 million single nucleotide polymorphisms across the 230 Mbp genome. We observe complex patterns of population structure and marked variations in local population size, some of which may be due at least in part to malaria control interventions. Insecticide resistance genes show strong signatures of recent selection associated with multiple independent mutations spreading over large geographical distances and between species. The genetic variability of natural populations substantially reduces the target space for novel gene-drive strategies for mosquito control. This large dataset provides a foundation for tracking the emergence and spread of insecticide resistance and developing new vector control tools.

Blood-sucking mosquitoes of the *Anopheles gambiae* species complex exert a heavy toll on human health, being the principal vectors of *Plasmodium falciparum* malaria in Africa. Increased use of insecticide-treated bed nets (ITNs) and other methods of vector control have led to substantial reductions in the burden of malaria in Africa over the past 15 years[1,2]. However, these gains could be reversed by insecticide resistance that is rapidly spreading across the continent[3,4] and by behavioural

---

[*] Lists of participants and their affiliations appear at the end of the paper

21   adaptations which cause mosquitoes to avoid contact with insecticides[5]. New insecticides are being

22   developed for use in public health[6,7] and there is growing support for gene drive technologies for

23   malaria vector control[8–10]. However, relatively little is known about natural genetic diversity of

24   *Anopheles* vector species, or the evolutionary and demographic processes that allow adaptive mutations

25   to emerge and spread through mosquito populations. This knowledge is needed to maximize the

26   efficacy and active lifespan of new insecticides and to design gene drive systems that work in the field.

27   The *Anopheles gambiae* 1000 Genomes Project[†] (Ag1000G) was established to discover natural genetic

28   variation within this species complex, and to provide a fundamental resource for applied research into

29   malaria vector control. Here we report on the first phase of the project, that has generated genome-

30   wide data on nucleotide variation in 765 wild-caught mosquitoes, sampled from 15 locations in 8

31   countries spanning a variety of ecological settings, including rainforest, inland savanna and coastal

32   biomes (Supplementary Fig. 1). We sampled the two major malaria vector species within the species

33   complex, *Anopheles gambiae sensu stricto* and *Anopheles coluzzii*, which are morphologically

34   indistinguishable and often sympatric but may differ in geographical range[11], larval ecology[12],

35   behaviour[13] and strategies for surviving the dry season[14]. *An. gambiae* and *An. coluzzii* have been

36   classified as different species[15] because they are genetically distinct[16–18]. However, although they

37   undergo assortative mating[19], reproductive isolation is incomplete: hybrids are viable and fertile, and

38   there is evidence for hybridization in nature varying over space[20–22] and time[23], creating opportunities

39   for gene flow between species[24,25]. The diversity of sampling in this project phase over geography,

40   ecology and species is not exhaustive, but does provide a broad platform from which to explore the

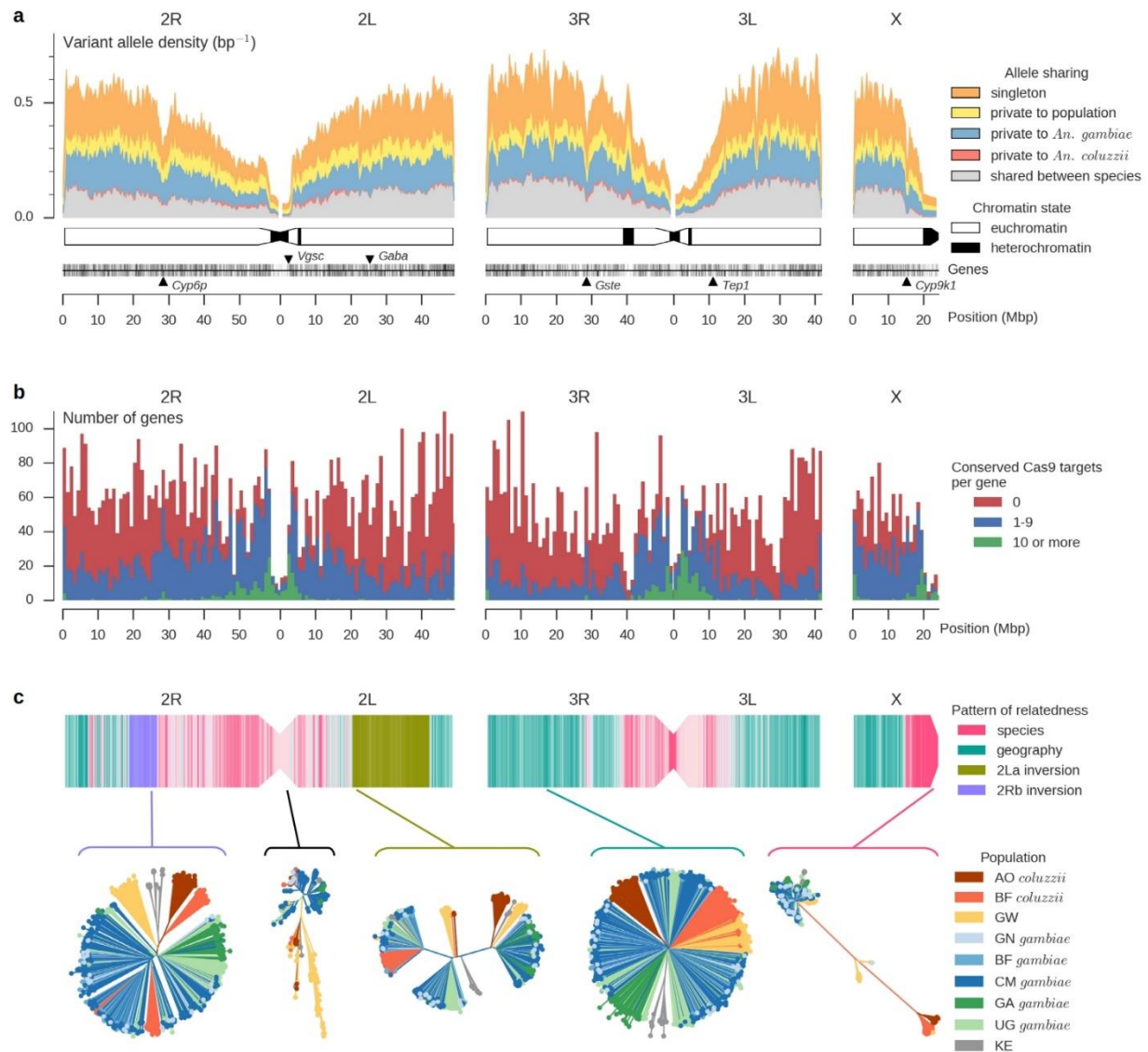41   factors shaping mosquito population variation, evolution and speciation.

[†] http://www.malariagen.net/ag1000g

Figure 1. Patterns of genomic variation. a, *Density of variant alleles in non-overlapping 200 kbp windows over the genome, computed as the number of variant alleles discovered in SNPs passing all quality filters divided by the number of accessible positions. Schematic of chromosomes below shows regions of heterochromatin[26].* ***b***, *Genomic distribution of genes containing conserved regions that could be targeted by CRISPR/Cas9 gene drive.* ***c***, *Variations in the pattern of relatedness between individual mosquitoes over the genome. The upper part of the plot shows a schematic of the three chromosomes painted using colours to represent the major pattern of relatedness within non-overlapping 100 kbp windows. Below, neighbour-joining trees are shown from a selection of genomic windows that are representative of the four major patterns of relatedness found, as well as for the window spanning the Vgsc gene on chromosome arm 2L which has a unique pattern of relatedness. The strength of colour indicates the strength of the correlation with the closest major pattern. AO = Angola; BF = Burkina Faso; GW = Guinea-Bissau; GN = Guinea; CM = Cameroon; GA = Gabon; UG = Uganda; KE = Kenya. Species status is uncertain for GW and KE populations.*

Genomic variation

42

43   We used the Illumina HiSeq platform to perform whole genome deep sequencing on individual

44   mosquitoes. After removing samples with low coverage (<14X) we analyzed data on 765 wild-caught

45   specimens and a further 80 specimens comprising parents and progeny from 4 lab crosses

46   (Supplementary Fig. 1). Sequence reads were aligned against the AgamP3 reference genome[27] and

47   putative single nucleotide polymorphisms (SNPs) were called from the alignments[28,29] (Supplementary

48   Text). The alignments were also used to identify genome regions accessible to SNP calling, where short

49   reads could be uniquely mapped and there was minimal evidence for structural variation[30,31]. We

50   classified 61% (141 Mbp) of the AgamP3 chromosomal reference sequence as accessible, including 91%

51   (18 Mbp) of coding and 59% (123 Mbp) of non-coding positions (Supplementary Fig. 2A). Mendelian

52   errors in the crosses were used to guide the design of filters to remove poor quality variant calls. In total

53   52,525,957 SNPs passed all quality filters. We then used statistical phasing, combined with information

54   from sequence reads[32], to estimate haplotypes for all wild-caught individuals. To assess the reliability of

55   this dataset, we performed capillary sequencing of 5 genes, from which we estimated a false discovery

56   rate of less than 1% and a sensitivity of 94% to detect SNPs within the accessible genome. We also

57   obtained >98% concordance of heterozygous genotype calls in comparisons with capillary sequence data

58   and >97% concordance in a second validation experiment using genotyping by primer-extension mass

59   spectrometry[33]. We assessed phasing performance for wild-caught individuals by comparison with

60   haplotypes generated from the crosses (Supplementary Fig. 3A) and from male X chromosome

61   haplotypes, obtaining results comparable to human sequencing studies[32] (Supplementary Fig. 3B).

62   Individual mosquitoes carried between 1.7 and 2.7 million variant alleles, with no systematic difference

63   observed between species (Supplementary Fig. 4A). SNPs were mostly biallelic, but 21% had three or

64   more alleles, and we discovered one variant allele every 2.2 bases of the accessible genome on average.

65   Variant allele density was similar on all chromosomes but markedly reduced in pericentromeric regions,

66   as expected due to linked selection in regions of low recombination[34–36] (Fig. 1A). Gene structure had a

67   strong influence on nucleotide diversity, with the lowest diversity observed at non-degenerate coding

68   positions and at the dinucleotide core of intron splice sites, as expected due to purifying selection on

69   deleterious functional mutations (Supplementary Fig. 4B). We also found that diversity at fourfold

70   degenerate codon positions and within short introns was twice the level found in longer introns and

71   intergenic regions, similar to studies in *Drosophila*[37] and *Heliconius*[38], indicating that most non-coding

72   sequence is under moderate selective constraint.

73   Since the advent of efficient genome editing using the CRISPR/Cas9 system[39], the push to implement

74   gene drive in *Anopheles* to carry out population suppression[10] or replacement[9] has intensified. However,

75   variants within the short ~21 bp Cas9 target site represent potential resistance alleles, and thus the

76   sheer density of SNPs could negatively impact successful deployment of gene drive in *Anopheles*. We

77   explored the accessible coding genome for CRISPR/Cas9 target sites and found viable targets in 10,711

78   of 12,901 annotated genes (Supplementary Text). However, only 5,012 genes retained at least one

79   viable target after accounting for variation within target sites, and this is likely to worsen with further

80   population sampling (Supplementary Text). These possible target genes were spread non-uniformly

81   across the genome, falling predominantly in pericentromeric regions, where levels of variation were

82   lower (Fig. 1B). The evolution of resistance to gene drive will be caused both by natural variation and by

83   the DNA repair machinery itself, and therefore drive-based methods are unlikely to work unless multiple

84   genes and multiple sites within each gene are targeted. To that end, we identified 544 genes that each

85   contain at least 10 non-overlapping conserved target sites, including 9 putative sterility genes[10]

86   (Supplementary Text). The genome sequences presented here are a valuable resource for prioritizing

87   genes and designing gene drive strategies that will be effective in natural populations.

Population structure and gene flow

Analysis of genetic structure provides a foundation for studying the evolutionary and demographic history of populations, and for understanding how genetic variants move between populations. We are particularly interested in gene flow across geographical ranges via migration, and gene flow between species via hybridization, as both can play a role in the spread of medically-important variants, including insecticide resistance mutations[24,25] and introduced genetic modifications. Previous studies of the *Anopheles gambiae* complex have shown that phylogenetic relationships can vary dramatically between different genomic regions[24,25,40–42]. We therefore began by computing genetic distances between individual mosquitoes and constructing neighbour-joining (NJ) trees within non-overlapping genomic windows of 100,000 accessible bases (Fig. 1C). By analyzing the correlation between genetic distances in different genomic windows, we identified four major patterns of relatedness, systematically associated with different genomic regions. Within pericentromeric regions of chromosomes X, 3, and arm 2R, mosquitoes segregated into two distinct and widely separated clusters, largely corresponding to the two species as determined by conventional molecular diagnostics[18,43]. Individuals from coastal Guinea-Bissau, where reproductive isolation between species is believed to have broken down[20–22], were an exception, being found in both clusters with poor correspondence to species assignments, as well as in an intermediate cluster. The large chromosomal inversions[44] 2La and 2Rb were each associated with a distinct pattern of relatedness, as expected if gene flow is limited by reduced recombination between inversion karyotypes[42,45]. Genetic structure was weak throughout most of the remainder of the genome, with some separation of populations at the extremes of the geographical range (Angola, Kenya), but no evidence of clustering by species. In addition to these four major patterns of relatedness, we found other distinct patterns within some isolated genome regions, including windows near the voltage-gated sodium channel (*Vgsc*) gene[46], a known locus of resistance to DDT and pyrethroid insecticides[24,25].
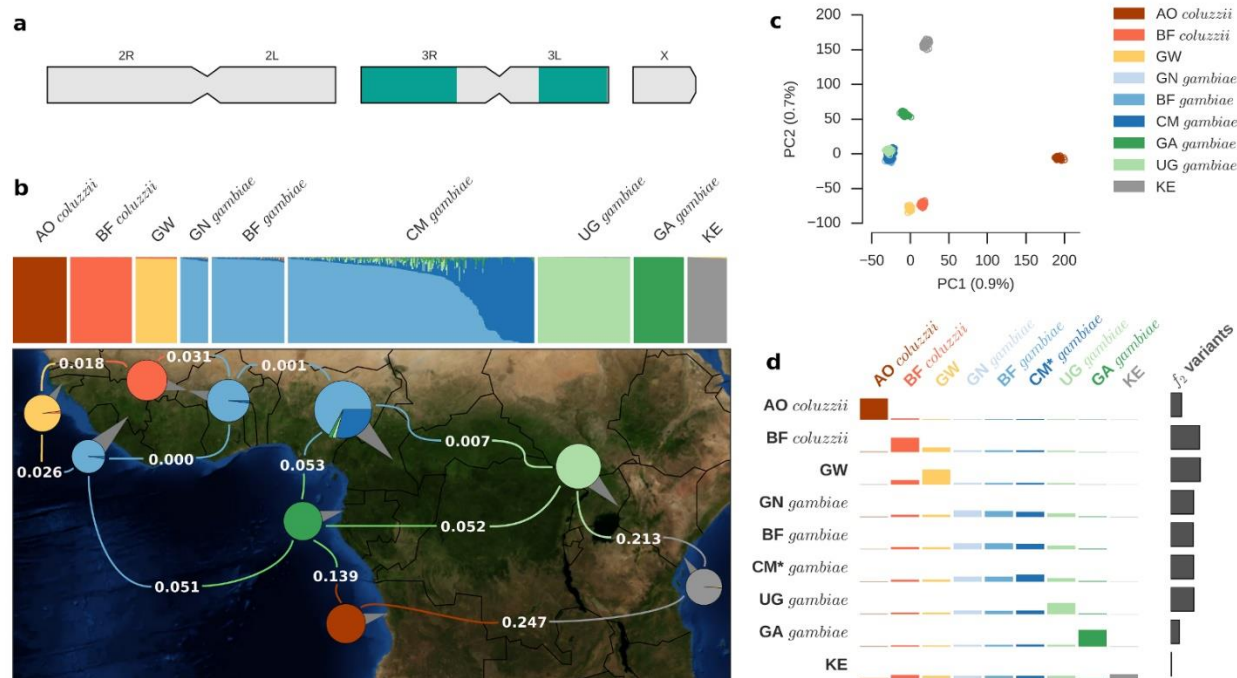
6

Figure 2. Geographical population structure. a, *Schematic showing regions of the genome used for analyses of geographical population structure highlighted in turquoise.* **b,** *ADMIXTURE and allele frequency differentiation (F_{ST}). The upper panel depicts each of the 765 wild-caught mosquitoes as a vertical bar, painted by the proportion of the genome inherited from each of K=8 inferred ancestral populations. Pie charts on the map depict the same inferred ancestral proportions summed over all individuals for each of 9 groups defined by species and country of origin; grey pointers attached to each pie chart show the sampling location. Average F_{ST} values are overlaid in white for selected pairs of populations.* **c,** *Principal components analysis. Each marker represents an individual mosquito, projected onto the first two principal components of genetic variation.* **d,** *Allele sharing in f_2 variants. The height of the coloured bars represent the probability of sharing a doubleton allele between two populations. Heights are normalized row-wise for each population. Grey bars at the end of each row depict the total number of doubletons found in individuals from the given population. CM* = Cameroon savanna sampling site only.*

111   These other patterns were characterized by short genetic distances between individuals from different

112   populations and species, indicating the influence of recent selective sweeps and adaptive gene flow.

113   To investigate the influence of geography on population structure, we analyzed data from Chromosome

114   3, which is free from high frequency polymorphic inversions[44] (Fig. 2A). We used ADMIXTURE to model

115   each individual as a mixture deriving from *K* ancestral populations[47] and compared with results from

116   principal components analysis (PCA) and allele frequency differentiation (F_{ST}) (Supplementary Text; Fig.

117   2B, 2C; Supplementary Figs. 5, 6). All analyses supported five major ancestral populations,

118   corresponding to: (i) Guinea, Burkina Faso, Cameroon and Uganda *An. gambiae*; (ii) Gabon *An. gambiae*;

119   (iii) Kenya; (iv) Angola *An. coluzzii*; (v) Burkina Faso *An. coluzzii* and Guinea-Bissau. These results are

7

120   consistent with previous evidence that the Congo Basin tropical rainforest and the East African Rift Zone

121   are natural barriers to gene flow[44,48–51]. Within each species, we found high F_{ST} across these barriers,

122   exceeding the level of differentiation between the two species at a single location (Fig. 2B;

123   Supplementary Fig. 6B), indicating that ecological discontinuities may have a stronger impact on gene

124   flow than assortative mating in sympatric populations.

125   The movement of mosquitoes affects not only the spread of genetic variants in vector populations, but

126   also the spatial and temporal dynamics of malaria parasite transmission. Previous studies have

127   suggested that purposeful movement of individual *Anopheles* mosquitoes is limited to short-range

128   dispersal up to 5km[52,53]; however, recent studies have provided evidence of long-distance seasonal

129   migration in *An. gambiae*[14]. If mosquitoes only travel short distances, we would expect to observe some

130   differentiation between mosquitoes sampled from different geographical locations. To complement

131   ADMIXTURE, PCA and F_{ST} results, we also studied the sharing of rare alleles (Fig. 2D), which should be

132   enriched for recent mutations and thus provide high resolution to detect subtle population structure. All

133   analyses provided evidence for differentiation between Uganda and *An. gambiae* populations to the

134   west, and between Guinea-Bissau and *An. coluzzii* from Burkina Faso (Fig. 2D; Supplementary Figs. 5, 6).

135   However, we found no evidence for differentiation between *An. gambiae* from Guinea and Burkina Faso

136   by any method. Some differentiation was detectable between *An. gambiae* from Burkina Faso and

137   Cameroon, but mosquitoes were sampled from multiple sites within Cameroon along an ecological cline

138   from savanna into rainforest, and there was evidence for some population structure and admixture

139   associated with these different ecosystems (Fig 2B; Supplementary Figs. 5A, 6A). Considering only the

140   Cameroon savanna site, differentiation between Cameroon and *An. gambiae* populations to the west

141   was extremely weak (Fig 2D; Supplementary Fig. 6B). These findings are consistent with substantial rates

142   of long-distance movement between savanna *An. gambiae* populations in West and Central Africa.

8

143   To examine gene flow between species in more detail, we analyzed a set of 506 SNPs previously found

144   to be highly differentiated between the two species in Mali[18]. These ancestry-informative markers

145   (AIMs) showed that a block of *An. gambiae* ancestry towards the centromere of chromosome arm 2L

146   has introgressed into *An. coluzzii* populations in both Burkina Faso and Angola (Supplementary Fig. 7).

147   This genomic region spans the *Vgsc* gene, where introgression of resistance mutations has previously

148   been reported in Ghana[24] and Mali[25], but this is the first evidence that introgressed mutations have

149   spread to *An. coluzzii* populations south of the Congo Basin rainforest. AIMs also showed that all

150   mosquitoes from Guinea-Bissau carried a mixture of *An. gambiae* and *An. coluzzii* alleles on all

151   chromosomes. These individuals were sampled from the coast, within a region of Far-West Africa that is

152   believed to be a zone of secondary contact between the two species, because mosquitoes have

153   frequently been found with a hybrid genotype at the species-diagnostic marker on the X chromosome,

154   and other genetic data have suggested extensive introgression[20,22,54–56]. Our AIM results are consistent

155   with this interpretation; however, PCA and ADMIXTURE analyses of chromosome 3 showed no evidence

156   of recent admixture in Guinea-Bissau, rather grouping all individuals together in a single population

157   separate from other West African populations of either species (Supplementary Figs. 5A, 6A). These

158   results suggest a distinct demographic history for this population, and caution against the use of any

159   single marker to infer species ancestry or recent hybridization. This point is reinforced by the

160   observation that all mosquitoes sampled from coastal Kenya also carried a mixture of species alleles at

161   AIMs on all chromosome arms, except for a 4 Mbp region of chromosome X spanning the location of the

162   conventional diagnostic marker, where only *An. gambiae* alleles were present (Supplementary Fig. 7).

163   This mixed ancestry was unexpected, as sympatry between *An. gambiae* and *An. coluzzii* does not

164   extend east of the Rift Zone, where it is generally assumed that *An. gambiae*, *An. arabiensis* and *An.*

165   *merus* are the only representatives of the *gambiae* complex[15]. There are several hypotheses that could

166   explain our AIM results for Kenya, including recent or historical admixture with *An. coluzzii* populations,

9

167   introgression with other species, or retention of ancestral variation. Further analyses and population

168   sampling will be required to resolve these questions; however, our data clearly demonstrate that a

169   simple *gambiae*/*coluzzii* species dichotomy is not sufficient to capture the rich diversity and complex

170   histories of contemporary populations.

Variations in population size

172   Demographic events in the history of a population, including expansions or contractions in effective

173   population size (N_e), can be inferred from the genomes of extant individuals[57]. For malaria vectors,

174   inferring changes in N_e has practical relevance, because it could provide a means to evaluate the impact

175   of vector control interventions. For each population, we computed summary statistics of genetic

176   diversity that are influenced by demographic history, including nucleotide diversity (π), site frequency

177   spectra (SFS) and decay of linkage disequilibrium (LD) (Fig. 3A). All populations north of the Congo Basin

178   rainforest and west of the Rift Zone had characteristics of large N_e and population expansion, with high

179   diversity (π = 1.5%), an excess of rare variants (Tajima's D < -1.5) and extremely rapid decay of LD (r^2 <

180   0.01 within < 1kbp). In Gabon and Angola, we found lower diversity, more extensive LD, and an SFS

181   closer to the null expectation under constant population size, indicating smaller N_e and different

182   demographic histories. In Kenya, we found the lowest level of diversity (π = 0.9%), a strong deficit of rare

183   variants (Tajima's D > 2), and much longer LD (r^2 > 0.01 at 10Mbp), suggesting a recent population

184   bottleneck.

185   We inferred the scale and timing of historical changes in N_e using two methods, Stairway Plot[58] and

186   ∂a∂i[59], both using site frequency spectra but taking different modelling approaches. Stairway Plot

187   inferred a major expansion in all populations north of the Congo Basin rainforest and west of the Rift

188   Zone (Fig. 3B; Supplementary Fig. 8A). Three-epoch ∂a∂i models also inferred expansions in these

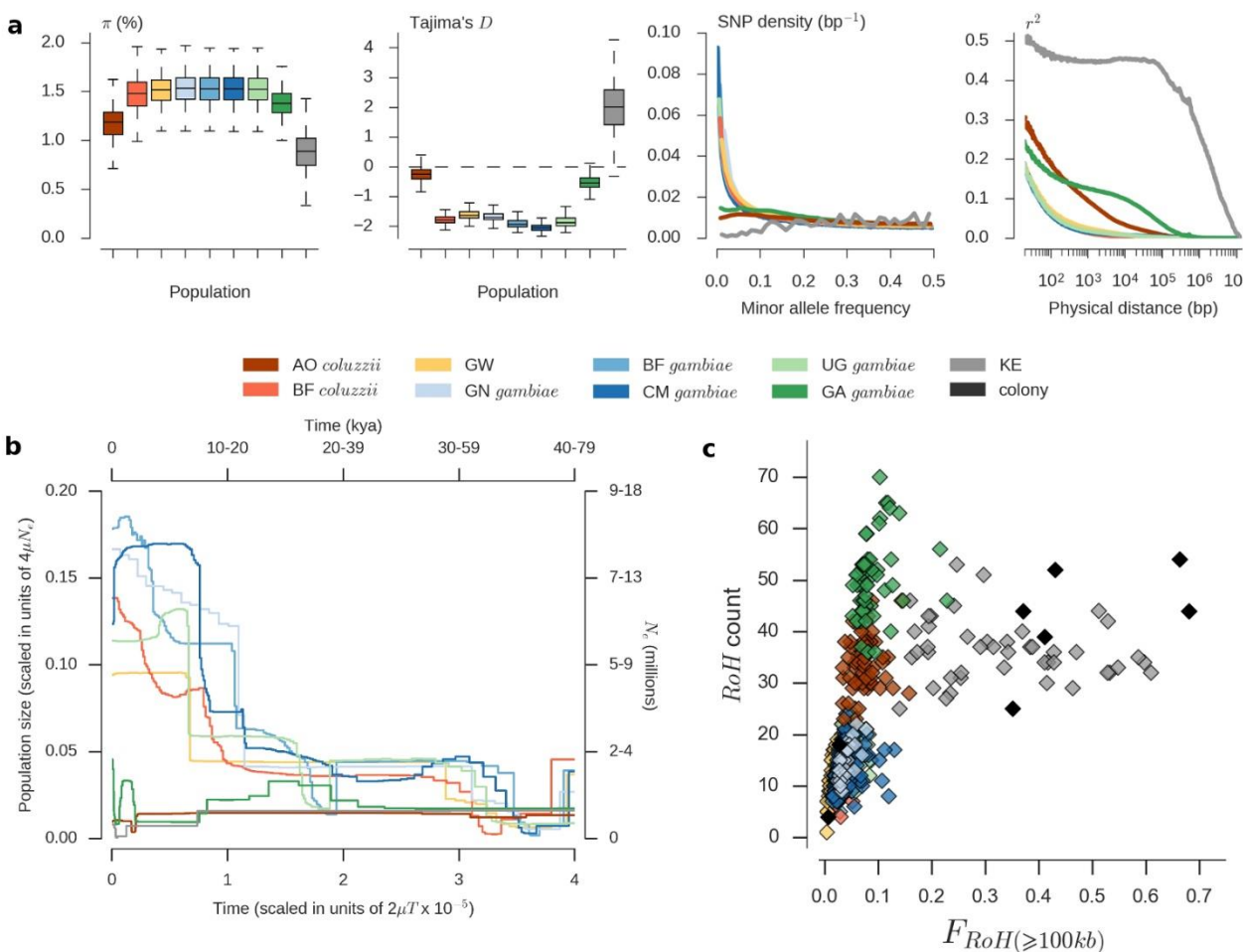189   populations, with comparable magnitudes and timings (Supplementary Fig. 8B). Translating these results

Figure 3. Genetic diversity and population size history. a, Statistics summarizing features of genetic diversity within each population. Nucleotide diversity (π) and Tajima's D are shown as the distribution of values calculated in non-overlapping 20kbp genomic windows. SNP density depicts the distribution of allele frequencies (site frequency spectrum) for each population, scaled such that a population with constant size over time is expected to have a constant SNP density over all allele frequencies. **b,** Stairway Plot of changes in population size over time, inferred from site frequency spectra. Absolute values of time and N_e are shown on alternative axes as a range of values, assuming lower and upper limits for the mutation rate as 2.8×10^{-9} and 5.5×10^{-9} respectively, and assuming T=11 generations per year. **c,** Runs of homozygosity in individual mosquitoes, highlighting evidence for recent inbreeding in Kenyan (grey) and colony (black) mosquitoes.

190   into absolute values for the timing and scale of expansion depends on the mutation rate, which has not

191   been estimated in *Anopheles*. Estimates in *Drosophila*[60,61] range from 2.8×10^{-9} to 5.5×10^{-9}, which would

192   date the onset of a major expansion in the range 7,000 to 25,000 years ago (Fig. 3B). *An. gambiae* and

193   *An. coluzzii* are both highly anthropophilic and so should have benefited from historical human

194   population growth, particularly the expansion of agricultural Bantu-speaking groups originating from

195   north of the Congo Basin beginning ~5,000 years ago[62–65]. The difference in timing suggests that either

11

196   the true mutation rate in *Anopheles* is higher than we have assumed, or that mosquito populations

197   benefited from some earlier human population growth or another factor. There have also been major

198   climatic changes since the last glacial maximum ~20,000 years ago, when overall environmental

199   conditions in Africa were much drier than present[66]. If a general reduction in aridity was the major

200   driver, then we might expect to see evidence for expansion in all mosquito populations sampled.

201   However, we inferred different demographic histories in Angola, Gabon and Kenya, although more

202   recent N_e fluctuations may be obscuring earlier events in these populations, particularly in Gabon and

203   Kenya (Fig. 3B; Supplementary Fig. 8).

204   In Kenya in 2006, free mass distribution of ITNs was carried out in multiple districts, resulting in a rapid

205   increase in ITN coverage, from less than 10% in 2004 to over 60% by the beginning of 2007[67].

206   Mosquitoes for this study were sampled from Kilifi County in 2012, and therefore originate from

207   populations experiencing sustained ITN pressure for several years. To investigate evidence for a very

208   recent bottleneck in this population, we analyzed runs of homozygosity (ROH). Kenyan mosquitoes had

209   between 10-60% of their genome within a long ROH, a level not seen in any other population (Fig. 3C).

210   This level of homozygosity is comparable to that found in isolated human populations[68] and domestic

211   animal breeds[69] due to recent inbreeding. We also observed similar ROH in mosquitoes originating from

212   lab colonies, which are typically maintained in cages of at most a few hundred individuals, and thus

213   where inbreeding is inevitable (Fig. 3C). Genetic signatures of recent inbreeding have previously been

214   observed in a mosquito population from Burkina Faso[70] and in a separate study of mosquitoes collected

215   from Kilifi in 2010[71]. However, there remains uncertainty as to whether ITN scale-up is the root cause of

216   mosquito population decline in Kilifi[71], particularly as other studies have found evidence for lower N_e[48]

217   and changes in species abundance[72] in the region pre-dating high levels of ITN coverage. Furthermore,

218   while ITNs have been effective in Kilifi, a substantial reduction in malaria prevalence had occurred prior

219   to free ITN distribution[73], thus multiple factors may be affecting vector and parasite populations in this

12

220   region. Sequencing mosquitoes and parasites before, during and after interventions, and across a range

221   of ecological and epidemiological settings, could help to resolve these questions, providing valuable

222   information about the impact and efficacy of different control strategies.

Evolution of insecticide resistance

224   Insecticide resistance is a polygenic trait with a broad phenotypic range, and several genes have

225   previously been associated with resistance in *Anopheles*, including genes encoding insecticide binding

226   targets and genes involved in insecticide metabolism[3]. It is not yet clear which of these genes, if any, are

227   responsible for epidemiologically relevant levels of resistance in the field. However, mutations that

228   confer an advantage under strong pressure from insecticide use will be positively selected, and so

229   evidence of recent selection in natural populations can help to identify and prioritize resistance genes

230   for further study. We used metrics of haplotype diversity[74] (H12) and haplotype homozygosity[75] (XP-

231   EHH) to scan the genome for genes with evidence of recent selection. Both metrics revealed strong

232   signals of selection in multiple populations at several genome locations containing genes associated with

233   insecticide resistance (Fig. 4; Supplementary Fig. 9). These included *Vgsc*, confirming evidence for

234   selection from population structure analyses described above; *Gste*, a cluster of glutathione S-

235   transferase genes including *Gste2,* previously implicated in metabolic resistance to DDT and

236   pyrethroids[76,77]; and *Cyp6p*, a cluster of genes encoding cytochrome P450 enzymes, including *Cyp6p3*

237   which is upregulated in permethrin and bendiocarb resistant mosquitoes[78,79].

238   Mutations in *An. gambiae Vgsc* codon 995 (orthologous to *Musca domestica Vgsc* codon 1014), known

239   as "*kdr*" due to their knock-down resistance phenotype, reduce susceptibility to DDT and pyrethroids by

240   altering binding-site conformation[46]. We found the Leucine→Phenylalanine (L995F) *kdr* mutation at high

241   frequency in West and Central Africa (Guinea 100%; Burkina Faso 93%; Cameroon 53%; Gabon 36%;
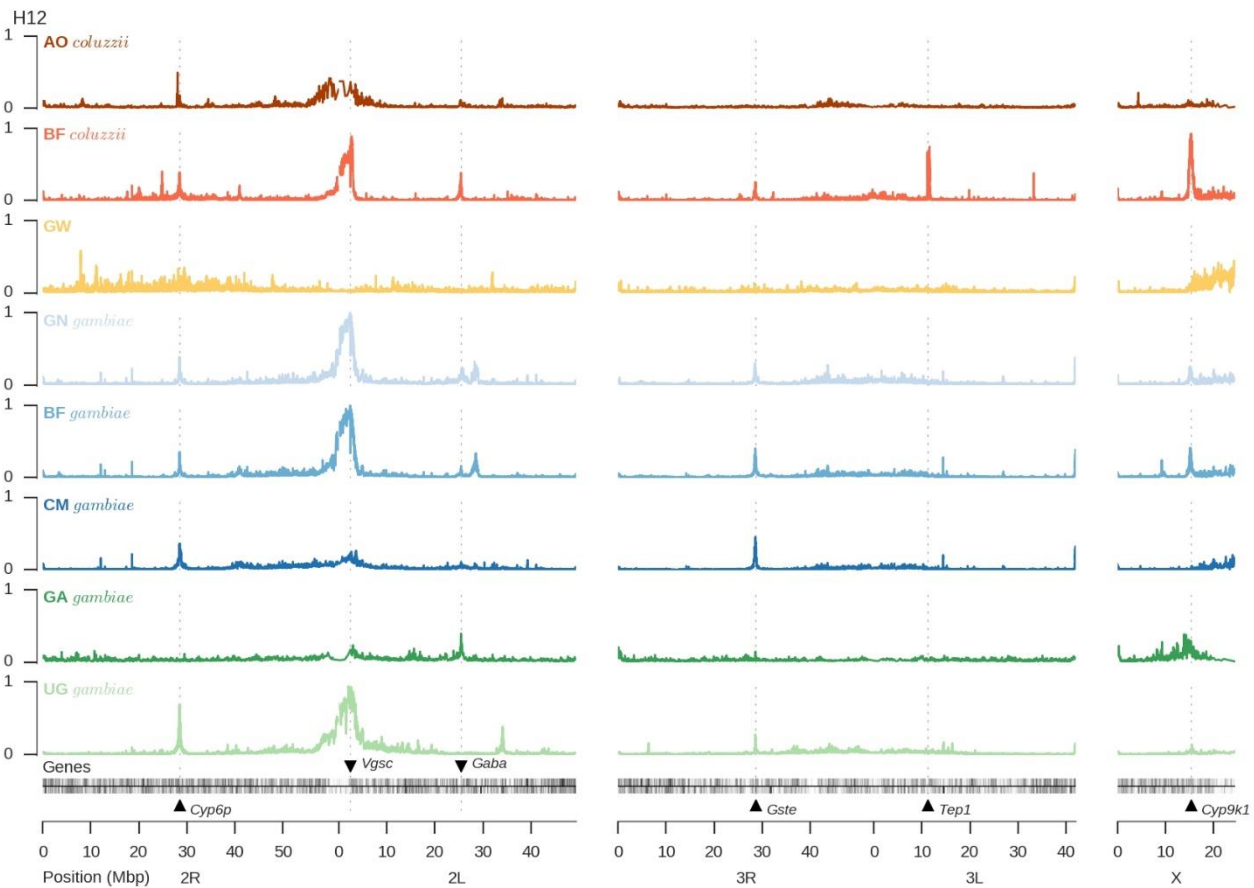
13

Figure 4. Genome scans for signatures of recent selection. Each track plots the H12 statistic in non-overlapping windows over the genome. A value of 1 indicates low haplotype diversity within a window, expected if one or two haplotypes have risen to high frequency due to recent selection. A value of 0 indicates high haplotype diversity, expected in neutral regions. Kenya is not shown because the high genome-wide levels of homozygosity mean reduced power to detect evidence of recent selection in specific genome regions.

242   Angola 86%). A second *kdr* allele, the Leucine→Serine (L995S) mutation, was present in Central and East

243   Africa (Cameroon 15%; Gabon 65%; Uganda 100%; Kenya 76%). To investigate the origins and

244   movements of these two distinct *kdr* mutations, we analyzed the genetic backgrounds on which they

245   were carried, using information from all 1,718 biallelic SNPs found across both coding and non-coding

246   regions of the *Vgsc* gene (Fig. 5). The L995F mutation occurred in five distinct haplotype clusters (labeled

247   F1-F5 in Fig. 5), while the L995S mutation was found in a further 5 haplotype clusters (labeled S1-S5 in

248   Fig. 5), indicating that the number of independent origins for each of these mutations is higher than

249   previously estimated[80–82]. Several *kdr* haplotypes have also spread between populations, despite

14

250   considerable geographic distance or ecological separation. For example, haplotype F1 is present in both

251   species and in 4 countries spanning the Congo Basin rainforest, and is the same haplotype previously

252   found to be introgressed from *An. gambiae* into *An. coluzzii* in Ghana[24], indicating strong selection

253   across a variety of ecological settings. Additionally, three *kdr* haplotypes (F4, F5, S2) were found in both

254   Cameroon and Gabon, providing multiple examples of recent adaptive gene flow between these two

255   otherwise highly differentiated populations. Finally, the S3 haplotype was found in both Uganda and

256   Kenya, showing that adaptive alleles can even cross the Rift Zone. While these remarkable patterns of

257   evolution and adaptive gene flow were primarily driven by the two *kdr* mutations, we found 16 other

258   non-synonymous mutations within *Vgsc* at a frequency above 1% (Fig. 5), of which 13 occurred

259   exclusively on haplotypes carrying the L995F *kdr* mutation, suggesting secondary selection acting on

260   mutations that enhance or compensate for the primary *kdr* phenotype.

261   Metabolic resistance is of particular concern as it has been implicated in extreme resistance phenotypes

262   observed in some *Anopheles* populations[79]. At both *Gste* and *Cyp6p* we found evidence that resistance

263   has emerged multiple times, and is also spreading between species and over considerable distances. At

264   the *Gste* locus we found at least four distinct haplotypes under selection (Supplementary Fig. 10A). One

265   of these haplotypes carried the *Gste*2-I114T mutation which enhances DDT metabolism[77,83], though the

266   other three haplotypes did not carry any known resistance mutations. At the *Cyp6p* locus we found at

267   least eight distinct haplotypes under selection (Supplementary Fig. 10B). Clearly there is much to learn

268   regarding the molecular basis of metabolic resistance, and our SNP data can be used to identify

269   candidate resistance mutations. For example, at both loci we found multiple non-synonymous SNPs that

270   were strongly associated with haplotypes under selection (Supplementary Fig. 10). These data provide a

271   starting point for new studies to characterize resistance phenotypes, and to develop improved tools for

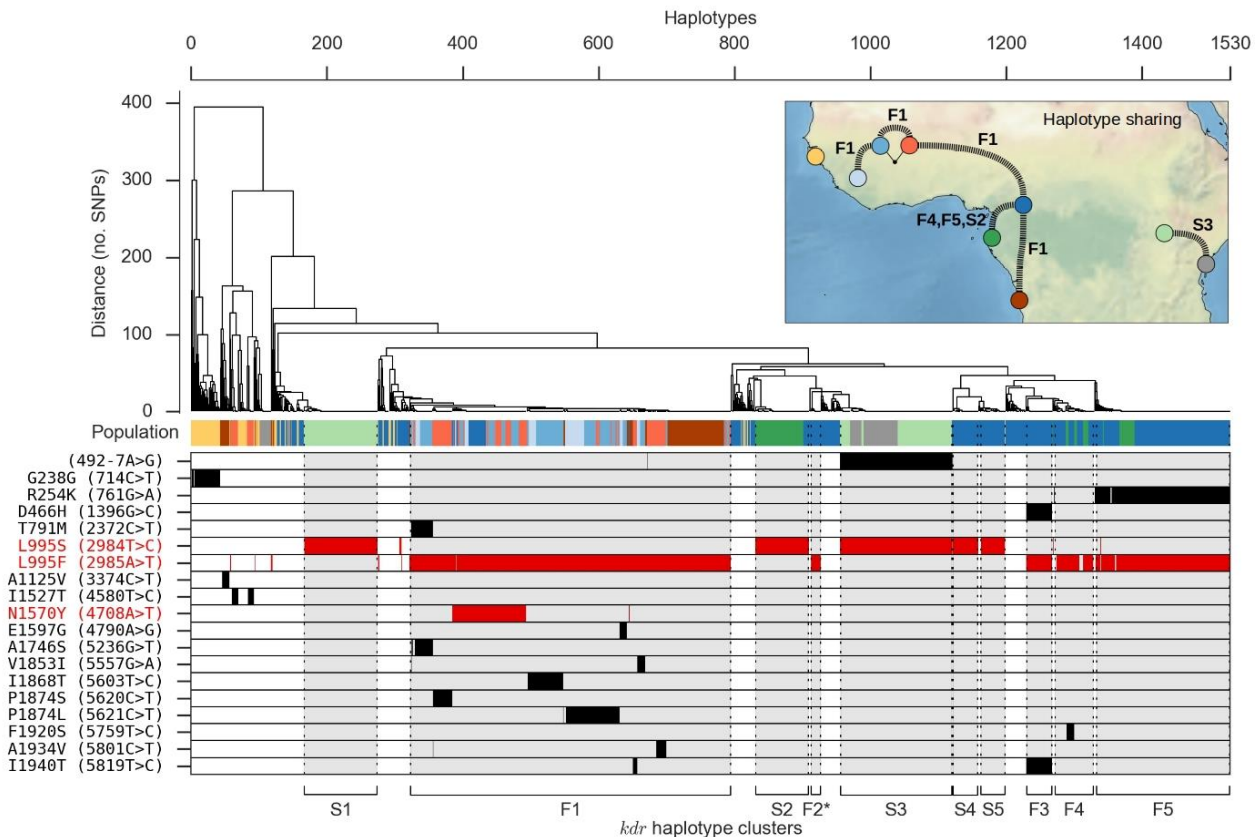272   monitoring and responding to the emergence and spread of resistance in natural populations.

Figure 5. Haplotype structure at the Vgsc gene. *The upper panel shows a dendrogram obtained by hierarchical clustering of haplotypes from wild-caught individuals. The colour bar immediately below shows the population of origin for each haplotype. Inset map depicts haplotypes shared between populations. The lower panel shows alleles carried by each haplotype at 19 SNPs with allele frequency > 1% that either change the amino acid sequence or occur within a splice region, and therefore may affect protein function (white = reference allele; black = alternate allele; red = previously known resistance-conferring allele). At the lower margin, we label 10 haplotype clusters carrying a kdr mutation (either L995F or L995S).*

Discussion

273   In this first phase of the Ag1000G project we have focused on nucleotide variation, revealing an

274   extraordinary reservoir of natural genetic diversity in mosquito populations. Nucleotide diversity is 1.5%

275   in most populations, twice that reported for African populations of *Drosophila melanogaster*[37,84] and ten

276   times greater than human populations[30], sustained by a network of large and highly interconnected

277   populations. The genomes that we have sequenced convey a rich mosaic of different ancestries, shaped

278   by geography, ecology, speciation, migration, selection, recombination and chromosomal inversions,

279   with different forces predominating in different genomic regions. Mosquito populations in different

280

16

281   parts of Africa have experienced major demographic changes, including expansions and contractions in

282   size, influenced at least in part by major events in the history of our own species. The introduction of

283   insecticides has led to intense selection pressure, repeatedly driving resistance mutations to high

284   frequency and demonstrating the potential for adaptive gene flow across the entire continent. The data

285   we have generated provide a resource for studying and responding to the ongoing evolution of malaria

286   vector populations. To facilitate access to this resource we have developed a novel web application‡ that

287   enables visual exploration of genomic data on populations and individual mosquitoes from the scale of a

288   whole chromosome down to individual nucleotides. Future project phases will increase both the

289   geographical and taxonomic representation of mosquito genomes sequenced, and will explore other

290   forms of genetic variation, including small insertion/deletion polymorphisms and large structural

291   variation. We will also continue to study fundamental population-genetic processes, including mutation,

292   recombination, natural selection, and the fine structure and history of gene flow between populations.

293   In 1899 Ronald Ross proposed that malaria could be controlled by destroying breeding sites of the

294   mosquitoes that transmit the disease[85]. *An. gambiae*, identified in the same year by Ross as a vector of

295   malaria in Africa[86], has proved resilient to a century of attempts to repress it. The vector control

296   armamentarium needs to be expanded, not only with new classes of insecticide and novel genetic

297   control strategies, but also with more effective tools for gathering intelligence, to enable those

298   responsible for planning and executing interventions to stay ahead of the mosquito's remarkable

299   capacity for evolutionary adaptation. There remain major knowledge gaps, e.g., concerning the rate and

300   range of long-distance migration, which are fundamental to understanding both malaria transmission

301   and the spread of insecticide resistance, and which will require detailed spatiotemporal analysis of

302   mosquito population structure. Most importantly, it is essential to start collecting population genomic

303   data prospectively as an integral part of major vector control interventions, to identify which strategies

---

‡ http://www.malariagen.net/apps/ag1000g

304   are most likely to cause increased resistance, or what it takes to cause a population crash of the

305   magnitude observed in our Kenyan data. By treating each major intervention as an experiment, and by

306   analyzing its impact on mosquito populations, we can aim to improve the efficacy and sustainability of

307   future interventions, while at the same time learning about basic processes of ecology and evolution.

Methods

309   Methods are described in Supplementary Text.

Data availability

311   All sequence reads from the Ag1000G project are available from the European Nucleotide Archive (ENA -

312   http://www.ebi.ac.uk/ena) under study PRJEB1670. Submission of sequence read alignments and

313   variant calls from Ag1000G phase 1 is in progress under ENA study PRJEB18691. Variant and haplotype

314   calls and associated data from Ag1000G phase 1 can be explored via an interactive web application or

315   downloaded via the MalariaGEN website (https://www.malariagen.net/projects/ag1000g#data).

Acknowledgments

18

The *Anopheles gambiae* 1000 Genomes Consortium

324

325 **Corresponding authors.** Alistair Miles[1,2], Mara K. N. Lawniczak[1], Martin Donnelly[3,1], Dominic

326 Kwiatkowski[1,2].

327 **Data analysis group.** Alistair Miles[1,2] (project lead), Nicholas J. Harding[2], Giordano Bottà[4,2], Chris S.

328 Clarkson[1], Tiago Antão[5], Krzysztof Kozak[1], Daniel R. Schrider[6], Andrew D. Kern[6], Seth Redmond[7], Igor

329 Sharakhov[8,9], Richard D. Pearson[1,2], Christina Bergey[10], Michael C. Fontaine[11], Martin Donnelly[3,1], Mara K.

330 N. Lawniczak[1], Dominic Kwiatkowski[1,2] (chair).

331 **Partner working group.** Martin Donnelly[3,1] (chair), Diego Ayala[12,13], Nora J. Besansky[10], Austin Burt[14],

332 Beniamino Caputo[4], Alessandra della Torre[4], Michael C. Fontaine[11], H. Charles J. Godfray[15], Matthew W.

333 Hahn[16], Andrew D. Kern[6], Dominic Kwiatkowski[1,2], Mara K. N. Lawniczak[1], Janet Midega[17], Daniel E.

---

[1] Malaria Programme, Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK
[2] MRC Centre for Genomics and Global Health, University of Oxford, Oxford OX3 7BN, UK
[3] Department of Vector Biology, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK
[4] Istituto Pasteur Italia – Fondazione Cenci Bolognetti, Dipartimento di Sanita Pubblica e Malattie Infettive, Università di Roma SAPIENZA, Rome, Italy
[5] University of Montana, Missoula, MT 59812, USA
[6] Department of Genetics, Rutgers University, 604 Alison Road, Piscataway, NJ 08854, USA
[7] Genome Sequencing and Analysis Program, Broad Institute, 415 Main Street, Cambridge, MA 02142, USA
[8] Department of Entomology, Virginia Tech, Blacksburg, VA 24061, USA
[9] Laboratory of Ecology, Genetics and Environmental Protection, Tomsk State University, Tomsk 634050, Russia
[10] Eck Institute for Global Health, Department of Biological Sciences, University of Notre Dame, IN 46556, USA
[11] Groningen Institute for Evolutionary Life Sciences (GELIFES), Nijenborgh 7, 9747 AG Groningen, The Netherlands
[12] Unité d'Ecologie des Systèmes Vectoriels, Centre International de Recherches Médicales de Franceville, Franceville, Gabon
[13] Institut de Recherche pour le Développement (IRD), UMR MIVEGEC (UM1, UM2, CNRS 5290, IRD 224), Montpellier, France
[14] Department of Life Sciences, Imperial College, Silwood Park, Ascot, Berkshire SL5 7PY, UK
[15] Department of Zoology, University of Oxford, The Tinbergen Building, South Parks Road, Oxford OX1 3PS, UK
[16] Department of Biology and School of Informatics and Computing, Indiana University, Bloomington, IN 47405, USA
[17] KEMRI-Wellcome Trust Research Programme, PO Box 230, Bofa Road, Kilifi, Kenya

334 Neafsey[7], Samantha O'Loughlin[14], João Pinto[18], Michelle Riehle[19], Igor Sharakhov[8,9], Kenneth D.

335 Vernick[20], David Weetman[3], Craig Wilding[21], Bradley White[22].

336 **Sample collections. Angola:** Arlete D. Troco[23], João Pinto[18]; **Burkina Faso:** Abdoulaye Diabaté[24],

337 Samantha O'Loughlin[14], Austin Burt[14]; **Cameroon:** Carlo Costantini[13,25], Kyanne R. Rohatgi[10], Nora J.

338 Besansky[10]; **Gabon:** Nohal Elissa[12], João Pinto[18]; **Guinea:** Boubacar Coulibaly[26], Michelle Riehle[19],

339 Kenneth D. Vernick[20]; **Guinea-Bissau:** João Pinto[18], João Dinis[27]; **Kenya:** Janet Midega[17], Charles Mbogo[17],

340 Philip Bejon[17]; **Uganda:** Craig Wilding[21], David Weetman[3], Henry Mawejje[28], Martin Donnelly[3,1]; **Crosses:**

341 David Weetman[3], Craig Wilding[21], Martin Donnelly[3,1].

342 **Sequencing and data production.** Jim Stalker[1], Kirk Rockett[2], Eleanor Drury[1], Daniel Mead[1], Anna

343 Jeffreys[2], Christina Hubbart[2], Kate Rowlands[2], Alison Isaacs[3], Dushyanth Jyothi[1], Cinzia Malangone[1].

344 **Web application development.** Paul Vauterin[2], Ben Jeffrey[2], Ian Wright[2], Lee Hart[2], Krzysztof

345 Kluczyński[2].

346 **Project coordination.** Victoria Cornelius[2], Bronwyn MacInnis[29], Christa Henrichs[2], Rachel

347 Giacomantonio[1], Dominic Kwiatkowski[1,2].

---

[18] Global Health and Tropical Medicine, GHTM, Instituto de Higiene e Medicina Tropical, IHMT, Universidade Nova de Lisboa, UNL, Rua da Junqueira 100, 1349-008 Lisbon, Portugal
[19] Department of Microbiology and Immunology, Microbial and Plant Genomics Institute, University of Minnesota, St. Paul, MN 55108
[20] Unit for Genetics and Genomics of Insect Vectors, Institut Pasteur, Paris, France
[21] School of Natural Sciences and Psychology, Liverpool John Moores University, Liverpool L3 3AF, UK
[22] Department of Entomology, University of California, Riverside, CA, USA
[23] Programa Nacional de Controle da Malária, Direcção Nacional de Saúde Pública, Ministério da Saúde, Luanda, Angola
[24] Institut de Recherche en Sciences de la Santé (IRSS), Bobo Dioulasso, Burkina Faso
[25] Laboratoire de Recherche sur le Paludisme, Organisation de Coordination pour la lutte contre les Endémies en Afrique Centrale (OCEAC), Yaoundé, Cameroon
[26] Malaria Research and Training Centre, Faculty of Medicine and Dentistry, University of Mali
[27] Instituto Nacional de Saúde Pública, Ministério da Saúde Pública, Bissau, Guiné-Bissau
[28] Infectious Diseases Research Collaboration, 2C Nakasero Hill Road, P.O. Box 7475, Kampala, Uganda
[29] The Broad Institute of Massachusetts Institute of Technology and Harvard, 415 Main Street, Cambridge, MA 02142, USA
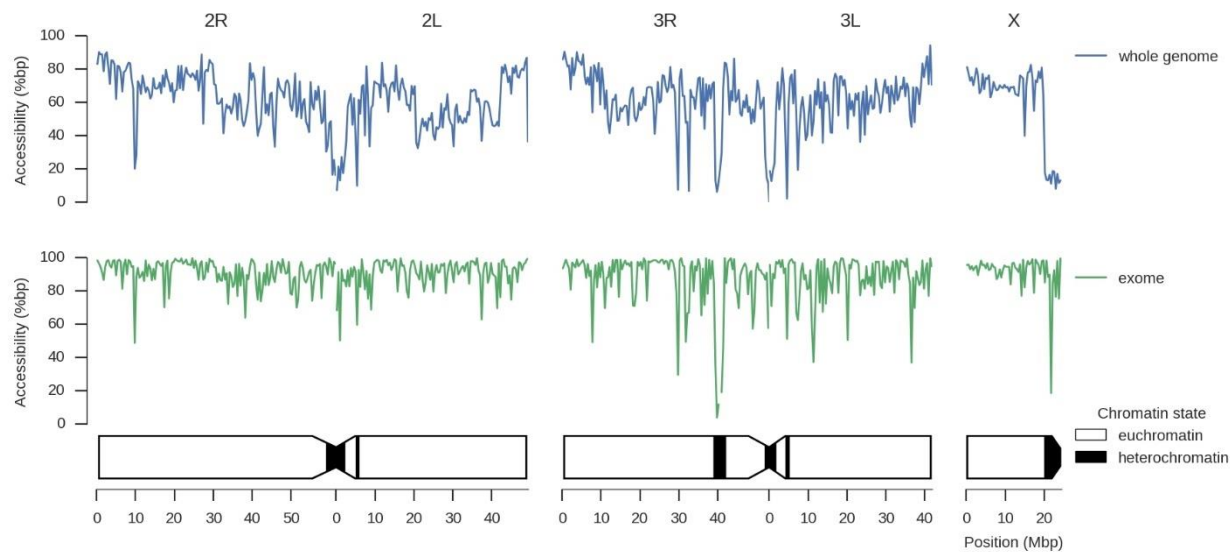
348   # Supplementary figures
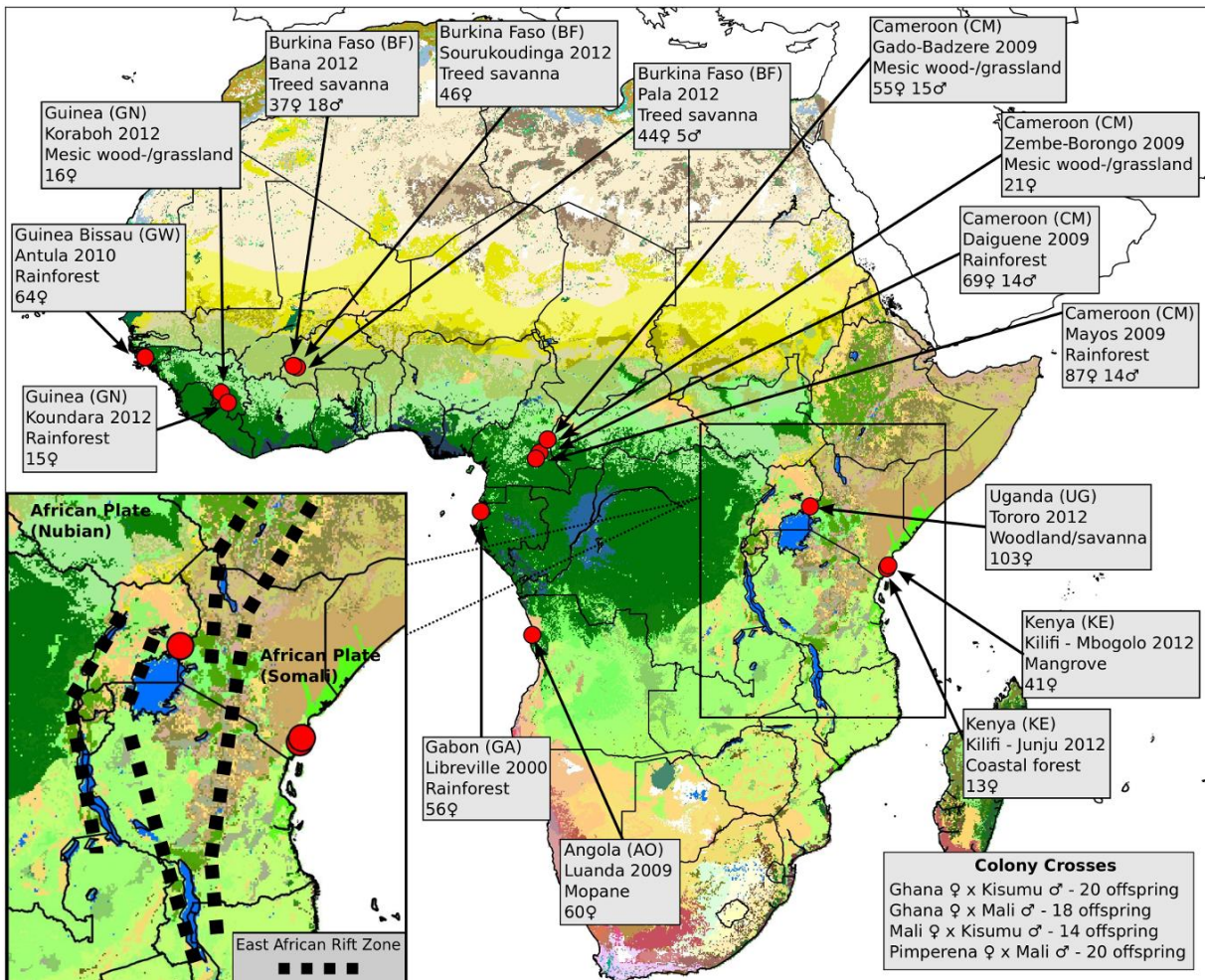
Supplementary Figure 1. Overview of population sampling. Red circles show sampling locations for wild-caught mosquitoes; black outlines show country borders. Colours in the map represent ecosystem classes; dark green represents forest ecosystems, see (87) Fig. 9 for a complete colour legend. The Congo Basin tropical rainforest is the large region of dark green in Central Africa, spanning parts of Cameroon, Equatorial Guinea, Gabon, Central African Republic, Republic of Congo and Democratic Republic of Congo. Sampling details for each site are shown in light grey boxes, including country (two-letter country code), name of sampling site, year of collection, predominant ecosystem classification[87] for the local region, and number and sex of individuals sequenced. Further details of sampling locations and methods are provided in Supplementary Text. For colony crosses, the direction of cross (colony of origin of mother and father) and number of offspring is shown. The inset map depicts geological fault lines in the East African Rift Zone[*].
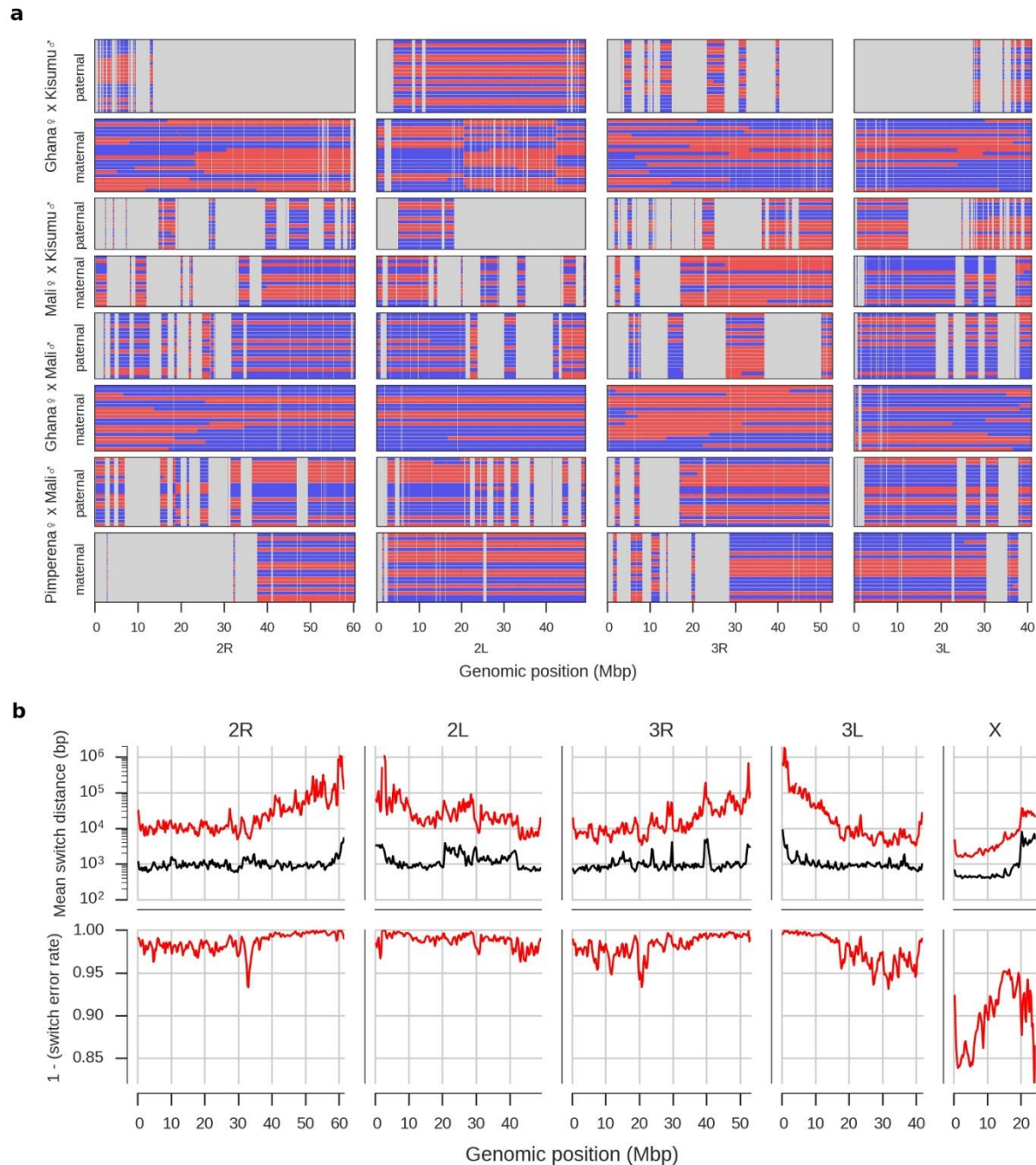
349

350

[*] http://pubs.usgs.gov/publications/text/East_Africa.html

Supplementary Figure 2. Genome accessibility. *Plots show the percentage of accessible bases in non-overlapping 400kbp windows. The schematic of chromosomes below shows chromatin state predictions from (26).*
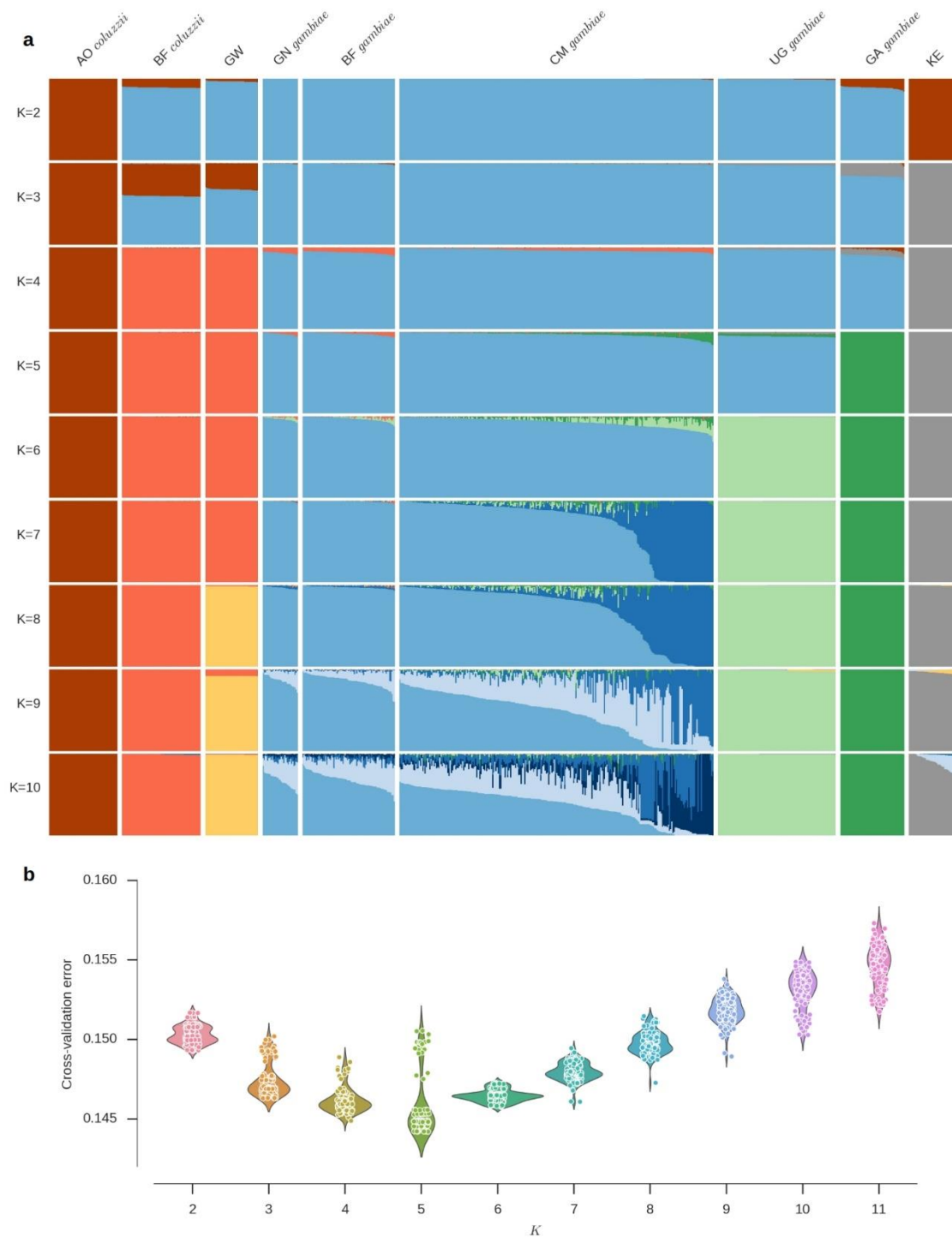
351

352

a

b

Supplementary Figure 3. Haplotype validation. a, Haplotypes inferred in the crosses. Each panel shows either maternal or paternal haplotypes from a single cross. Each row within a panel represents a single progeny haplotype. Haplotypes are coloured by parental inheritance (blue = allele from parent's first chromosome, red = allele from parent's second chromosome). Switches between colours along a haplotype indicate putative recombination events. Regions that were within a run of homozygosity in the parent and thus not informative for haplotype validation are masked in grey. b, Error rate estimates for haplotypes inferred in wild-caught individuals. Upper plots show estimates for the mean switch distance (red line) in windows over the genome, compared to the mean switch distance if heterozygotes were phased randomly (black line). Lower plots show the switch error rate, which estimates the probability of a switch error occurring between two adjacent heterozygous genotype calls.
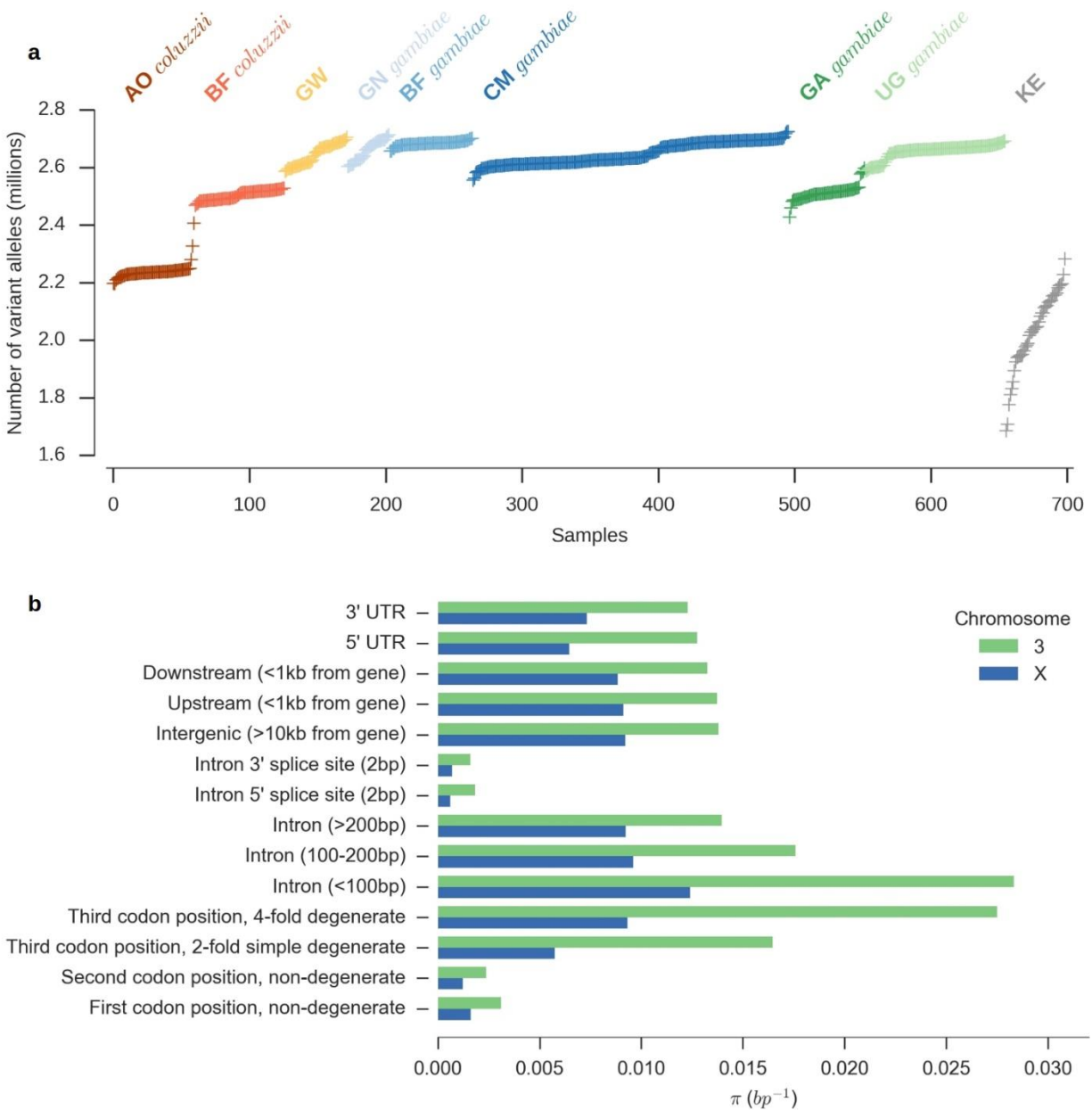
Supplementary Figure 4. Variant discovery and nucleotide diversity. a, *Total number of variant alleles discovered per individual mosquito sequenced. Only females are plotted.* **b**, *Average nucleotide diversity (π) in relation to gene architecture.*

353

354

Supplementary Figure 5. ADMIXTURE analysis. a, *Ancestry proportions within individual mosquitoes for ADMIXTURE models from K=2 to K=10 ancestral populations. Each vertical bar represents the proportion of ancestry within a single individual, with colours corresponding to ancestral populations. These data are the average of the major q-matrix clusters derived by CLUMPAK analysis.* **b**, *Violin plot of cross-validation error for each of 100 replicates for each K values.*

355

Supplementary Figure 6. Population structure and allele frequency differentiation. a, *Principal components analysis of the 765 wild-caught mosquitoes, showing the first 10 components of genetic variation. The final panel shows the variance explained by each component.* **b**, *Average allele frequency differentiation (F_{ST}) between pairs of populations. The lower left triangle shows average F_{ST} between each population pair. The upper right triangle shows the Z score for each F_{ST} value estimated via a block-jackknife procedure.*

356
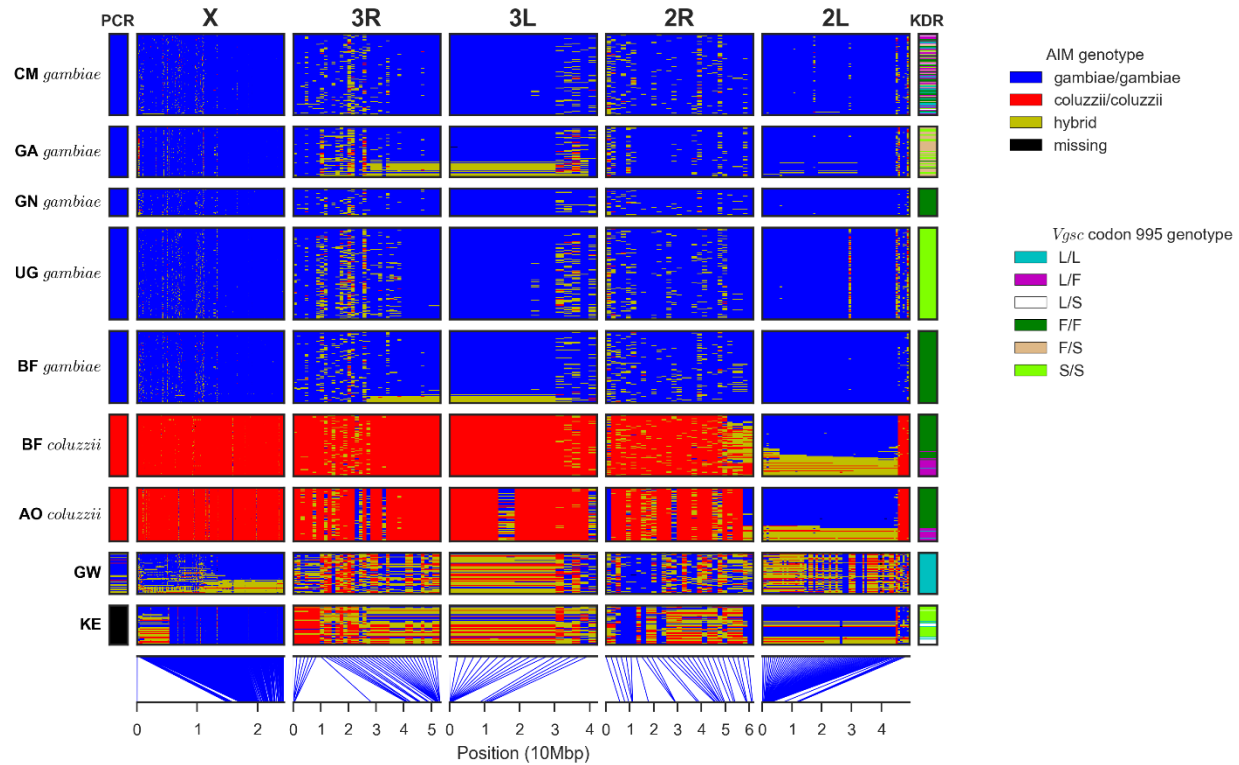
Supplementary Figure 7. Ancestry informative markers (AIMs). *Rows represent individual mosquitoes (grouped by population) and columns represent SNPs (grouped by chromosome arm). Colours represent genotype. The column at the far left shows the species assignment according to the conventional molecular test based on a single marker on the X chromosome, which was performed for all individuals except Kenya (KE). The column at the far right shows the genotype for kdr mutations in Vgsc codon 995. Lines at the lower edge show the physical locations of the AIM SNPs.*

357

358

359

360
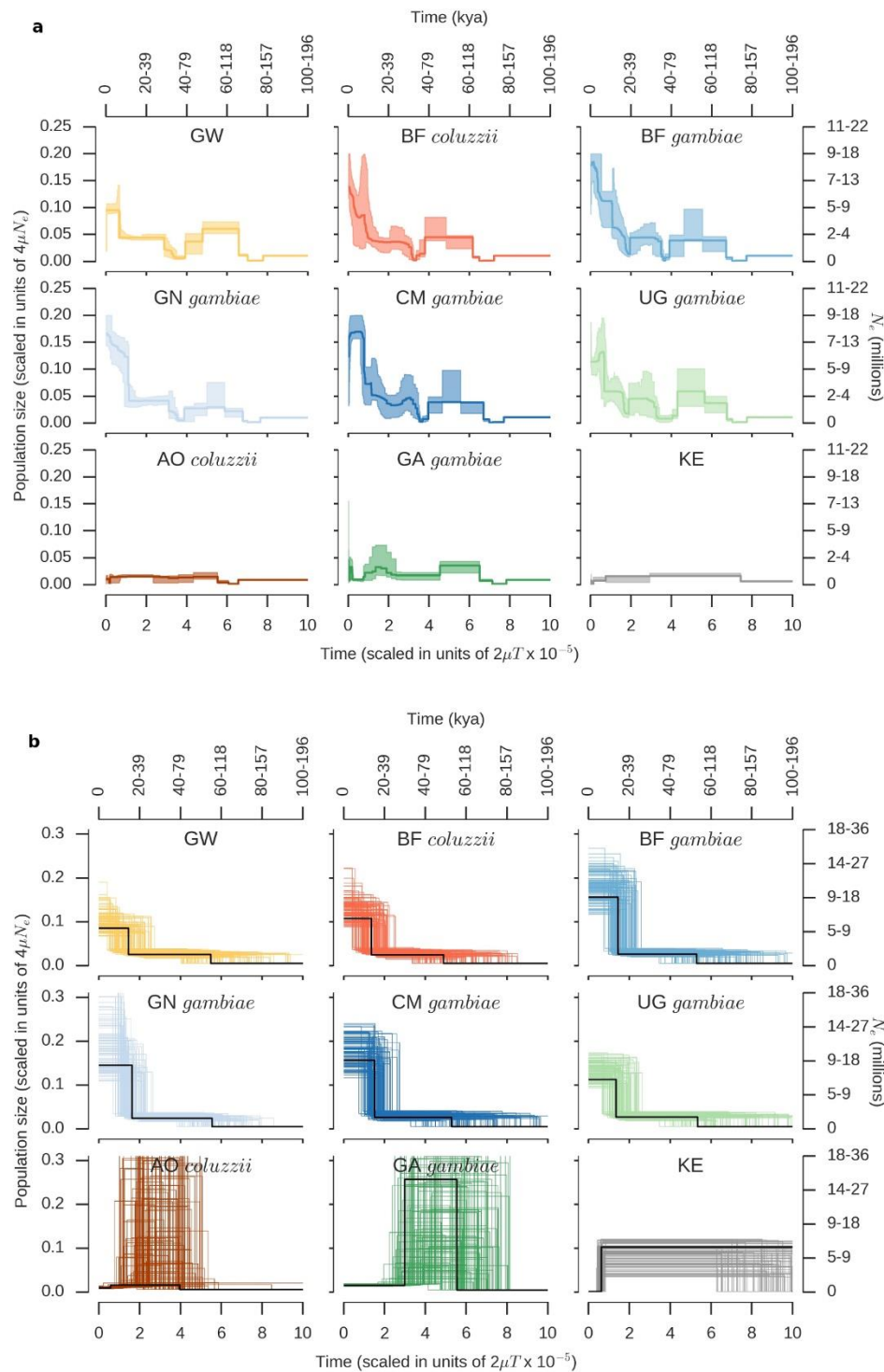
Supplementary Figure 8. Inferred population size histories. a, Stairway Plot inferred histories for each population. The shaded area shows the 95% confidence interval from 199 bootstrap replicates. **b**, Inferred histories from ∂a∂i three epoch models. Black line shows the history with the highest likelihood found by optimization; coloured lines show 100 histories with the highest likelihoods from even sampling of the model parameter space (Supplementary Text). Absolute time and N_e are shown as a range assuming 11 generations per year and a mutation rate of between 2.8×10^{-9} and 5.5×10^{-9}.

361
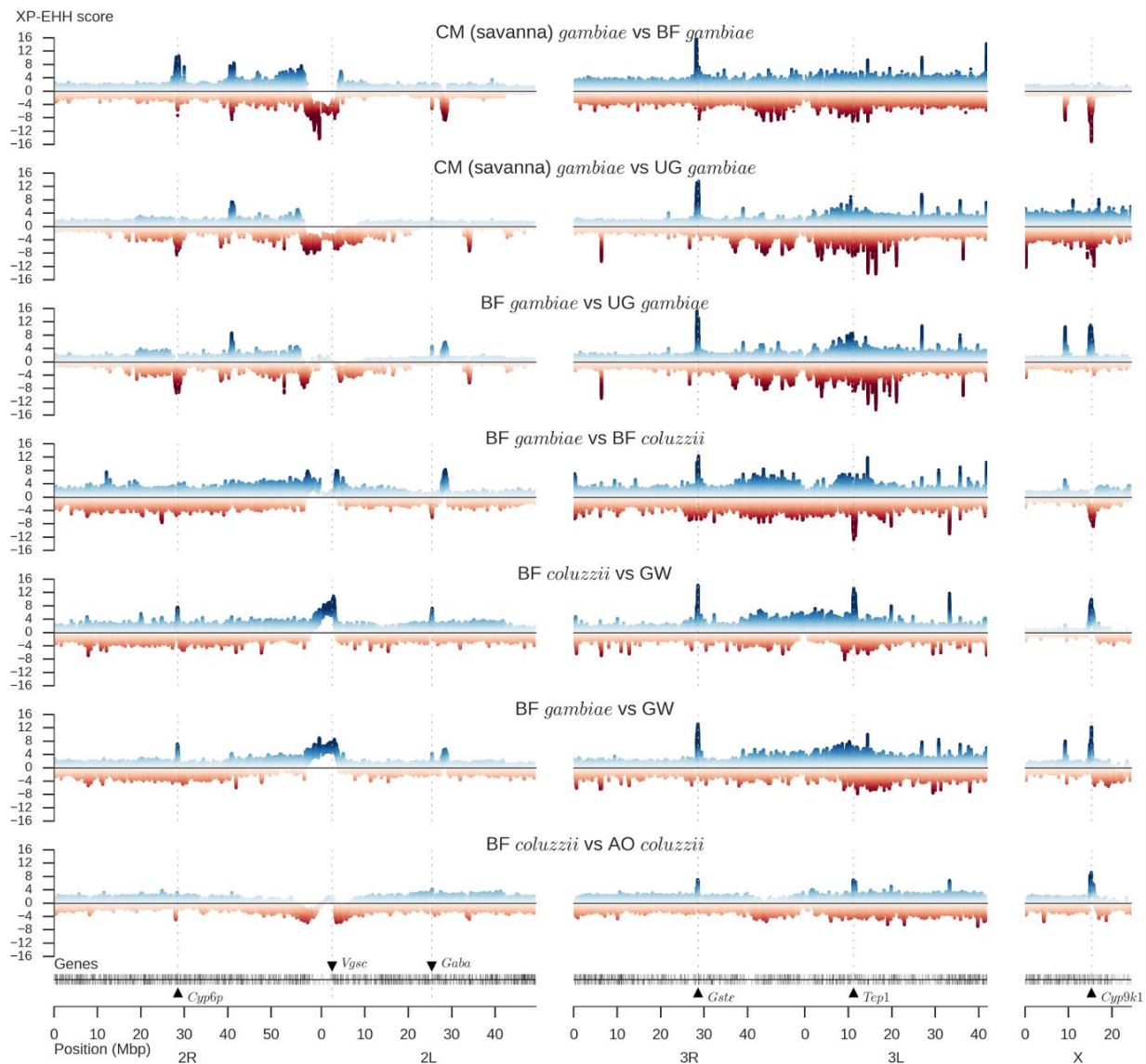
Supplementary Figure 9. Cross-population genome scans for signatures of recent selection. *For each population comparison (e.g., BF gambiae versus BF coluzzii), positive XP-EHH values indicate longer haplotypes and therefore recent selection in the first population (e.g., BF gambiae), and negative XP-EHH values indicate selection in the second population (e.g., BF coluzzii).*

362
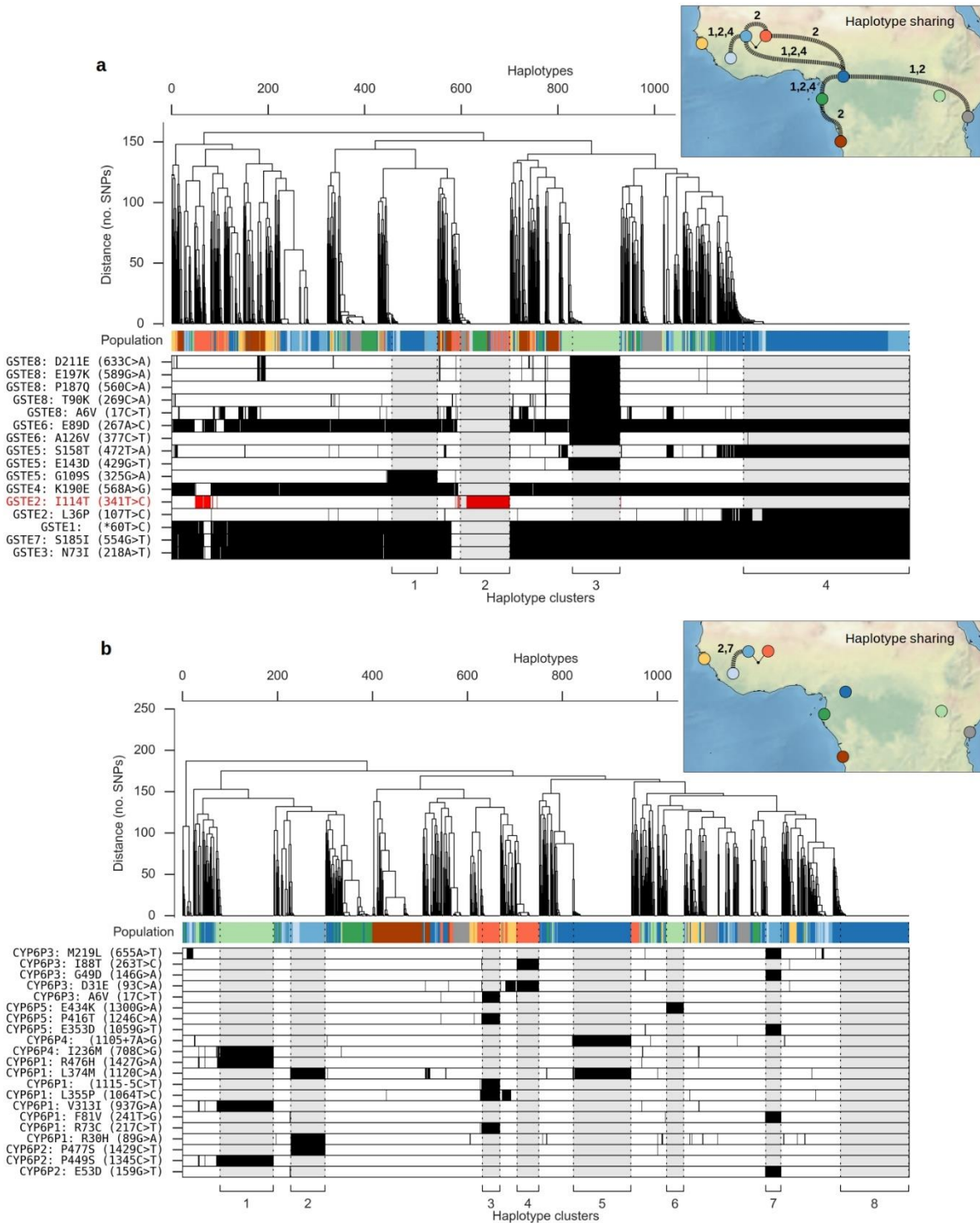
Supplementary Figure 10. Haplotype structure at metabolic insecticide resistance loci. *Plot components are as described for Fig. 5. For both loci, SNPs shown in the lower panel are all either non-synonymous or splice site variants, and are associated with one or more haplotypes under selection. **a**, Haplotype clustering using 1,375 SNPs within the region 3R:28,591,663-28,602,280 spanning 8 genes (Gste1-Gste8). **b**, Haplotype clustering using 1,844 SNPs within the region 2R:28,491,415-28,502,910 spanning 5 genes (Cyp6p1-Cyp6p5).*

363 References

364   1.   Noor, A. M. *et al.* The changing risk of Plasmodium falciparum malaria infection in Africa: 2000-

365   10: a spatial and temporal analysis of transmission intensity. *Lancet (London, England)* **383,**

366   1739–47 (2014).

367   2.   Bhatt, S. *et al.* The effect of malaria control on Plasmodium falciparum in Africa between 2000

368   and 2015. *Nature* **526,** 207–211 (2015).

369   3.   Hemingway, J. *et al.* Averting a malaria disaster: will insecticide resistance derail malaria control?

370   *Lancet* (2016). doi:10.1016/S0140-6736(15)00417-1

371   4.   Churcher, T. S. *et al.* The impact of pyrethroid resistance on the efficacy and effectiveness of

372   bednets for malaria control in Africa. *Elife* **5,** 352 (2016).

373   5.   Gatton, M. L. *et al.* The importance of mosquito behavioural adaptations to malaria control in

374   Africa. *Evolution (N. Y).* **67,** (2013).

375   6.   Raghavendra, K. *et al.* Chlorfenapyr: a new insecticide with novel mode of action can control

376   pyrethroid resistant malaria vectors. *Malar. J.* **10,** 16 (2011).

377   7.   Oxborough, R. M. *et al.* A new class of insecticide for malaria vector control: evaluation of

378   mosquito nets treated singly with indoxacarb (oxadiazine) or with a pyrethroid mixture against

379   Anopheles gambiae and Culex quinquefasciatus. *Malar. J.* **14,** 353 (2015).

380   8.   Windbichler, N. *et al.* A synthetic homing endonuclease-based gene drive system in the human

381   malaria mosquito. *Nature* **473,** 212–215 (2011).

382   9.   Gantz, V. M. *et al.* Highly efficient Cas9-mediated gene drive for population modification of the

383   malaria vector mosquito Anopheles stephensi. *Proc. Natl. Acad. Sci.* 201521077 (2015).

384   doi:10.1073/pnas.1521077112

385   10.   Hammond, A. *et al.* A CRISPR-Cas9 gene drive system targeting female reproduction in the

386   malaria mosquito vector *Anopheles gambiae*. *Nat. Biotechnol.* 1–8 (2015). doi:10.1038/nbt.3439

387   11.   Tene Fossog, B. *et al.* Habitat segregation and ecological character displacement in cryptic African

388   malaria mosquitoes. *Evol. Appl.* **8,** 326–345 (2015).

389   12.   Diabate, A. *et al.* Larval development of the molecular forms of Anopheles gambiae (Diptera:

390   Culicidae) in different habitats: a transplantation experiment. *J Med Entomol* **42,** 548–553 (2005).

391   13.   Gimonneau, G. *et al.* A behavioral mechanism underlying ecological divergence in the malaria

392   mosquito Anopheles gambiae. *Behav. Ecol.* **21,** 1087–1092 (2010).

393   14.   Dao, A. *et al.* Signatures of aestivation and migration in Sahelian malaria mosquito populations.

394   *Nature* **516,** 387–90 (2014).

395   15.   Coetzee, M. *et al.* Anopheles coluzzii and anopheles amharicus, new members of the anopheles

396   gambiae complex. *Zootaxa* **3619,** (2013).

397   16.   Torre, A. della *et al.* Molecular evidence of incipient speciation within Anopheles gambiae s.s. in

398   West Africa. *Insect Mol. Biol.* **10,** 9–18 (2001).

399   17.   Lawniczak, M. K. N. *et al.* Widespread divergence between incipient Anopheles gambiae species

400   revealed by whole genome sequences. *Science* **330,** 512–4 (2010).

401   18.   Neafsey, D. E. *et al.* SNP genotyping defines complex gene-flow boundaries among African

402   malaria vector mosquitoes. *Science* **330,** 514–7 (2010).

403   19.   Aboagye-Antwi, F. *et al.* Experimental Swap of Anopheles gambiae's Assortative Mating

404   Preferences Demonstrates Key Role of X-Chromosome Divergence Island in Incipient Sympatric

405   Speciation. *PLoS Genet.* **11,** e1005141 (2015).

406   20.   Oliveira, E. *et al.* High Levels of Hybridization between Molecular Forms of Anopheles gambiae

407   from Guinea Bissau. *J. Med. Entomol.* **45,** 1057–1063 (2008).

408   21.   Caputo, B. *et al.* The 'far-west' of Anopheles gambiae molecular forms. *PLoS One* **6,** (2011).

409   22.   Weetman, D., Wilding, C. S., Steen, K., Pinto, J. & Donnelly, M. J. Gene flow-dependent genomic

410   divergence between Anopheles gambiae M and S forms. *Mol. Biol. Evol.* **29,** 279–91 (2012).

411   23.   Lee, Y. *et al.* Spatiotemporal dynamics of gene flow and hybrid fitness between the M and S

412   forms of the malaria mosquito, Anopheles gambiae. *Proc. Natl. Acad. Sci. U. S. A.* **110,** (2013).

413   24.   Clarkson, C. S. *et al.* Adaptive introgression between Anopheles sibling species eliminates a major

414   genomic island but not reproductive isolation. *Nat. Commun.* **5,** 4248 (2014).

415   25.   Norris, L. C. *et al.* Adaptive introgression in an African malaria mosquito coincident with the

416   increased usage of insecticide-treated bed nets. *Proc. Natl. Acad. Sci.* 201418892 (2015).

417   doi:10.1073/pnas.1418892112

418   26.   Sharakhova, M. V *et al.* Genome mapping and characterization of the Anopheles gambiae

419   heterochromatin. *BMC Genomics* **11,** 459 (2010).

420   27.   Sharakhova, M. V *et al.* Update of the Anopheles gambiae PEST genome assembly. *Genome Biol.*

421   **8,** R5 (2007).

422   28.   Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform.

423   *Bioinformatics* **25,** 1754–60 (2009).

424   29.   DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation

425   DNA sequencing data. *Nat. Genet.* **43,** 491–8 (2011).

426   30.   Abecasis, G. R. *et al.* A map of human genome variation from population-scale sequencing.

427   *Nature* **467,** 1061–73 (2010).

428   31.   The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092

429   human genomes. *Nature* **491,** 56–65 (2012).

430   32.   Delaneau, O., Howie, B., Cox, A. J., Zagury, J.-F. & Marchini, J. Haplotype estimation using

431   sequencing reads. *Am. J. Hum. Genet.* **93,** 687–96 (2013).

432   33.   Gabriel, S., Ziaugra, L. & Tabbaa, D. SNP genotyping using the Sequenom MassARRAY iPLEX

433   platform. *Curr. Protoc. Hum. Genet.* **Chapter 2,** Unit 2.12 (2009).

434   34.   Vincenten, N. *et al.* The kinetochore prevents centromere-proximal crossover recombination

435   during meiosis. *Elife* **4,** 923–937 (2015).

436   35.   Burri, R. *et al.* Linked selection and recombination rate variation drive the evolution of the

437   genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers.

438   *Genome Res.* **25,** 1656–65 (2015).

439   36.   Chan, A. H., Jenkins, P. A. & Song, Y. S. Genome-wide fine-scale recombination rate variation in

440   Drosophila melanogaster. *PLoS Genet.* **8,** e1003090 (2012).

441   37.   Lack, J. B. *et al.* The Drosophila Genome Nexus: A Population Genomic Resource of 623

442   Drosophila melanogaster Genomes, Including 197 from a Single Ancestral Range Population.

443   *Genetics* genetics.115.174664- (2015). doi:10.1534/genetics.115.174664

444   38.   Martin, S. H. *et al.* Natural Selection and Genetic Diversity in the Butterfly Heliconius

445   Melpomene. *Genetics* (2016).

446   39.   Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial

34

447   immunity. *Science* **337,** 816–21 (2012).

448 40. Turner, T. L., Hahn, M. W. & Nuzhdin, S. V. Genomic islands of speciation in Anopheles gambiae.

449   *PLoS Biol.* **3,** 1572–1578 (2005).

450 41. White, B. J., Cheng, C., Simard, F., Costantini, C. & Besansky, N. J. Genetic association of physically

451   unlinked islands of genomic divergence in incipient species of Anopheles gambiae. *Mol. Ecol.* **19,**

452   925–39 (2010).

453 42. Fontaine, M. C. *et al.* Extensive introgression in a malaria vector species complex revealed by

454   phylogenomics. *Science* science.1258524- (2014). doi:10.1126/science.1258524

455 43. Fanello, C., Santolamazza, F. & Della Torre, A. Simultaneous identification of species and

456   molecular forms of the Anopheles gambiae complex by PCR-RFLP. *Med. Vet. Entomol.* **16,** 461–

457   465 (2002).

458 44. Coluzzi, M., Sabatini, A., della Torre, A., Di Deco, M. A. & Petrarca, V. A polytene chromosome

459   analysis of the Anopheles gambiae species complex. *Science* **298,** 1415–8 (2002).

460 45. Stump, A. D. *et al.* Genetic exchange in 2La inversion heterokaryotypes of Anopheles gambiae.

461   *Insect Mol. Biol.* **16,** 703–9 (2007).

462 46. Davies, T. G. E., Field, L. M., Usherwood, P. N. R. & Williamson, M. S. A comparative study of

463   voltage-gated sodium channels in the Insecta: Implications for pyrethroid resistance in

464   Anopheline and other Neopteran species. *Insect Mol. Biol.* (2007). doi:10.1111/j.1365-

465   2583.2007.00733.x

466 47. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated

467   individuals. *Genome Res.* **19,** 1655–64 (2009).

468   48.   Lehmann, T. *et al.* The Rift Valley Complex as a Barrier to Gene Flow for Anopheles gambiae in

469   Kenya. *J. Hered.* **91,** 165–168 (1999).

470   49.   Lehmann, T. Population Structure of Anopheles gambiae in Africa. *J. Hered.* **94,** 133–147 (2003).

471   50.   Slotman, M. A. *et al.* Evidence for subdivision within the M molecular form of Anopheles

472   gambiae. *Mol. Ecol.* **16,** 639–649 (2006).

473   51.   Pinto, J. *et al.* Geographic population structure of the African malaria vector Anopheles gambiae

474   suggests a role for the forest-savannah biome transition as a barrier to gene flow. *Evol. Appl.* **6,**

475   910–24 (2013).

476   52.   Service, M. W. Mosquito (Diptera: Culicidae) dispersal--the long and short of it. *J Med Entomol*

477   **34,** 579–588 (1997).

478   53.   Costantini, C. *et al.* Density, survival and dispersal of Anopheles gambiae complex mosquitoes in

479   a West African Sudan savanna village. *Med. Vet. Entomol.* **10,** 203–219 (1996).

480   54.   Marsden, C. D. *et al.* Asymmetric introgression between the M and S forms of the malaria vector,

481   Anopheles gambiae, maintains divergence despite extensive hybridization. *Mol. Ecol.* **20,** (2011).

482   55.   Nwakanma, D. C. *et al.* Breakdown in the process of incipient speciation in Anopheles gambiae.

483   *Genetics* **193,** (2013).

484   56.   Caputo, B. *et al.* The last bastion? X chromosome genotyping of Anopheles gambiae species pair

485   males from a hybrid zone reveals complex recombination within the major candidate 'genomic

486   island of speciation'. *Mol. Ecol.* **25,** 5719–5731 (2016).

487   57.   Schraiber, J. G. & Akey, J. M. Methods and models for unravelling human evolutionary history.

488   *Nat. Rev. Genet.* **16,** 727–740 (2015).

489   58.   Liu, X. & Fu, Y.-X. Exploring population size changes using SNP frequency spectra. *Nat. Genet.* **47,**

490   555–559 (2015).

491   59.   Gutenkunst, R. N., Hernandez, R. D., Williamson, S. H. & Bustamante, C. D. Inferring the joint

492   demographic history of multiple populations from multidimensional SNP frequency data. *PLoS*

493   *Genet.* **5,** e1000695 (2009).

494   60.   Keightley, P. D., Ness, R. W., Halligan, D. L. & Haddrill, P. R. Estimation of the spontaneous

495   mutation rate per nucleotide site in a Drosophila melanogaster full-sib family. *Genetics* **196,** 313–

496   20 (2014).

497   61.   Schrider, D. R., Houle, D., Lynch, M. & Hahn, M. W. Rates and genomic consequences of

498   spontaneous mutational events in Drosophila melanogaster. *Genetics* **194,** 937–54 (2013).

499   62.   Grollemund, R. *et al.* Bantu expansion shows that habitat alters the route and pace of human

500   dispersals. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 13296–13301 (2015).

501   63.   Bostoen, K. *et al.* Middle to Late Holocene Paleoclimatic Change and the Early Bantu Expansion in

502   the Rain Forests of Western Central Africa. *Curr. Anthropol.* **56,** 354–384 (2015).

503   64.   Li, S., Schlebusch, C. & Jakobsson, M. Genetic variation reveals large-scale population expansion

504   and migration during the expansion of Bantu-speaking peoples. *Proc. R. Soc. London B Biol. Sci.*

505   **281,** (2014).

506   65.   Gignoux, C. R., Henn, B. M. & Mountain, J. L. Rapid, global demographic expansions after the

507   origins of agriculture. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 6044–9 (2011).

508   66.   Anhuf, D. in *Southern Hemisphere Paleo- and Neoclimates* 225–248 (Springer Berlin Heidelberg,

509   2000). doi:10.1007/978-3-642-59694-0_15

510   67.   Noor, A. M. *et al.* Increasing Coverage and Decreasing Inequity in Insecticide-Treated Bed Net

511   Use among Rural Kenyan Children. *PLoS Med.* **4,** e255 (2007).

512   68.   McQuillan, R. *et al.* Runs of homozygosity in European populations. *Am. J. Hum. Genet.* **83,** 359–

513   72 (2008).

514   69.   Purfield, D. C., Berry, D. P., McParland, S. & Bradley, D. G. Runs of homozygosity and population

515   history in cattle. *BMC Genet.* **13,** 70 (2012).

516   70.   Crawford, J. E. *et al.* Evolution of GOUNDRY, a cryptic subgroup of Anopheles gambiae s.l., and its

517   impact on susceptibility to Plasmodium infection. *Mol. Ecol.* (2016). doi:10.1111/mec.13572

518   71.   O'Loughlin, S. M. *et al.* Genomic signatures of population decline in the malaria mosquito

519   Anopheles gambiae. *Malar. J.* **15,** 182 (2016).

520   72.   Mwangangi, J. M. *et al.* Shifts in malaria vector species composition and transmission dynamics

521   along the Kenyan coast over the past 20 years. *Malar. J.* **12,** 13 (2013).

522   73.   Mogeni, P. *et al.* Age, Spatial, and Temporal Variations in Hospital Admissions with Malaria in

523   Kilifi County, Kenya: A 25-Year Longitudinal Observational Study. *PLOS Med.* **13,** e1002047

524   (2016).

525   74.   Garud, N. R., Messer, P. W., Buzbas, E. O. & Petrov, D. A. Recent Selective Sweeps in North

526   American Drosophila melanogaster Show Signatures of Soft Sweeps. *PLoS Genet.* **11,** 1–32 (2015).

527   75.   Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in human

528   populations. *Nature* **449,** 913–8 (2007).

529   76.   Mitchell, S. N. *et al.* Metabolic and target-site mechanisms combine to confer strong DDT

530   resistance in Anopheles gambiae. *PLoS One* **9,** e92662 (2014).

531   77.   Opondo, K. O. *et al.* Does insecticide resistance contribute to heterogeneities in malaria

532   transmission in The Gambia? *Malar. J.* **15,** 166 (2016).

533   78.   Müller, P. *et al.* Field-caught permethrin-resistant Anopheles gambiae overexpress CYP6P3, a

534   P450 that metabolises pyrethroids. *PLoS Genet.* **4,** e1000286 (2008).

535   79.   Edi, C. V *et al.* CYP6 P450 enzymes and ACE-1 duplication produce extreme and multiple

536   insecticide resistance in the malaria mosquito Anopheles gambiae. *PLoS Genet.* **10,** e1004236

537   (2014).

538   80.   Pinto, J. *et al.* Multiple Origins of Knockdown Resistance Mutations in the Afrotropical Mosquito

539   Vector Anopheles gambiae. *PLoS One* **2,** e1243 (2007).

540   81.   ETANG, J. *et al.* Polymorphism of intron-1 in the voltage-gated sodium channel gene of

541   Anopheles gambiae s.s. populations from Cameroon with emphasis on insecticide knockdown

542   resistance mutations. *Mol. Ecol.* **18,** 3076–3086 (2009).

543   82.   Lynd, A. *et al.* Field, genetic, and modeling approaches show strong positive selection acting upon

544   an insecticide resistance mutation in Anopheles gambiae s.s. *Mol. Biol. Evol.* **27,** 1117–25 (2010).

545   83.   Mitchell, S. N. *et al.* Metabolic and Target-Site Mechanisms Combine to Confer Strong DDT

546   Resistance in Anopheles gambiae. *PLoS One* **9,** e92662 (2014).

547   84.   Langley, C. H. *et al.* Genomic variation in natural populations of Drosophila melanogaster.

548   *Genetics* **192,** 533–98 (2012).

549   85.   Ross, R. Inaugural Lecture on the Possibility of Extirpating Malaria from Certain Localities by a

550   New Method. *Br. Med. J.* **2,** 1–4 (1899).

551   86.   de Souza, D. K. *et al.* Filling the Gap 115 Years after Ronald Ross: The Distribution of the

552   Anopheles coluzzii and Anopheles gambiae s.s from Freetown and Monrovia, West Africa. *PLoS*

553   *One* **8,** e64939 (2013).

554   87.   Sayre, R. G. *et al.* A new map of standardized terrestrial ecosystems of Africa. *African Geogr. Rev.*

555   (2013).

556