1 Pollutants and insecticides drive local adaptation in African malaria
2 mosquitoes
3
4
5
6
7 Colince Kamdem[1*], Caroline Fouet[1], Stephanie Gamez[1], Bradley J. White[1,2,*]
8
9 [1]Department of Entomology, University of California, Riverside, CA 92521
10
11 [2]Center for Disease Vector Research, Institute for Integrative Genome Biology,
12 University of California, Riverside, CA 92521
13
14 [*]Corresponding author: e-mail: colincek@ucr.edu; bwhite@ucr.edu
15
16
17

18   ABSTRACT

19   The *Anopheles gambiae* complex contains a number of highly anthropophilic

20   mosquito species that have acquired exceptional ability to thrive in complex

21   human habitats. Thus, examining the evolutionary history of this Afrotropical

22   mosquito may yield vital information on the selective processes that occurred

23   during the adaptation to human-dominated environments. We performed reduced

24   representation sequencing on 941 mosquitoes of the *Anopheles gambiae*

25   complex collected across four ecogeographic zones in Cameroon. We find

26   evidence for genetic and geographic subdivision within *An. coluzzii* and *An.*

27   *gambiae* sensu *stricto* – the two most significant malaria vectors in the region.

28   Importantly, in both species, rural and urban populations are genetically

29   differentiated. Genome scans reveal pervasive signatures of selection centered

30   on genes involved in xenobiotic resistance. Notably, a selective sweep containing

31   detoxification enzymes is prominent in urban mosquitoes that exploit polluted

32   breeding sites. Overall, our study suggests that recent anthropogenic

33   environmental modifications and widespread use of insecticides are driving

34   population differentiation and local adaptation in vectors with potentially

35   significant consequences for malaria epidemiology.

36

37   INTRODUCTION

38   Humans and their actions on ecosystems can profoundly alter the

39   evolutionary trajectory of other organisms in various ways (Palumbi 2001; Bull

40   and Maron 2016; Hendry et al. 2017). Signatures of human-driven evolutionary

41   changes are particularly evident in cities and urbanizing regions, which often

42   necessitate a rapid adaptation to complex environments (Alberti 2015; Alberti et

43   al. 2017; Hendry et al. 2017). A wealth of data exists on genetic and phenotypic

44   signals associated with adaptation to anthropogenic changes in multiple taxa, but

45   the extent of signatures across genomes remain largely unexplored. The

46   widespread use of xenobiotics in agriculture and public health also increases

47   selective pressures on pathogens and their vectors, prompting significant

48   evolutionary adjustments that have yet to be fully elucidated (Georghiou 1972;

49   Mallet 1989; Palumbi 2001). The application of powerful genomic tools now

50   available will help identify the genes and genomic regions underlying adaptation

51   to human-induced selective pressures.

52   The most effective malaria mosquitoes of the Afrotropical region are

53   remarkably anthropophilic and closely associated with human habitats (White

54   1974; Coluzzi et al. 1979; Antonio-Nkondjio et al. 2006; Sinka et al. 2010; Dia et

55   al. 2013). Therefore, dissecting the genome of contemporary populations of

56   these species may provide significant insights into the genetic targets of recent

57   selection driven by anthropogenic disturbance. Adaptation to human-dominated

58   environments is challenging and very selective even among highly specialized

59   mosquitoes that feed almost exclusively on human hosts. For example, tens of

3

60   *Anopheles* species transmit human malaria parasites across the continent, but

61   only four members of the *Anopheles gambiae* complex thrive in dense urban

62   areas (Antonio-Nkondjio et al. 2005; Oyewole and Awolola 2006; Mourou et al.

63   2010; Dabiré et al. 2012; Kamdem et al. 2012). The recent scaling up of

64   measures using insecticides and insecticide-treated materials to combat malaria

65   expose mosquitoes to an even more disruptive adaptive challenge (Sokhna et al.

66   2013; WHO 2013). Although the genetic basis remains unknown, studies suggest

67   that both physiological and ecological limits and species ranges are profoundly

68   altered in *Anopheles* mosquitoes as a result of rapid urbanization and

69   widespread use of insecticides (Jones et al. 2012; Kamdem et al. 2012;

70   Mwangangi et al. 2013; Tene Fossog et al. 2013; Antonio-Nkondjio et al. 2015).

71   The *Anopheles gambiae* complex is a group of at least nine isomorphic

72   mosquito species exhibiting varying degrees of geographic and reproductive

73   isolation (Riehle et al. 2011; Coetzee et al. 2013; Fontaine et al. 2015). Owing to

74   ongoing adaptive speciation, the principal vectors of the complex have an

75   amazing plasticity in adjusting to diverse ecological conditions, allowing them to

76   track humans across most of tropical Africa (Davidson 1964; Coluzzi et al. 1979;

77   Coluzzi et al. 2002). This adaptive flexibility and the underlying behavioral

78   complexity have undermined at least one major effort at vector control

79   (Molineaux and Gramiccia 1980). Importantly, the epidemiological consequences

80   of adaptive radiation will be even more challenging to tackle if a part of the

81   divergence recently described within the particularly efficient vectors *An.*

82   *gambiae* sensu stricto (hereafter *An. gambiae*) and *An. coluzzii* is driven by

4

83 human-induced changes (Touré et al. 1998; Wondji et al. 2005; Slotman et al.

84 2007; Pinto et al. 2013; Caputo et al. 2014). This hypothesis has been formulated

85 based on ecological observations and population structure, but the genetic

86 underpinnings still need to be clearly assessed by examining patterns of variation

87 across the genome (Coluzzi et al. 2002; Kamdem et al. 2012; Caputo et al.

88 2014).

89 To delineate the population structure and the role of recent processes –

90 mainly urbanization and the rapid scaling up of insecticide-treated bed net

91 coverage – in the genetic differentiation, we genotyped 941 mosquitoes of the

92 *An. gambiae* complex collected from diverse environments in Cameroon at

93 >8,000 SNP. We found strong evidence for complex geographic and genetic

94 structuring giving rise to seven subpopulations distributed along a continuum of

95 genomic differentiation within both *An. gambiae* and *An. coluzzii*. By combining

96 genomic evidence and ecological knowledge, we revealed that the strongest

97 correlates of genetic differentiation in this region are intense suburban

98 agriculture, urbanization, insecticides and forest-savannah subdivisions. These

99 findings add to a growing body of data indicating that the inevitable rise of

100 insecticide resistance together with ongoing local adaptation of *Anopheles*

101 species in urban areas are the major challenges for current vector control

102 strategies.

103 RESULTS

104 *Population structure of* An. gambiae *sensu lato (s.l.) sibling species*

105 We performed extensive sampling of human-associated *Anopheles* across

106 the main ecological zones in Cameroon (Table S1, Figure S1) to collect diverse

107 populations belonging to the four species of the *An. gambiae* complex that are

108 known to occur in the country (Simard et al. 2009). Certain subpopulations or

109 species can be overlooked when sampling is biased toward one type of

110 population (Riehle et al. 2011). To maximize the chances that our samples best

111 represent the genetic diversity within each species, we used several sampling

112 methods (Service 1993) to collect both larvae and adult populations. In addition,

113 populations of *An. gambiae* and *An. coluzzii* segregate along urbanization

114 gradients, which seem to be the most important driver of ecological divergence in

115 the forest zone (Kamdem et al. 2012). To validate this hypothesis and to

116 investigate the genomic targets of local adaptation in urban environments, we

117 surveyed several neighborhoods representing the urban and suburban

118 ecosystems in the two biggest cities of the forest area: Douala and Yaoundé

119 (Figure S1).

120   To assess the genetic relatedness among individuals, we subjected all

121 941 mosquitoes that were morphologically identified as *An. gambiae* s.l. to

122 population genomic analysis. Individual mosquitoes were genotyped in parallel at

123 a dense panel of markers using double-digest restriction associated DNA

124 sequencing (ddRADseq), which enriches for a representative and reproducible

125 fraction of the genome that can be sequenced on the Illumina platform (Peterson

126 et al. 2012). After aligning ddRADseq reads to the *An. gambiae* reference

127 genome, we used STACKS (Catchen et al. 2011; Catchen et al. 2013) to identify

128 8,476 SNPs (~1 SNP every 30kb across the genome) within consensus RAD loci

6

129 present in more than 80% of individuals, and we inferred the population structure

130 based on these variations. First, we performed principal component analysis

131 (PCA) across all 941 individuals (Figure 1A). The top three components explain

132 28.4% of the total variance and group individuals into five main clusters.

133 Likewise, a neighbor-joining (NJ) tree, based on Euclidian distance of allele

134 frequencies, shows five distinct clades of mosquitoes (Figure 1B). We

135 hypothesized that these groups at least partially correspond to the four sibling

136 species – *An. gambiae*, *An. coluzzii*, *An. arabiensis*, and *An. melas* – known to

137 occur in Cameroon. To confirm, we typed a subset of 288 specimens using

138 species identification PCRs (Fanello et al. 2002; Santolamazza et al. 2004) and

139 found that each cluster comprised a single species. In agreement with previous

140 surveys (Wondji et al. 2005; Simard et al. 2009), our collections indicate that the

141 brackish water breeding *An. melas* is limited to coastal regions, while the arid-

142 adapted *An. arabiensis* is restricted to the savannah. In contrast, *An. gambiae*

143 and *An. coluzzii* are distributed across the four eco-geographic zones of

144 Cameroon (Figure 1D). Lee et al. 2013 recently reported frequent bouts of

145 hybridization between *An. gambiae* and *An. coluzzii* in Cameroon. While both the

146 PCA and NJ trees clearly separate the two species, the PCA does show

147 intermixing of some rare individuals consistent with semi-permeable species

148 boundaries.

149 In support of population structuring below the species level, Bayesian

150 clustering analysis with fastSTRUCTURE (Raj et al. 2014) finds that 7 population

151 clusters (*k*) best explain the genetic variance present in our data (Figure 1C,

152   Figure S2). Indeed, grouping of samples within *An. gambiae* and *An. coluzzii*

153   clades suggests that additional subdivision may exist within each species (Figure

154   1A, 1B). Ancestry plots further support inference from the PCA and NJ tree: at

155   least two subgroups compose *An. coluzzii* and *An. gambiae* individuals show

156   admixed ancestry, while *An. arabiensis* and *An. melas* are panmictic (Figure 1C,

157   Figure S2). A cryptic subgroup of *An. gambiae* has been discovered recently by

158   comparing indoor and outdoor fauna from the same village in Burkina Faso

159   (Riehle et al. 2011). Visual inspections of our PCA, NJ and fastSTRUCTURE

160   clustering results do not indicate any genetic subdivision based on the collection

161   methods or the developmental stage. To explicitly test for the effects of the

162   sampling methods and the geographic origin of samples on the genetic variance

163   among individuals, we applied a hierarchical Analysis of Molecular Variance

164   (AMOVA) (Excoffier et al. 1992), which confirmed the absence of genetic

165   structuring based on microhabitats or temporal segregations in both species. The

166   large majority of the genetic variation is attributable to differences among

167   individuals. While the geographic origin accounts for 13.2 % ($p < 0.001$) and

168   9.4% ($p < 0.001$) of the variance in *An. coluzzii* and *An. gambiae* respectively, the

169   effect of the types of sample is marginal and not significant.

170

171   *Population Structure within* An. gambiae

172   To further resolve the population structure within 357 *An. gambiae*

173   specimens, we performed population genetic analysis with a set of 9,345 filtered

174   SNPs. Using a combination of PCA, NJ trees, and ancestry assignment, we

8

175   consistently identify three distinct subgroups within *An. gambiae* (Figure 2, Figure

176   S2). The first and largest subgroup (termed *GAM1*) comprises the vast majority

177   of all *An. gambiae* specimens including individuals collected in all four eco-

178   geographic regions (Table S1). A total of 17 individuals make up a second small

179   subgroup (termed *GAM2*). Interestingly, individuals assigned to this cluster

180   include both larvae and adults collected in 3 different villages spread across 2

181   eco-geographic regions. In the absence of any obvious evidence of niche

182   differentiation between *GAM1* and *GAM2,* it is unclear what is driving and/or

183   maintaining divergence between the two sympatric subgroups. Specimens

184   collected from Nkolondom, a suburban neighborhood of Yaoundé where larval

185   sites associated with small-scale agricultural irrigation are common (Fossog

186   Tene et al. 2013; Nwane et al. 2013), form a genetically distinct third subgroup

187   (termed *Nkolondom*) that appears to be a locally-adapted ecotype.

188

189   *Population Structure within* An. coluzzii

190   To examine population structure within 521 *An. coluzzii* specimens, we

191   utilized 9,822 SNPs that passed stringent filtration. All analyses show a clear split

192   between individuals from the northern savannah region and the southern three

193   forested regions of Cameroon (Coastal, Forest, Forest-Savannah) (Figure 3A-C).

194   In principle, the north-south structuring could be caused solely by differences in

195   chromosome 2 inversion frequencies, which form a cline from near absence in

196   the humid south to fixation in the arid north (Simard et al. 2009; Fouet et al.

197   2012). However, we find SNPs from all five chromosomal arms consistently

9

198 separate northern and southern mosquitoes, indicating a substantial genome-

199 wide divergence between the two populations (Figure S3).

200   Southern populations of *An. coluzzii* were collected from three different

201 areas: Douala (the largest city of Cameroon), Yaoundé (the second largest city)

202 and the rural coastal region. PCA, NJ trees, and fastSTRUCTURE show clear

203 clustering of southern samples by collection site (Figure 3D-F). Mosquitoes from

204 Douala, situated on the coastal border, contain a mixture of urban and coastal

205 polymorphisms as illustrated by their intermediate position along PC3 (Figure

206 3D). Despite considerable geographic segregation, clusters are not fully discrete,

207 likely owing to substantial migration between the three sites. Taken together, the

208 data suggest a dynamic and ongoing process of local adaptation within southern

209 *An. coluzzii*. In contrast, no similar geographic clustering is observed in northern

210 populations (Figure S2, S4). Slight variations can be observed between genetic

211 clustering patterns suggested by different methods. We have only considered

212 subdivisions that were consistent across the three methods we employed. For

213 example, individuals from Yaoundé have been treated as a single subgroup,

214 although PCA showed that a few specimens were relatively detached from the

215 main cluster. This putative subdivision was not supported by the NJ tree and

216 fastSTRUCTURE analyses (Figure 3D-F). All populations described as "urban"

217 were collected from the most urbanized areas of the city, which are characterized

218 by a high frequency of built environments as described in Kamdem *et al*. 2012.

219

220 *Relationships Between Species and Subgroups*

10

221 Population genomic analysis identified four different *An. gambiae* s.l.

222 species present within our samples. Within *An. gambiae* and *An. coluzzii* we

223 identified seven potential subgroups with apparently varying levels of divergence.

224 To further explore the relationships between different populations, we built an

225 unrooted NJ tree using pairwise levels of genomic divergence (F_{ST}) between all

226 species and subgroups (Figure 4, Table S2). As previously observed in a

227 phylogeny based on whole genome sequencing (Fontaine et al. 2015), we find

228 that *An. melas* is highly divergent from all other species (F_{ST} ~ 0.8), while *An.*

229 *arabiensis* shows intermediate levels of divergence (F_{ST} ~ 0.4) from *An. gambiae*

230 and *An. coluzzii*. As expected, the sister species *An. gambiae* and *An. coluzzii*

231 are more closely related to each other (F_{ST} ~ 0.2) than any other species. When

232 examining differentiation between subgroups within *An. coluzzii*, we find that the

233 southern and northern subgroups are relatively divergent (F_{ST} > 0.1), while

234 differentiation between local ecotypes within the south is much lower (F_{ST} <

235 0.04), corroborating patterns of population structure described within this species

236 (Figure 3D-F). Similarly, as suggested by the genetic structuring, the level of

237 differentiation between *An. gambiae* subgroups *GAM1* and *GAM2* is relatively

238 significant (F_{ST} ~ 0.1), while the suburban ecotype from Nkolondom shows a low

239 level of divergence from *GAM1* (F_{ST} ~ 0.05). To further examine the degree of

240 isolation of subgroups within species, we assessed the reductions in observed

241 heterozygosity with respect to that expected under Hardy–Weinberg Equilibrium

242 among *An. coluzzii* and *An. gambiae* populations, by computing the average

243 Wright's inbreeding coefficient, F_{IS}, across genome-wide SNPs. Values of F_{IS}

11

244 close to 1 indicate a deviation from Hardy–Weinberg Equilibrium and the

245 existence of cryptic subdivisions while F_{IS} close to 0 suggest that there are no

246 barriers to gene flow. In spite of the strong population genetic structure observed

247 within *An. coluzzii* and *An. gambiae*, we found surprisingly low genome-wide F_{IS}

248 values (less than 0.0003) in both species, suggesting a lack of assortative

249 mating. Overall, there is no evidence for reproductive isolation within *An.*

250 *gambiae* and *An. coluzzii* despite ongoing local adaptation and significant

251 population differentiation.

252

253 *Signatures of selection*

254 To find potential targets of selection within subgroups, we performed

255 scans of nucleotide diversity (θ_w, θ_π) and allele frequency spectrum (*Tajima's D*)

256 in non-overlapping 150-kb windows across the genome. Scans of θ_w, θ_π, *and*

257 *Tajima's* D were conducted by importing aligned, but otherwise unfiltered, reads

258 directly into ANGSD, which uses genotype likelihoods to calculate summary

259 statistics (Korneliussen et al. 2014). Natural selection can increase the frequency

260 of an adaptive variant within a population, leading to localized reductions in

261 genetic diversity as the haplotype containing the adaptive variant(s) sweeps

262 towards fixation (Maynard Smith and Haigh 1974; Tajima 1989). Selective

263 processes can also promote the coexistence of multiple alleles in the gene pool

264 of a population (balancing selection). Thus, genomic regions harboring targets of

265 selection should exhibit extreme values of diversity and allele frequency spectra

266 relative to genome-wide averages (Storz 2005).

267 We also performed genome scans using both a relative (F_{ST}) and absolute

268 (d_{xy}) measure of divergence calculated with STACKS and *ngsTools* (Fumagalli et

269 al. 2014), respectively. For both diversity and divergence scans we used a

270 maximum of 40 mosquitoes per population, prioritizing individuals with the

271 highest coverage in populations when sample size exceeded 40. In contrast to

272 Tajima's *D* and F_{ST}, the genome-wide distribution of d_{xy} and nucleotide diversity

273 in 150-kb sliding windows yielded relatively noisy patterns due to a large variance

274 (Wakeley 1996) (Figure 5 and 6). As a result, the identification of signatures of

275 selection was based primarily on outliers of Tajima's *D* and F_{ST}. Estimates of d_{xy}

276 and nucleotide diversity were used to confirm genomic locations that were

277 pinpointed as candidate selective sweep on the basis of values of Tajima's *D* and

278 F_{ST}. Precisely, genomic regions were considered as targets of selection if they

279 mapped to significant peaks or depressions of diversity and d_{xy}, and their values

280 of F_{ST} and Tajima's *D* were among the top 1% of the empirical distribution in at

281 least one population. Significantly negative values of Tajima's *D* relative to the

282 genome-wide average suggest an increase in low-frequency mutations due to

283 negative or positive selection whereas significantly positive values of Tajima's *D*

284 indicate a balancing selection.

285

286 *Targets of Selection within* An. gambiae *subgroups*

287 Our estimates of genome-wide diversity levels (Table S3) within *An.*

288 *gambiae* subgroups are comparable to previous values based on RAD

289 sequencing of East African *An. gambiae* s.l. populations (O'Loughlin et al. 2014).

290   As expected, the large *GAM1* population harbors more genetic diversity than the

291   apparently rare *GAM2* population or the geographically restricted Nkolondom

292   ecotype (Table S3, Figure 5A-C). *Tajima's D* is consistently negative across the

293   entire genome of all three subgroups, indicating an excess of low-frequency

294   variants that are likely the result of recent population expansion (Figure 5A-C).

295   Indeed, demographic models infer relatively recent bouts of population expansion

296   in all three subgroups (Table S4).

297   Genome scans of each subgroup reveal genomic regions that show

298   concordant reductions in diversity and allele frequency spectrum consistent with

299   selection (highlighted in Figure 5 A-C). Based on the 1% cutoff of Tajima's *D*

300   (upper and lower bounds) and F_{ST}, we identified 4 candidate regions exhibiting

301   strong signatures of selection. It should be noted that due to the reduced

302   representation sequencing approach we used, our analysis may be highlighting

303   only clear instances of selection, which are likely both recent and strong (Tiffin

304   and Ross-Ibarra 2014). An apparent selective event on the left arm of

305   chromosome 2 near the centromere is found in all populations. This selective

306   sweep is characterized by a prominent depression of Tajima's *D* (~1.5Mb in

307   width) and contains ~80 genes including the pyrethroid knockdown resistance

308   gene (*kdr*). Although it is difficult to precisely identify the specific gene that has

309   been influenced by selection in this region, the voltage-gated sodium channel

310   gene (*kdr*), which is a pervasive target of selection in *Anopheles* mosquitoes

311   (Clarkson et al. 2014; Norris et al. 2015) represents the strongest candidate in

312   our populations. Three mutations of the para-type sodium channel conferring

14

313   knockdown resistance (1014F, 1014L and 1014S) are widespread in Cameroon

314   (Nwane et al. 2011). The drop in Tajima's D in the putative kdr sweep is sharpest

315   in the Nkolondom population. Both 1014F and 1014L kdr alleles coexist at high

316   frequencies in mosquitoes from this suburban neighborhood of Yaoundé where

317   larvae can be readily collected from irrigated garden plots that likely contain

318   elevated levels of pesticides directed at agricultural pests (Nwane et al. 2011;

319   Fossog Tene et al. 2013; Nwane et al. 2013).

320   Another region exhibiting consistent signatures of selection in all

321   populations is found around the centromere on the X chromosome. Functional

322   analyses of gene ontology (GO) terms (Table S5) revealed a significant

323   representation of chitin binding proteins in this sweep (p = 1.91e-4). A strong

324   genetic divergence among the three subgroups of $An.$ $gambiae$ characterized by

325   significant F_{ST} and d_{xy} peaks is also observed at ~30 Mb and ~40 Mb on

326   chromosome 3R. Positive outliers of both Tajima's D and nucleotide diversity

327   suggest that this genetic divergence is due to balancing selection on multiple

328   alleles among populations. Functional analyses of GO terms indicate that the

329   region at ~40 Mb on 3R is enriched in cell membrane proteins, genes involved in

330   olfaction and epidermal growth factors (EGF) genes (Table S5). Finally, a striking

331   depression in Tajima's D supported by a marked dip in nucleotide diversity

332   occurs on chromosome 2L from ~33-35 Mb in the Nkolondom population

333   exclusively (Figure 5A-C). Despite the lack of genetic differentiation, this region –

334   enriched in six EGFs (Table S5) – is probably a recent selective sweep, which

15

335 could facilitate larval development of this subgroup in pesticide-laced agricultural

336 water.

337

338 *Targets of Selection within* An. coluzzii *subpopulations*

339   As above, we used diversity, allele frequency spectra and genetic

340 differentiation metrics to scan for targets of selection in the four subgroups of *An.*

341 *coluzzii*. Overall, genetic diversity is higher in the northern savannah population

342 than either of three southern populations, which all exhibit similar levels of

343 diversity (Table S3). Just as in *An. gambiae*, all subgroups have consistently

344 negative Tajima's *D* values confirming demographic models of population

345 expansion (Table S4).

346   We detected 33 regions where either F_{ST} or Tajima's *D* was above the 99th

347 percentile of the empirical distribution in at least one population. This included at

348 least 10 hotspots of F_{ST} clustered within the 2L*a* inversion, which segregates

349 between forest and savannah populations, as well as a region centered on the

350 resistance to dieldrin (*rdl*) locus at ~25 Mb on 2L (Figure 6H). However, based on

351 the 1% threshold of both Tajima's *D* and F_{ST}, we found only 5 candidate selective

352 sweeps including the *kdr* region, the sweep on the X chromosome, and the two

353 hot spots of balancing selection detected on the chromosome 3R in *An. gambiae*

354 (Figure 6A-D). The fifth putative selective sweep characterized by a sharp drop in

355 both diversity and Tajima's *D* occurs on 3R from ~28.5-29.0 Mb, with the decline

356 being more significant in urban populations. The particularly reduced diversity at

357 this sweep in urban mosquitoes strongly suggests it may contain variant(s) that

16

358 confer adaptation to human-induced changes. Indeed, this genomic region

359 harbors a cluster of both Glutathione S-transferase (*GSTE1-GSTE7*) and

360 cytochrome P450 (*CYP4C27*, *CYP4C35*, *CYP4C36*) genes, and functional

361 analyses of GO terms reveals an overrepresentation of terms containing

362 "Glutathione S-transferase" (Table S5, *p* = 5.14e-10). Both the *GSTE* and

363 cytochrome P450 gene families are known to confer metabolic resistance to

364 insecticides and pollutants in mosquitoes (Enayati et al. 2005; David et al. 2013).

365 In particular, *GSTE5* and *GSTE6* are intriguing candidate targets of selection as

366 each is up-regulated in highly insecticide resistant *An. arabiensis* populations that

367 recently colonized urban areas of Bobo-Dioulasso, Burkina Faso (Jones et al.

368 2012).

369 Both relative and absolute divergences between populations at the two

370 selective sweeps involved in xenobiotic resistance probably reflect a spatial

371 variation in selection along gradients of anthropogenic disturbance. In particular,

372 the *kdr* locus exhibits minimal divergence in all pairwise comparisons, suggesting

373 that the same resistance haplotype is under selection in each population (Figure

374 6E-H). This hypothesis is corroborated by data on the distribution of *kdr*

375 mutations in Cameroon (Nwane et al. 2011). As opposed to *An. gambiae*

376 populations, which are broadly polymorphic for three *kdr* alleles, the 1014L

377 mutation is the major allele present at very high frequencies in *An. coluzzii*

378 throughout the country. The GSTE/CYP450 sweep on 3R shows a peak in F_{ST},

379 but surprisingly low absolute divergence between Yaoundé and Coastal

380 mosquitoes – a pattern consistent with genomic regions that underwent a

17

381   selective sweep over a broad geographic area followed by episodes of within-

382   population selection leading to reduced diversity across certain locations (Irwin et

383   al. 2016). Comparisons between Douala and Coastal populations show a more

384   moderate increase in F_{ST}, presumably due to high rates of migration between

385   these nearby sites. To further explore the 3R GSTE/CYP450 sweep, we

386   reconstructed haplotypes for all 240 *An. coluzzii* southern chromosomes across

387   the 28 SNPs found within the sweep. In the Yaoundé population, a single

388   haplotype is present on 44 out of 80 (55%) chromosomes (all grey SNPs), while

389   an additional 11 haplotypes are within one mutational step of this common

390   haplotype (Figure 7A). In Douala, the same haplotype is the most common, but

391   present at a lower frequency (31%) than in Yaoundé (Figure 7B). This haplotype

392   is found on only 6/80 (7.5%) coastal chromosomes (Figure 7C). Population

393   genomic analyses of the same 28 SNPs (Figure 7D-F) mirror results of the

394   haplotype analysis, confirming that haplotype inference did not bias the results.

395

396   DISCUSSION

397   *Genetic differentiation within species of the* An. gambiae *complex*

398   The *An. gambiae* complex is a model of adaptive radiation with a puzzling

399   evolutionary history (Coluzzi et al. 2002; Ayala and Coluzzi 2005). Due to the

400   lack of observable phenotypes that segregate between populations, it is almost

401   impossible to tackle the genetic basis of divergence between and within species

402   of the complex using traditional genetic mapping methods. Consequently, the

403   examination of sequence divergence in the wild with high-throughput sequencing

404   and genotyping approaches is particularly adapted to address the environmental

405   and genomic targets of ecological divergence in *An. gambiae* s.l. In that respect,

406   patterns of genomic divergence have started to be dissected and significant

407   insights have been gained into the signatures of selection among continental

408   populations (Lawniczak et al. 2010; Neafsey et al. 2010; White et al. 2011).

409   However, ecological divergence and local adaptation within species of the *An.*

410   *gambiae* complex sometimes occur at lower spatial scales (often a few

411   kilometres) (Sogoba et al. 2008; Kamdem et al. 2012) that are ignored by

412   continental-scale analyses. Here we have applied a population genomic

413   approach to investigate the genomic architecture of selection across a more

414   restricted geographic area characterized by a great environmental diversity. We

415   first showed that reduced representation sequencing of 941 *An. gambiae* s.l.

416   collected in or near human settlements in 33 sites scattered across Cameroon

417   facilitated rapid identification of known sibling species and revealed a complex

418   hierarchical population structure within *An. gambiae* and *An. coluzzii*. This result

19

419   is opposite to that found in East Africa, where RADseq markers detected no

420   population differentiation within species of the *An. gambiae* complex (O'Loughlin

421   et al. 2014). Indeed, contrary to their East African counterparts, West African

422   populations have a long history of genetic differentiation that is still unresolved

423   despite the application of several generations of genetic markers (Coluzzi et al.

424   1979; Touré et al. 1998; Coluzzi et al. 2002; Wondji et al. 2002). Polymorphic

425   chromosomal inversions and microsatellite markers have suggested the

426   presence of additional subdivision within *An. coluzzii* in Cameroon (Wondji et al.

427   2005; Slotman et al. 2007). Using genome-wide SNPs, we have shown that this

428   subdivision results in at least three genetically distinct clusters including

429   populations from the savannah, coastal and urban areas. Little is known about

430   the phenotypic differentiation of these populations, but recent studies suggest

431   that mosquitoes from the coastal region tolerate higher concentrations of salt

432   (Tene Fossog et al. 2015) while urban populations are highly resistant to

433   insecticides and environmental pollutants (Antonio-Nkondjio et al. 2011; Fossog

434   Tene et al. 2013; Tene Fossog et al. 2013; Antonio-Nkondjio et al. 2015). Models

435   of ecological niches of *An. gambiae* s.l. in Cameroon predict that favourable

436   habitats of *An. coluzzii* populations are fragmented and marginal in contrast to

437   *An. gambiae*, which occupies a broad range across the country (Simard et al.

438   2009). In line with this prediction, we have found a weak differentiation and

439   higher diversity in *An. gambiae* over significant geographic areas despite the

440   presence of an emerging suburban cluster in Nkolondom.

441

20

442   *Genomic signatures of selection*

443   When approached with powerful tools, sensitive enough to capture

444   shallow and recent processes, populations depicting increasing levels of genetic

445   differentiation along a speciation continuum may provide insights into the targets

446   of selection at early stages of ecological divergence (Savolainen et al. 2013). In

447   principle, weakly differentiated populations such as the subgroups we described

448   typically display a marginal genomic divergence because there has not been

449   enough time for neutral and selective forces to shape the genomic architecture

450   (Nosil and Feder 2012; Andrew and Rieseberg 2013). Indeed, we found

451   signatures consistent with selection only across a handful of loci in structured

452   populations of *An. gambiae* and *An. coluzzii*. Although some targets of selection

453   are likely missing due to the limited genomic coverage of the RADseq approach

454   we used (Arnold et al. 2013; Tiffin and Ross-Ibarra 2014), the few selective

455   sweeps we found are relevant based on the genes they contain. Specifically, the

456   five main targets of selection in our populations are enriched in genes whose

457   functions include insecticide resistance and detoxification, epidermal growth,

458   cuticle formation and olfaction. The cuticle plays a major role at the interface

459   between several biological functions in insects and recent data support the

460   hypothesis that the coevolution of multiple mechanisms, including cuticular

461   barriers, has occurred in highly pyrethroid-resistant *An*. *gambiae* (Balabanidou et

462   al. 2016). Olfaction also mediates a wide range of both adult and larval

463   behaviours including feeding, host preference, and mate selection in blood-

464   feeding insects (Bowen 1991; Carey et al. 2010; Takken and Verhulst 2013).

465   Finally, due to the extreme association between the most effective vectors of the

466   *An. gambiae* complex and human habitats, a long-standing hypothesis has

467   suggested that human-driven selection is the main modulator of ecological

468   divergence between and within species (Coluzzi et al. 1979; Coluzzi et al. 2002;

469   Kamdem et al. 2012; Caputo et al. 2014). Consistent with this hypothesis, we

470   have found that urban and suburban mosquitoes are genetically differentiated

471   from rural populations and exhibit strong signatures of recent selection at loci

472   containing xenobiotic resistance genes.

473   *Anopheles* mosquitoes currently experience strong selective pressures

474   across the African continent due to the increased use of pesticides/insecticides in

475   agriculture and vector control over the last few decades (WHO 2012; Reid and

476   McKenzie 2016). The idea that insecticide resistance can affect the population

477   structure of mosquitoes as a result of a limited gene flow and/or local adaptation

478   between resistant and sensitive populations was already evoked in the 1990s

479   (Silvestrini et al. 1998; Lenormand et al. 1999). In addition to increased

480   resistance, large-scale exposure of vector populations to insecticide-treated bed

481   nets in Africa is also correlated with changes in species distribution and

482   behavioral shifts (Bøgh et al. 1998; Derua et al. 2012; Moiroux et al. 2012;

483   Mwangangi et al. 2013; Sokhna et al. 2013). However, so far, there is no

484   evidence that these changes can lead to genetic clustering or the creation of

485   cryptic populations within species (Main et al. 2016). Our data also support that

486   the genetic differentiation within the *An. gambiae* complex in Cameroon is due to

487   regional-scale subdivisions that do not correspond to known distributional or

22

488 behavioural shifts. Previous studies have documented significant reductions in

489 the population size of *An. gambiae* s.l. after introduction of long lasting

490 insecticide treated nets, but did not determine the influence of exposure on

491 population or genomic structuring (Athrey et al. 2012). The demographic history

492 of subgroups of *An. gambiae* and *An. coluzzii* inferred from allele frequency

493 spectra does not indicate signals of recent population bottlenecks that can be

494 associated with insecticide exposure (supplementary information, Table S4). It

495 has also been shown that adaptive introgression, allowing genomic regions

496 containing insecticide resistance genes to spread among sibling species of *An.*

497 *gambiae* s.l., coincides with major insecticide-treated net distribution campaigns

498 (Clarkson et al. 2014; Norris et al. 2015). Although the frequency of natural

499 hybrids is low (less than 2%), episodic gene flow occurs between species of the

500 complex in Cameroon (Lee et al. 2013), but adaptive introgression at insecticide

501 resistance loci have yet to be explicitly demonstrated in this region. Insecticide

502 resistance, involving both multiple alleles of *kdr* and significant levels of

503 metabolic resistance, is widespread among *An. gambiae* and *An. coluzzii*

504 populations in Cameroon especially in urban areas and locations where intensive

505 farming practices are common (Ndjemaï et al. 2009; Antonio-Nkondjio et al.

506 2011; Nwane et al. 2011; Fossog Tene et al. 2013; Nwane et al. 2013; Antonio-

507 Nkondjio et al. 2015). Alongside a mass distribution of insecticide-treated nets

508 throughout the country that was held in 2011, indoor residual insecticide spraying

509 has been tested in a few pilot locations, but this approach is not yet applied on a

510 large scale (Etang et al. 2011; Bowen 2013). We hypothesized that regional

23

511   variations in insecticide exposure exist across the country and contribute to the

512   patterns observed at the two selective sweeps associated with xenobiotic

513   resistance. Knockdown and metabolic resistance involving some of the genes

514   present in the 3R GSTE/CYP450 sweep are the most common resistance

515   mechanisms against the classes of insecticides currently used on a large scale

516   that are found in *An. gambiae* s.l (Nwane et al. 2013). As expected, urban and

517   suburban populations that are exposed to higher levels of insecticides/pollutants

518   exhibit stronger signals of selection (Kamdem et al. 2012; Fossog Tene et al.

519   2013; Tene Fossog et al. 2013). Indeed, the synergistic effects of the two types

520   of xenobiotics could be exerting intense selection pressure for pleiotropic

521   resistance in urban mosquitoes (Nwane et al. 2013). Further analysis of the 3R

522   GSTE/CYP450 sweep using a combination of whole genome resequencing and

523   functional genomic approaches should help resolve the specific genes that are

524   targets of local adaptation in this region of the genome. Additional studies are

525   also needed to further characterize the spatial and temporal variations as well as

526   the haplotypes correlated with insecticide-induced selection within the sweeps.

527   Effective strategies for managing insecticide resistance that has reached a

528   critical tipping point among malaria vectors on the continent are urgently needed

529   (Hemingway et al. 2016; Ranson and Lissenden 2016). In addition, considerable

530   evidence suggests that insecticide resistance has much deeper consequences

531   on the mosquito biology by interfering with other key biological functions such as

532   the immune system, which in turn profoundly affects the fitness of the vector and

533   its efficiency to transmit the pathogen (vectorial capacity) (Alout et al. 2013; Alout

534   et al. 2014; Alout et al. 2016). As a result, detailed knowledge of genomic

535   signatures of insecticide resistance and their pleiotropic effects on vectorial

536   capacity is necessary to design effective resistance management. Surprisingly,

537   despite intense research on insecticide resistance, this information is lacking in

538   all malaria vectors partly because efforts have so far focused on a few resistance

539   genes and have failed to capture the widespread effects of insecticide resistance

540   across the genome. The population genomic approaches we have used are

541   among the new tools that are being applied to investigate the genomic targets of

542   insecticide resistance at a fine scale in mosquitoes and will facilitate the

543   development of rational resistance management strategies.

544

545 *MATERIALS AND METHODS*

546

547 *Mosquito collections*

548   We collected *Anopheles* mosquitoes from 33 locations spread across the

549 four major ecogeographic regions of Cameroon from August to November 2013

550 (Table S1). All collections were performed by researchers who were given

551 malaria prophylaxis before and after the collection period. Indoor resting adult

552 mosquitoes were collected by pyrethrum spray catch, while host-seeking adults

553 were obtained via indoor/outdoor human-baited landing catch. Larvae were

554 sampled using standard dipping procedures (Service 1993). Individual

555 mosquitoes belonging to the *An. gambiae* complex were identified using

556 morphological keys (Gillies and De Meillon 1968; Gillies and Coetzee 1987).

557

558 *ddRADseq Library Construction*

559   Genomic DNA was extracted from adults using the ZR-96 Quick-gDNA kit

560 (Zymo Research) and from larvae using the DNeasy Extraction kit (Qiagen). A

561 subset of individuals were assigned to sibling species using PCR-RFLP assays

562 that type fixed SNP differences in the rDNA (Fanello et al. 2002; Santolamazza

563 et al. 2004). Preparation of ddRAD libraries largely followed (Turissini et al.

564 2014). Briefly, ~1/3rd of the DNA extracted from an individual mosquito (10µl) was

565 digested with *MluC1* and *NlaIII* (New England Biolabs). Barcoded adapters (1 of

566 48) were ligated to overhangs and 400 bp fragments were selected using 1.5%

567 gels on a BluePippin (Sage Science). One of six indices was added during PCR

26

568 amplification. Each library contained 288 individuals and was subjected to single

569 end, 100 bp sequencing across one or two flow cells lanes run on an Illumina

570 HiSeq2500.

571 Raw sequence reads were demultiplexed and quality filtered using the

572 STACKS v 1.29 process_radtags pipeline (Catchen et al. 2011; Catchen et al.

573 2013). After removal of reads with ambiguous barcodes, incorrect restriction

574 sites, and low sequencing quality (mean Phred < 33), GSNAP (Wu and Nacu

575 2010) was used to align reads to the *An. gambiae* PEST reference genome

576 (AgamP3) allowing up to five mismatches per read. The STACKS pipeline was

577 then used to identify unique RAD tags and construct consensus assemblies for

578 each and to call individual SNP genotypes using a maximum-likelihood statistical

579 model.

580

581 *Population Genomic Analysis*

582 Population genetic structure was assessed using the SNP dataset output

583 by the *populations* program of STACKS. We used PLINK v 1.19 to retrieve

584 subsets of genome-wide SNPs as needed (Purcell et al. 2007). PCA, neighbor-

585 joining tree analyses, and Bayesian information criterion (BIC) were implemented

586 using the packages *adegenet* and *ape* in R (Paradis et al. 2004; Jombart 2008; R

587 Development Core Team 2014). Ancestry analyses were conducted in

588 fastSTRUCTURE v 1.0 (Raj et al. 2014) using the logistic method. The

589 choosek.py script was used to find the appropriate number of populations (k); in

590 cases where a range of k was suggested, the BIC-inferred number of clusters

591   was chosen. CLUMPP v1.1.2 (Jakobsson and Rosenberg 2007) was used to

592   summarize assignment results across independent runs and DISTRUCT v1.1

593   (Rosenberg 2004) was used to visualize ancestry assignment of individual

594   mosquitoes. We used a subset of 1,000 randomly chosen SNPs to calculate

595   average pairwise F_{ST} between populations in GENODIVE v 2.0 using up to 40

596   individuals – prioritized by coverage – per population (Meirmans and Van

597   Tienderen 2004). Using this same subset of 1,000 SNPs, we conducted an

598   AMOVA (Excoffier et al. 1992) to quantify the effect of the sampling method and

599   the geographic origin on the genetic variance among individuals in GENODIVE.

600   We used 10,000 permutations to assess significance of F_{ST} values and AMOVA.

601   We input pairwise F_{ST} values into the program Fitch from the Phylip (Plotree and

602   Plotgram 1989) suite to create the population-level NJ tree. F_{IS} values were

603   computed with the *populations* program in STACKS.

604

605   *Genome Scans for Selection*

606   We used ANGSD v 0.612 (Korneliussen et al. 2014) to calculate

607   nucleotide diversity (θ_w and θ_π) and Tajima's *D* in 150-kb non-overlapping

608   windows. Unlike most genotyping algorithms, ANGSD does not perform hard

609   SNP calls, instead taking genotyping uncertainty into account when calculating

610   summary statistics. Similarly, absolute divergence (d_{xy}) was calculated using

611   *ngsTools* (Fumagalli et al. 2014) based on genotype likelihoods generated by

612   ANGSD. Kernel smoothed values for 150-kb windows for all four metrics ($\theta_w, \theta_\pi,$

613   *D, d_{xy}*) were obtained with the R package *KernSmooth*. F_{ST} (based on AMOVA)

28

614   was calculated with the *populations* program in STACKS using only loci present

615   in 80% of individuals. A Kernel smoothing procedure implemented in STACKS

616   was used to obtain F_{ST} values across 150-kb windows. Because regions with

617   unusually high or low read depth can yield unreliable estimates of diversity and

618   divergence parameters due to the likelihood of repeats and local misassembly,

619   we checked that the average per-locus sequencing coverage was consistent

620   throughout the genome (Figure S5). To determine if selective sweeps were

621   enriched for specific functional annotation classes, we used the program DAVID

622   6.7 with default settings (Huang et al. 2009). We delimitated the physical limits of

623   the selective sweep as the region corresponding to the base of the peak or the

624   depression of Tajima's *D*. Haplotypes across the GSTE/CYP450 sweep were

625   reconstructed by PHASE v 2.1.1 using the default recombination model

626   (Stephens et al. 2001; Stephens and Scheet 2005).

627

637

638

639   AUTHOR CONTRIBUTIONS

640   Conceived and designed the experiments: CK BJW. Performed the experiments:

641   CK BJW SG. Analyzed the data: CK CF BJW. Wrote the paper: CK CF BJW.

642

643   REFERENCES

644   Alberti M. 2015. Eco-evolutionary dynamics in an urbanizing planet. Trends Ecol.

645   Evol. 30:114–126.

646   Alberti M, Marzluff J, Hunt VM. 2017. Urban driven phenotypic changes:

647   empirical observations and theoretical implications for eco-evolutionary

648   feedback. Philos. Trans. R. Soc. B Biol. Sci. 372:20160029.

649   Alout H, Dabiré RK, Djogbénou L, Abate L, Corbel V, Chandre F, Cohuet A.

650   2016. Interactive cost of *Plasmodium* infection and insecticide resistance in

651   the malaria vector *Anopheles gambiae*. Sci. Rep.:1–11.

652   Alout H, Djègbè I, Chandre F, Djogbénou LS, Dabiré RK, Corbel V, Cohuet A.

653   2014. Insecticide exposure impacts vector– parasite interactions in

654   insecticide- resistant malaria vectors. Proc. R. Soc. B Biol. Sci.

655   281:20140389.

656   Alout H, Ndam NT, Sandeu MM, Djégbe I, Chandre F, Dabiré RK, Djogbénou LS,

657   Corbel V, Cohuet A. 2013. Insecticide Resistance Alleles Affect Vector

658   Competence of Anopheles gambiae s.s. for Plasmodium falciparum Field

659   Isolates. PLoS One 8.

660   Andrew RL, Rieseberg LH. 2013. Divergence is focused on few genomic regions

661   early in speciation: incipient speciation of sunflower ecotypes. Evolution

662   67:2468–2482.

663   Antonio-Nkondjio C, Fossog BT, Ndo C, Djantio BM, Togouet SZ, Awono-

664   Ambene P, Costantini C, Wondji CS, Ranson H. 2011. Anopheles gambiae

665   distribution and insecticide resistance in the cities of Douala and Yaoundé

666   (Cameroon): influence of urban agriculture and pollution. Malar. J. 10:154.

667 Antonio-Nkondjio C, Kerah CH, Simard F, Awono-Ambene P, Chouaibou M,

668   Tchuinkam T, Fontenille D. 2006. Complexity of the malaria vectorial system

669   in Cameroon: contribution of secondary vectors to malaria transmission. J.

670   Med. Entomol. 43:1215–1221.

671 Antonio-Nkondjio C, Simard F, Awono-Ambene P, Ngassam P, Toto JC,

672   Tchuinkam T, Fontenille D. 2005. Malaria vectors and urbanization in the

673   equatorial forest region of south Cameroon. Trans. R. Soc. Trop. Med. Hyg.

674   99:347–354.

675 Antonio-Nkondjio C, Tene Fossog B, Kopya E, Poumachu Y, Menze Djantio B,

676   Ndo C, Tchuinkam T, Awono-Ambene P, Wondji CS. 2015. Rapid evolution

677   of pyrethroid resistance prevalence in Anopheles gambiae populations from

678   the cities of Douala and Yaoundé (Cameroon). Malar. J. 14.

679 Arnold B, Corbett-Detig RB, Hartl D, Bomblies K. 2013. RADseq underestimates

680   diversity and introduces genealogical biases due to nonrandom haplotype

681   sampling. Mol. Ecol. 22:3179–3190.

682 Athrey G, Hodges TK, Reddy MR, Overgaard HJ, Matias A, Ridl FC,

683   Kleinschmidt I, Caccone A, Slotman M a. 2012. The effective population size

684   of malaria mosquitoes: large impact of vector control. PLoS Genet.

685   8:e1003097.

686 Ayala FJ, Coluzzi M. 2005. Chromosome speciation: humans, Drosophila, and

687   mosquitoes. Proc. Natl. Acad. Sci. U. S. A. 102 Suppl:6535–6542.

688   Balabanidou V, Kampouraki A, MacLean M, Blomquist GJ, Tittiger C, Juárez MP,

689   Mijailovsky SJ, Chalepakis G, Anthousi A, Lynd A, et al. 2016. Cytochrome

690   P450 associated with insecticide resistance catalyzes cuticular hydrocarbon

691   production in *Anopheles gambiae*. Proc. Natl. Acad. Sci. 113:201608295.

692   Bøgh C, Pedersen EM, Mukoko D a, Ouma JH. 1998. Permethrin-impregnated

693   bednet effects on resting and feeding behaviour of lymphatic filariasis vector

694   mosquitoes in Kenya. Med. Vet. Entomol. 12:52–59.

695   Bowen HL. 2013. Impact of a mass media campaign on bed net use in

696   Cameroon. Malar. J. 12:36.

697   Bowen MF. 1991. The sensory physiology of host-seeking behavior in

698   mosquitoes. Annu. Rev. Entomol. 36:139–158.

699   Bull JW, Maron M. 2016. How humans drive speciation as well as extinction.

700   Proc. R. Soc. B 283:1–10.

701   Caputo B, Nwakanma D, Caputo FP, Jawara M, Oriero EC, Hamid-Adiamoh M,

702   Dia I, Konate L, Petrarca V, Pinto J, et al. 2014. Prominent intraspecific

703   genetic divergence within Anopheles gambiae sibling species triggered by

704   habitat discontinuities across a riverine landscape. Mol. Ecol. 23:4574–4589.

705   Carey AF, Wang G, Su CY, Zwiebel LJ, Carlson JR. 2010. Odorant reception in

706   the malaria mosquito Anopheles gambiae. Nature 464:66–71.

707   Catchen J, Hohenlohe P a., Bassham S, Amores A, Cresko W. 2013. Stacks: An

708   analysis tool set for population genomics. Mol. Ecol. 22:3124–3140.

34

709   Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks:

710   building and genotyping Loci de novo from short-read sequences. G3

711   (Bethesda). 1:171–182.

712   Clarkson CS, Weetman D, Essandoh J, Yawson AE, Maslen G, Manske M, Field

713   SG, Webster M, Antão T, MacInnis B, et al. 2014. Adaptive introgression

714   between Anopheles sibling species eliminates a major genomic island but

715   not reproductive isolation. Nat. Commun. 5:4248.

716   Coetzee M, Hunt RH, Wilkerson R, Torre A Della, Coulibaly MB, Besansky NJ.

717   2013. Anopheles coluzzii and Anopheles amharicus, new members of the

718   Anopheles gambiae complex. 3619:246–274.

719   Coluzzi M, Sabatini A, Petrarca V, Di Deco M. 1979. Chromosomal differentiation

720   and adaptation to human environments in the Anopheles gambiae complex.

721   Trans R Soc Trop Med Hyg 73:483–497.

722   Coluzzi M, Sabatini A, Torre A, Angela M, Deco D, Petrarca V. 2002. A Polytene

723   Chromosome Analysis of the Anopheles gambiae Species Complex.

724   Science (80-.). 298:1415–1419.

725   Dabiré RK, Namountougou M, Sawadogo SP, Yaro LB, Toé HK, Ouari A,

726   Gouagna L-C, Simard F, Chandre F, Baldet T, et al. 2012. Population

727   dynamics of Anopheles gambiae s.l. in Bobo-Dioulasso city : bionomics,

728   infection rate and susceptibility to insecticides. Parasit. Vectors 5:127.

729   David J, Ismail HM, Chandor-proust A, Ingraham MJ, John M, Paine I. 2013.

730   Role of cytochrome P450s in insecticide resistance : impact on the control of

731   mosquito-borne diseases and use of insecticides on Earth Role of

732   cytochrome P450s in insecticide resistance : impact on the control of

733   mosquito-borne diseases and use of insecti. Philos. Trans. R. Soc. Lond. B.

734   Biol. Sci.

735   Davidson G. 1964. The five mating types of the Anopheles gambiae complex.

736   Riv. Malariol. 13:167–183.

737   Derua Y a, Alifrangis M, Hosea KM, Meyrowitsch DW, Magesa SM, Pedersen

738   EM, Simonsen PE. 2012. Change in composition of the Anopheles gambiae

739   complex and its possible implications for the transmission of malaria and

740   lymphatic filariasis in north-eastern Tanzania. Malar. J. 11:188.

741   Dia I, Guelbeogo M, Ayala D. 2013. Advances and Perspectives in the Study of

742   the Malaria Mosquito Anopheles funestus.

743   Enayati AA, Ranson H, Hemingway J. 2005. Insect glutathione transferases and

744   insecticide resistance. Insect Mol. Biol. 14:3–8.

745   Etang J, Nwane P, Mbida JA, Piameu M, Manga B, Souop D, Awono-Ambene P,

746   Ahmed AM, Braimah N, Drakeley C, et al. 2011. Variations of insecticide

747   residual bio-efficacy on different types of walls: results from a community-

748   based trial in south Cameroon. Malar. J. 10:333.

749   Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance

750   inferred from metric distances among DNA haplotypes: Application to human

751   mitochondrial DNA restriction data. Genetics 131:479–491.

752   Fanello C, Santolamazza F, DellaTorre A. 2002. Simultaneous identification of

753   species and molecular forms of the Anopheles gambiae complex by PCR-
754   RFLP. Med. Vet. Entomol. 16:461–464.

755 Fontaine MC, Pease JB, Steele a., Waterhouse RM, Neafsey DE, Sharakhov I
756   V., Jiang X, Hall a. B, Catteruccia F, Kakani E, et al. 2015. Extensive
757   introgression in a malaria vector species complex revealed by
758   phylogenomics. Science (80-.). 347:1258524–1258524.

759 Fossog Tene B, Poupardin R, Costantini C, Awono-Ambene P, Wondji CS,
760   Ranson H, Antonio-Nkondjio C. 2013. Resistance to DDT in an Urban
761   Setting: Common Mechanisms Implicated in Both M and S Forms of
762   Anopheles gambiae in the City of Yaoundé Cameroon. PLoS One 8.

763 Fouet C, Gray E, Besansky NJ, Costantini C. 2012. Adaptation to aridity in the
764   malaria mosquito Anopheles gambiae: chromosomal inversion
765   polymorphism and body size influence resistance to desiccation. PLoS One
766   7:e34841.

767 Fumagalli M, Vieira FG, Linderoth T. 2014. ngsTools : methods for population
768   genetics analyses from Next-Generation Sequencing data. Bioinformatics.

769 Georghiou GP. 1972. The evolution of resistance to pesticides. Annu. Rev. Ecol.
770   Syst. 3:133–168.

771 Gillies MT, Coetzee M. 1987. A supplement to the Anophelinae of Africa south of
772   the Sahara. Johannesburg: The South African Institute for Medical Research

773 Gillies MT, De Meillon B. 1968. The Anophelinae of Africa South of the Sahara.
774   Second Edi. Johannesburg: Publications of the South African Institute for

775   Medical Research

776   Hemingway J, Ranson H, Magill A, Kolaczinski J, Fornadel C, Gimnig J, Coetzee

777   M, Simard F, Roch DK, Hinzoumbe CK, et al. 2016. Averting a malaria

778   disaster: Will insecticide resistance derail malaria control? Lancet 387:1785–

779   1788.

780   Hendry AP, Gotanda KM, Svensson EI. 2017. Human influences on evolution,

781   and the ecological a nd societal consequences.

782   Huang DW, Sherman BT, Lempicki RA. 2009. Systematic and integrative

783   analysis of large gene lists using DAVID bioinformatics resources. Nat.

784   Protoc 4:44–57.

785   Irwin DE, Alcaide M, Delmore KE, Irwin JH, Owens GL. 2016. Recurrent

786   selection explains genomic regions of high relative but low absolute

787   differentiation in the greenish warbler ring species. Mol. Ecol. 25:4488–4507.

788   Jakobsson M, Rosenberg N. 2007. CLUMPP: a cluster matching and

789   permutation program for dealing with multimodality in analysis of population

790   structure. Bioinformatics 23:1801– 1806.

791   Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic

792   markers. Bioinformatics 24:1403–1405.

793   Jones CM, Toé HK, Sanou A, Namountougou M, Hughes A, Diabaté A, Dabiré

794   R, Simard F, Ranson H. 2012. Additional Selection for Insecticide

795   Resistance in Urban Malaria Vectors: DDT Resistance in Anopheles

796   arabiensis from Bobo-Dioulasso, Burkina Faso. PLoS One 7:e45995.

797 Kamdem C, Tene Fossog B, Simard F, Etouna J, Ndo C, Kengne P, Boussès P,

798   Etoa F-X, Awono-Ambene P, Fontenille D, et al. 2012. Anthropogenic

799   Habitat Disturbance and Ecological Divergence between Incipient Species of

800   the Malaria Mosquito Anopheles gambiae. PLoS One 7:e39453.

801 Korneliussen T, Albrechtsen A, Nielsen R. 2014. ANGSD: Analysis of Next

802   Generation Sequencing Data. BMC Bioinformatics 15:356.

803 Lawniczak MKN, Emrich SJ, Holloway a K, Regier a P, Olson M, White B,

804   Redmond S, Fulton L, Appelbaum E, Godfrey J, et al. 2010. Widespread

805   divergence between incipient Anopheles gambiae species revealed by

806   whole genome sequences. Science 330:512–514.

807 Lee Y, Marsden CD, Norris LC, Collier TC, Main BJ, Fofana A, Cornel AJ,

808   Lanzaro GC. 2013. Spatiotemporal dynamics of gene flow and hybrid fitness

809   between the M and S forms of the malaria mosquito, Anopheles gambiae.

810   Proc. Natl. Acad. Sci. U. S. A. 110:19854–19859.

811 Lenormand T, Bourguet D, Guillemaud T, Raymond M. 1999. Tracking the

812   evolution of insecticide resistance in the mosquito Culex pipiens. Nature

813   400:861–864.

814 Main BJ, Lee Y, Ferguson HM, Kreppel KS, Kihonda A, Govella NJ, Collier TC,

815   Cornel AJ, Eskin E, Kang EY, et al. 2016. The Genetic Basis of Host

816   Preference and Resting Behavior in the Major African Malaria Vector,

817   Anopheles arabiensis. PLoS Genet. 12:1–17.

818 Mallet J. 1989. The evolution of insecticide resistance: Have the insects won?

819   Trends Ecol. Evol. 4:336–340.

820   Maynard Smith J, Haigh J. 1974. The hitch-hiking effect of a favorable gene.

821   Genet Res 23:22–35.

822   Meirmans P, Van Tienderen P. 2004. GENOTYPE and GENODIVE: two

823   programs for the analysis of genetic diversity of asexual organisms. Mol.

824   Ecol. Notes 4:792–794.

825   Moiroux N, Gomez MB, Pennetier C, Elanga E, Djènontin A, Chandre F, Djègbé

826   I, Guis H, Corbel V. 2012. Changes in anopheles funestus biting behavior

827   following universal coverage of long-lasting insecticidal nets in benin. J.

828   Infect. Dis. 206:1622–1629.

829   Molineaux L, Gramiccia G. 1980. The Garki Project. Research on the

830   epidemiology and control of malaria in the sudan savanna of West Africa.

831   World Health Organization

832   Mourou J-R, Coffinet T, Jarjaval F, Pradines B, Amalvict R, Rogier C, Kombila M,

833   Pagès F. 2010. Malaria transmission and insecticide resistance of

834   Anopheles gambiae in Libreville and Port-Gentil, Gabon. Malar. J. 9:321.

835   Mwangangi JM, Mbogo CM, Orindi BO, Muturi EJ, Midega JT, Nzovu J, Gatakaa

836   H, Githure J, Borgemeister C, Keating J, et al. 2013. Shifts in malaria vector

837   species composition and transmission dynamics along the Kenyan coast

838   over the past 20 years. Malar. J. 12:13.

839   Ndjemaï HNM, Patchoké S, Atangana J, Etang J, Simard F, Bilong CFB, Reimer

840   L, Cornel A, Lanzaro GC, Fondjo E. 2009. The distribution of insecticide

841 resistance in Anopheles gambiae s.l. populations from Cameroon: an
842 update. Trans. R. Soc. Trop. Med. Hyg. 103:1127–1138.

843 Neafsey DE, Lawniczak MKN, Park DJ, Redmond SN, Coulibaly MB, Traoré SF,
844 Sagnon N, Costantini C, Johnson C, Wiegand RC, et al. 2010. SNP
845 genotyping defines complex gene-flow boundaries among African malaria
846 vector mosquitoes. Science 330:514–517.

847 Norris LC, Main BJ, Lee Y, Collier TC, Fofana A, Cornel AJ, Lanzaro GC. 2015.
848 Adaptive introgression in an African malaria mosquito coincident with the
849 increased usage of insecticide-treated bed nets. Proc. Natl. Acad.
850 Sci.:201418892.

851 Nosil P, Feder JL. 2012. Widespread yet heterogeneous genomic divergence.
852 Mol. Ecol. 21:2829–2832.

853 Nwane P, Etang J, Chouaïbou M, Toto J, Mimpfoundi R, Simard F. 2011. Kdr-
854 based insecticide resistance in Anopheles gambiae s.s populations in. BMC
855 Res. Notes 4:463.

856 Nwane P, Etang J, Chouaïbou M, Toto JC, Koffi A, Mimpfoundi R, Simard F.
857 2013. Multiple insecticide resistance mechanisms in Anopheles gambiae s.l.
858 populations from Cameroon, Central Africa. Parasit. Vectors 6:41.

859 O'Loughlin SM, Magesa S, Mbogo C, Mosha F, Midega J, Lomas S, Burt A.
860 2014. Genomic Analyses of Three Malaria Vectors Reveals Extensive
861 Shared Polymorphism but Contrasting Population Histories. Mol. Biol.
862 Evol.:1–14.

863 Oyewole IO, Awolola TS. 2006. Impact of urbanisation on bionomics and

864   distribution of malaria vectors in Lagos, southwestern Nigeria. J. Vector

865   Borne Dis. 43:173–178.

866 Palumbi SR. 2001. Evolution - Humans as the world's greatest

867   evolutionary force. Science (80-.). 293:1786–1790.

868 Paradis E, Claude J, Strimmer K. 2004. Analyses of Phylogenetics and Evolution

869   in R language. Bioinformatics 20:289–290.

870 Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double Digest

871   RADseq: An Inexpensive Method for De Novo SNP Discovery and

872   Genotyping in Model and Non-Model Species. PLoS One 7:e37135.

873 Pinto J, Egyir-Yawson A, Vicente J, Gomes B, Santolamazza F, Moreno M,

874   Charlwood J, Simard F, Elissa N, Weetman D, et al. 2013. Geographic

875   population structure of the African malaria vector Anopheles gambiae

876   suggests a role for the forest-savannah biome transition as a barrier to gene

877   flow. Evol. Appl. 6:910–924.

878 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, Maller J,

879   Sklar P, de Bakker P, Daly M, et al. 2007. PLINK: a toolset for whole-

880   genome association and population-based linkage analysis. Am. J. Hum.

881   Genet. 81.

882 R Development Core Team. 2014. R: A language and environment for statistical

883   computing. R Foundation for Statistical Computing, Vienna, Austria

884 Raj A, Stephens M, Pritchard JK. 2014. FastSTRUCTURE: Variational inference

42

885   of population structure in large SNP data sets. Genetics 197:573–589.

886   Ranson H, Lissenden N. 2016. Insecticide Resistance in African Anopheles

887   Mosquitoes: A Worsening Situation that Needs Urgent Action to Maintain

888   Malaria Control. Trends Parasitol. 32:187–196.

889   Reid MC, McKenzie FE. 2016. The contribution of agricultural insecticide use to

890   increasing insecticide resistance in African malaria vectors. Malar. J. 15:107.

891   Riehle MM, Guelbeogo WM, Gneme A, Eiglmeier K, Holm I, Bischoff E, Garnier

892   T, Snyder GM, Li X, Markianos K, et al. 2011. A cryptic subgroup of

893   Anopheles gambiae is highly susceptible to human malaria parasites.

894   Science 331:596–598.

895   Rosenberg N. 2004. DISTRUCT: a program for the graphical display of

896   population structure. Mol. Ecol. Resour. 4:137–138.

897   Santolamazza F, Torre A Della, Caccone A. 2004. Short Report : a New

898   Polymerase Chain Reaction − Restriction Fragment Length Polymorphism

899   Method To Identify Anopheles Arabiensis From an . Gambiae and Its Two

900   Molecular Forms From Degraded Dna Templates or Museum Samples.

901   Trop. Med. 70:604–606.

902   Savolainen O, Lascoux M, Merilä J. 2013. Ecological genomics of local

903   adaptation. Nat. Rev. Genet. 14:807–820.

904   Service MW. 1993. Mosquito ecology: field sampling methods. (Elsevier Applied

905   Science, London UK, editor.).

906   Silvestrini F, Severini C, di Pardo V, Romi R, de Matthaeis E, Raymond M. 1998.

907 Population structure and dynamics of insecticide resistance genes in Culex

908 pipiens populations from Italy. Heridity 81:342–348.

909 Simard F, Ayala D, Kamdem GC, Pombi M, Etouna J, Ose K, Fotsing J-M,

910 Fontenille D, Besansky NJ, Costantini C. 2009. Ecological niche partitioning

911 between Anopheles gambiae molecular forms in Cameroon: the ecological

912 side of speciation. BMC Ecol. 9:17.

913 Sinka ME, Bangs MJ, Manguin S, Coetzee M, Mbogo CM, Hemingway J, Patil

914 AP, Temperley WH, Gething PW, Kabaria CW, et al. 2010. The dominant

915 Anopheles vectors of human malaria in Africa, Europe and the Middle East:

916 occurrence data, distribution maps and bionomic précis. Parasit. Vectors

917 3:117.

918 Slotman M a, Tripet F, Cornel a J, Meneses CR, Lee Y, Reimer LJ, Thiemann

919 TC, Fondjo E, Fofana a, Traoré SF, et al. 2007. Evidence for subdivision

920 within the M molecular form of Anopheles gambiae. Mol. Ecol. 16:639–649.

921 Sogoba N, Vounatsou P, Bagayoko MM, Doumbia S, Dolo G, Gosoniu L, Traoré

922 SF, Smith T a, Touré YT. 2008. Spatial distribution of the chromosomal

923 forms of anopheles gambiae in Mali. Malar. J. 7:205.

924 Sokhna C, Ndiath MO, Rogier C. 2013. The changes in mosquito vector

925 behaviour and the emerging resistance to insecticides will challenge the

926 decline of malaria. Clin. Microbiol. Infect. 19:902–907.

927 Stephens M, Scheet P. 2005. Accounting for decay of linkage disequilibrium in

928 haplotype inference and missing-data imputation. Am. J. Hum. Genet.

929   76:449–462.

930   Stephens M, Smith NJ, Donnelly P. 2001. A new statistical method for haplotype

931   reconstruction from population data. Am. J. Hum. Genet. 68:978–989.

932   Storz JF. 2005. Using genome scans of DNA polymorphism to infer adaptive

933   population divergence. Mol. Ecol. 14:671–688.

934   Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by

935   DNA polymorphism. Genetics 123:585–595.

936   Takken W, Verhulst NO. 2013. Host Preferences of Blood-Feeding Mosquitoes.

937   Annu. Rev. Entomol.:433–453.

938   Tene Fossog B, Antonio-Nkondjio C, Kengne P, Njiokou F, Besansky NJ,

939   Costantini C. 2013. Physiological correlates of ecological divergence along

940   an urbanization gradient: differential tolerance to ammonia among molecular

941   forms of the malaria mosquito Anopheles gambiae. BMC Ecol. 13:1.

942   Tene Fossog B, Ayala D, Acevedo P, Kengne P, Ngomo Abeso Mebuy I,

943   Makanga B, Magnus J, Awono-Ambene P, Njiokou F, Pombi M, et al. 2015.

944   Habitat segregation and ecological character displacement in cryptic African

945   malaria mosquitoes. Evol. Appl. 8:326–345.

946   Tiffin P, Ross-Ibarra J. 2014. Advances and limits of using population genetics to

947   understand local adaptation. Trends Ecol. Evol. 29:673–680.

948   Touré Y, Petrarca V, Traoré S, Coulibaly A, Maiga H, Sankaré O, Sow M, Di

949   Deco M, Coluzzi M. 1998. The distribution and inversion polymorphism of

950   chromosomally recognized taxa of the Anopheles gambiae complex in Mali,

45

951   West Africa. Parassitologia 40:477–511.

952   Turissini D, Gamez S, White BJ. 2014. Genome-Wide Patterns of Polymorphism

953   in an Inbred Line of the African Malaria Mosquito Anopheles gambiae.

954   Genome Biol. Evol. 6:3094–3104.

955   Wakeley J. 1996. The variance of pairwise nucleotide differences in two

956   populations with migration. Theor. Popul. Biol. 49:39–57.

957   White B, Lawniczak M, Cheng C, Coulibaly MB, Wilson MD, Sagnon N,

958   Costantini C, Simard F, Christophides GK, Besansky NJ. 2011. Adaptive

959   divergence between incipient species of Anopheles gambiae increases

960   resistance to Plasmodium. Proc. Natl. Acad. Sci. 108:244–249.

961   White GB. 1974. Anopheles gambiae complex and disease transmission in

962   Africa. Trans R Soc Trop Med Hyg 68:278–298.

963   WHO. 2012. Global Plan for Insecticide Resistance Management in Malaria

964   Vectors. :24.

965   WHO. 2013. World malaria report 2013.

966   Wondji C, Simard F, Fontenille D. 2002. Evidence for genetic differentiation

967   between the molecular forms M and S within the Forest chromosomal form

968   of *Anopheles gambiae* in an area of sympatry. Insect Mol. Biol. 11:11–19.

969   Wondji C, Simard F, Petrarca V. 2005. Species and populations of the Anopheles

970   gambiae complex in Cameroon with special emphasis on chromosomal and

971   molecular forms of Anopheles gambiae ss. J. Med. Entomol.

972   Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and

973   splicing in short reads. Bioinformatics 26:873–881.

974

975

976 FIGURE LEGENDS

977 **Figure 1. *Anopheles gambiae* complex sibling species are genetically**

978 **distinct.** A) PCA B) NJ Tree, and C) fastSTRUCTURE analyses clearly separate

979 the four *An. gambiae* complex species that occur in Cameroon into discrete

980 genetic clusters. Additional subdivision below the species level is apparent within

981 *An. coluzzii* and *An. gambiae.* D) Species composition varies strongly between

982 eco-geographic regions. Sampling sites are denoted by black dots.

983

984 **Figure 2. *Anopheles gambiae* is divided into three subpopulations.** A) PCA,

985 B) NJ Tree, and C) fastSTRUCTURE analyses reveal subdivisions within *An.*

986 *gambiae.* We term the most abundant group *GAM1*, while a second small, but

987 widely distributed group, is termed *GAM2*. Finally, most individuals from the

988 village of Nkolondom are genetically distinct from other *An. gambiae* suggestive

989 of local adaptation.

990

991 **Figure 3. *Anopheles coluzzii* is divided into four subgroups.** A) PCA, NJ

992 Tree, and C) fastSTRUCTURE reveal major population structuring between *An.*

993 *coluzzii* from the northern Savannah eco-geographic region and *An. coluzzii* from

994 the southern three forested regions. Within the south, D) PCA, E) NJ Tree, and

995 F) fastSTRUCTURE analyses separate mosquitoes based on geographic origin,

996 although clustering is not fully discrete indicating a dynamic interplay between

997 local adaptation and migration.

998

999 **Figure 4. The genetic distance (F_{ST}) between populations show recent**

1000 **radiation within _An. gambiae_ and _An. coluzzii_ clades**. In an unrooted, F_{ST}-

1001 based NJ tree, _An. melas_ is most distant from all other species, while _An._

1002 _gambiae_ and _An. coluzzii_ are sister species. Southern populations of _An. coluzzii_

1003 are more closely related to each other than to the northern savannah population.

1004 In contrast to geographic distance, the Douala subpopulation is genetically closer

1005 to Yaoundé rather than Coastal mosquitoes. Within _An. gambiae_, a relatively

1006 deep split is present between _GAM2_ and _GAM1_, while _Nkolondom_ appears to

1007 have recently diverged from _GAM1_.

1008

1009 **Figure 5. Genome scans reveal footprints of global and local adaptation in**

1010 **_An. gambiae_ subpopulations.** A-C) Diversity and Tajima's _D_ are plotted for

1011 each of the three subpopulations. Brown asterisks denote windows above the

1012 99th percentile or below the 1st percentile of empirical distribution of Tajima's _D_.

1013 D-E) Both absolute (d_{xy}) and relative (F_{ST}) divergence between populations are

1014 plotted across 150-kb windows. Red asterisks indicate windows above the 99th

1015 percentile of empirical distribution of F_{ST}. In all populations, concordant dips in

1016 diversity and Tajima's _D_ are evident near the pericentromeric region of 2L where

1017 the _para_ sodium channel gene is located. Three other selective sweeps located

1018 on the X and 3R chromosomes are highlighted (grey boxes).

1019

1020 **Figure 6. Strong signatures of selection are centered on xenobiotic**

1021 **resistance loci in subpopulations of _An. coluzzii_.** A-H) Grey boxes highlight

49

1022   five selective sweeps characterized by extreme values of F_{ST} and Tajima's D. As

1023   in *An. gambiae*, sharp declines in diversity and allele frequency spectrum at the

1024   *para* sodium channel gene are present in all populations. A sweep encompassing

1025   a cluster of detoxification genes on 3R is sharpest in urban mosquitoes.

1026

1027   **Figure 7. Spatially varying selection between urban and coastal**

1028   **populations.** For each of the three southern *An. coluzzii* subpopulations, 80

1029   reconstructed haplotypes are visualized by color-coding 28 bi-allelic SNPs in the

1030   3R GSTE/CYP450 sweep either grey or white. A single invariant haplotype -- all

1031   grey SNPs -- is common in (A) Yaoundé, less so in (B) Douala, and very rare in

1032   (C) coastal populations. D-E) Similarly, in PCA and NJ Tree analysis of the same

1033   28 SNPs, coastal individuals (navy blue) are diffuse across genotypic space,

1034   while Yaoundé mosquitoes (purple) are tightly clustered. As expected, Douala

1035   (pink) exhibits an intermediate degree of variation. F) STRUCTURE analysis

1036   based solely on the 28 SNPs within the sweep shows clear distinctions between

1037   the three populations.

1038

1039

1040   SUPPLEMENTAL FIGURE LEGENDS

1041   **Figure S1.** Detailed map of sampling sites in Cameroon.

1042

1043   **Figure S2.** Bayesian information criterion was used to determine the most likely

1044   number of clusters/populations for A) all 941 samples, B) all *An. coluzzii*, C) *An.*

1045   *coluzzii* from the northern savannah, D) *An. arabiensis*, E) the 309 individuals

1046   used for genome scans F) all *An. gambiae*, G) southern *An. coluzzii*, and H) *An.*

1047   *melas.* BIC scores for 1 to 50 clusters are plotted. The lower the BIC score the

1048   better the model fits the observed genetic diversity.

1049

1050   **Figure S3.** Southern (pale pink) and northern populations (deep pink) of *An.*

1051   *coluzzii* are readily separated in PCA (top) and NJ trees (bottom) using SNPs

1052   exclusively from any of the five chromosomal arms.

1053

1054   **Figure S4.** No population substructure is detectable within northern *An. coluzzii*

1055   using PCA and NJ tree analysis.

1056

1057   **Figure S5.** Mean sequencing coverage per individual is plotted in 300kb non-

1058   overlapping windows across the genome. Only individuals used in the genome

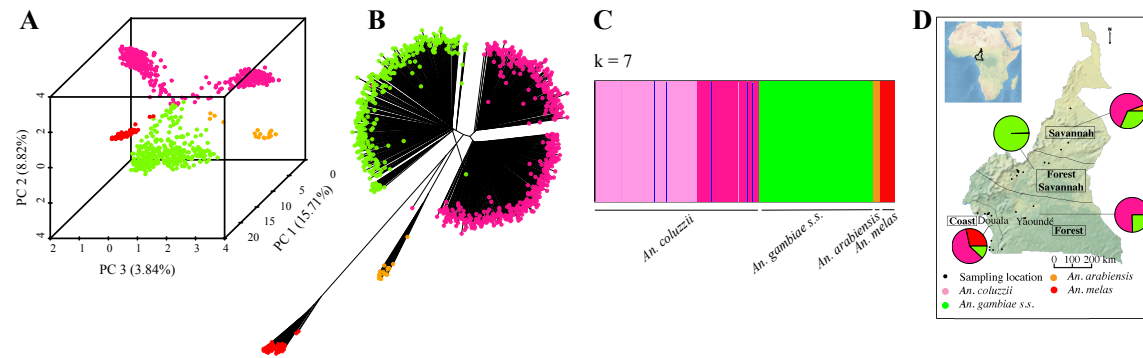1059   scans are included in the coverage calculation.

Figure 1.

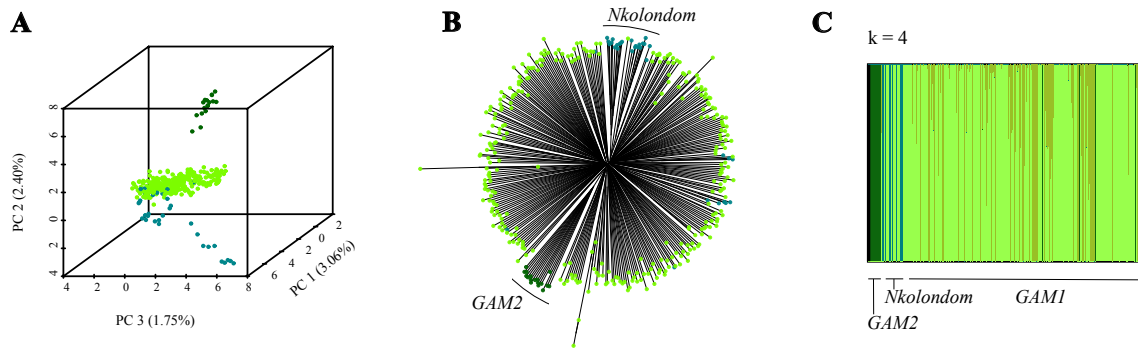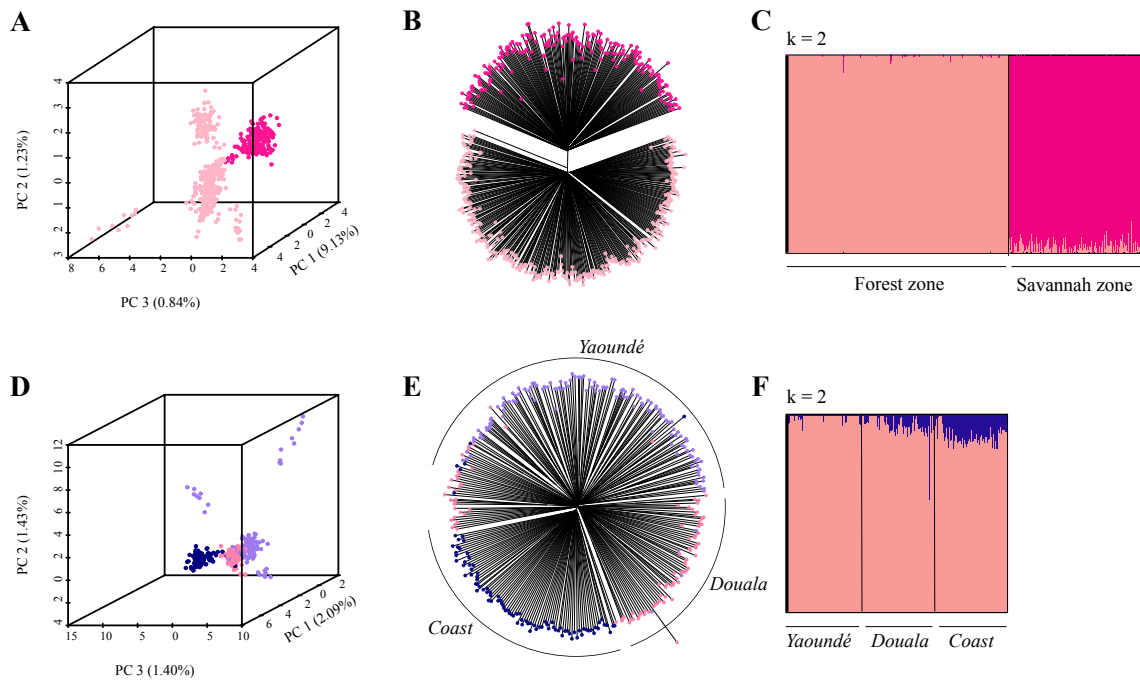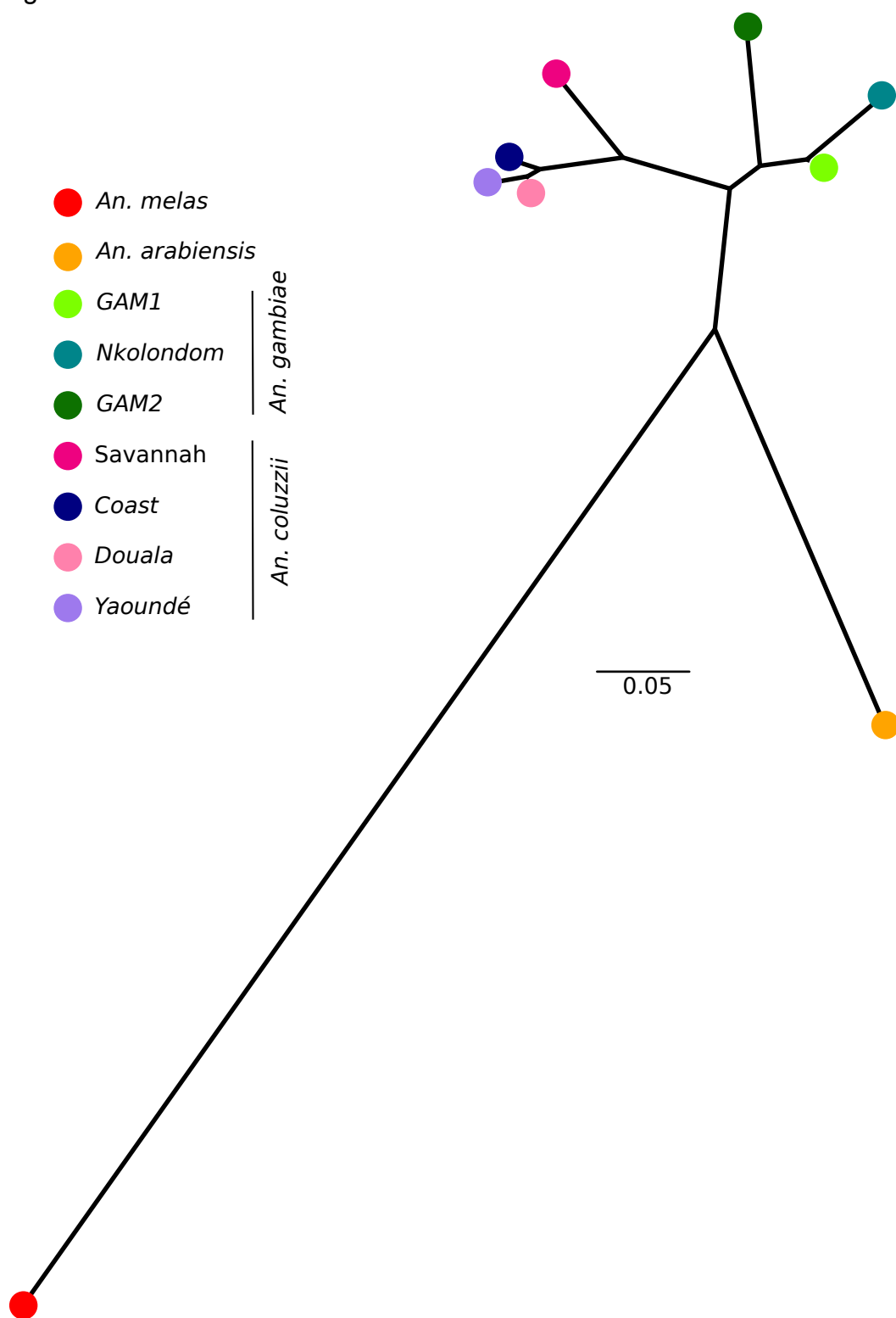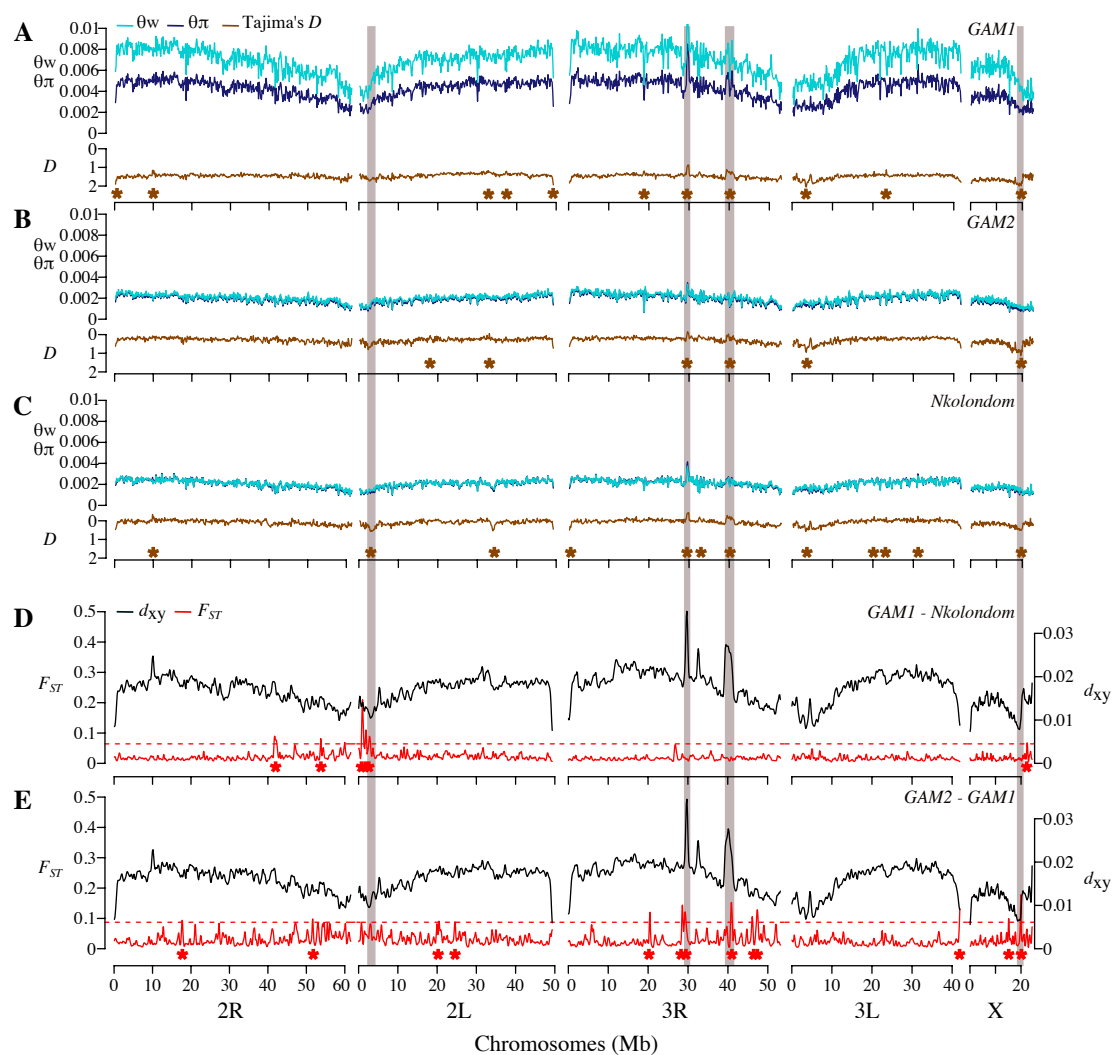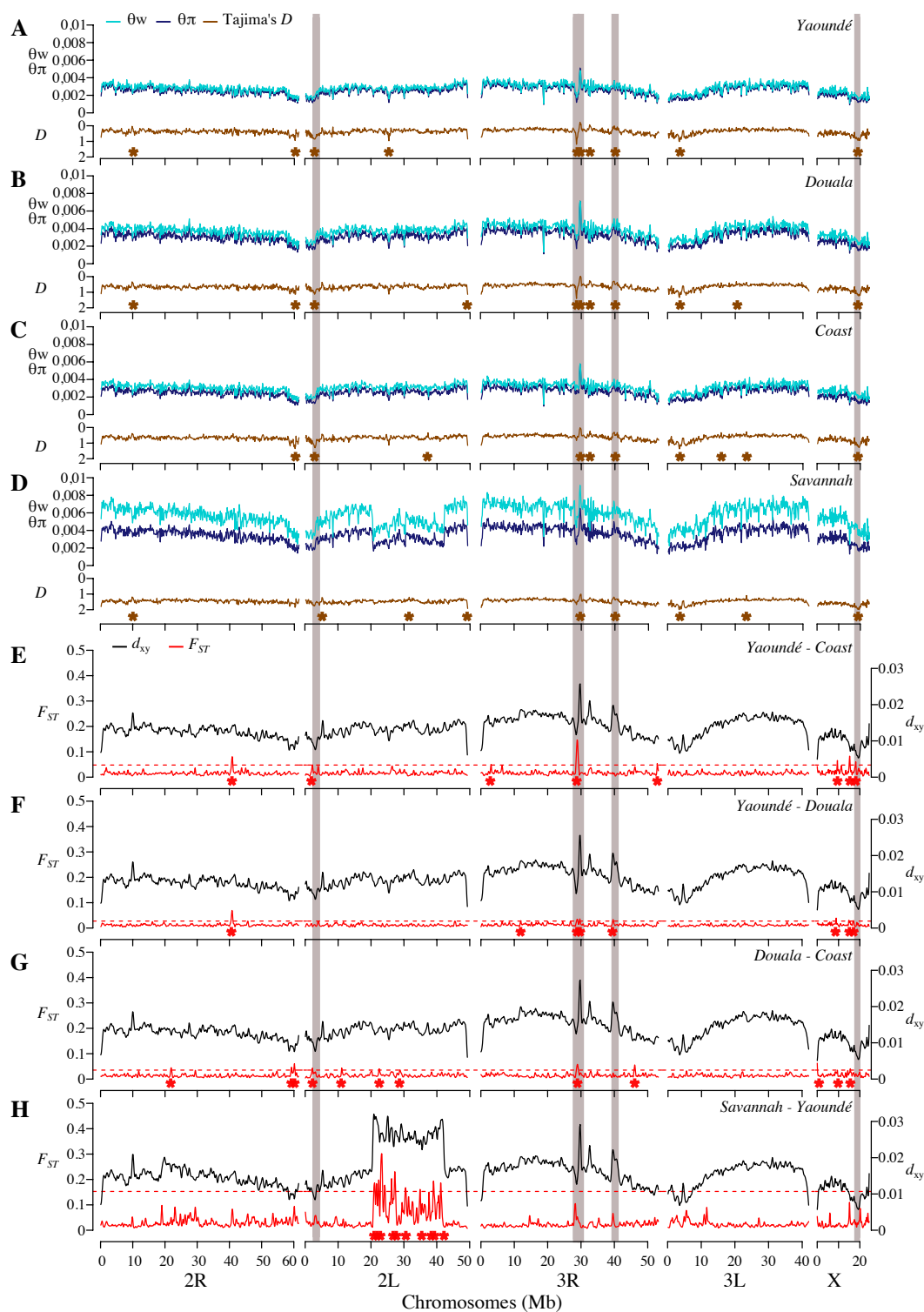Figure 2.
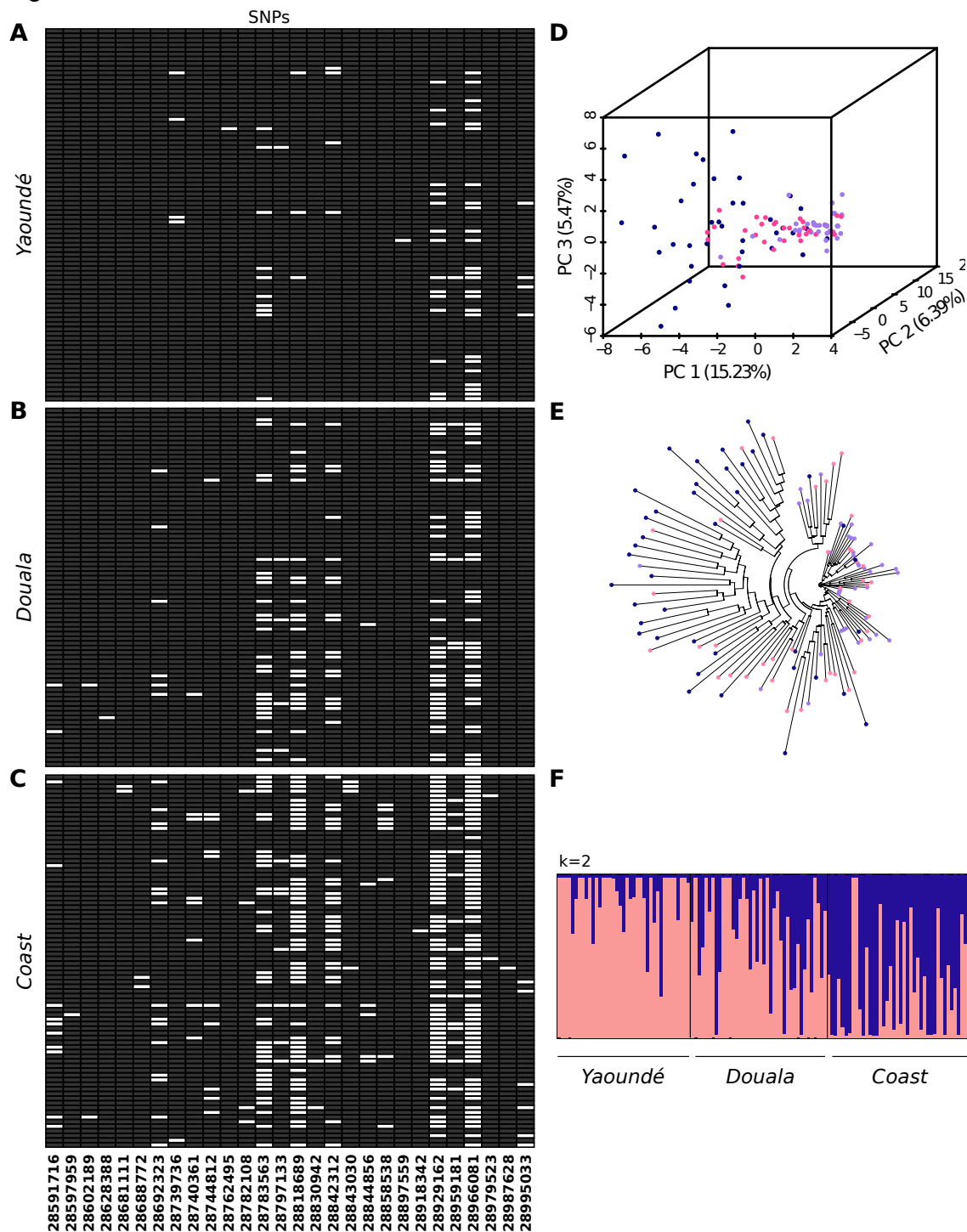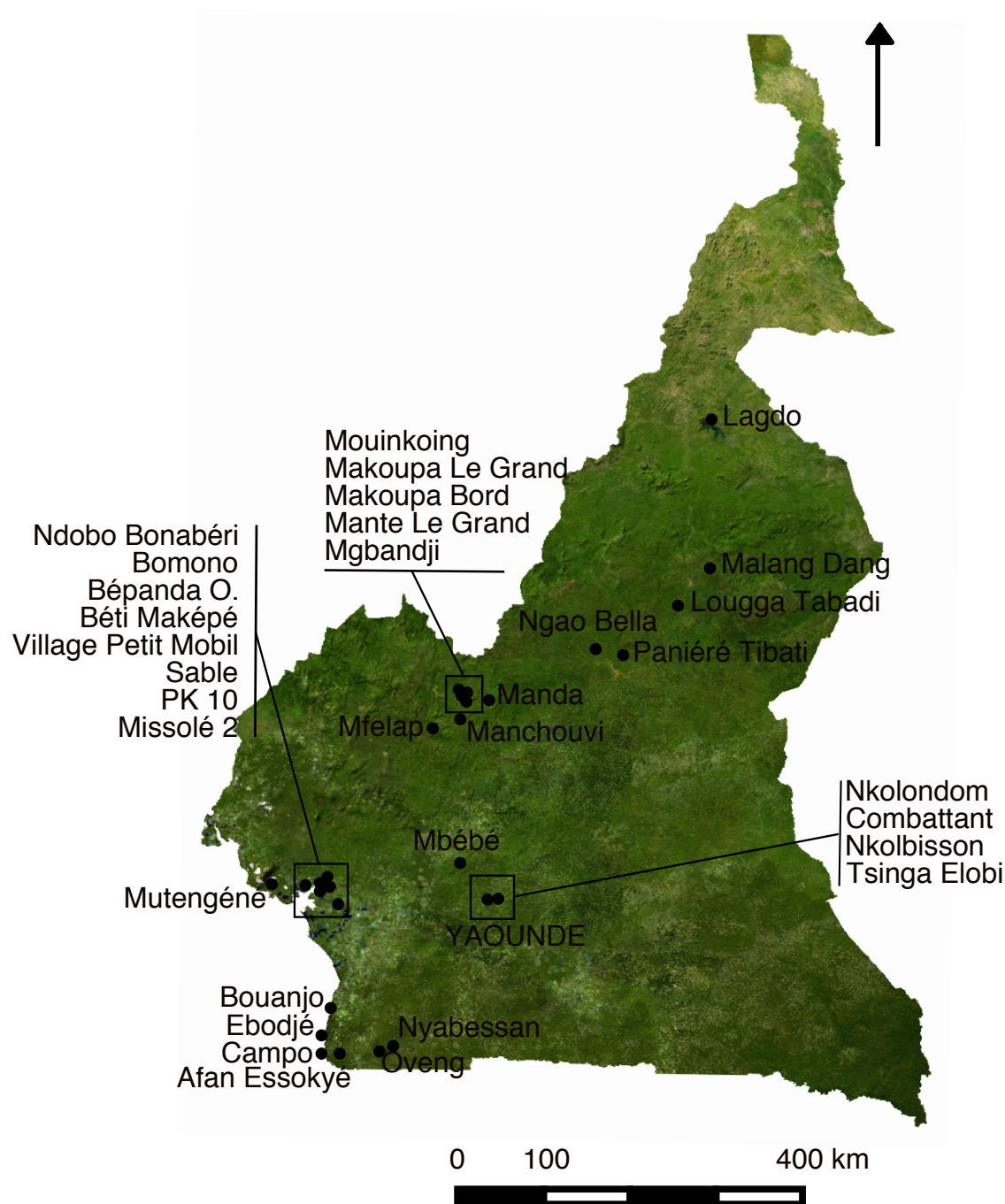
Figure 3.

54

Figure 4.

Figure 5.

Figure 6.

Figure 7.

58

Figure S1.

Figure S2.

60
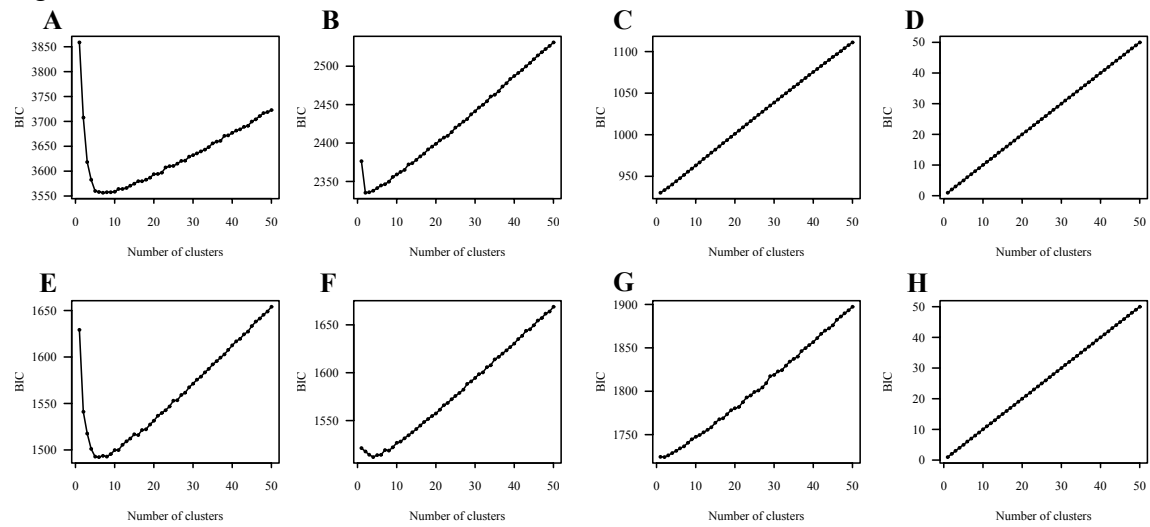
Figure S3.
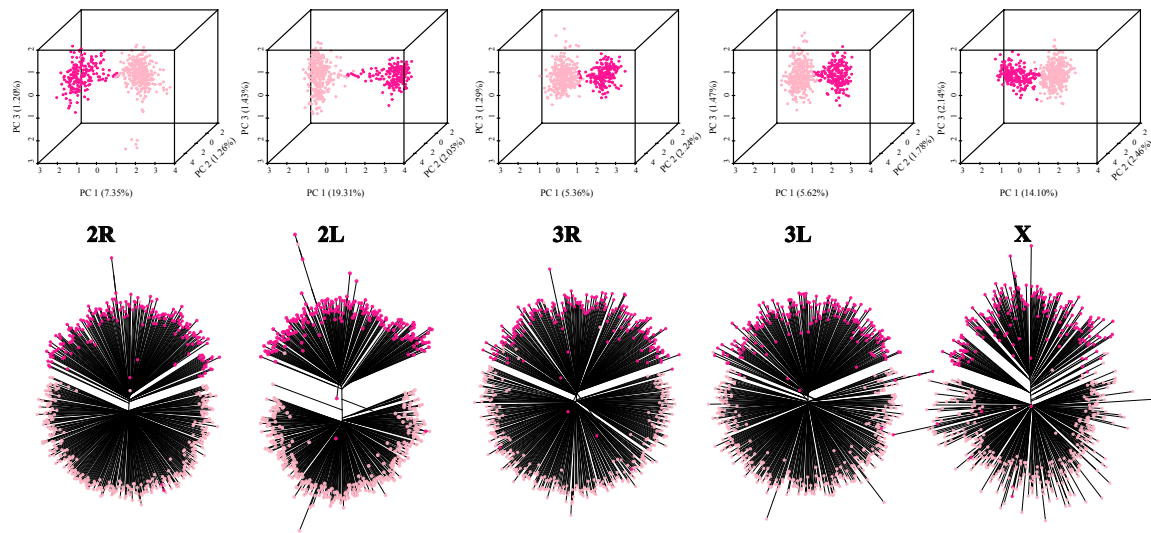
Figure S4.

A
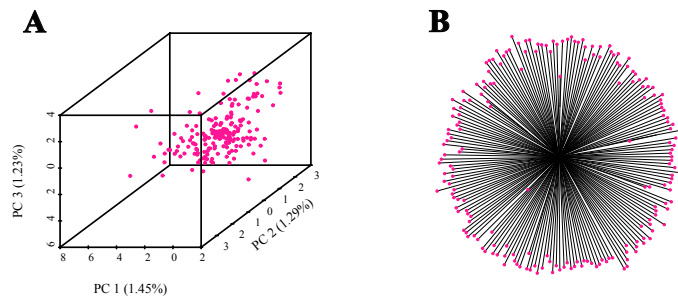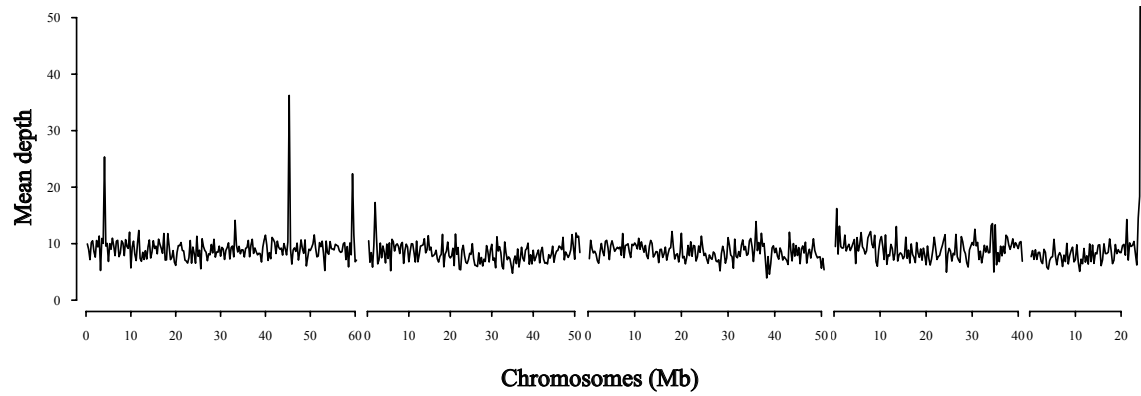
B

Figure S5.

1060

64

1061

1062   SUPPLEMENTAL INFORMATION

1063   Consistently negative Tajima's *D* across all subgroups may reflect recent

1064   population expansions. To further address this hypothesis we modeled the

1065   demographic history of each population using a diffusion-based approach

1066   implemented in the software package ∂a∂i v 1.6.3 (Gutenkunst et al. 2009). We

1067   fit four alternative demographic models (*neutral, growth, two-epoch, bottle-*

1068   *growth*), without migration or recombination, to the folded allele frequency

1069   spectrum of each cryptic subgroup of *An. gambiae s.s.* and *An. coluzzii*. The best

1070   model was selected based on the highest composite log likelihood, the lowest

1071   Akaike Information Criterion (AIC), and visual inspection of residuals. As the

1072   choice of model can be challenging in recently diverged populations, we

1073   prioritized the simplest model when we found it difficult to discriminate between

1074   conflicting models. To obtain uncertainty estimates for the demographic

1075   parameters we used the built-in bootstrap function implemented in ∂a∂i to derive

1076   95% bootstrap confidence intervals.

1077   Results indicate that GAM1, GAM2, and Savannah populations have

1078   experienced recent size increases. However, for the southern populations of

1079   *Yaoundé, Coast, Douala, and Nkolondom* the best demographic model is a

1080   *bottle-growth* (Table S4). While most classical studies report *An. gambiae s.l.*

1081   populations that are in expansion (Donnelly et al. 2001), a more recent study

1082   employing RAD markers revealed that some East African populations have more

1083   complex demographic histories, often involving several changes in effective

65

1084   population size (N_e) as we observed in southern forest populations of both *An.*

1085   *coluzzii* and *An. gambiae* (O'Loughlin et al. 2014).

1086

1087   SUPPLEMENTAL REFERENCES

1088   Donnelly MJ, Licht MC, Lehmann T. 2001. Evidence for recent population

1089   expansion in the evolutionary history of the malaria vectors Anopheles

1090   arabiensis and Anopheles gambiae. Mol. Biol. Evol. 18:1353–1364.

1091   Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring

1092   the joint demographic history of multiple populations from multidimensional

1093   SNP frequency data. PLoS Genet. 5:e1000695.

1094   O'Loughlin SM, Magesa S, Mbogo C, Mosha F, Midega J, Lomas S, Burt A.

1095   2014. Genomic Analyses of Three Malaria Vectors Reveals Extensive

1096   Shared Polymorphism but Contrasting Population Histories. Mol. Biol.

1097   Evol.:1–14.

1098

Table S1. Locations of *An. gambiae s.l.* mosquitoes sequenced in Cameroon

Ecogeographic regions	Sampling locations	Geographic coordinates	HLC-OUT	HLC-IN	LC	SPRAY	Total	YDE	DLA	CST	SAV	GAM2	NKO	GAM1	A	ME
Savannah	Lagdo	9°02'56"N, 13°39'22"E	52	27	21	43	143				116			9	18	
	Lougga Tabadi	7°06'36"N, 13°12'36"E	2				2							2		
	Malang Dang	7°26'06"N, 13°33'14"E				11	11				2			9		
	Ngao Bella	6°30'04"N, 12°25'34"E	28	17		69	114				65	14		35		
	Paniéré Tibati	6°28'08"N, 12°37'44"E	5		22	16	43				13	1		29		
	Total		87	44	43	139	313				196	15		84	18	
Forest-Savannah Transition	Makoupa Bord	6°06'06"N, 11°11'29"E	2			16	18							18		
	Makoupa Le Grand	6°02'28"N, 11°10'39"E			4	10	14							14		
	Manchoutvi	5°52'48"N, 11°06'36"E			70		70					2		68		
	Manda	5°43'32"N, 10°52'06"E	1	2		33	36							36		
	Mante Le Grand	6°03'44"N, 11°12'18"E				8	8				1			7		
	Mfelap	5°43'30"N, 10°52'00"E				8	8							8		
	Mgbandji	6°05'49"N, 11°08'26"E			3	5	8							8		
	Mouinkoing	6°02'56"N, 11°24'36"E				3	3							3		
	Total		3	2	77	83	165				1	2		162		
Forest (urban)	Bepanda Omnisport (Douala)	4°03'18"N, 9°43'16"E			7		7		7							
	Beti Maképé (Douala)	4°03'54"N, 9°45'40"E			25		25		6					19		
	Bomono Gare (Douala)	4°04'55"N, 9°35'35"E			13		13		9					4		
	PK10 (Douala)	4°02'49"N, 9°46'47"E			20		20		16					4		
	Missolé 2 (Douala)	3°59'17"N, 9°54'22"E			8		8		8							
	Ndobo Bonabéri (Douala)	4°04'39"N, 9°40'12"E			1		1		1							
	Sable (Douala)	4°04'52"N, 9°43'34"E	28		5		33		33							
	Village Petit Mobil (Douala)	4°00'16"N, 9°45'18"E	1		16		17		17							
	Nkolbisson (Yaoundé)	3°52'29"N, 11°26'58"E			9		9	9								
	Nkolondom (Yaoundé)	3°58'20"N, 11°30'56"E			48		48	12					33	3		
	Tsinga Elobi (Yaoundé)	3°52'49"N, 11°30'23"E			46		46	46								
	Combattant (Yaoundé)	3°52'36"N, 11°30'46"E	18		48		66	65					1			
	Total		47	0	246	0	293	132	97				34	30		
Forest (rural)	Mbébé	4°10'00"N, 11°04'00"E				5	5							5		
	Nyabessan Centre	2°24'00"N, 10°24'00"E	4	1			5							5		
	Oveng	2°44'00"N, 11°27'00"E	1				1							1		
	Total		5	1	0	5	11							11		
Coast	Afan Essokyé	2°22'01"N, 9°58'59"E	1			13	14			3				11		
	Bouanjo	2°48'00"N, 9°54'00"E			5		5			5						
	Campo	2°22'01"N, 9°49'01"E	48		16	27	91			79				1		11
	Ebodjé	2°30'00"N, 9°49'05"E	35	4		3	42			5				3		34
	Mutengéné	4°06'53"N, 9°14'51"E			7		7		1	2				4		
	Total		84	4	28	43	159		1	94				19		45
Total			226	51	394	270	941	132	98	94	197	17	34	306	18	45

HLC-OUT, human landing catches performed outdoor; HLC-IN, human landing catches performed indoor; LC, larval collection; SPRAY, spray catches.

68

Table S2. Pairwise comparison of genetic distance (F_{ST}) among cryptic subgroups and sibling species of *An. gambiae s.l.*

F_{ST}	Yaoundé	Douala	Coast	Savannah	GAM2	Nkolondom	GAM1	An. arabiensis
Yaoundé	-							
Douala	0.016	-						
Coast	0.035	0.023	-					
Savannah	0.127	0.109	0.103	-				
GAM2	0.244	0.199	0.230	0.168	-			
Nkolondom	0.197	0.201	0.200	0.247	0.161	-		
GAM1	0.183	0.153	0.178	0.188	0.090	0.050	-	
An. arabiensis	0.451	0.406	0.498	0.383	0.478	0.372	0.327	-
An. melas	0.851	0.836	0.830	0.792	0.841	0.818	0.781	0.872

Table S3. Average nucleotide diversity in seven cryptic subgroups of *An. coluzzii* and *An. gambiae s.s.*

Population	$\theta\pi$ (bp^{-1})	θw (bp^{-1})
Yaoundé	0.0025	0.0028
Douala	0.0031	0.0037
Coast	0.0025	0.0031
Savannah	0.0035	0.0057
GAM2	0.0019	0.0020
Nkolondom	0.0020	0.0021
GAM1	0.0042	0.0069

Table S4. Parameters of demographic models inferred from folded Site Frequency Spectrum (SFS) of autosomal SNPs in seven cryptic subpopulations of *An. gambiae*

Species	Population	Best Model	Log Likelihood	Final Pop Size[a] (95% CI)	Bottleneck Size[b] (95% CI)	Time[c] (95% CI)
Anopheles coluzzii						
	Yaoundé	*Bottle-growth*	-83.45	1.18 (0.92 – 1.69)	24.85 (11.94 - 125.77)	0.58 (0.43 - 1.01)
	Douala	*Bottle-growth*	-84.50	2.20 (1.52 - 3.77)	5.66 (2.80 - 142.42)	0.59 (0.41 - 1.77)
	Coast	*Bottle-growth*	-82.95	1.72 (1.28 - 2.33)	23.83 (10.95 - 128.52)	0.67 (0.50 - 1.10)
Anopheles gambiae s.s.	*Savannah*	*Two-epoch*	-102.65	6.73 (6.35 - 7.21)		0.62 (0.54 - 0.72)
	GAM2	*Two-epoch*	-37.48	3.60 (1.74 - 7.70)		3.60 (0.56 - 9.30)
	Nkolodom	*Bottle-growth*	-74.92	1.08 (0.82 - 1.12)	24.95 (47.91 - 99.84)	0.63 (0.47 - 0.91)
	GAM1	*Two-epoch*	-103.33	6.75 (6.33 - 7.24)		0.68 (0.60 - 0.78)

[a] Ratio of contemporary to ancient population size.
[b] Ratio of population size after instantaneous change to ancient population size.
[c] Time in the past at which instantaneous change happened and growth began (in units of 2*Na generations).

72

Table S5. Functional analysis of gene ontology terms in candidate regions.
Supplementary online information