POLLUTANTS AND INSECTICIDES DRIVE LOCAL ADAPTATION IN AFRICAN MALARIA MOSQUITOES

Colince Kamdem[1*], Caroline Fouet[1], Stephanie Gamez[1], Bradley J. White[1,2,*]

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

[2]Center for Disease Vector Research, Institute for Integrative Genome Biology, University of California, Riverside, CA 92521

[*]Corresponding authors: guy.kamdemdjoko@ucr.edu; bwhite@ucr.edu

ABSTRACT

Urbanization presents unique environmental challenges to human commensal species. The Afrotropical *Anopheles gambiae* complex contains a number of synanthropic mosquito species that are major vectors of malaria. To examine ongoing cryptic diversification within the complex, we performed reduced representation sequencing on 941 mosquitoes collected across four ecogeographic zones in Cameroon. We find evidence for clear subdivision within *An. coluzzii* and *An. gambiae s.s*. – the two most significant malaria vectors in the region. Importantly, in both species rural and urban populations of mosquitoes were genetically differentiated. Genome scans of cryptic subgroups reveal pervasive signatures of selection centered on genes involved in xenobiotic resistance. Notably, a selective sweep containing eight detoxification enzymes is unique to urban mosquitoes that exploit polluted breeding sites. Overall, our study reveals that anthropogenic environmental modification is driving population differentiation and local adaptation in African malaria mosquitoes with potentially significant consequences for malaria epidemiology.

INTRODUCTION

In species occupying ecologically heterogeneous ranges, natural selection can drive local adaptation, increasing mean individual fitness and promoting biological diversification [1, 2]. Contemporary anthropogenic alteration of the landscape may be increasing pressure for local adaptation in diverse taxa [3]. For example, the rise of urban centers over the past two centuries has presented unique challenges to human-commensal species, often necessitating the rapid evolution of resistance to pollutants and pesticides [4-6]. Successful local adaptation requires that selection overcome the homogenizing effect of gene flow from nearby populations [7]. Theoretical simulations suggest that such divergence with gene flow can occur under a range of conditions [8], although the most likely genomic distribution of the underlying adaptive variants remains unclear [9, 10]. Studies of populations in the earliest stages of ecological divergence should help elucidate the conditions needed for local adaptation and relevant targets of natural selection [11].

The Afrotropical *Anopheles gambiae* complex is a group of at least nine isomorphic mosquito species exhibiting varying degrees of geographic and reproductive isolation [12-14]. Owing to the critical role its members play in sustaining malaria transmission, a wealth of genomic data exists on the complex including whole genome assemblies for most of the species [15, 16]. Because radiation of the complex began ~1.85 mya, interspecific genetic comparisons will yield little insight into the establishment of divergence with gene flow. However, both ecological and genetic evidence suggests that contemporary local

60 adaptation and diversification is occurring within *Anopheles gambiae s.s.*
61 (hereafter *An. gambiae*) and *Anopheles coluzzii*, two of the most widespread and
62 important vectors of malaria within the complex [17-20]. Up to now, shallow
63 sampling and/or or the use of low-resolution genetic markers limited the ability to
64 delineate new cryptic subgroups within either species.

65 We genotyped 941 mosquitoes collected from diverse environments in
66 Cameroon at >8,000 SNPs and find strong evidence for ongoing diversification
67 within both *An. gambiae* and *An. coluzzii*. In total, the two species harbor seven
68 cryptic subgroups distributed along a continuum of genomic differentiation. While
69 *An. gambiae* exhibits relatively high levels of panmixia, we did identify an ecotype
70 associated with intense suburban agriculture and a second subgroup that
71 appears partially reproductively isolated but exhibits no obvious
72 ecological/geographical distinctions. In contrast, *An. coluzzii* is separated into
73 multiple ecotypes exploiting different regional-scale habitats, including highly
74 urbanized landscapes. In most cryptic subgroups, selective sweeps contain an
75 excess of detoxification enzymes and insecticide resistance genes, suggesting
76 that human activity mediates spatially varying natural selection in both species.
77 Worrisomely, ongoing local adaptation to urban sites with high prevalence of
78 xenobiotics is both increasing malaria transmission spatially and further
79 compromising the effectiveness of chemical control tools. Moreover, the
80 extensive population structure within both species will present a formidable
81 challenge to novel vector control strategies that rely on population replacement.

82

RESULTS

We performed extensive, largely unbiased collections of human-associated *Anopheles* across four ecological zones in the central African country of Cameroon (Table S1). To robustly detect any cryptic populations, we subjected all 941 mosquitoes that were morphologically identified as *An. gambiae* s.l. to population genomic analysis. Individual mosquitoes were genotyped in parallel at a dense panel of markers using double-digest restriction associated DNA sequencing (ddRADseq), which enriches for a representative and reproducible fraction of the genome that can be sequenced on the Illumina platform [21].

Identification of An. gambiae s.l. sibling species

After aligning ddRADseq reads to the *An. gambiae* reference genome, we used STACKS to remove loci present in <80% of individuals, leaving 8,476 SNPs (~1 SNP every 30kb across the genome) for population structure inference [22, 23]. First, we performed principal component analysis (PCA) on genetic diversity across all 941 individuals (Figure 1A). The top three components explain 28.4% of the total variance and group individuals into five main clusters. Likewise, a neighbor-joining (NJ) tree, based on Euclidian distance of allele frequencies, shows five distinct clades of mosquitoes (Figure 1B). We hypothesized that these groups at least partially correspond to the four sibling species – *An. gambiae*, *An. coluzzii*, *An. arabiensis*, and *An. melas* – known to occur in Cameroon. To confirm, we typed a subset of 288 specimens using validated species ID PCRs and found that each cluster comprised a single species [24, 25]. In agreement

with previous surveys [26, 27], our collections indicate that the brackish water breeding *An. melas* is limited to coastal regions, while the arid-adapted *An. arabiensis* is restricted to the savannah. In contrast, *An. gambiae* and *An. coluzzii* are distributed across the four eco-geographic zones of Cameroon (Figure 1D). Lee and colleagues [28] recently reported frequent bouts of hybridization between *An. gambiae* and *An. coluzzii* in Cameroon. While both the PCA and NJ trees clearly separate the two species, the PCA does show intermixing of some rare individuals consistent with semi-permeable species boundaries.

In support of population structuring below the species level, Bayesian clustering analysis finds that 7 population clusters (*k*) best explain the genetic variance present in our sample set (Figure S1). Indeed, grouping of samples within *An. gambiae* and *An. coluzzii* clades suggests that additional subdivision may exist within each species (Figure 1A, 1B). Ancestry plots made with fastSTRUCTURE [29] further support inference from the PCA and NJ tree: at least two subgroups compose *An. coluzzii* and admixture is present within *An. gambiae*, while *An. arabiensis* and *An. melas* are largely panmictic (Figure 1C, Figure S1).

Cryptic Population Structure within An. gambiae s.s.

To further resolve the population structure within 357 *An. gambiae* specimens, we performed population genetic analysis with a set of 9,345 filtered SNPs. Using a combination of PCA, NJ trees, and ancestry assignment, we

132 consistently identify three distinct subgroups within *An. gambiae* (Figure 2A-C).
133 The first and largest subgroup (termed *GAM1*) comprises the vast majority of all
134 *An. gambiae* specimens including individuals collected in all four eco-geographic
135 regions. A total of 17 individuals make up a second small subgroup (termed
136 *GAM2*). Interestingly, individuals assigned to this cluster include both larvae and
137 adults collected in 3 different villages spread across 2 eco-geographic regions. In
138 the absence of any obvious evidence of niche differentiation between *GAM1* and
139 *GAM2,* it is unclear what is driving and/or maintaining divergence between the
140 two sympatric subgroups. Specimens collected from Nkolondom, a suburban
141 neighborhood of Yaoundé where larval sites associated with small-scale
142 agricultural irrigation are common [30, 31], form a genetically distinct third
143 subgroup (termed *Nkolondom*) that appears to be a locally-adapted ecotype.
144
145 *Cryptic Population Structure within An. coluzzii*
146 To examine population structure within 521 *An. coluzzii* specimens, we
147 utilized 9,822 SNPs that passed stringent filtration. All analyses show a clear split
148 between individuals from the northern savannah region and the southern three
149 forested regions of Cameroon (Coastal, Forest, Forest-Savannah) (Figure 3A-C).
150 In principle, the north-south structuring could be caused solely by differences in
151 chromosome 2 inversion frequencies, which form a cline from near absence in
152 the humid south to fixation in the arid north. However, we find SNPs from all five
153 chromosomal arms robustly and consistently separate northern and southern

154 mosquitoes, suggesting incipient reproductive isolation between the two
155 populations (Figure S2).
156 Southern populations of *An. coluzzii* were collected from three different
157 areas: Douala (the largest city of Cameroon), Yaoundé (the second largest city)
158 and the rural coastal region. PCA, NJ trees, and fastSTRUCTURE show clear
159 clustering of southern samples by collection site (Figure 3D-F). Mosquitoes from
160 Douala, situated on the coastal border, contain a mixture of urban and coastal
161 polymorphisms as illustrated by their intermediate position along PC3 (Figure
162 3D). Despite considerable geographic segregation, clusters are not fully discrete,
163 likely owing to substantial migration between the three sites. Taken together, the
164 data suggest a dynamic and ongoing process of local adaptation within southern
165 *An. coluzzii*. In contrast, no similar geographic clustering is observed in northern
166 populations (Figure S1, S3).
167
168 *Relationships Between Species and Subgroups*
169 Population genomic analysis identified four different *An. gambiae* s.l.
170 species present within our samples. Within *An. gambiae* and *An. coluzzii* we
171 identified seven potential subgroups with apparently varying levels of isolation.
172 To further explore the relationships between different populations, we built an
173 unrooted NJ tree using pairwise levels of genomic divergence (F_{ST}) between all
174 species and subgroups (Figure 4, Table S2). As previously observed in a
175 phylogeny based on whole genome sequencing [16], we find that *An. melas* is
176 highly divergent from all other species (F_{ST} ~ 0.8), while *An. arabiensis* shows

177 intermediate levels of divergence (F_{ST} ~ 0.4) from *An. gambiae* and *An. coluzzii*.
178 As expected, the sister species *An. gambiae* and *An. coluzzii* are more closely
179 related to each other (F_{ST} ~ 0.2) than any other species. When examining
180 differentiation between subgroups within *An. coluzzii*, we find that the southern
181 and northern subgroups are highly divergent (F_{ST} > 0.1), while differentiation
182 between local ecotypes within the south is much lower (F_{ST} < 0.04). The *An.*
183 *gambiae* subgroups *GAM1* and *GAM2* are highly diverged (F_{ST} ~ 0.1) from each
184 other suggesting genuine barriers to gene flow despite sympatry, while the
185 suburban ecotype from Nkolondom shows a low level of divergence from *GAM1*
186 (F_{ST} ~ 0.05), characteristic of ongoing local adaptation. In sum, we find a gradient
187 of differentiation between species and subgroups ranging from complete (or
188 nearly complete) reproductive isolation down to the initial stages of divergence
189 with gene flow.
190
191 *Using genome scans to identify selective sweeps*
192
193 To find potential targets of selection within subgroups we performed scans
194 of nucleotide diversity (θ_w, θ_π) and allele frequency spectrum (*Tajima's D*) using
195 non-overlapping 150-kb windows across the genome. Scans of θ_w, θ_π, *and*
196 *Tajima's* D were conducted by importing aligned, but otherwise unfiltered, reads
197 directly into ANGSD, which uses genotype likelihoods to calculate summary
198 statistics [32]. Natural selection can increase the frequency of an adaptive variant
199 within a population, leading to localized reductions in genetic diversity (and a
200 skew towards rare polymorphisms) as the haplotype containing the adaptive
201 variant(s) sweeps towards fixation [33, 34]. Thus, genomic regions harboring

202 targets of recent selection should exhibit extreme reductions in these metrics
203 relative to genome-wide averages [35].
204 We also performed genome scans using both a relative (F_{ST}) and absolute
205 (d_{xy}) measure of divergence calculated with STACKS and ANGSD, respectively
206 [36]. If positive selection is acting on alternative haplotypes of the same locus in
207 two populations, values of F_{ST} and d_{xy} should increase at the target of selection.
208 Whereas spatially varying selection that acts on one population, but not the
209 other, should produce a spike in F_{ST} between populations and no change in d_{xy}.
210 Finally, parallel selection on the same haplotype in two populations should lead
211 to a decrease in both metrics. For both diversity and divergence scans we used a
212 maximum of 40 mosquitoes per population, prioritizing individuals with the
213 highest coverage in populations where sample size exceeded 40.
214
215 *Targets of Selection within An. gambiae subgroups*
216 Our calculations of genome-wide diversity levels (Table S3) within *An.*
217 *gambiae* subgroups are comparable to previous estimates based on RAD
218 sequencing of Kenyan *An. gambiae* s.l. populations [37]. As expected, the large
219 *GAM1* population harbors more genetic diversity than the apparently rare *GAM2*
220 population or the geographically restricted Nkolondom ecotype (Table S3, Figure
221 5A-C). *Tajima's D* is consistently negative across the entire genome of all three
222 subgroups, indicating an excess of low-frequency variants that are likely the
223 result of recent population expansion (Figure 5A-C) [34]. Indeed, formal

224 demographic models infer relatively recent bouts of population expansion in all
225 three subgroups (Table S4).
226 Genome scans of each subgroup reveal genomic regions that show
227 concordant dips in diversity and allele frequency spectrum consistent with recent
228 positive selection (highlighted in Figure 5 A-C). An apparent selective sweep on
229 the left arm of chromosome 2 near the centromere is found in all populations.
230 Notably, the *para* sodium channel gene is embedded within this sweep. The *kdr*
231 allele of *para* confers knockdown resistance to pyrethroids and selective sweeps
232 in the same genomic location have been previously identified in many *An.*
233 *gambiae* s.l. populations [38-42].
234 Other sweeps are population-specific. For example, a major sweep unique
235 to the Nkolondom population is found on chromosome 2L from ~33-35 Mb.
236 Nkolondom is suburban region of Yaoundé where larvae can be readily collected
237 from irrigated garden plots that likely contain elevated levels of pesticides
238 directed at agricultural pests [30, 31]. Intriguingly, the region contains six
239 epidermal growth factor (*EGF*) genes, which could facilitate larval development of
240 this subgroup in pesticide-laced agricultural water. It should be noted that due to
241 the reduced representation sequencing approach we used, our analysis is
242 necessarily conservative, highlighting only clear instances of positive selection,
243 which are likely both recent and strong [10]. Additional smaller sweeps outside
244 our power of detection are likely prevalent in all subgroups.
245 We next performed divergence based genome scans. At the putative *kdr*
246 sweep, we observe contrasting patterns in local values of d_{xy} and F_{ST} (Figure 5D-

247 E). In both the GAM1-GAM2 and GAM1-Nkolondom comparisons, d_{xy} dips at the
248 *kdr* sweep, while local values of F_{ST} actually increase. While not definitive, the
249 dramatic drop in d_{xy} suggests that the same resistant haplotype is sweeping
250 through each population. Localized increases in F_{ST} could owe to differences in
251 *kdr* allele frequencies between populations; despite parallel selection, the sweep
252 may be closer to fixation in certain populations relative to others, perhaps due to
253 differences in selection intensity. At the Nkolondom specific 2L *EGF* sweep, we
254 observe an increase in F_{ST} between *GAM1-Nkolondom* and no clear change in
255 $d_{xy,}$ indicative of ongoing local adaptation in Nkolondom.
256
257 *Targets of Selection within An. coluzzii subpopulations*
258
259 As above, we first used diversity metrics to scan for targets of selection in
260 the four subgroups of *An. coluzzii*. Overall, genetic diversity is higher in the
261 northern savannah population than either of three southern populations, which all
262 exhibit similar levels of diversity (Table S3). Just as in *An. gambiae*, all
263 subgroups have consistently negative *Tajima's D* values confirming demographic
264 models of population expansion (Table S4). Examination of genome-wide plots
265 reveals that all subgroups exhibit dips in diversity and *Tajima's D* at the *kdr* locus
266 (Figure 6A-D). A large sweep at ~25 Mb on 2L centered on the resistance to
267 dieldrin (*rdl*) locus is clearly present in all southern groups. In the northern
268 savannah population, a pronounced dip in diversity occurs at *rdl*, but *Tajima's D*
269 stays constant. While we can confidently infer that a resistant *rdl* allele has
270 recently swept through all southern populations of *An. coluzzii,* evidence for a

271 sweep in the northern population is inconclusive. Regional and population-
272 specific sweeps are also evident including a sweep on 3R is nearly exclusive to
273 mosquitoes collected in the urban areas of Yaoundé and Douala. A sharp,
274 dramatic drop in both diversity and Tajima's *D* occurs on 3R from ~28.5-29.0 Mb
275 with the decline being more dramatic in Yaoundé than Douala. Geographical
276 limitation of the sweep to urban mosquitoes strongly suggests it may contain
277 variant(s) that confer adaptation to extreme levels of anthropogenic disturbance.
278 Compellingly, the selective sweep contains a cluster of both Glutathione S-
279 transferase (*GSTE1-GSTE7*) and cytochrome P450 (*CYP4C27*, *CYP4C35*,
280 *CYP4C36*) genes. Both gene families can confer metabolic resistance to
281 insecticides and pollutants in mosquitoes [43-45]. In particular, *GSTE5* and
282 *GSTE6* are intriguing candidate targets of selection as each is up-regulated in
283 highly insecticide resistant *An. arabiensis* populations that recently colonized
284 urban areas of Bobo-Dioulasso, Burkina Faso [46, 47].
285 Genome scans of both relative and absolute divergence between
286 populations reveal intriguing patterns at sites of putative selective sweeps. The
287 *kdr* locus exhibits minimal divergence in all pairwise comparisons, suggesting
288 that the same resistance haplotype is under selection in each population (Figure
289 6E-H). The region surrounding the *rdl* gene shows low F_{ST} and a pronounced dip
290 in d_{xy} between all southern populations, confirming that the same haplotype is
291 sweeping through these three populations. Differentiation between southern and
292 northern populations at *rdl* is obscured by the high divergence between
293 alternative arrangements of the 2L*a* inversion. Finally, the urban-centric

294 GSTE/CYP450 sweep on 3R shows a peak in F_{ST} between Yaoundé and Coastal
295 mosquitoes and minimal change in d_{xy} – a pattern consistent with local
296 adaptation. Comparisons between Douala and Coastal populations show a more
297 moderate increase in F_{ST}, presumably due to high rates of mosquito migration
298 between these nearby sites. Finally, when comparing mosquitoes from the two
299 urban areas (Douala and Yaoundé), we observe only a slight bump in F_{ST} and a
300 dramatic dip in d_{xy} indicative of an ongoing, shared selective sweep.
301 To further explore the 3R GSTE/CYP450 sweep, we reconstructed
302 haplotypes for all 240 *An. coluzzii* southern chromosomes across the 28 SNPs
303 found within the sweep. In the Yaoundé population, a single haplotype is present
304 on 44 out of 80 (55%) chromosomes (all grey SNPs), while an additional 11
305 haplotypes are within one mutational step of this common haplotype (Figure 7A).
306 In Douala, the same haplotype is the most common, but present at a lower
307 frequency (31%) than in Yaoundé (Figure 7B). Strikingly, this haplotype is found
308 on only 6/80 (7.5%) coastal chromosomes (Figure 7C). The overall low
309 nucleotide variation and high frequency of a single haplotype in Yaoundé is
310 consistent with positive selection acting on a de novo variant(s) to generate the
311 3R GSTE/CYP450 sweep. Less intense selection pressure in Douala, and
312 particularly the Coast, would explain the markedly higher haplotype diversity in
313 these two populations relative to Yaoundé. It is also possible that Douala
314 mosquitoes experience similar selection pressures to Yaoundé mosquitoes, but
315 frequent migrant haplotypes from the nearby rural Coast populations decrease
316 the efficiency of local adaptation. Importantly, multiple population genomic

317 analyses of the same 28 SNPs (Figure 7D-F) mirror results of the haplotype
318 analysis, confirming that haplotype inference did not bias the results. In sum, we
319 hypothesize that divergence in xenobiotic levels between urban and rural larval
320 habitats is the main ecological force driving spatially variable selection at this
321 locus.
322

DISCUSSION

Reduced representation sequencing of 941 *An. gambiae* s.l. collected in or near human settlements facilitated rapid identification of known sibling species and revealed multiple instances of novel cryptic diversification within *An. gambiae* and *An. coluzzii*. Genome scans in subgroups highlighted numerous selective sweeps centered on detoxification enzymes suggesting that xenobiotic exposure, particularly in urban and suburban areas, is driving local adaptation within both species. More generally, our study highlights the utility of reduced representation sequencing as a 'discovery tool' for finding hidden population structure within species. Further, such dense genotyping data also permits localization of recent selective sweeps, providing insight into the ecological forces driving diversification.

Anthropogenic Mediated Selection

Human activity has altered the evolutionary trajectory of diverse taxa. In insects, spatially varying intensity of insecticide application can drive divergence between populations, potentially leading to reproductive isolation. While a plausible scenario, scant empirical evidence supports the hypothesis [48]. Previous studies have documented dramatic reductions in the population size of *An. gambiae* s.l. after introduction of long lasting insecticide treated nets (LLINs), but did not determine the influence of exposure on population structuring [49, 50]. Among Cameroonian populations of both *An. gambiae and An. coluzzii*, we find a pervasive signature of selection at the *para* sodium channel gene. We infer that this sweep confers globally beneficial resistance to LLINs, which are ubiquitous

347 in Cameroon and treated with pyrethroids that target *para* [51]. In contrast,
348 selective sweeps centered on other insecticide resistance genes are restricted to
349 specific geographic locations/populations. For example, a sweep at the *rdl* locus
350 is limited to southern populations of *An. coluzzii*. Initial selection for dieldrin
351 resistance likely occurred during massive indoor residual spraying campaigns
352 conducted by the WHO in southern Cameroon during the 1950s. Indeed, the
353 spraying was so intense that it temporarily eliminated *An. gambiae* s.l. from
354 Yaoundé (and likely other locations in the forest region) [52]. However, due to
355 high human toxicity, dieldrin has been banned for use in mosquito control since
356 the mid-1970s. In the absence of insecticide exposure, resistant *rdl* mosquitoes
357 are significantly less fit than wild type mosquitoes [53-55], making the continued
358 persistence of resistant alleles in southern *An. coluzzii* populations puzzling. One
359 plausible explanation is that other cyclodienes targeting *rdl*, such as fipronil and
360 lindane, are still commonly used in agriculture and may frequently runoff into *An.*
361 *coluzzii* larval habitats, imposing strong selection for resistant mosquitoes. A
362 similar phenomenon was recently proposed to explain the maintenance of
363 resistance *rdl* alleles in both *Culex* and *Aedes* mosquitoes [56].

364 Mosquitoes inhabiting Cameroon's two major cities, Yaoundé and Douala,
365 provide a clearer example of how xenobiotic exposure can directly influence
366 population structure. Both cities have seen exponential human population growth
367 over the past 50 years, creating a high concentration of hosts for anthropohilic
368 mosquitoes. Despite elevated levels of organic pollutants and insecticides in
369 urban relative to rural larval sites, surveys show substantial year-round

370 populations of *An. gambiae* and *An. coluzzii* in both cities [57-59]. Bioassays of
371 insecticide resistance demonstrate that urban mosquitoes have significantly
372 higher levels of resistance to multiple insecticides compared to rural mosquitoes
373 [30, 31, 57, 60, 61]. In support of human mediated local adaptation, we find a
374 selective sweep in urban *An. coluzzii* mosquitoes centered on a cluster of
375 GSTE/CYP450 detoxification genes. While the specific ecological driver of the
376 selective sweep is unknown, GSTE and P450 enzymes detoxify both organic
377 pollutants and insecticides [62-64]. Indeed, the synergistic effects of the two
378 types of xenobiotics could be exerting intense selection pressure for pleiotropic
379 resistance in urban mosquitoes [44, 45, 65]. Regardless of the underlying targets
380 of selection, it is clear that mosquitoes inhabiting highly disturbed urban and
381 suburban landscapes are genetically differentiated from rural populations.
382 Further analysis of specific sweeps using a combination of whole genome
383 resequencing and emerging functional genetics approaches (e.g. CRISPR/Cas9)
384 should help resolve the specific targets of local adaptation in urban mosquitoes,
385 while also shedding light on the evolutionary history of the enigmatic subgroup
386 *GAM2*.

387 *Impacts on Vector Control*

388 Just five decades ago, there was not a single city in Sub-Saharan African
389 with a population over 1 million; today there are more than 40. Population shifts
390 to urban areas will only continue to increase with the United Nations estimating
391 that 60% of Africans will live in large cities by 2050 [66]. When urbanization
392 commenced, it was widely assumed that malaria transmission would be minimal

393 because rural *Anopheles* vectors would not be able to complete development in
394 the polluted larval habitats present in cities [67]. However, increasingly common
395 reports of endemic malaria transmission in urban areas across Sub-Saharan
396 Africa unequivocally demonstrate that anophelines are exploiting the urban niche
397 [68-70]. Specifically, our study shows that *An. gambiae* s.l. from the urban and
398 suburban centers of southern Cameroon form genetically distinct subgroups
399 relative to rural populations. Local adaptation to urban environments is
400 accompanied by strong selective sweeps centered on putative xenobiotic
401 resistance genes, which are likely driven by a combination of exposure to organic
402 pollutants and insecticides in larval habitats. The rapid adaptation of *Anopheles*
403 to the urban landscape poses a growing health risk to Africans as levels of
404 resistance in these populations negate the effectiveness of almost all commonly
405 used insecticides. Moreover, repeated instances of beneficial alleles
406 introgressing between *An. gambiae* s.l. species make the emergence of highly
407 resistant subgroups even more troubling [14, 16, 41, 42, 71]. In essence, urban
408 populations can serve as a reservoir for resistance alleles, which have the
409 potential to rapidly move between species/populations as needed. Clearly,
410 sustainable malaria vector control, urban or otherwise, requires not only more
411 judicious use of insecticides, but also novel strategies not reliant on chemicals.
412 Towards this goal, various vector control methods that aim to replace or
413 suppress wild mosquito populations using genetic drive are currently under
414 development (e.g. [72]). While promising, the complexities of ongoing cryptic

415 diversification within African *Anopheles* must be explicitly planned for prior to the

416 release of transgenic mosquitoes.

417

418 *MATERIALS AND METHODS*

419
420 *Mosquito collections*
421
422 In 2013, we collected *Anopheles* from 30 locations spread across the four
423 major ecogeographic regions of Cameroon (Table S1). Indoor resting adult
424 mosquitoes were collected by pyrethrum spry catch, while host-seeking adults
425 were obtained via indoor/outdoor human-baited landing catch. Larvae were
426 collected using standard dipping procedures [73]. All researchers were provided
427 with malaria chemoprophylaxis throughout the collection period following [74].
428 Individual mosquitoes belonging to the *An. gambiae* s.l. complex were identified
429 by morphology [75].
430

431 *ddRADseq Library Construction*

432 Genomic DNA was extracted from adults using the ZR-96 Quick-gDNA kit
433 (Zymo Research) and from larvae using the DNeasy Extraction kit (Qiagen). A
434 subset of individuals were assigned to sibling species using PCR-RFLP assays
435 that type fixed SNP differences in the rDNA [76]. Preparation of ddRAD libraries
436 largely followed [77]. Briefly, ~1/3rd of the DNA extracted from an individual
437 mosquito (10ul) was digested with *MluC1* and *NlaIII* (New England Biolabs).
438 Barcoded adapters (1 of 48) were ligated to overhangs and 400 bp fragments
439 were selected using 1.5% gels on a BluePippin (Sage Science). One of six
440 indices was added during PCR amplification. Each library contained 288
441 individuals and was subjected to single end, 100 bp sequencing across one or

442 two flow cells lanes run on an Illumina HiSeq2500. A detailed library preparation
443 protocol is available at mosquitogenomics.org/protocols.
444 Raw sequence reads were demultiplexed and quality filtered using the
445 STACKS v 1.29 process_radtags pipeline [22, 23]. After removal of reads with
446 ambiguous barcodes, incorrect restriction sites, and low sequencing quality
447 (mean Phred < 33), GSNAP was used to align reads to the *An. gambiae* PEST
448 reference genome (AgamP4.2) allowing up to five mismatches per read. After
449 discarding reads that perfectly aligned to more than one genomic position, we
450 used STACKS to identify unique RAD tags and construct consensus assemblies
451 for each. Individual SNP genotypes were called using default setting in the
452 maximum-likelihood statistical model implemented in the STACKS genotypes
453 pipeline.
454
455 *Population Genomic Analysis*
456 Population genetic structure was assessed using the SNP dataset output
457 by the *populations* program of STACKS. We used PLINK v 1.19 to retrieve
458 subsets of genome-wide SNPs as needed [78]. PCA, neighbor-joining tree
459 analyses, and Bayesian information criterion (BIC) were implemented using the
460 packages *adegenet* and *ape* in R [79, 80]. Ancestry analyses were conducted in
461 fastSTRUCTURE v 1.0 [29] using the logistic method. The choosek.py script was
462 used to find the appropriate number of populations (k); in cases where a range of
463 k was suggested, the BIC-inferred number of clusters was chosen [81]. Ancestry
464 assignment of individual mosquitoes was then visualized with DISTRUCT v 1.1

465 [82]. We input pairwise F_{ST} values into the program Fitch from the Phylip [83]
466 suite to create the population-level NJ tree.
467
468 *Genome Scans for Selection*
469 We used ANGSD v 0.612 [32] to calculate nucleotide diversity (θ_w and θ_π)
470 and Tajima's *D* in 150-kb non-overlapping windows. Unlike most genotyping
471 algorithms, ANGSD does not perform hard SNP calls, instead taking genotyping
472 uncertainty into account when calculating summary statistics. Similarly, absolute
473 divergence (d_{xy}) was calculated using *ngsTools* [84] based on genotype
474 likelihoods generated by ANGSD. Kernel smoothed values for 150-kb windows
475 for all four metrics (θ_w, θ_π, D, d_{xy}) were obtained with the R package *KernSmooth*.
476 F_{ST} (based on AMOVA) was calculated with the *populations* program in STACKS
477 using only loci present in 80% of individuals. A Kernel smoothing procedure
478 implemented in STACKS was used to obtain F_{ST} values across 150-kb windows.
479 Selective sweeps were identified by locating regions where Tajima's *D* was in the
480 top percentile of empirical values in at least one population. Regions with low
481 Tajima's *D*, but unusually high or low read depth (Figure S4) were deemed
482 unreliable due to the likelihood of repeats and local misassembly. We used a
483 subset of 1,000 randomly chosen SNPs to calculate average pairwise F_{ST}
484 between populations in GENODIVE using up to 40 individuals – prioritized by
485 coverage – per population [85]. To determine if selective sweeps were enriched
486 for specific functional annotation classes, we used the program DAVID 6.7 with

487 default settings [86]. Haplotypes across the GSTE/CYP450 sweep were

488 reconstructed by PHASE v 2.1.1 using the default recombination model (87).

489

ACKNOWLEDGEMENTS

This work was supported by the University of California Riverside and National Institutes of Health (1R01AI113248, 1R21AI115271 to BJW). We thank Elysée Nchoutpouem and Raymond Fokom for assistance collecting mosquitoes and Sina Hananian for assisting in DNA extraction.

AUTHOR CONTRIBUTIONS

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

502 REFERENCES
503

504 1. Nosil, P., Vines, T.H., and Funk, D.J. (2005). Reproductive isolation
505 caused by natural selection against immigrants from divergent habitats.
506 Evolution *59*, 705-719.
507 2. Hereford, J. (2009). A quantitative survey of local adaptation and fitness
508 trade-offs. The American Naturalist *173*, 579-588.
509 3. Gaston, K.J. (2010). Urban ecology, (Cambridge University Press).
510 4. Pelz, H.-J., Rost, S., Hünerberg, M., Fregin, A., Heiberg, A.-C., Baert, K.,
511 MacNicoll, A.D., Prescott, C.V., Walker, A.-S., and Oldenburg, J. (2005).
512 The genetic basis of resistance to anticoagulants in rodents. Genetics
513 *170*, 1839-1847.
514 5. Song, Y., Endepols, S., Klemann, N., Richter, D., Matuschka, F.-R., Shih,
515 C.-H., Nachman, M.W., and Kohn, M.H. (2011). Adaptive introgression of
516 anticoagulant rodent poison resistance by hybridization between old world
517 mice. Current Biology *21*, 1296-1301.
518 6. Davies, T.G.E., Field, L.M., and Williamson, M.S. (2012). The re-
519 emergence of the bed bug as a nuisance pest: implications of resistance
520 to the pyrethroid insecticides. Medical and Veterinary Entomology *26*, 241-
521 254.
522 7. Kawecki, T.J., and Ebert, D. (2004). Conceptual issues in local adaptation.
523 Ecology Letters *7*, 1225-1241.

524 8. Berdahl, A., Torney, C.J., Schertzer, E., and Levin, S.A. (2015). On the
525 evolutionary interplay between dispersal and local adaptation in
526 heterogeneous environments. Evolution.
527 9. Le Corre, V., and Kremer, A. (2012). The genetic differentiation at
528 quantitative trait loci under local adaptation. Molecular Ecology *21*, 1548-
529 1566.
530 10. Tiffin, P., and Ross-Ibarra, J. (2014). Advances and limits of using
531 population genetics to understand local adaptation. Trends in ecology &
532 evolution *29*, 673-680.
533 11. Feder, J.L., Flaxman, S.M., Egan, S.P., Comeault, A.A., and Nosil, P.
534 (2013). Geographic mode of speciation and genomic divergence. Annual
535 Review of Ecology, Evolution, and Systematics *44*, 73-97.
536 12. White, B.J., Collins, F.H., and Besansky, N.J. (2011). Evolution of
537 Anopheles gambiae in Relation to Humans and Malaria. Annu Rev Ecol
538 Evol S *42*, 111-132.
539 13. Coetzee, M., Hunt, R.H., Wilkerson, R., Della Torre, A., Coulibaly, M.B.,
540 and Besansky, N.J. (2013). Anopheles coluzzii and Anopheles amharicus,
541 new members of the Anopheles gambiae complex. Zootaxa *3619*, 246-
542 274.
543 14. Crawford, J., Riehle, M.M., Guelbeogo, W.M., Gneme, A., Sagnon, N.f.,
544 Vernick, K.D., Nielsen, R., and Lazzaro, B.P. (2014). Reticulate speciation
545 and adaptive introgression in the Anopheles gambiae species complex.
546 bioRxiv, 009837.

547 15. Holt, R.A., Subramanian, G.M., Halpern, A., Sutton, G.G., Charlab, R.,
548 Nusskern, D.R., Wincker, P., Clark, A.G., Ribeiro, J.M., Wides, R., et al.
549 (2002). The genome sequence of the malaria mosquito *Anopheles*
550 *gambiae*. Science *298*, 129-149.
551 16. Fontaine, M.C., Pease, J.B., Steele, A., Waterhouse, R.M., Neafsey, D.E.,
552 Sharakhov, I., Jiang, X., Hall, A.B., Catteruccia, F., Kakani, E., et al.
553 (2014). Extensive introgression in a malaria vector species complex
554 revealed by phylogenomics. Science.
555 17. Slotman, M.A., Tripet, F., Cornel, A.J., Meneses, C.R., Lee, Y., Reimer,
556 L.J., Thiemann, T.C., Fondjo, E., Fofana, A., Traore, S.F., et al. (2007).
557 Evidence for subdivision within the M molecular form of Anopheles
558 gambiae. Mol Ecol *16*, 639-649.
559 18. Wang-Sattler, R., Blandin, S., Ning, Y., Blass, C., Dolo, G., Toure, Y.T.,
560 della Torre, A., Lanzaro, G.C., Steinmetz, L.M., Kafatos, F.C., et al.
561 (2007). Mosaic Genome Architecture of the Anopheles gambiae Species
562 Complex. PLoS One *2*, -.
563 19. Lee, Y., Marsden, C.D., Norris, L.C., Collier, T.C., Main, B.J., Fofana, A.,
564 Cornel, A.J., and Lanzaro, G.C. (2013). Spatiotemporal dynamics of gene
565 flow and hybrid fitness between the M and S forms of the malaria
566 mosquito, Anopheles gambiae. Proceedings of the National Academy of
567 Sciences *110*, 19854-19859.
568 20. Caputo, B., Nwakanma, D., Caputo, F., Jawara, M., Oriero, E.,
569 Hamid-Adiamoh, M., Dia, I., Konate, L., Petrarca, V., and Pinto, J. (2014).

570 Prominent intraspecific genetic divergence within Anopheles gambiae
571 sibling species triggered by habitat discontinuities across a riverine
572 landscape. Molecular Ecology *23*, 4574-4589.
573 21. Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., and Hoekstra, H.E.
574 (2012). Double digest RADseq: an inexpensive method for de novo SNP
575 discovery and genotyping in model and non-model species. PLoS One *7*,
576 e37135.
577 22. Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., and Cresko,
578 W.A. (2013). Stacks: an analysis tool set for population genomics.
579 Molecular Ecology *22*, 3124-3140.
580 23. Catchen, J.M., Amores, A., Hohenlohe, P., Cresko, W., and Postlethwait,
581 J.H. (2011). Stacks: building and genotyping loci de novo from short-read
582 sequences. G3: Genes, Genomes, Genetics *1*, 171-182.
583 24. Scott, J.A., Brogdon, W.G., and Collins, F.H. (1993). Identification of
584 single specimens of the Anopheles gambiae complex by the polymerase
585 chain reaction. Am J Trop Med Hyg *49*, 520-529.
586 25. Santolamazza, F., Della Torre, A., and Caccone, A. (2004). Short report: A
587 new polymerase chain reaction-restriction fragment length polymorphism
588 method to identify *Anopheles arabiensis* from *An. gambiae* and its two
589 molecular forms from degraded DNA templates or museum samples. Am
590 J Trop Med Hyg *70*, 604-606.
591 26. Wondji, C., Frederic, S., Petrarca, V., Etang, J., Santolamazza, F., Della
592 Torre, A., and Fontenille, D. (2005). Species and populations of the

593 *Anopheles gambiae* complex in Cameroon with special emphasis on
594 chromosomal and molecular forms of *Anopheles gambiae s.s*. J Med
595 Entomol *42*, 998-1005.
596 27. Simard, F., Ayala, D., Kamdem, G.C., Etouna, J., Ose, K., Fotsing, J.-M.,
597 Fontenille, D., Besansky, N.J., and Costantini, C. (2009). Ecological niche
598 partitioning between the M and S molecular forms of *Anopheles gambiae*
599 in Cameroon: the ecological side of speciation. BMC Ecology *9*, 17.
600 28. Lee, Y., Marsden, C.D., Norris, L.C., Collier, T.C., Main, B.J., Fofana, A.,
601 Cornel, A.J., and Lanzaro, G.C. (2013). Spatiotemporal dynamics of gene
602 flow and hybrid fitness between the M and S forms of the malaria
603 mosquito, Anopheles gambiae. Proceedings of the National Academy of
604 Sciences of the United States of America *110*, 19854-19859.
605 29. Raj, A., Stephens, M., and Pritchard, J.K. (2014). fastSTRUCTURE:
606 variational inference of population structure in large SNP data sets.
607 Genetics *197*, 573-589.
608 30. Nwane, P., Etang, J., Chouaibou, M., Toto, J.C., Koffi, A., Mimpfoundi, R.,
609 and Simard, F. (2013). Multiple insecticide resistance mechanisms in
610 Anopheles gambiae sl populations from Cameroon, Central Africa. Parasit
611 Vectors *6*, 41.
612 31. Tene, B.F., Poupardin, R., Costantini, C., Awono-Ambene, P., Wondji,
613 C.S., Ranson, H., and Antonio-Nkondjio, C. (2013). Resistance to DDT in
614 an urban setting: common mechanisms implicated in both M and S forms

615 of Anopheles gambiae in the city of Yaoundé Cameroon. PLoS One *8*,
616 e61408.
617 32. Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD:
618 analysis of next generation sequencing data. BMC bioinformatics *15*, 356.
619 33. Maynard Smith, J., and Haigh, j. (1974). The hitch-hiking effect of a
620 favorable gene. Genetical research *23*, 22-35.
621 34. Tajima, F. (1989). Statistical method for testing the neutral mutation
622 hypothesis by DNA polymorphism. Genetics *123*, 585-595.
623 35. Storz, J.F. (2005). Using genome scans of DNA polymorphism to infer
624 adaptive population divergence. Mol Ecol *14*, 671-688.
625 36. Cruickshank, T.E., and Hahn, M.W. (2014). Reanalysis suggests that
626 genomic islands of speciation are due to reduced diversity, not reduced
627 gene flow. Molecular Ecology *23*, 3133-3157.
628 37. O'Loughlin, S.M., Magesa, S., Mbogo, C., Mosha, F., Midega, J., Lomas,
629 S., and Burt, A. (2014). Genomic analyses of three malaria vectors reveals
630 extensive shared polymorphism but contrasting population histories.
631 Molecular Biology and Evolution *31*, 889-902.
632 38. Donnelly, M.J., Corbel, V., Weetman, D., Wilding, C.S., Williamson, M.S.,
633 and Black, W.C. (2009). Does kdr genotype predict insecticide-resistance
634 phenotype in mosquitoes? Trends in Parasitology *25*, 213-219.
635 39. Lynd, A., Weetman, D., Barbosa, S., Yawson, A.E., Mitchell, S., Pinto, J.,
636 Hastings, I., and Donnelly, M.J. (2010). Field, genetic, and modeling
637 approaches show strong positive selection acting upon an insecticide

638 resistance mutation in Anopheles gambiae ss. Molecular Biology and
639 Evolution *27*, 1117-1125.
640 40. Jones, C.M., Liyanapathirana, M., Agossa, F.R., Weetman, D., Ranson,
641 H., Donnelly, M.J., and Wilding, C.S. (2012). Footprints of positive
642 selection associated with a mutation (N1575Y) in the voltage-gated
643 sodium channel of Anopheles gambiae. Proceedings of the National
644 Academy of Sciences *109*, 6614-6619.
645 41. Clarkson, C.S., Weetman, D., Essandoh, J., Yawson, A.E., Maslen, G.,
646 Manske, M., Field, S.G., Webster, M., Antão, T., and MacInnis, B. (2014).
647 Adaptive introgression between Anopheles sibling species eliminates a
648 major genomic island but not reproductive isolation. Nature
649 communications *5*.
650 42. Norris, L.C., Main, B.J., Lee, Y., Collier, T.C., Fofana, A., Cornel, A.J., and
651 Lanzaro, G.C. (2015). Adaptive introgression in an African malaria
652 mosquito coincident with the increased usage of insecticide-treated bed
653 nets. Proceedings of the National Academy of Sciences *112*, 815-820.
654 43. Enayati, A.A., Ranson, H., and Hemingway, J. (2005). Insect glutathione
655 transferases and insecticide resistance. Insect Molecular Biology *14*, 3-8.
656 44. Nkya, T.E., Akhouayri, I., Kisinza, W., and David, J.-P. (2013). Impact of
657 environment on mosquito response to pyrethroid insecticides: facts,
658 evidences and prospects. Insect Biochemistry and Molecular Biology *43*,
659 407-416.

660 45. David, J.-P., Ismail, H.M., Chandor-Proust, A., and Paine, M.J.I. (2013).
661 Role of cytochrome P450s in insecticide resistance: impact on the control
662 of mosquito-borne diseases and use of insecticides on Earth.
663 Philosophical Transactions of the Royal Society of London B: Biological
664 Sciences *368*, 20120429.
665 46. Dabiré, R.K., Namountougou, M., Sawadogo, S.P., Yaro, L.B., Toé, H.K.,
666 Ouari, A., Gouagna, L.-C., Simard, F., Chandre, F., and Baldet, T. (2012).
667 Population dynamics of Anopheles gambiae sl in Bobo-Dioulasso city:
668 bionomics, infection rate and susceptibility to insecticides. Parasit Vectors
669 *5*, 127.
670 47. Jones, C.M., Toé, H.K., Sanou, A., Namountougou, M., Hughes, A.,
671 Diabaté, A., Dabiré, R., Simard, F., and Ranson, H. (2012). Additional
672 selection for insecticide resistance in urban malaria vectors: DDT
673 resistance in Anopheles arabiensis from Bobo-Dioulasso, Burkina Faso.
674 PLoS One *7*, e45995.
675 48. Chen, H., Wang, H., and Siegfried, B.D. (2012). Genetic differentiation of
676 western corn rootworm populations (Coleoptera: Chrysomelidae) relative
677 to insecticide resistance. Ann Entomol Soc Am *105*, 232-240.
678 49. Athrey, G., Hodges, T.K., Reddy, M.R., Overgaard, H.J., Matias, A., Ridl,
679 F.C., Kleinschmidt, I., Caccone, A., and Slotman, M.A. (2012). The
680 effective population size of malaria mosquitoes: large impact of vector
681 control. PLoS genetics *8*, e1003097.

682 50. Bayoh, M.N., Mathias, D.K., Odiere, M.R., Mutuku, F.M., Kamau, L.,
683 Gimnig, J.E., Vulule, J.M., Hawley, W.A., Hamel, M.J., and Walker, E.D.
684 (2010). Anopheles gambiae: historical population decline associated with
685 regional distribution of insecticide-treated bed nets in western Nyanza
686 Province, Kenya.
687 51. Bowen, H.L. (2013). Impact of a mass media campaign on bed net use in
688 Cameroon. Malar J *12*, 10.1186.
689 52. Livadas, G., Mouchet, J., Gariou, J., and Chastang, R. (1958). Can one
690 foresee the eradication of malaria in wooded areas of South Cameroun.
691 Rivista di Malariologia *37*, 229.
692 53. Rowland, M. (1991). Activity and mating competitiveness of gamma
693 HCH/dieldrin resistant and susceptible male and virgin female Anopheles
694 gambiae and An. stephensi mosquitoes, with assessment of an
695 insecticide-rotation strategy. Medical and Veterinary Entomology *5*, 207.
696 54. Rowland, M. (1991). Behaviour and fitness of gamma HCH/dieldrin
697 resistant and susceptible female Anopheles gambiae and An. stephensi
698 mosquitoes in the absence of insecticide. Medical and Veterinary
699 Entomology *5*, 193.
700 55. Platt, N., Kwiatkowska, R., Irving, H., Diabaté, A., Dabire, R., and Wondji,
701 C. (2015). Target-site resistance mutations (kdr and RDL), but not
702 metabolic resistance, negatively impact male mating competiveness in the
703 malaria vector Anopheles gambiae. Heredity.

704 56. Tantely, M.L., Tortosa, P., Alout, H., Berticat, C., Berthomieu, A., Rutee,
705 A., Dehecq, J.-S., Makoundou, P., Labbé, P., and Pasteur, N. (2010).
706 Insecticide resistance in Culex pipiens quinquefasciatus and Aedes
707 albopictus mosquitoes from La Reunion Island. Insect Biochemistry and
708 Molecular Biology *40*, 317-324.
709 57. Antonio-Nkondjio, C., Fossog, B.T., Ndo, C., Djantio, B.M., Togouet, S.Z.,
710 Awono-Ambene, P., Costantini, C., Wondji, C.S., and Ranson, H. (2011).
711 Anopheles gambiae distribution and insecticide resistance in the cities of
712 Douala and Yaounde(Cameroon): influence of urban agriculture and
713 pollution. Malaria Journal *10*, 154-154.
714 58. Kamdem, C., Fossog, B.T., Simard, F., Etouna, J., Ndo, C., Kengne, P.,
715 Boussès, P., Etoa, F.-X., Awono-Ambene, P., and Fontenille, D. (2012).
716 Anthropogenic habitat disturbance and ecological divergence between
717 incipient species of the malaria mosquito Anopheles gambiae. PLoS One
718 *7*, e39453.
719 59. Antonio-Nkondjio, C., Youmsi-Goupeyou, M., Kopya, E., Tene-Fossog, B.,
720 Njiokou, F., Costantini, C., and Awono-Ambene, P. (2014). Exposure to
721 disinfectants (soap or hydrogen peroxide) increases tolerance to
722 permethrin in Anopheles gambiae populations from the city of Yaoundé,
723 Cameroon. Malaria Journal *13*, 296.
724 60. Nwane, P., Etang, J., Chouaibou, M., Toto, J.C., Kerah-Hinzoumbé, C.,
725 Mimpfoundi, R., Awono-Ambene, H.P., and Simard, F. (2009). Trends in
726 DDT and pyrethroid resistance in Anopheles gambiae ss populations from

727 urban and agro-industrial settings in southern Cameroon. BMC infectious
728 diseases *9*, 163.
729 61. Antonio-Nkondjio, C., Fossog, B.T., Kopya, E., Poumachu, Y., Djantio,
730 B.M., Ndo, C., Tchuinkam, T., Awono-Ambene, P., and Wondji, C.S.
731 (2015). Rapid evolution of pyrethroid resistance prevalence in Anopheles
732 gambiae populations from the cities of Douala and Yaoundé (Cameroon).
733 Malaria Journal *14*, 1-9.
734 62. Suwanchaichinda, C., and Brattsten, L. (2001). Effects of exposure to
735 pesticides on carbaryl toxicity and cytochrome P450 activities in Aedes
736 albopictus larvae (Diptera: Culicidae). Pesticide Biochemistry and
737 Physiology *70*, 63-73.
738 63. David, J.-P., Coissac, E., Melodelima, C., Poupardin, R., Riaz, M.A.,
739 Chandor-Proust, A., and Reynaud, S. (2010). Transcriptome response to
740 pollutants and insecticides in the dengue vector Aedes aegypti using next-
741 generation sequencing technology. BMC Genomics *11*, 216.
742 64. Poupardin, R., Riaz, M.A., Jones, C.M., Chandor-Proust, A., Reynaud, S.,
743 and David, J.-P. (2012). Do pollutants affect insecticide-driven gene
744 selection in mosquitoes? Experimental evidence from transcriptomics.
745 Aquatic Toxicology *114*, 49-57.
746 65. Mueller, P., Chouaibou, M., Pignatelli, P., Etang, J., Walker, E.D.,
747 Donnelly, M.J., Simard, F., and Ranson, H. (2008). Pyrethroid tolerance is
748 associated with elevated expression of antioxidants and agricultural

749 practice in Anopheles arabiensis sampled from an area of cotton fields in
750 Northern Cameroon. Molecular Ecology *17*, 1145-1155.
751 66. United Nations, D.o.E.a.S.A., Population Division (2014). World
752 Urbanization Prospects: The 2014 Revision, Highlights.
753 (ST/ESA/SER.A/352).
754 67. Donnelly, M.J., McCall, P., Lengeler, C., Bates, I., D'Alessandro, U.,
755 Barnish, G., Konradsen, F., Klinkenberg, E., Townson, H., and Trape, J.-
756 F. (2005). Malaria and urbanization in sub-Saharan Africa. Malar J *4*, 12.
757 68. Robert, V., Macintyre, K., Keating, J., Trape, J.-F., Duchemin, J.-B.,
758 Warren, M., and Beier, J.C. (2003). Malaria transmission in urban sub-
759 Saharan Africa. The American Journal of Tropical Medicine and Hygiene
760 *68*, 169-176.
761 69. Keiser, J., Utzinger, J., De Castro, M.C., Smith, T.A., Tanner, M., and
762 Singer, B.H. (2004). Urbanization in sub-saharan Africa and implication for
763 malaria control. The American Journal of Tropical Medicine and Hygiene
764 *71*, 118-127.
765 70. De Silva, P.M., and Marshall, J.M. (2012). Factors contributing to urban
766 malaria transmission in sub-Saharan Africa: a systematic review. Journal
767 of tropical medicine *2012*.
768 71. Weill, M., Chandre, F., Brengues, C., Manguin, S., Akogbeto, M., Pasteur,
769 N., Guillet, P., and Raymond, M. (2000). The kdr mutation occurs in the
770 Mopti form of Anopheles gambiae s.s. through introgression. Insect Mol
771 Biol *9*, 451-455.

772 72. Windbichler, N., Menichelli, M., Papathanos, P.A., Thyme, S.B., Li, H.,
773 Ulge, U.Y., Hovde, B.T., Baker, D., Monnat, R.J., Jr., Burt, A., et al.
774 (2011). A synthetic homing endonuclease-based gene drive system in the
775 human malaria mosquito. Nature *473*, 212-215.
776 73. Service, M.W. (1993). Mosquito Ecology. Field Sampling Methods, 2nd
777 ed., (London: Elsevier Applied Science Publishers, Ltd.).
778 74. Gimnig, J.E., Walker, E.D., Otieno, P., Kosgei, J., Olang, G., Ombok, M.,
779 Williamson, J., Marwanga, D., Abong'o, D., and Desai, M. (2013).
780 Incidence of malaria among mosquito collectors conducting human
781 landing catches in western Kenya. The American Journal of Tropical
782 Medicine and Hygiene *88*, 301-308.
783 75. Gillies, M.T., and De Meillon, B. (1968). The Anophelinae of Africa South
784 of the Sahara, 2nd Edition, (Johannesburg: South African Institute for
785 Medical Research).
786 76. Fanello, C., Santolamazza, F., and della Torre, A. (2002). Simultaneous
787 identification of species and molecular forms of the *Anopheles gambiae*
788 complex by PCR-RFLP. Med Vet Entomol *16*, 461-464.
789 77. Turissini, D.A., Gamez, S., and White, B.J. (2014). Genome-wide patterns
790 of polymorphism in an inbred line of the African malaria mosquito
791 Anopheles gambiae. Genome biology and evolution *6*, 3094-3104.
792 78. Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.,
793 Bender, D., Maller, J., Sklar, P., De Bakker, P.I., and Daly, M.J. (2007).

794 PLINK: a tool set for whole-genome association and population-based
795 linkage analyses. The American Journal of Human Genetics *81*, 559-575.
796 79. Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of
797 phylogenetics and evolution in R language. Bioinformatics *20*, 289-290.
798 80. Jombart, T. (2008). adegenet: a R package for the multivariate analysis of
799 genetic markers. Bioinformatics *24*, 1403-1405.
800 81. Jakobsson, M., and Rosenberg, N.A. (2007). CLUMPP: a cluster matching
801 and permutation program for dealing with label switching and
802 multimodality in analysis of population structure. Bioinformatics *23*, 1801-
803 1806.
804 82. Rosenberg, N.A. (2004). DISTRUCT: a program for the graphical display
805 of population structure. Mol Ecol Notes *4*, 137-138.
806 83. Plotree, D., and Plotgram, D. (1989). PHYLIP-phylogeny inference
807 package (version 3.2). cladistics *5*, 163-166.
808 84. Fumagalli, M., Vieira, F.G., Linderoth, T., and Nielsen, R. (2014).
809 ngsTools: methods for population genetics analyses from next-generation
810 sequencing data. Bioinformatics *30*, 1486-1487.
811 85. Meirmans, P.G., and Van Tienderen, P.H. (2004). GENOTYPE and
812 GENODIVE: two programs for the analysis of genetic diversity of asexual
813 organisms. Mol Ecol Notes *4*, 792-794.
814 86. Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2008). Systematic and
815 integrative analysis of large gene lists using DAVID bioinformatics
816 resources. Nature protocols *4*, 44-57.
817

818 FIGURE LEGENDS

819 **Figure 1. *Anopheles gambiae* complex sibling species are genetically**
820 **distinct.** A) PCA B) NJ Tree, and C) fastSTRUCTURE analyses clearly separate
821 the four *An. gambiae* complex species that occur in Cameroon into discrete
822 genetic clusters. Additional subdivison below the species level is apparent within
823 *An. coluzzii* and *An. gambiae.* D) Species composition varies dramatically
824 between eco-geographic regions. Sampling sites are denoted by black dots.

825

826 **Figure 2. *Anopheles gambiae* is divided into three cryptic subpopulations.**
827 A) PCA, B) NJ Tree, and C) fastSTRUCTURE analyses reveal subdivision within
828 *An. gambiae.* We term the most abundant group *GAM1*, while a second small,
829 but widely distributed group, is termed *GAM2*. Finally, most individuals from the
830 village of Nkolondom are genetically distinct from other *An. gambiae* suggestive
831 of local adaptation.

832

833 **Figure 3. *Anopheles coluzzii* is divided into four subgroups.** A) PCA, NJ
834 Tree, and C) fastSTRUCTURE reveal major population structuring between *An.*
835 *coluzzii* from the northern Savannah eco-geographic region and *An. coluzzii* from
836 the southern three forested regions. Within the south, D) PCA, E) NJ Tree, and
837 F) fastSTRUCTURE analyses separate mosquitoes based on geographic origin,
838 although clustering is not fully discrete indicating a dynamic interplay between
839 local adaptation and migration.

840

841 **Figure 4. Phylogenetic relationships between populations show recent**
842 **radiation within *An. gambiae* and *An. coluzzii* clades**. In an unrooted, F_{ST}-
843 based NJ tree, *An. melas* is most distant from all other species, while *An.*
844 *gambiae* and *An. coluzzii* are sister species. Southern populations of *An. coluzzii*
845 are more closely related to each other than to the northern savannah population.
846 In contrast to geographic distance, the Douala subpopulation is genetically closer
847 to Yaoundé rather than Coastal mosquitoes. Within *An. gambiae*, a relatively
848 deep split is present between *GAM2* and *GAM1*, while *Nkolondom* appears to
849 have recently diverged from *GAM1*.
850
851 **Figure 5. Genome scans reveal footprints of global and local adaptation in**
852 ***An. gambiae* subpopulations.** A-C) Diversity and Tajima's *D* are plotted for
853 each of three subpopulations. A dashed line marks the 1st percentile empirical
854 distribution of Tajima's *D*. In all populations, concordant dips in diversity and
855 Tajima's *D* are evident near the pericentromeric region of 2L where the *para*
856 sodium channel gene is located. A population-specific sweep in Nkolondom is
857 present at ~34 Mb on 2L. D-E) Both absolute (d_{xy}) and relative (F_{ST}) divergence
858 between populations are plotted with a dashed line marking the 99th percentile
859 distribution of F_{ST}. An increase in F_{ST} between *GAM1-Nkolondom* occurs at the
860 *Nkolondom*-specific 2L sweep.
861
862 **Figure 6. Strong positive selection acts on xenobiotic resistance loci in**
863 **subpopulations of *An. coluzzii.*** A-D) Sharp declines in diversity and allele

frequency spectrum at the *para* sodium channel gene are present in all populations. A sweep centered on the resistance to dieldrin (*rdl*) gene is present in southern subpopulations, while a dramatic sweep encompassing a cluster of detoxification genes on 3R is limited to urban mosquitoes. E-G) No evidence for locally elevated divergence is observed at the *para* or *rdl* loci suggesting a shared sweep amongst populations. In contrast, urban-rural mosquitoes show extreme levels of divergence at the detoxification-enriched sweep on 3R.

Figure 7. Spatially varying selection between urban and coastal populations. For each of the three southern *An. coluzzii* subpopulations, 80 reconstructed haplotypes are visualized by color-coding 28 bi-allelic SNPs in the 3R GSTE/CYP450 sweep either grey or white. A single invariant haplotype -- all grey SNPs -- is common in (A) Yaoundé, less so in (B) Douala, and very rare in (C) coastal populations. D-E) Similarly, in PCA and NJ Tree analysis of the same 28 SNPs, coastal individuals (navy blue) are diffuse across genotypic space, while Yaounde mosquitoes (purple) are tightly clustered. As expected, Douala (pink)exhibits an intermediate degree of variation. F) STRUCTURE analysis based solely on the 28 SNPs within the sweep shows clear distinctions between the three populations.

887 SUPPLEMENTAL FIGURE LEGENDS

888 **Figure S1.** Bayesian information criterion was used to determine the most likely
889 number of clusters/populations for A) all 941 samples, B) all *An. coluzzii*, C) *An.*
890 *coluzzii* from the northern savannah, D) *An. arabiensis*, E) the 309 individuals
891 used for genome scans F) all *An. gambiae*, G) southern *An. coluzzii*, and H) *An.*
892 *melas.* BIC scores for 1 to 50 clusters are plotted. The lower the BIC score the
893 better the model fits the observed genetic diversity.

894

895 **Figure S2.** Southern (pale pink) and northern populations (deep pink) of *An.*
896 *coluzzii* are readily separated in PCA (top) and NJ trees (bottom) using SNPs
897 exclusively from any of the five chromosomal arms.

898

899 **Figure S3.** No population substructure is detectable within northern *An. coluzzii*
900 using PCA and NJ tree analysis.

901

902 **Figure S4.** Mean sequencing coverage per individual is plotted in 300kb non-
903 overlapping windows across the genome. Only individuals used in the genome
904 scans are included in the coverage calculation.

Figure 1.

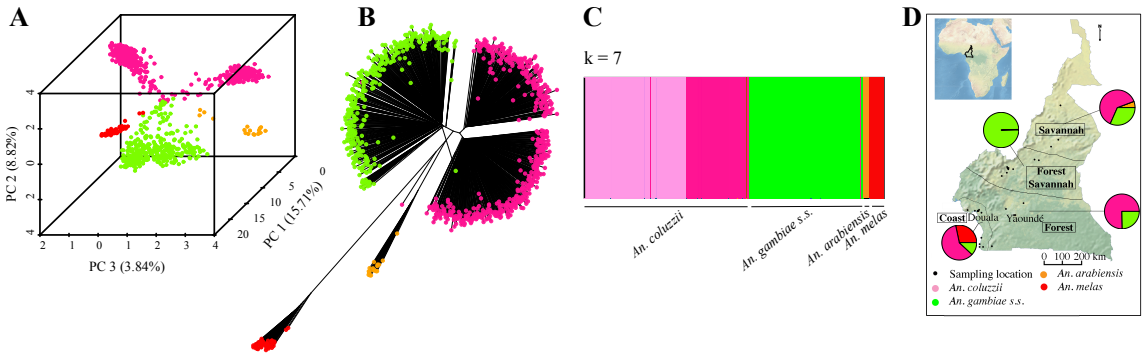

A
PC 2 (8.82%)
PC 3 (3.84%)
PC 1 (15.71%)
B
C
k = 7
An. coluzzii
An. gambiae s.s.
An. arabiensis
An. melas
D
Savannah
Forest Savannah
Coast
Douala
Yaoundé
Forest
0 100 200 km
Sampling location
An. coluzzii
An. gambiae s.s.
An. arabiensis
An. melas

Figure 2.

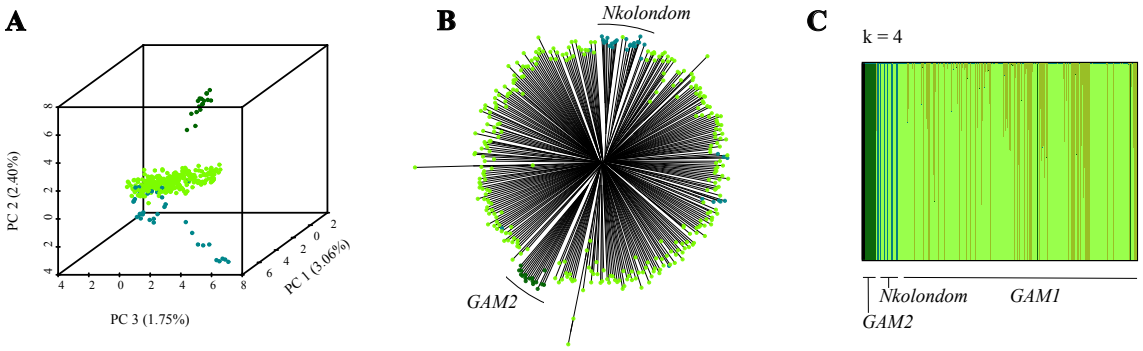

A
PC 2 (2.40%)
PC 3 (1.75%)
PC 1 (3.06%)
B
Nkolondom
GAM2
C
k = 4
Nkolondom
GAM1
GAM2

Figure 3.

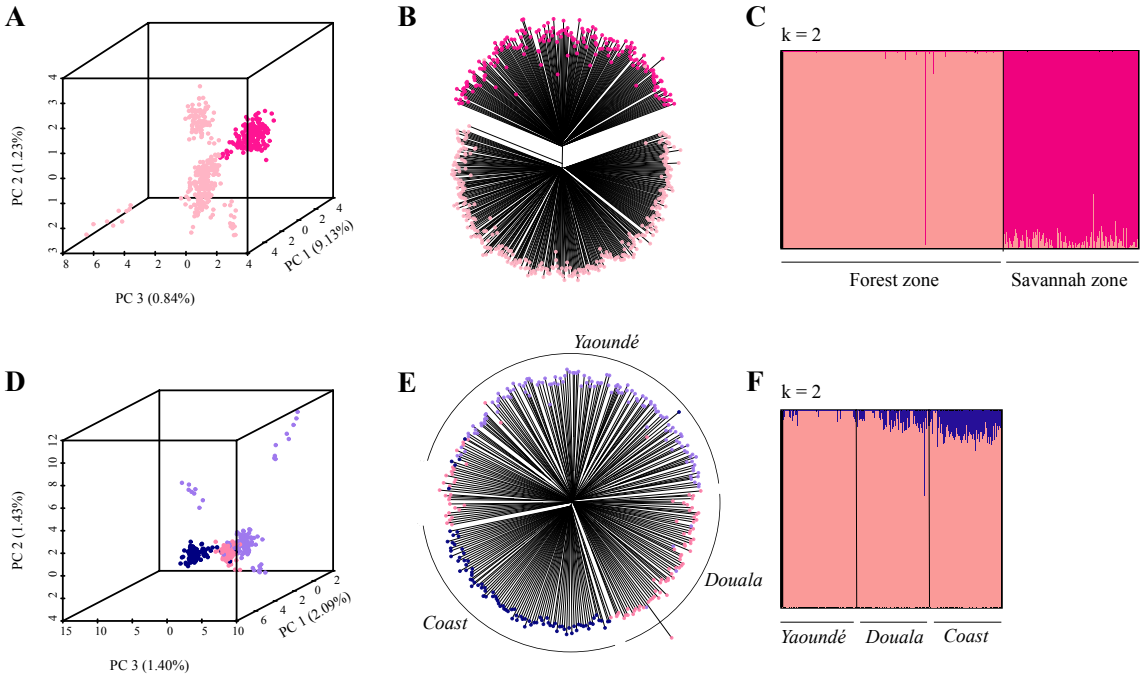

A
PC 2 (1.23%)
PC 3 (0.84%)
PC 1 (9.13%)
B
C
k = 2
Forest zone
Savannah zone
D
PC 2 (1.43%)
PC 3 (1.40%)
PC 1 (2.09%)
E
Yaoundé
Douala
Coast
F
k = 2
Yaoundé
Douala
Coast

Figure 4.

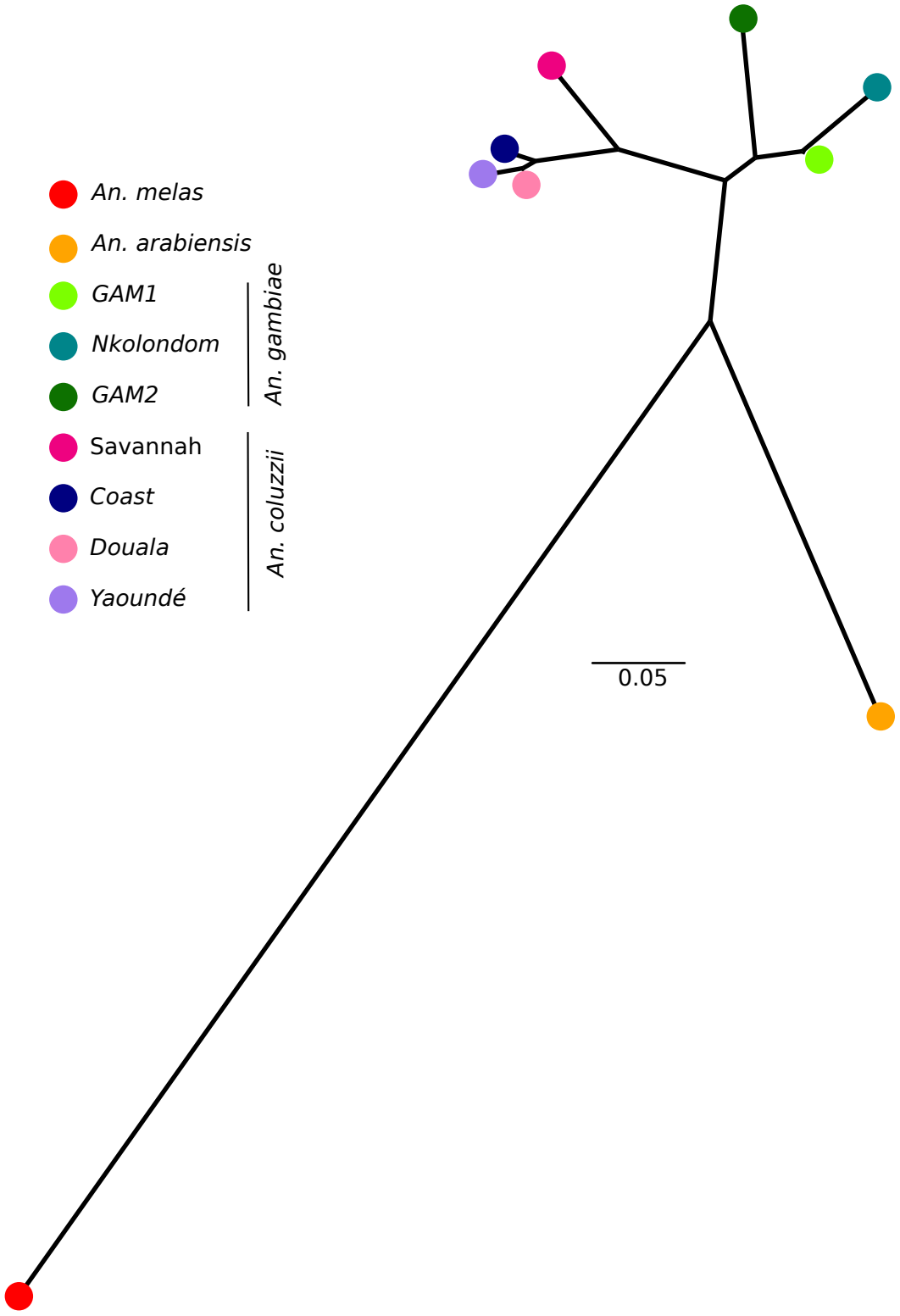

An. melas
An. arabiensis
GAM1
Nkolondom
GAM2
Savannah
Coast
Douala
Yaoundé
An. gambiae
An. coluzzii
0.05

Figure 5.

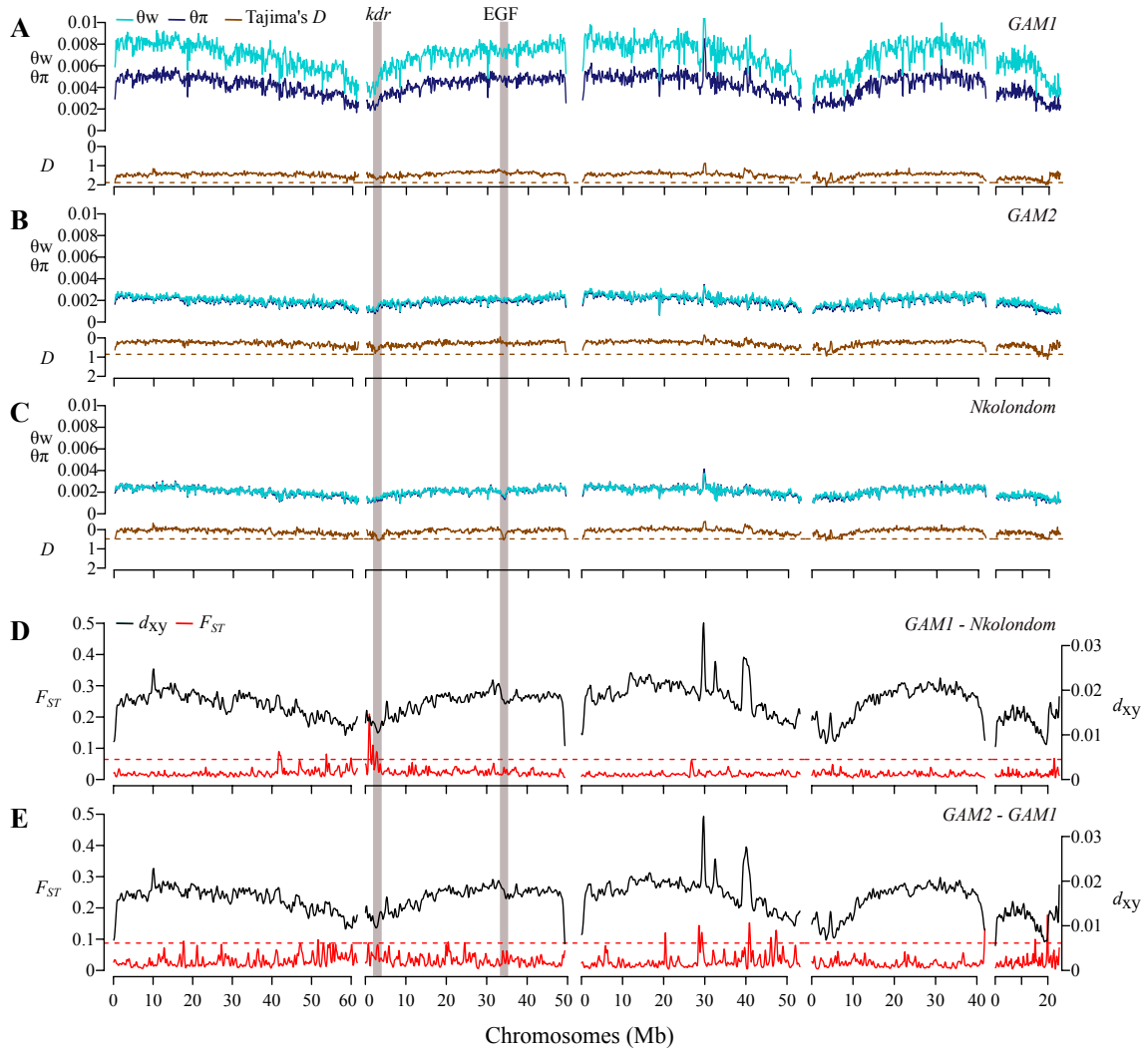

A
B
C
D
E
θw
θπ
Tajima's D
kdr
EGF
GAM1
GAM2
Nkolondom
GAM1 - Nkolondom
GAM2 - GAM1
dxy
FST
Chromosomes (Mb)

Figure 6.

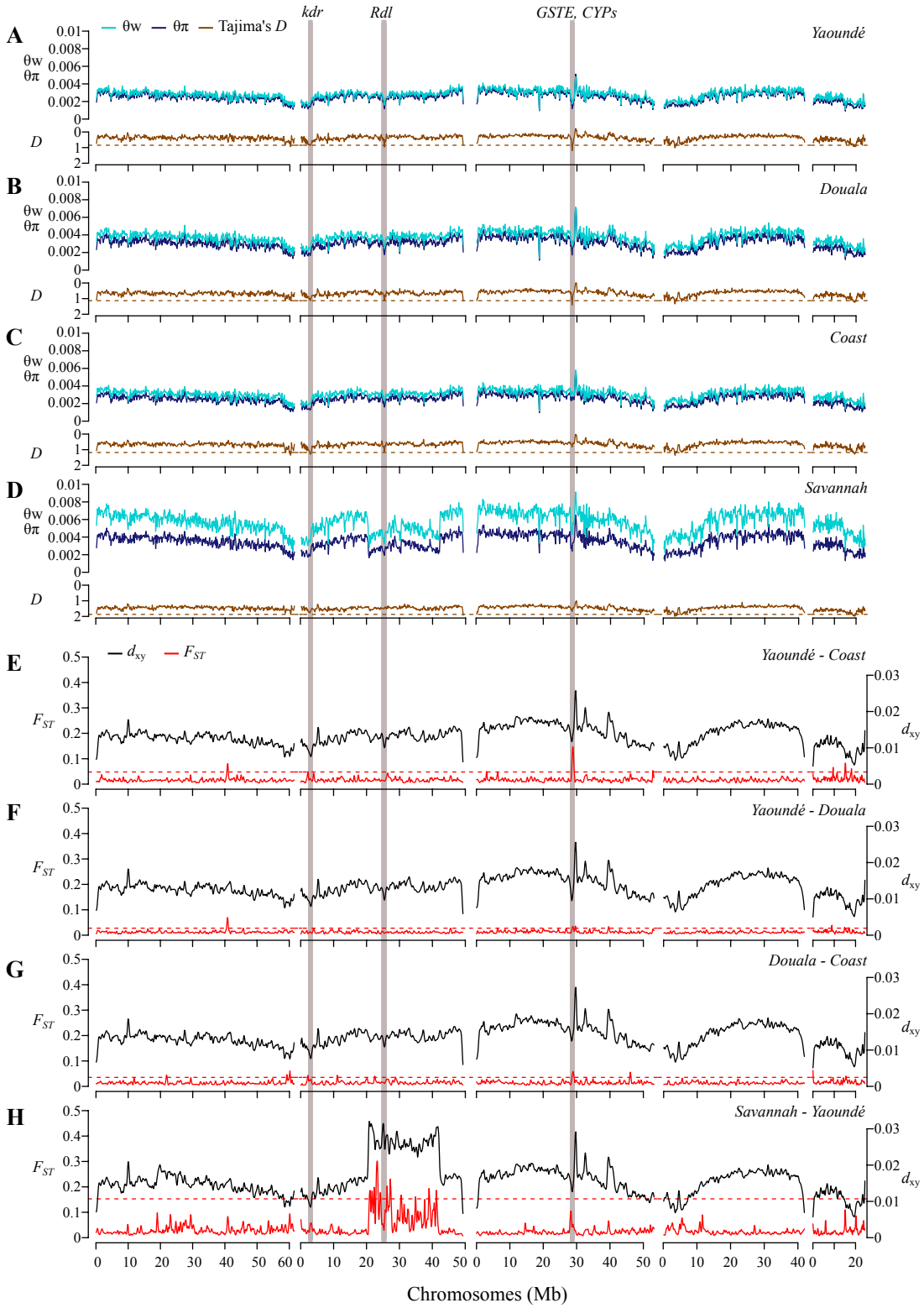

kdr
Rdl
GSTE, CYPs
θw
θπ
Tajima's D
A
Yaoundé
B
Douala
C
Coast
D
Savannah
E
Yaoundé - Coast
F
Yaoundé - Douala
G
Douala - Coast
H
Savannah - Yaoundé
d_{xy}
F_{ST}
Chromosomes (Mb)

Figure 7.

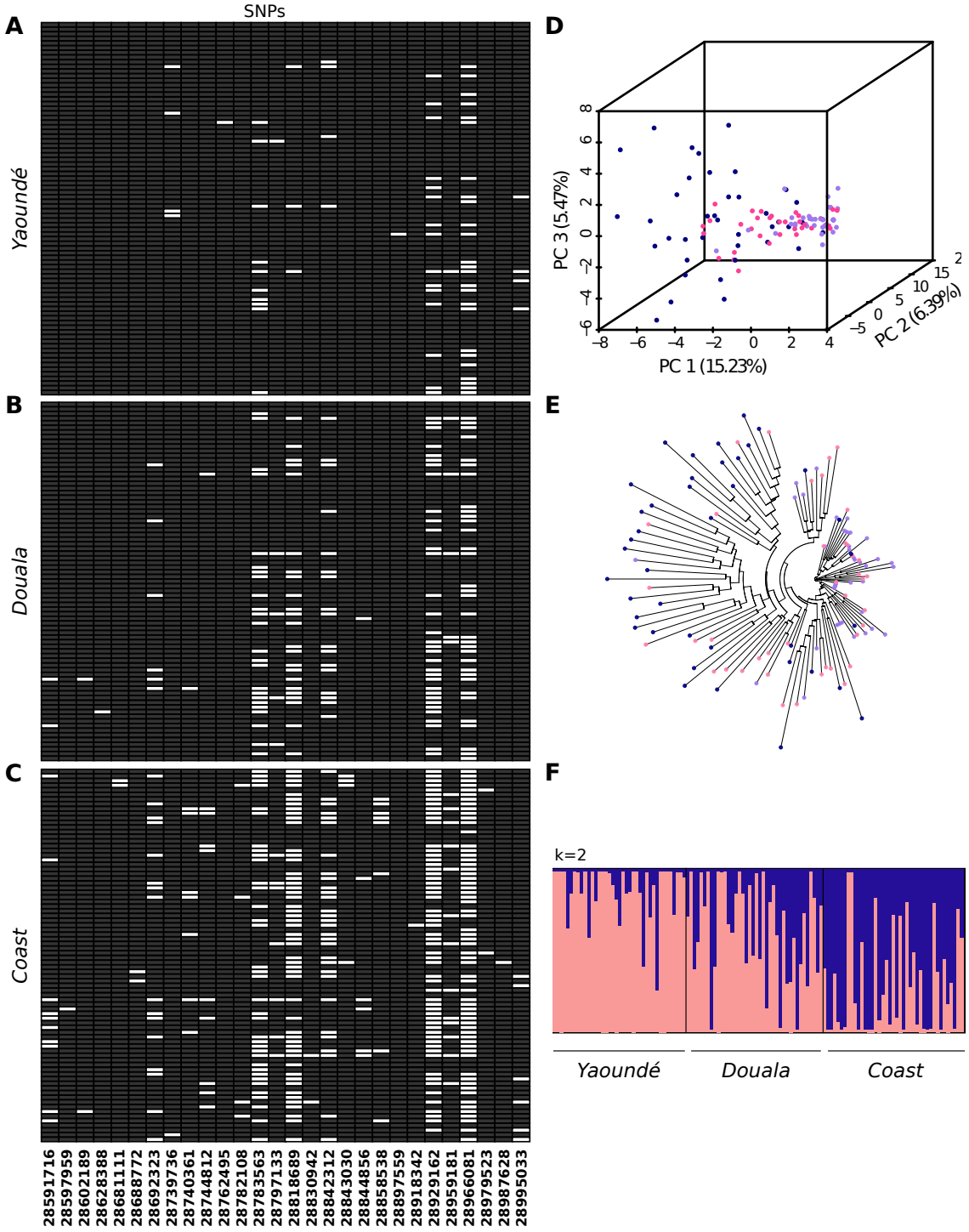

SNPs
A
Yaoundé
B
Douala
C
Coast
28591716
28597959
28602189
28628388
28681111
28688772
28692323
28739736
28740361
28744812
28762495
28782108
28783563
28797133
28818689
28830942
28842312
28843030
28844856
28858538
28897559
28918342
28929162
28959181
28966081
28979523
28987628
28995033
D
PC 3 (5.47%)
PC 1 (15.23%)
PC 2 (6.39%)
E
F
k=2
Yaoundé
Douala
Coast

Figure S1.

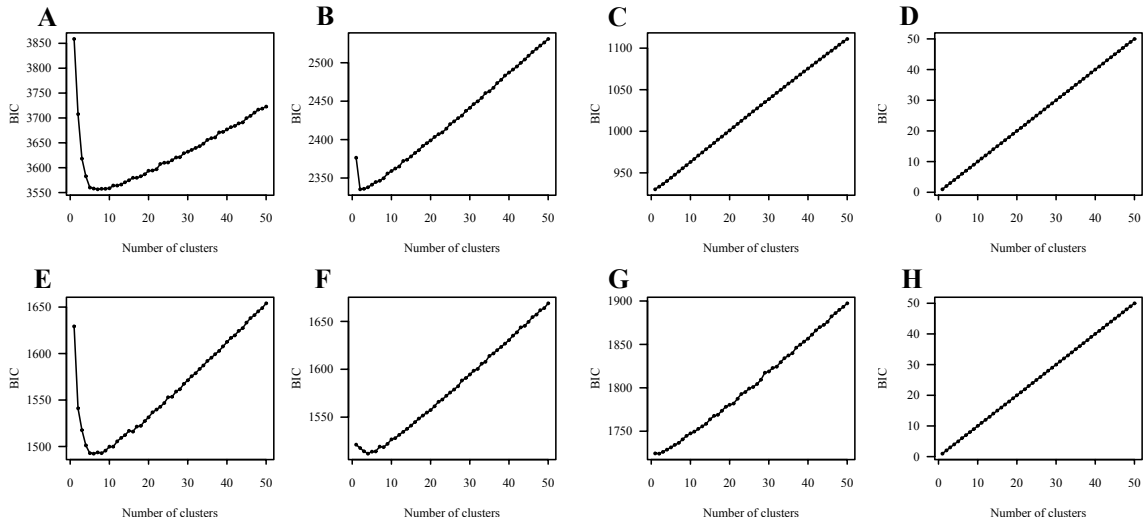


Figure S2.

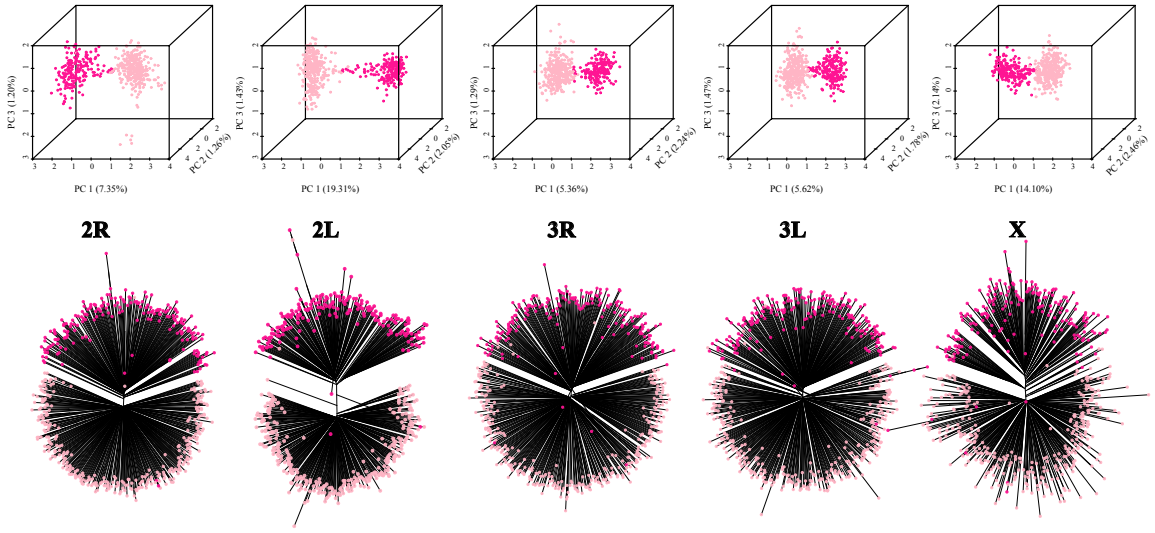

PC 1 (7.35%)
PC 2 (1.26%)
PC 3 (1.20%)
PC 1 (19.31%)
PC 2 (2.05%)
PC 3 (1.43%)
PC 1 (5.36%)
PC 2 (2.24%)
PC 3 (1.29%)
PC 1 (5.62%)
PC 2 (1.78%)
PC 3 (1.47%)
PC 1 (14.10%)
PC 2 (2.46%)
PC 3 (2.14%)
2R
2L
3R
3L
X

Figure S3.

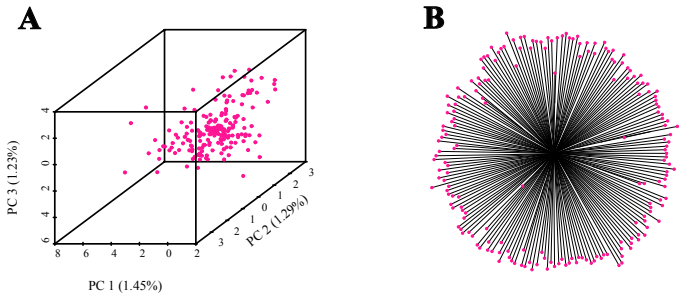
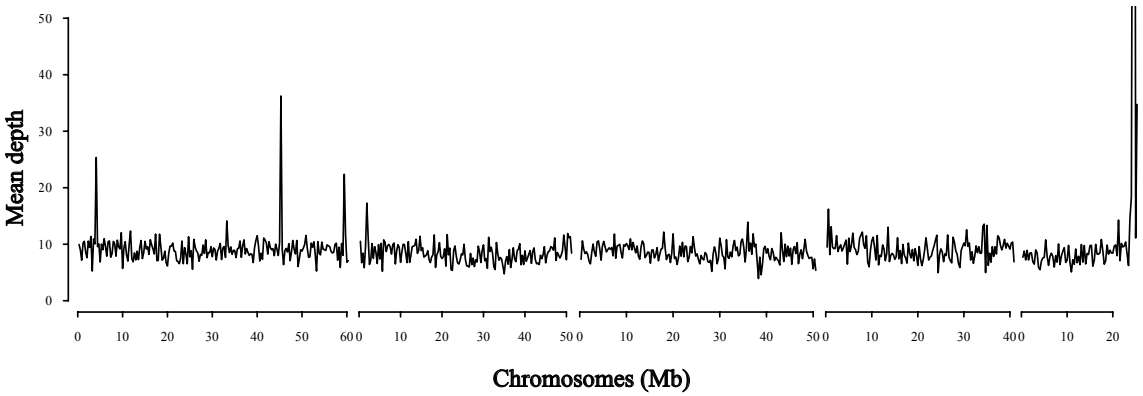


Figure S4.


Mean depth
50
40
30
20
10
0
0 10 20 30 40 50 60 0 10 20 30 40 50 0 10 20 30 40 50 0 10 20 30 40 0 10 20
Chromosomes (Mb)

SUPPLEMENTAL INFORMATION

Consistently negative Tajima's *D* across all subgroups may reflect recent population expansions. To further address this hypothesis we modeled the demographic history of each population using a diffusion-based approach implemented in the software package ∂a∂i v 1.6.3 (1). We fit four alternative demographic models (*neutral, growth, two-epoch, bottle-growth*), without migration or recombination, to the folded allele frequency spectrum of each cryptic subgroup of *An. gambiae s.s.* and *An. coluzzii*. The best model was selected based on the highest composite log likelihood, the lowest Akaike Information Criterion (AIC), and visual inspection of residuals. As the choice of model can be challenging in recently diverged populations, we prioritized the simplest model when we found it difficult to discriminate between conflicting models. To obtain uncertainty estimates for the demographic parameters we used the built-in bootstrap function implemented in ∂a∂i to derive 95% bootstrap confidence intervals.

Results indicate that GAM1, GAM2, and Savannah populations have experienced recent size increases. However, for the southern populations of *Yaoundé, Coast, Douala, and Nkolondom* the best demographic model is a *bottle-growth* (Table S4). While most classical studies report *An. gambiae s.l.* populations that are in expansion (e.g. 2), a more recent study employing RAD markers revealed that some East African populations have more complex demographic histories, often involving several changes in effective population size (N_e) as we observed in southern forest populations of both *An. coluzzii* and

An. gambiae. It has also been shown that *Anopheles* mosquitoes can experience drastic declines in N_e due to insecticidal campaigns (3). Such events affect demographic parameters and could be a plausible explanation for the difficulty we encountered in distinguishing between *bottle-growth* and *two-epoch* models in some populations.

31 SUPPLEMENTAL REFERENCES

32 1. Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009)
33 Inferring the joint demographic history of multiple populations from
34 multidimensional SNP frequency data. *PLoS Genet* 5(10):e1000695.

35 2. Donnelly MJ, Licht MC, Lehmann T (2001) Evidence for recent population
36 expansion in the evolutionary history of the malaria vectors Anopheles
37 arabiensis and Anopheles gambiae. *Mol Biol Evol* 18(7):1353–1364.

38 3. Athrey G, et al. (2012) The effective population size of malaria mosquitoes:
39 large impact of vector control. *PLoS Genet* 8(12):e1003097.

Table S1. Locations of *An. gambiae s.l.* mosquitoes sequenced in Cameroon

Ecogeographic regions	Village Name	Geographic coordinates	Sampling methods				Total
			HLC-OUT	HLC-IN	LC	SPRAY	
Savannah	Lagdo	9°02'56"N, 13°39'22"E	52	27	21	43	143
	Lougga Tabadi	7°06'36"N, 13°12'36"E	2	0	0	0	2
	Malang Dang	7°26'06"N, 13°33'14"E	0	0	0	11	11
	Ngao Bella	6°30'04"N, 12°25'34"E	28	17	0	69	114
	Paniéré Tibati	6°28'08"N, 12°37'44"E	5	0	22	16	43
	Total		87	44	43	139	313
F-S Transition	Makoupa Bord	6°06'06"N, 11°11'29"E	2	0	0	16	18
	Makoupa Le Grand	6°02'28"N, 11°10'39"E	0	0	4	10	14
	Manchoutvi	5°52'48"N, 11°06'36"E	0	0	70	0	70
	Manda	5°43'32"N, 10°52'06"E	1	2	0	33	36
	Mante Le Grand	6°03'44"N, 11°12'18"E	0	0	0	8	8
	Mfelap	5°43'32"N, 10°52'06"E	0	0	0	8	8
	Mgbandji	6°05'49"N, 11°08'26"E	0	0	3	5	8
	Mouinkoing	6°02'56"N, 11°24'36"E	0	0	0	3	3
	Total		3	2	77	83	165
Forest (urban)	Bepanda Omnisport (Douala)	4°03'18"N, 9°43'16"E	0	0	7	0	7
	Beti Maképé (Douala)	4°03'54"N, 9°45'40"E	0	0	25	0	25
	Bomono Gare (Douala)	4°04'55"N, 9°35'35"E	0	0	13	0	13
	PK10 (Douala)	4°02'49"N, 9°46'47"E	0	0	20	0	20
	Missolé 2 (Douala)	3°59'17"N, 9°54'22"E	0	0	8	0	8
	Ndobo Bonabéri (Douala)	4°04'39"N, 9°40'12"E	0	0	1	0	1
	Sable (Douala)	4°04'52"N, 9°43'34"E	28	0	5	0	33
	Village Petit Mobil (Douala)	4°00'16"N, 9°45'18"E	1	0	16	0	17
	Nkolbisson (Yaoundé)	3°52'29"N, 11°26'58"E	0	0	9	0	9
	Nkolondom (Yaoundé)	3°58'20"N, 11°30'56"E	0	0	48	0	48
	Tsinga Elobi (Yaoundé)	3°52'49"N, 11°30'23"E	0	0	46	0	46
	Combattant (Yaoundé)	3°52'36"N, 11°30'46"E	18	0	48	0	66
	Total		47	0	246	0	293
Forest (rural)	Mbébé	4°10'00"N, 11°04'00"E	0	0	0	5	5
	Nyabessan Centre	2°24'00"N, 10°24'00"E	4	1	0	0	5
	Oveng	2°44'00"N, 11°27'00"E	1	0	0	0	1
	Total		5	1	0	5	11

Coast	Afan Essokyé	2°22'01"N, 9°58'59"E	1	0	0	13	14
	Bouanjo	2°48'00"N, 9°54'00"E	0	0	5	0	5
	Campo	2°22'01"N, 9°49'01"E	48		16	27	91
	Ebodjé	2°30'00"N, 9°49'05"E	35	4	0	3	42
	Mutengéné	4°06'53"N, 9°14'51"E	0	0	7	0	7
	Total		84	4	28	43	159
Total			226	51	394	270	941

HLC-OUT, human landing catches performed outdoor; HLC-IN, human landing catches performed indoor; LC, larval collection; SPRAY, spray catches.

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	θπ (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).