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 author: e-mail: colincek@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

Natural selection can drive local adaptation, increasing mean individual fitness and promoting biological diversification (Nosil, et al. 2005; Hereford 2009). Contemporary anthropogenic alteration of the landscape may be increasing pressure for local adaptation in diverse taxa (Gaston 2010). 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 (Pelz, et al. 2005; Song, et al. 2011; Davies, et al. 2012). Successful local adaptation requires that selection overcome the homogenizing effect of gene flow from nearby populations (Kawecki and Ebert 2004). Theoretical simulations suggest that such divergence with gene flow can occur under a range of conditions (Berdahl, et al. 2015), although the most likely genomic distribution of the underlying adaptive variants remains unclear (Le Corre and Kremer 2012; Tiffin and Ross-Ibarra 2014). 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 (Feder, et al. 2013).

The Afrotropical *Anopheles gambiae* complex is a group of at least nine isomorphic mosquito species exhibiting varying degrees of geographic and reproductive isolation (White, et al. 2011; Coetzee, et al. 2013; Crawford, et al. 2014). 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 (Holt, et al. 2002; Fontaine, et al.

2014). 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 suggest that contemporary local adaptation and diversification is occurring within *Anopheles gambiae s.s.* (hereafter *An. gambiae*) and *Anopheles coluzzii*, two of the most widespread and important vectors of malaria within the complex (Slotman, et al. 2007; Wang-Sattler, et al. 2007; Lee, et al. 2013b; Caputo, et al. 2014). Up to now, shallow sampling and/or or the use of low-resolution genetic markers limited the ability to delineate new cryptic subgroups within either species.

We genotyped 941 mosquitoes collected from diverse environments in Cameroon at >8,000 SNPs and find strong evidence for ongoing diversification within both *An. gambiae* and *An. coluzzii*. In total, the two species harbor seven cryptic subgroups distributed along a continuum of genomic differentiation. While *An. gambiae* exhibits relatively high levels of panmixia, we did identify an ecotype associated with intense suburban agriculture and a second subgroup that appears partially reproductively isolated but exhibits no obvious ecological/geographical distinctions. In contrast, *An. coluzzii* is separated into multiple ecotypes exploiting different regional-scale habitats, including highly urbanized landscapes. In most cryptic subgroups, selective sweeps contain an excess of detoxification enzymes and insecticide resistance genes, suggesting that human activity mediates spatially varying natural selection in both species. The extensive population structure within both species represent an additional challenge to vector control strategies. Moreover, ongoing local adaptation and

cryptic diversification of *Anopheles* species in human-dominated environments may contribute to increase malaria transmission.

RESULTS

Identification of An. gambiae s.l. *sibling species*

We performed extensive sampling of human-associated *Anopheles* across the main ecological zones in the central African country of Cameroon (Table S1) with the objective to collect diverse populations belonging to the four species of the *An. gambiae* complex that are present in the country (Simard et al. 2009). As recently shown (Riehle et al. 2011), certain cryptic subgroups can be overlooked when sampling is focused on the collection of only one type of population. Therefore, to maximize the chances that our samples best represent the genetic diversity within each species and to identify cryptic groups, we used several sampling methods (Service 1993) to collect both larvae and adult populations. In Addition, populations of *An. gambiae* and *An. coluzzii* segregate along urbanization gradients, which seem to be the most important driver of ecological divergence in the forest zone (Kamdem et al. 2012). To validate this hypothesis and to investigate the genomic targets of local adaptation in urban environments, we surveyed several neighborhoods representing the urban and suburban ecotypes in the two biggest cities of the forest area: Douala and Yaoundé (Figure S1).

To investigate the genetic relatedness among individuals and to 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 (Peterson, et al. 2012).

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 (Catchen, et al. 2011; Catchen, et al. 2013). 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 (Scott, et al. 1993; Santolamazza, et al. 2004). In agreement with previous surveys (Wondji, et al. 2005; Simard, et al. 2009), 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 (Lee, et al. 2013a) 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 with fastSTRUCTURE (Raj, et al. 2014) finds that 7 population clusters (*k*) best explain the genetic variance present in our sample set (Figure 1C, Figure S2). 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 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 panmictic (Figure 1C, Figure S2). Riehle et al. 2011 recently discovered a cryptic subgroup by comparing indoor and outdoor fauna from the same village in Burkina Faso. Visual inspections of our PCA, NJ and fastSTRUCTURE clustering results do not indicate any genetic subdivision based on the collection methods or the developmental stage. To explicitly test for the effects of the sampling methods and the geographic origin of samples on the genetic variance among individuals, we applied a hierarchical Analysis of Molecular Variance (AMOVA) (Excoffier et al. 1992). This methods partition the genetic variance among individuals in order to quantify the effects of several variables taken at different hierarchical levels on the genetic diversity. We noted that the large majority of the genetic variation was attributable to differences among individuals (86.3% ($p < 0.001$) in *An. coluzzii* and 90.4% in *An. gambiae s.s.* ($p < 0.001$)). However, the geographic origin of

individuals retains a significant component of the genetic variation in the two species. Respectively 13.2 % ($p < 0.001$) and 9.4% ($p < 0.001$) of the genetic variation were partitioned across collection regions nested in sample type in *An. gambiae s.s.* and *An. coluzzii*. The amount of variance due to the types of sample was very low and not significant (less than 1%, $p = 0.29$ for *An. gambiae s.s.* and $p = 0.20$ for *An. coluzzii*) implying that – as suggested by PCA, NJ and fastSTRUCTURE analyses – no genetic structuring based on microhabitats or temporal segregations is apparent in both species in Cameroon.

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 consistently identify three distinct subgroups within *An. gambiae* (Figure 2, Figure S2). The first and largest subgroup (termed *GAM1*) comprises the vast majority of all *An. gambiae* specimens including individuals collected in all four eco-geographic regions (Table S1). A total of 17 individuals make up a second small subgroup (termed *GAM2*). Interestingly, individuals assigned to this cluster include both larvae and adults collected in 3 different villages spread across 2 eco-geographic regions. In the absence of any obvious evidence of niche differentiation between *GAM1* and *GAM2,* it is unclear what is driving and/or maintaining divergence between the two sympatric subgroups. Specimens collected from Nkolondom, a suburban neighborhood of Yaoundé where larval

sites associated with small-scale agricultural irrigation are common (Nwane, et al. 2013; Tene, et al. 2013), form a genetically distinct third subgroup (termed *Nkolondom*) that appears to be a locally-adapted ecotype.

Cryptic Population Structure within An. coluzzii

To examine population structure within 521 *An. coluzzii* specimens, we utilized 9,822 SNPs that passed stringent filtration. All analyses show a clear split between individuals from the northern savannah region and the southern three forested regions of Cameroon (Coastal, Forest, Forest-Savannah) (Figure 3A-C). In principle, the north-south structuring could be caused solely by differences in chromosome 2 inversion frequencies, which form a cline from near absence in the humid south to fixation in the arid north. However, we find SNPs from all five chromosomal arms consistently separate northern and southern mosquitoes, indicating a substantial genome-wide divergence between the two populations (Figure S3).

Southern populations of *An. coluzzii* were collected from three different areas: Douala (the largest city of Cameroon), Yaoundé (the second largest city) and the rural coastal region. PCA, NJ trees, and fastSTRUCTURE show clear clustering of southern samples by collection site (Figure 3D-F). Mosquitoes from Douala, situated on the coastal border, contain a mixture of urban and coastal polymorphisms as illustrated by their intermediate position along PC3 (Figure 3D). Despite considerable geographic segregation, clusters are not fully discrete, likely owing to substantial migration between the three sites. Taken together, the

196 data suggest a dynamic and ongoing process of local adaptation within southern
197 *An. coluzzii*. In contrast, no similar geographic clustering is observed in northern
198 populations (Figure S2, S4). Light variations can occur between genetic
199 clustering patterns suggested by different methods. We have only considered
200 subdivisions that were consistent across the three methods we employed. For
201 example, PCA showed that a few individuals from Yaoundé were relatively
202 detached from the main cluster, but this putative subdivision was not supported
203 by the NJ tree and fastSTRUCTURE analyses (Figure 3D-F). We have therefore
204 treated the Yaoundé subgroup as a single population. All populations described
205 as "urban" were collected from the most urbanized areas of the city, which are
206 characterized by a high density of built environments as described in Kamdem et
207 al. 2012.
208
209 *Relationships Between Species and Subgroups*
210 Population genomic analysis identified four different *An. gambiae* s.l.
211 species present within our samples. Within *An. gambiae* and *An. coluzzii* we
212 identified seven potential subgroups with apparently varying levels of isolation.
213 To further explore the relationships between different populations, we built an
214 unrooted NJ tree using pairwise levels of genomic divergence (F_{ST}) between all
215 species and subgroups (Figure 4, Table S2). As previously observed in a
216 phylogeny based on whole genome sequencing (Fontaine, et al. 2014), we find
217 that *An. melas* is highly divergent from all other species (F_{ST} ~ 0.8), while *An.*
218 *arabiensis* shows intermediate levels of divergence (F_{ST} ~ 0.4) from *An. gambiae*

and *An. coluzzii*. As expected, the sister species *An. gambiae* and *An. coluzzii* are more closely related to each other (F_{ST} ~ 0.2) than any other species. When examining differentiation between subgroups within *An. coluzzii*, we find that the southern and northern subgroups are highly divergent (F_{ST} > 0.1), while differentiation between local ecotypes within the south is much lower (F_{ST} < 0.04). The *An. gambiae* subgroups *GAM1* and *GAM2* are highly diverged (F_{ST} ~ 0.1) from each other suggesting genuine barriers to gene flow despite sympatry, while the suburban ecotype from Nkolondom shows a low level of divergence from *GAM1* (F_{ST} ~ 0.05), characteristic of ongoing local adaptation. In sum, we find a gradient of differentiation between species and subgroups ranging from complete (or nearly complete) reproductive isolation down to the initial stages of divergence with gene flow. To further examine the degree of isolation of subgroups within species, we assessed the reductions in observed heterozygosity with respect to that expected under Hardy–Weinberg Equilibrium among *An. coluzzii* and *An. gambiae s.s.* populations, by computing the average Wright's inbreeding coefficient, F_{IS}, across genome-wide SNPs. Values of F_{IS} close to 1 indicate a deviation from Hardy–Weinberg Equilibrium and the existence of cryptic subdivisions while F_{IS} close to 0 suggest that there are no barriers to gene flow. In spite of the strong population genetic structure observed within *An. coluzzii* and *An. gambiae*, we found surprisingly low genome-wide F_{IS} values (less than 0.0003, $p < 0.005$) in both species, suggesting a lack of assortative mating. Overall, in *An. gambiae* and *An. coluzzii*, ongoing local

241 adaptation and genetic differentiation are parallel to high levels of admixture and
242 extensive shared polymorphisms among individuals.
243
244 *Using genome scans to identify selective sweeps*
245 To find potential targets of selection within subgroups we performed scans
246 of nucleotide diversity (θ_w, θ_π) and allele frequency spectrum (*Tajima's D*) using
247 non-overlapping 150-kb windows across the genome. Scans of θ_w, θ_π, *and*
248 *Tajima's* D were conducted by importing aligned, but otherwise unfiltered, reads
249 directly into ANGSD, which uses genotype likelihoods to calculate summary
250 statistics (Korneliussen, et al. 2014). Natural selection can increase the
251 frequency of an adaptive variant within a population, leading to localized
252 reductions in genetic diversity as the haplotype containing the adaptive variant(s)
253 sweeps towards fixation (Maynard Smith and Haigh 1974; Tajima 1989).
254 Selective processes can also promote the coexistence of multiple alleles in the
255 gene pool of a population (balancing selection). Thus, genomic regions harboring
256 targets of recent selection should exhibit extreme values of diversity and allele
257 frequency spectra relative to genome-wide averages (Storz 2005).
258 We also performed genome scans using both a relative (F_{ST}) and absolute
259 (d_{xy}) measure of divergence calculated with STACKS and *ngsTools*, respectively.
260 If positive selection is acting on alternative haplotypes of the same locus in two
261 populations, values of F_{ST} and d_{xy} should increase at the target of selection.
262 Whereas spatially varying selection that acts on one population, but not the
263 other, should produce a spike in F_{ST} between populations and no change in d_{xy}.

264 Finally, parallel selection on the same haplotype in two populations should lead
265 to a decrease in both metrics (Cruickshank and Hahn 2014). For both diversity
266 and divergence scans we used a maximum of 40 mosquitoes per population,
267 prioritizing individuals with the highest coverage in populations where sample
268 size exceeded 40. In contrast to Tajima's *D* and F_{ST}, the genome-wide
269 distribution of d_{xy} and nucleotide diversity in 150-kb sliding windows yielded
270 relatively noisy patterns (Figure 5 and 6). As a result, we based the identification
271 of signatures of selection primarily on outliers of Tajima's *D* and F_{ST}. Estimates of
272 d_{xy} and nucleotide diversity were used only to confirm genomic locations that
273 were pinpointed as candidate selective sweep on the basis of values of Tajima's
274 *D* and F_{ST}. Precisely, genomic regions were considered as targets of selection if
275 they mapped to significant peaks or depressions of diversity and d_{xy}, and their
276 values of F_{ST} and Tajima's *D* were among the top 1% of the empirical distribution
277 in at least one population. Significantly negative values of Tajima's *D* relative to
278 the genome-wide average suggest an increase in low-frequency mutations due
279 to negative or positive selection whereas significantly positive values of Tajima's
280 *D* indicate a balancing selection.
281
282 *Targets of Selection within An. gambiae subgroups*
283 Our estimates of genome-wide diversity levels (Table S3) within *An.*
284 *gambiae* subgroups are comparable to previous estimates based on RAD
285 sequencing of East African *An. gambiae* s.l. populations (O'Loughlin, et al. 2014).
286 As expected, the large *GAM1* population harbors more genetic diversity than the

apparently rare *GAM2* population or the geographically restricted Nkolondom ecotype (Table S3, Figure 5A-C). *Tajima's D* is consistently negative across the entire genome of all three subgroups, indicating an excess of low-frequency variants that are likely the result of recent population expansion (Figure 5A-C) (Tajima 1989). Indeed, demographic models infer relatively recent bouts of population expansion in all three subgroups (Table S4).

Genome scans of each subgroup reveal genomic regions that show concordant dips in diversity and allele frequency spectrum consistent with recent positive selection (highlighted in Figure 5 A-C). Based on the 1% cutoff of Tajima's *D* (upper and lower bounds) and F_{ST}, we identified 4 candidate regions exhibiting strong signatures of selection. It should be noted that due to the reduced representation sequencing approach we used, our analysis is necessarily conservative, highlighting only clear instances of selection, which are likely both recent and strong (Tiffin and Ross-Ibarra 2014). An apparent selective event on the left arm of chromosome 2 near the centromere is found in all populations. This selective sweep is characterized by a prominent depression of Tajima's *D* (~1.5Mb in width) and contains ~80 genes including the pyrethroid knockdown resistance gene (*kdr*). Although it is difficult to precisely identify the specific gene that has been influenced by selection in this region, the voltage-gated sodium channel gene (*kdr*) involved in insecticide resistance in insects, which is a pervasive target of selection in *Anopheles* mosquitoes (Clarkson et al. 2014; Norris et al. 2015) represents the strongest candidate in our populations. Interestingly, the drop in Tajima's *D* in the putative *kdr* sweep is sharpest in the

Nkolondom population, which suggests that insecticide resistance is shaping the genomic differentiation and local adaptation in *An. gambiae*. Indeed, Nkolondom is a suburban neighborhood of Yaoundé where larvae can be readily collected from irrigated garden plots that likely contain elevated levels of pesticides directed at agricultural pests (Nwane, et al. 2013; Tene, et al. 2013). The *kdr* allele of *para* confers knockdown resistance to pyrethroids and selective sweeps in the same genomic location have been previously identified in many *An. gambiae* s.l. populations (Donnelly, et al. 2009; Lynd, et al. 2010; Jones, Liyanapathirana, et al. 2012; Clarkson, et al. 2014; Norris, et al. 2015). At the *kdr* sweep, we observe contrasting patterns in local values of d_{xy} and F_{ST} (Figure 5D-E). In both the *GAM1-GAM2* and *GAM1-Nkolondom* comparisons, d_{xy} dips while local values of F_{ST} actually increase. While not definitive, the significant drop in d_{xy} suggests that the same resistant haplotype is sweeping through each population. Localized increases in F_{ST} could owe to differences in *kdr* allele frequencies between populations; despite parallel selection, the sweep may be closer to fixation in certain populations relative to others, perhaps due to differences in selection intensity.

Another region exhibiting consistent signatures of selection in all populations is found around the centromere on the X chromosome. Functional analyses of gene ontology (GO) terms (Table S5) revealed a significant representation of chitin binding proteins in this region (p = 1.91e-4). A strong genetic divergence among the three subgroups of *An. gambiae* characterized by significant F_{ST} and d_{xy} peaks is also observed at ~30 Mb and ~40 Mb on

chromosome 3R. Positive outliers of both Tajima's *D* and nucleotide diversity suggest that this genetic divergence is due to balancing selection on multiple alleles among populations. Functional analyses of GO terms indicate that the region at ~40 Mb on 3R is enriched in cell membrane proteins, genes involved in olfaction and epidermal growth factors (EGF) genes (Table S5). Finally, a striking depression in Tajima's *D* supported by a marked dip in nucleotide diversity occurs on chromosome 2L from ~33-35 Mb in the Nkolondom population exclusively (Figure 5A-C). Despite the lack of genetic differentiation, this region – enriched in six EGFs (Table S5) – is probably a recent selective sweep, which could facilitate larval development of this subgroup in pesticide-laced agricultural water.

Targets of Selection within An. coluzzii subpopulations

As above, we used diversity, allele frequency spectra and genetic differentiation metrics to scan for targets of selection in the four subgroups of *An. coluzzii*. Overall, genetic diversity is higher in the northern savannah population than either of three southern populations, which all exhibit similar levels of diversity (Table S3). Just as in *An. gambiae*, all subgroups have consistently negative Tajima's *D* values confirming demographic models of population expansion (Table S4). Based on the 1% threshold of Tajima's *D* and F_{ST}, we found 5 putative selective sweeps in *An. coluzzii* populations including the *kdr* region, the sweep on the X chromosome, and the two hot spots of balancing selection detected on the chromosome 3R in *An. gambiae* (Figure 6A-D). The

fifth putative selective sweep characterized by a sharp drop in both diversity and Tajima's *D* occurs on 3R from ~28.5-29.0 Mb, with the decline being more significant in urban populations. Geographical limitation of the sweep to urban mosquitoes strongly suggests it may contain variant(s) that confer adaptation to extreme levels of anthropogenic disturbance. Indeed, this genomic region harbors a cluster of both Glutathione S-transferase (*GSTE1*-*GSTE7*) and cytochrome P450 (*CYP4C27*, *CYP4C35*, *CYP4C36*) genes, and functional analyses of GO terms reveals an overrepresentation of terms containing "Glutathione S-transferase" (Table S5, p = 5.14e-10). Both the *GSTE* and cytochrome P450 gene families are known to confer metabolic resistance to insecticides and pollutants in mosquitoes (Enayati, et al. 2005; David, et al. 2013; Nkya, et al. 2013). In particular, *GSTE5* and *GSTE6* are intriguing candidate targets of selection as each is up-regulated in highly insecticide resistant *An. arabiensis* populations that recently colonized urban areas of Bobo-Dioulasso, Burkina Faso (Jones et al. 2012).

As in *An. gambiae*, we also detected multiple regions that could be targets of selection, but were less well supported because only one of the metrics (F_{ST} or Tajima's *D*) was above the 99th percentile of the empirical distribution. We found ~30 regions where significant F_{ST} peaks were not correlated to exceptional values of Tajima's *D* and vice versa. This included at least 10 hotspots of F_{ST} clustered within the 2L*a* inversion, which segregates between forest and savannah populations (Figure 6H). Most notably, at ~25 Mb on 2L, a region centered on the resistance to dieldrin (*rdl*) locus, large dips in Tajima's *D* are

379 evident in all southern groups. In the northern savannah population, a
380 pronounced dip in diversity occurs at this putative sweep, but Tajima's *D* stays
381 constant. This region contains ~40 genes, but just as with the *kdr* gene, the *rdl*
382 locus is arguably the prime candidate target of selection. This gene plays a key
383 role in insensitivity to insecticides (Ffrench-Constant et al. 2004), and studies
384 have confirmed the presence of footprints of selection around this locus in *An.*
385 *gambiae* s.l. populations (Lawniczak et al. 2010; Crawford et al. 2015).

386 The increased use of pesticides/insecticides in agriculture and vector
387 control imposes an unprecedented adaptive challenge to mosquito populations
388 (Bøgh et al. 1998; Moiroux et al. 2012; Mwangangi et al. 2013; Clarkson et al.
389 2014; Norris et al. 2015). As a result, both selection and adaptive introgression
390 are acting at the scale of a few decades around loci that provide selective
391 advantage against pesticides (Clarkson et al. 2014; Norris et al. 2015). Our
392 findings indicate that the genetic response to this challenge is also spreading
393 across multiple loci leaving sharp signatures of selection around clusters of
394 detoxification enzymes and major insecticide resistance genes in *An. gambiae*
395 and *An. coluzzii* populations. As expected, *An. coluzzii* populations that are
396 exposed to particularly high levels of insecticides/pollutants in human-dominated
397 environments (Kamdem et al. 2012; Fossog Tene et al. 2013; Tene Fossog et al.
398 2013) are more enriched in genomic regions bearing signatures of human-driven
399 selection. Both relative and absolute divergences between populations at the
400 three selective sweeps involved in xenobiotic resistance reflect the spatial
401 variation of selection along gradients of anthropogenic disturbance. In particular,

the *kdr* locus exhibits minimal divergence in all pairwise comparisons, suggesting that the same resistance haplotype is under selection in each population (Figure 6E-H). In contrast, the region surrounding the *rdl* gene shows low F_{ST} and a pronounced dip in d_{xy} between all southern populations, confirming that the same haplotype is sweeping through these three populations. However, differentiation between southern and northern populations at *rdl* may be obscured by the high divergence between alternative arrangements of the 2L*a* inversion. Finally, the urban-centric GSTE/CYP450 sweep on 3R shows a peak in F_{ST} between Yaoundé and Coastal mosquitoes and minimal change in d_{xy} – a pattern consistent with local adaptation. Comparisons between Douala and Coastal populations show a more moderate increase in F_{ST}, presumably due to high rates of mosquito migration between these nearby sites. The slight bump in F_{ST} is coupled to a large dip in d_{xy} indicative of an ongoing, shared selective sweep between the two cities.

To further explore the 3R GSTE/CYP450 sweep, we reconstructed haplotypes for all 240 *An. coluzzii* southern chromosomes across the 28 SNPs found within the sweep. In the Yaoundé population, a single haplotype is present on 44 out of 80 (55%) chromosomes (all grey SNPs), while an additional 11 haplotypes are within one mutational step of this common haplotype (Figure 7A). In Douala, the same haplotype is the most common, but present at a lower frequency (31%) than in Yaoundé (Figure 7B). Strikingly, this haplotype is found on only 6/80 (7.5%) coastal chromosomes (Figure 7C). The overall low nucleotide variation and high frequency of a single haplotype in Yaoundé is

consistent with positive selection acting on a de novo variant(s) to generate the 3R GSTE/CYP450 sweep. Less intense selection pressure in Douala, and particularly the Coast, would explain the markedly higher haplotype diversity in these two populations relative to Yaoundé. It is also possible that Douala mosquitoes experience similar selection pressures to Yaoundé mosquitoes, but frequent migrant haplotypes from the nearby rural Coast populations decrease the efficiency of local adaptation. Importantly, multiple population genomic analyses of the same 28 SNPs (Figure 7D-F) mirror results of the haplotype analysis, confirming that haplotype inference did not bias the results. In sum, we hypothesize that divergence in xenobiotic levels between urban and rural larval habitats is the main ecological force driving spatially variable selection at this locus.

DISCUSSION

Population genetic structure and cryptic subdivisions within the An. gambiae *complex*

The *Anopheles gambiae* complex, as a model of adaptive radiation with a puzzling evolutionary history, has been recognized as a unique portal into the genetic architecture of ecological speciation (Coluzzi et al. 2002; Ayala and Coluzzi 2005). This system has however been refractory to traditional genetic mapping methods, due mainly to the lack of observable phenotypes that segregate between populations. Recently, patterns of genomic divergence have started to be dissected thanks to the application of high-throughput sequencing and genotyping methods, and significant insights have been gained into the genomic targets of selection among populations at continental scale (Lawniczak et al. 2010; Neafsey et al. 2010; White et al. 2011). Here we have applied a population genomic approach to investigate the genomic architecture of selection at the scale of one country. We showed that reduced representation sequencing of 941 *An. gambiae* s.l. collected in or near human settlements in 33 sites scattered across Cameroon facilitated rapid identification of known sibling species and revealed multiple instances of novel cryptic diversification within *An. gambiae* and *An. coluzzii*. This result is opposite to that found in East Africa, where RADseq markers revealed no population genetic structure within species of the *An. gambiae* complex (O'Loughlin et al. 2014). Historically, West African populations have a greater tendency to be differentiated (Coluzzi et al. 1985; Coluzzi et al. 2002), and the presence of cryptic subdivisions within *An. coluzzii*

461 has already been suspected in Cameroon (Wondji et al. 2005; Slotman et al.
462 2007). As revealed by genome-wide SNPs, these subdivisions result in at least
463 three genetically distinct clusters: a strongly differentiated population confined to
464 the arid savannah area, a coastal subgroup presumably adapted to tolerate high
465 concentrations of salt (Tene Fossog et al. 2015) and a cluster encompassing
466 urban populations that are known to thrive in breeding sites containing high
467 levels of organic waste and present a more complex insecticide resistance profile
468 (Antonio-Nkondjio et al. 2011; Fossog Tene et al. 2013; Tene Fossog et al. 2013;
469 Antonio-Nkondjio et al. 2015). The modelling of ecological niches of *An. gambiae*
470 s.l. in Cameroon predicts that favourable habitats of *An. coluzzii* populations are
471 fragmented and much more marginal landscapes in contrast to *An. gambiae*,
472 which occupies a broader environmental niche across the country (Simard et al.
473 2009). In line with this prediction, genome-wide SNPs indicate a weak genetic
474 differentiation and high genetic diversity in *An. gambiae* over large geographic
475 areas. Nevertheless, despite this broad geographic connectivity and extensive
476 gene flow among populations, ongoing local adaption results in the emergence of
477 geographic clusters in suburban areas.

478 It has been hypothesized that the actual number and the diversity of
479 malaria vector species across the African continent are largely underestimated
480 because of the limited power of morphological and genetic markers employed so
481 far (Stevenson et al. 2012). As a result, genome-wide studies and extensive
482 sampling are supposed to lead to the discovery of unknown species. Among the
483 941 *An. gambiae* s.l. we sequenced, we instead found a clear match between the

484 species identified by both morphological observations and PCR of the ribosomal
485 DNA and those suggested by thousands of SNPs scattered throughout the
486 genome. Furthermore, the intensive use of insecticide-treated bed nets has
487 triggered complex behavioural adaptations and changes in species distribution
488 that may ultimately lead to splits and the creation of cryptic populations within
489 species (Bøgh et al. 1998; Derua et al. 2012; Moiroux et al. 2012; Mwangangi et
490 al. 2013; Sokhna et al. 2013). For example, populations that are evolving to bite
491 outdoor or earlier at night to escape bed nets have an increased likelihood to
492 differentiated into distinct gene pools (Bøgh et al. 1998; Riehle et al. 2011;
493 Moiroux et al. 2012). Using a comprehensive sampling of larvae and adults at
494 different time points during diurnal and nocturnal activities, in or near human
495 settlements, we can conclude that no genetic clustering beyond the regional-
496 scale subdivisions we described is apparent over the scale of our study.
497
498 *Genomic signatures of selection*
499 Populations depicting increasing levels of genetic differentiation along a
500 speciation continuum are ideal to investigate the targets of selection at early
501 stages of ecological divergence (Savolainen et al. 2013). We have scanned
502 genomes of more or less divergent populations of *An. gambiae* and *An. coluzzii*
503 using several divergence and diversity metrics in order to identify outlier regions,
504 which likely contain factors involved in ecological divergence and/or reproductive
505 isolation. In principle, in weakly differentiated populations such as the subgroups
506 we described for which neutral and selective processes have yet to shape the

genomic architecture, signatures of selection are often clustered within a few regions of the genome (Nosil and Feder 2012; Andrew and Rieseberg 2013). Indeed, we found that footprints of natural selection in structured populations of *An. gambiae* and *An. coluzzii* occur across a few loci enriched in genes whose functions include insecticide resistance and detoxification, epidermal growth, cuticle formation and olfaction. Although some targets of selection are likely missing because of our experimental approach (Arnold et al. 2013; Tiffin and Ross-Ibarra 2014), previous studies on ecological and phenotypic divergence in *An. gambiae* s.l. suggest that footprints of selection we identified are particularly relevant. First, the cuticle plays a major role at the interface between several biological functions in insects and cuticular proteins are extremely diverse within and among species (Vannini et al. 2014). Importantly, certain cuticular proteins are associated with resistance to insecticides by contributing to a thicker cuticle in *Anopheles* (Wood et al. 2010; Vannini et al. 2014). Olfaction also mediates a wide range of both adult and larval behaviors in blood-feeding mosquitoes (Bowen 1991; Takken and Verhulst 2013). A large family of olfactory receptors has been characterized including candidate chemosensory genes directly involved in the response to cues that are required for feeding, host preference, and mate selection in *An. gambiae* s.l (Carey et al. 2010; Liu et al. 2010; Rinker et al. 2013). Finally, malaria vectors of the *An. gambiae* complex are among the most synanthropic insects in the world, which has led to the hypothesis that human-driven selection is one of the main modulators of ecological divergence between and within species (Coluzzi et al. 2002; Kamdem et al. 2012).

Consistent with this hypothesis, we have found that genomic regions harbouring genes involved in resistance to insecticides and pollutants are the dominant targets of selection in emerging subgroups adapted to urban areas. Although more direct implications of genes within these selective sweeps will ultimately be necessary to validate the role of human mediated selection in local adaptation, our data provide a genomic perspective on the interactions between human actions and the contemporary evolution of a mosquito species.

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 (Chen, et al. 2012). Previous studies have documented significant 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 (Bayoh, et al. 2010; Athrey, et al. 2012). 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 in Cameroon and treated with pyrethroids that target *para* (Bowen 2013). In contrast, selective sweeps centered on other insecticide resistance genes are restricted to specific geographic locations/populations. For example, a sweep at

553 the *rdl* locus is limited to southern populations of *An. coluzzii*. Initial selection for
554 dieldrin resistance likely occurred during massive indoor residual spraying
555 campaigns conducted by the WHO in southern Cameroon during the 1950s.
556 Indeed, the spraying was so intense that it temporarily eliminated *An. gambiae*
557 s.l. from Yaoundé (and likely other locations in the forest region) (Livadas, et al.
558 1958). However, due to high human toxicity, dieldrin has been banned for use in
559 mosquito control since the mid-1970s. In the absence of insecticide exposure,
560 resistant *rdl* mosquitoes are significantly less fit than wild type mosquitoes
561 (Rowland 1991a, b; Platt, et al. 2015), making the continued persistence of
562 resistant alleles in southern *An. coluzzii* populations puzzling. One plausible
563 explanation is that other cyclodienes targeting *rdl*, such as fipronil and lindane,
564 are still commonly used in agriculture and may frequently runoff into *An. coluzzii*
565 larval habitats, imposing strong selection for resistant mosquitoes. A similar
566 phenomenon was recently proposed to explain the maintenance of resistance *rdl*
567 alleles in both *Culex* and *Aedes* mosquitoes (Tantely, et al. 2010).

568 Mosquitoes inhabiting Cameroon's two major cities, Yaoundé and Douala,
569 provide a clearer example of how xenobiotic exposure can directly influence
570 population structure. Both cities have seen exponential human population growth
571 over the past 50 years, creating a high concentration of hosts for anthropohilic
572 mosquitoes. Despite elevated levels of organic pollutants and insecticides in
573 urban relative to rural larval sites, surveys show substantial year-round
574 populations of *An. gambiae* and *An. coluzzii* in both cities (Antonio-Nkondjio, et
575 al. 2011; Kamdem, et al. 2012; Antonio-Nkondjio, et al. 2014). Bioassays of

576 insecticide resistance demonstrate that urban mosquitoes have significantly
577 higher levels of resistance to multiple insecticides compared to rural mosquitoes
578 (Nwane, et al. 2009; Antonio-Nkondjio, et al. 2011; Nwane, et al. 2013; Tene, et
579 al. 2013; Antonio-Nkondjio, et al. 2015). In support of human mediated local
580 adaptation, we find a selective sweep in urban *An. coluzzii* mosquitoes centered
581 on a cluster of GSTE/CYP450 detoxification genes. While the specific ecological
582 driver of the selective sweep is unknown, GSTE and P450 enzymes detoxify both
583 organic pollutants and insecticides (Suwanchaichinda and Brattsten 2001; David,
584 et al. 2010; Poupardin, et al. 2012). Indeed, the synergistic effects of the two
585 types of xenobiotics could be exerting intense selection pressure for pleiotropic
586 resistance in urban mosquitoes (Mueller, et al. 2008; David, et al. 2013; Nkya, et
587 al. 2013). Regardless of the underlying targets of selection, it is clear that
588 mosquitoes inhabiting highly disturbed urban and suburban landscapes are
589 genetically differentiated from rural populations. Further analysis of specific
590 sweeps using a combination of whole genome resequencing and emerging
591 functional genetics approaches (e.g. CRISPR/Cas9) should help resolve the
592 specific targets of local adaptation in urban mosquitoes, while also shedding light
593 on the evolutionary history of the enigmatic subgroup *GAM2*.
594
595 *Impacts on Vector Control*
596 Just five decades ago, there was not a single city in Sub-Saharan African
597 with a population over 1 million; today there are more than 40. Population shifts
598 to urban areas will only continue to increase with the United Nations estimating

599 that 60% of Africans will live in large cities by 2050 (United Nations 2014). When
600 urbanization commenced, it was widely assumed that malaria transmission would
601 be minimal because rural *Anopheles* vectors would not be able to complete
602 development in the polluted larval habitats present in cities (Donnelly, et al.
603 2005). However, increasingly common reports of endemic malaria transmission
604 in urban areas across Sub-Saharan Africa unequivocally demonstrate that
605 anophelines are exploiting the urban niche (Robert, et al. 2003; Keiser, et al.
606 2004; De Silva and Marshall 2012). Specifically, our study shows that *An.*
607 *gambiae* s.l. from the urban and suburban centers of southern Cameroon form
608 genetically distinct subgroups relative to rural populations. Local adaptation to
609 urban environments is accompanied by strong selective sweeps centered on
610 putative xenobiotic resistance genes, which are likely driven by a combination of
611 exposure to organic pollutants and insecticides in larval habitats. The rapid
612 adaptation of *Anopheles* to the urban landscape poses a growing health risk as
613 levels of resistance in these populations negate the effectiveness of almost all
614 commonly used insecticides. Moreover, repeated instances of beneficial alleles
615 introgressing between *An. gambiae* s.l. species make the emergence of highly
616 resistant subgroups even more troubling (Weill, et al. 2000; Clarkson, et al. 2014;
617 Crawford, et al. 2014; Fontaine, et al. 2014; Norris, et al. 2015). In essence,
618 urban populations can serve as a reservoir for resistance alleles, which have the
619 potential to rapidly move between species/populations as needed. Clearly,
620 sustainable malaria vector control, urban or otherwise, requires not only more
621 judicious use of insecticides, but also novel strategies not reliant on chemicals.

622 Towards this goal, various vector control methods that aim to replace or
623 suppress wild mosquito populations using genetic drive are currently under
624 development (e.g. (Windbichler, et al. 2011)). While promising, the complexities
625 of ongoing cryptic diversification within African *Anopheles* must be explicitly
626 planned for prior to the release of transgenic mosquitoes.
627

MATERIALS AND METHODS

Mosquito collections

In 2013, we collected *Anopheles* from 33 locations spread across the four major ecogeographic regions of Cameroon (Table S1). Indoor resting adult mosquitoes were collected by pyrethrum spray catch, while host-seeking adults were obtained via indoor/outdoor human-baited landing catch. Larvae were collected using standard dipping procedures (Service 1993). All researchers were provided with malaria chemoprophylaxis throughout the collection period. Individual mosquitoes belonging to the *An. gambiae* s.l. complex were identified by morphology (Gillies and De Meillon 1968; Gillies and Coetzee 1987).

ddRADseq Library Construction

Genomic DNA was extracted from adults using the ZR-96 Quick-gDNA kit (Zymo Research) and from larvae using the DNeasy Extraction kit (Qiagen). A subset of individuals were assigned to sibling species using PCR-RFLP assays that type fixed SNP differences in the rDNA (Fanello, et al. 2002). Preparation of ddRAD libraries largely followed (Turissini, et al. 2014). Briefly, ~1/3rd of the DNA extracted from an individual mosquito (10µl) was digested with *MluC1* and *NlaIII* (New England Biolabs). Barcoded adapters (1 of 48) were ligated to overhangs and 400 bp fragments were selected using 1.5% gels on a BluePippin (Sage Science). One of six indices was added during PCR amplification. Each library

650 contained 288 individuals and was subjected to single end, 100 bp sequencing
651 across one or two flow cells lanes run on an Illumina HiSeq2500.

652 Raw sequence reads were demultiplexed and quality filtered using the
653 STACKS v 1.29 process_radtags pipeline (Catchen, et al. 2011; Catchen, et al.
654 2013). After removal of reads with ambiguous barcodes, incorrect restriction
655 sites, and low sequencing quality (mean Phred < 33), GSNAP was used to align
656 reads to the *An. gambiae* PEST reference genome (AgamP4.2) allowing up to
657 five mismatches per read. After discarding reads that perfectly aligned to more
658 than one genomic position, we used STACKS to identify unique RAD tags and
659 construct consensus assemblies for each. Individual SNP genotypes were called
660 using default setting in the maximum-likelihood statistical model implemented in
661 the STACKS genotypes pipeline.

662

663 *Population Genomic Analysis*

664 Population genetic structure was assessed using the SNP dataset output
665 by the *populations* program of STACKS. We used PLINK v 1.19 to retrieve
666 subsets of genome-wide SNPs as needed (Purcell, et al. 2007). PCA, neighbor-
667 joining tree analyses, and Bayesian information criterion (BIC) were implemented
668 using the packages *adegenet* and *ape* in R (Paradis et al. 2004; Jombart 2008; R
669 Development Core Team 2014). Ancestry analyses were conducted in
670 fastSTRUCTURE v 1.0 (Raj, et al. 2014) using the logistic method. The
671 choosek.py script was used to find the appropriate number of populations (k); in
672 cases where a range of k was suggested, the BIC-inferred number of clusters

673 was chosen. CLUMPP v1.1.2 (Jakobsson & Rosenberg 2007) was used to
674 summarize assignment results across independent runs and DISTRUCT v1.1
675 (Rosenberg 2004) was used to visualize ancestry assignment of individual
676 mosquitoes. We used a subset of 1,000 randomly chosen SNPs to calculate
677 average pairwise F_{ST} between populations in GENODIVE v 2.0 using up to 40
678 individuals – prioritized by coverage – per population (Meirmans and Van
679 Tienderen 2004). Using this same subset of 1,000 SNPs, we conducted an
680 AMOVA to quantify the effect of the sampling method and the geographic origin
681 on the genetic variance among individuals in GENODIVE. We used 10,000
682 permutations to assess significance of F_{ST} values and AMOVA. We input
683 pairwise F_{ST} values into the program Fitch from the Phylip (Plotree and Plotgram
684 1989) suite to create the population-level NJ tree. F_{IS} values were computed with
685 the *populations* program in STACKS.
686
687 *Genome Scans for Selection*
688 We used ANGSD v 0.612 (Korneliussen, et al. 2014) to calculate
689 nucleotide diversity (θ_w and θ_π) and Tajima's D in 150-kb non-overlapping
690 windows. Unlike most genotyping algorithms, ANGSD does not perform hard
691 SNP calls, instead taking genotyping uncertainty into account when calculating
692 summary statistics. Similarly, absolute divergence (d_{xy}) was calculated using
693 *ngsTools* (Fumagalli, et al. 2014) based on genotype likelihoods generated by
694 ANGSD. Kernel smoothed values for 150-kb windows for all four metrics (θ_w, θ_π,
695 D, d_{xy}) were obtained with the R package *KernSmooth*. F_{ST} (based on AMOVA)

was calculated with the *populations* program in STACKS using only loci present in 80% of individuals. A Kernel smoothing procedure implemented in STACKS was used to obtain F_{ST} values across 150-kb windows. Because regions with unusually high or low read depth can yield unreliable estimates of diversity and divergence parameters due to the likelihood of repeats and local misassembly, we checked that the average per-locus sequencing coverage was consistent throughout the genome (Figure S5). To determine if selective sweeps were enriched for specific functional annotation classes, we used the program DAVID 6.7 with default settings (Huang, et al. 2008). We physically delimitated the selective sweep as the region corresponding to the base of the peak or the depression of Tajima's *D*. Haplotypes across the GSTE/CYP450 sweep were reconstructed by PHASE v 2.1.1 using the default recombination model (Stephens, et al. 2001; Stephens and Scheet 2005).

710 ACKNOWLEDGEMENTS

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

715

716

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.

721 REFERENCES

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

725 Antonio-Nkondjio C, Fossog BT, Kopya E, Poumachu Y, Djantio BM, Ndo C,
726 Tchuinkam T, Awono-Ambene P, Wondji CS. 2015. Rapid evolution of
727 pyrethroid resistance prevalence in Anopheles gambiae populations from the
728 cities of Douala and Yaoundé (Cameroon). Malaria Journal 14:1-9.

729 Antonio-Nkondjio C, Fossog BT, Ndo C, Djantio BM, Togouet SZ, Awono-Ambene P,
730 Costantini C, Wondji CS, Ranson H. 2011. Anopheles gambiae distribution and
731 insecticide resistance in the cities of Douala and Yaoundé (Cameroon):
732 influence of urban agriculture and pollution. Malaria Journal 10:154-154.

733 Antonio-Nkondjio C, Youmsi-Goupeyou M, Kopya E, Tene-Fossog B, Njiokou F,
734 Costantini C, Awono-Ambene P. 2014. Exposure to disinfectants (soap or
735 hydrogen peroxide) increases tolerance to permethrin in Anopheles gambiae
736 populations from the city of Yaoundé, Cameroon. Malaria Journal 13:296.

737 Arnold B, Corbett-Detig RB, Hartl D, Bomblies K. 2013. RADseq underestimates
738 diversity and introduces genealogical biases due to nonrandom haplotype
739 sampling. Mol. Ecol. 22:3179–3190.

740 Athrey G, Hodges TK, Reddy MR, Overgaard HJ, Matias A, Ridl FC, Kleinschmidt I,
741 Caccone A, Slotman MA. 2012. The effective population size of malaria
742 mosquitoes: large impact of vector control. PLoS genetics 8:e1003097.

743 Bayoh MN, Mathias DK, Odiere MR, Mutuku FM, Kamau L, Gimnig JE, Vulule JM,
744 Hawley WA, Hamel MJ, Walker ED. 2010. Anopheles gambiae: historical
745 population decline associated with regional distribution of insecticide-treated
746 bed nets in western Nyanza Province, Kenya. Malaria journal 9:62.

747 Berdahl A, Torney CJ, Schertzer E, Levin SA. 2015. On the evolutionary interplay
748 between dispersal and local adaptation in heterogeneous environments.
749 Evolution. 69-6: 1390–1405

750 Bøgh C, Pedersen EM, Mukoko D a, Ouma JH. 1998. Permethrin-impregnated bednet
751 effects on resting and feeding behaviour of lymphatic filariasis vector
752 mosquitoes in Kenya. Med. Vet. Entomol. 12:52–59.

753 Bowen MF. 1991. The sensory physiology of host-seeking behavior in mosquitoes.
754 Annu. Rev. Entomol. 36:139–158.

755 Bowen HL. 2013. Impact of a mass media campaign on bed net use in Cameroon.
756 Malaria Journal 12:10.1186.

757 Caputo B, Nwakanma D, Caputo F, Jawara M, Oriero E, Hamid ‐ Adiamoh M, Dia I,
758 Konate L, Petrarca V, Pinto J. 2014. Prominent intraspecific genetic divergence
759 within Anopheles gambiae sibling species triggered by habitat discontinuities
760 across a riverine landscape. Molecular Ecology 23:4574-4589.

761 Carey AF, Wang G, Su CY, Zwiebel LJ, Carlson JR. 2010. Odorant reception in the
762 malaria mosquito Anopheles gambiae. Nature 464:66–71.

763 Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: an
764 analysis tool set for population genomics. Molecular Ecology 22:3124-3140.

765 Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks:
766 building and genotyping loci de novo from short-read sequences. G3: Genes,
767 Genomes, Genetics 1:171-182.

768 Chen H, Wang H, Siegfried BD. 2012. Genetic differentiation of western corn
769 rootworm populations (Coleoptera: Chrysomelidae) relative to insecticide
770 resistance. Annals of the Entomological Society of America 105:232-240.

771 Clarkson CS, Weetman D, Essandoh J, Yawson AE, Maslen G, Manske M, Field SG,
772 Webster M, Antão T, MacInnis B. 2014. Adaptive introgression between
773 Anopheles sibling species eliminates a major genomic island but not
774 reproductive isolation. Nature communications 5:4248

775 Coetzee M, Hunt RH, Wilkerson R, Della Torre A, Coulibaly MB, Besansky NJ. 2013.
776 Anopheles coluzzii and Anopheles amharicus, new members of the Anopheles
777 gambiae complex. Zootaxa 3619:246-274.

778 Coluzzi M, Petrarca V, Di Deco MA. 1985. Chromosomal inversion intergradation and
779 incipient speciation in Anopheles gambiae. Bolletino di Zool. 52:45–63.

780 Coluzzi M, Sabatini A, Torre A, Angela M, Deco D, Petrarca V. 2002. A Polytene
781 Chromosome Analysis of the Anopheles gambiae Species Complex. Science
782 298:1415–1419.

783 Crawford JE, Riehle MM, Guelbeogo WM, Gneme A, Sagnon N, Vernick KD, Nielsen R,
784 Lazzaro BP. 2015. Reticulate speciation and barriers to introgression in the
785 anopheles gambiae species complex. Genome Biol. Evol. 7:3116–3131.

786 Cruickshank TE, Hahn MW. 2014. Reanalysis suggests that genomic islands of
787 speciation are due to reduced diversity, not reduced gene flow. Molecular
788 Ecology 23:3133-3157.

789 Dabiré RK, Namountougou M, Sawadogo SP, Yaro LB, Toé HK, Ouari A, Gouagna L-C,
790 Simard F, Chandre F, Baldet T. 2012. Population dynamics of Anopheles
791 gambiae sl in Bobo-Dioulasso city: bionomics, infection rate and susceptibility
792 to insecticides. Parasit Vectors 5:127.

793 David J-P, Coissac E, Melodelima C, Poupardin R, Riaz MA, Chandor-Proust A,
794 Reynaud S. 2010. Transcriptome response to pollutants and insecticides in the
795 dengue vector Aedes aegypti using next-generation sequencing technology.
796 BMC Genomics 11:216.

797 David J-P, Ismail HM, Chandor-Proust A, Paine MJI. 2013. Role of cytochrome P450s
798 in insecticide resistance: impact on the control of mosquito-borne diseases and
799 use of insecticides on Earth. Philosophical Transactions of the Royal Society of
800 London B: Biological Sciences 368:20120429.

801 Davies TGE, Field LM, Williamson MS. 2012. The re-emergence of the bed bug as a
802 nuisance pest: implications of resistance to the pyrethroid insecticides. Medical
803 and Veterinary Entomology 26:241-254.

804 De Silva PM, Marshall JM. 2012. Factors contributing to urban malaria transmission
805 in sub-Saharan Africa: a systematic review. Journal of tropical medicine 2012.

806 Derua Y a, Alifrangis M, Hosea KM, Meyrowitsch DW, Magesa SM, Pedersen EM,
807 Simonsen PE. 2012. Change in composition of the Anopheles gambiae complex

808 and its possible implications for the transmission of malaria and lymphatic
809 filariasis in north-eastern Tanzania. Malaria Journal 11:188.

810 Donnelly MJ, Corbel V, Weetman D, Wilding CS, Williamson MS, Black WC. 2009.
811 Does kdr genotype predict insecticide-resistance phenotype in mosquitoes?
812 Trends in Parasitology 25:213-219.

813 Donnelly MJ, McCall P, Lengeler C, Bates I, D'Alessandro U, Barnish G, Konradsen F,
814 Klinkenberg E, Townson H, Trape J-F. 2005. Malaria and urbanization in sub-
815 Saharan Africa. Malaria Journal 4:12.

816 Enayati AA, Ranson H, Hemingway J. 2005. Insect glutathione transferases and
817 insecticide resistance. Insect Molecular Biology 14:3-8.

818 Excoffier L, Smouse PE, Quattro JM. 1992. Analysis of molecular variance inferred
819 from metric distances among DNA haplotypes: Application to human
820 mitochondrial DNA restriction data. Genetics 131:479–491.

821 Fanello C, Santolamazza F, della Torre A. 2002. Simultaneous identification of
822 species and molecular forms of the *Anopheles gambiae* complex by PCR-RFLP.
823 Med Vet Entomol 16:461-464.

824 Feder JL, Flaxman SM, Egan SP, Comeault AA, Nosil P. 2013. Geographic mode of
825 speciation and genomic divergence. Annual Review of Ecology, Evolution, and
826 Systematics 44:73-97.

827 Ffrench-Constant RH, Daborn PJ, Le Goff G. 2004. The genetics and genomics of
828 insecticide resistance. Trends Genet. 20:163–170.

829 Fontaine MC, Pease JB, Steele A, Waterhouse RM, Neafsey DE, Sharakhov I, Jiang X,
830 Hall AB, Catteruccia F, Kakani E, et al. 2014. Extensive introgression in a
831 malaria vector species complex revealed by phylogenomics. Science. 347:
832 10.1126/science.1258524

833 Fossog Tene B, Poupardin R, Costantini C, Awono-Ambene P, Wondji CS, Ranson H,
834 Antonio-Nkondjio C. 2013. Resistance to DDT in an Urban Setting: Common
835 Mechanisms Implicated in Both M and S Forms of Anopheles gambiae in the
836 City of Yaoundé Cameroon. PLoS One 8:e61408.

837 Fumagalli M, Vieira FG, Linderoth T, Nielsen R. 2014. ngsTools: methods for
838 population genetics analyses from next-generation sequencing data.
839 Bioinformatics 30:1486-1487.

840 Gaston KJ. 2010. Urban ecology: Cambridge University Press, Cambridge, UK

841 Gillies MT, Coetzee MA. 1987. Supplement to the Anophelinae of Africa south of the
842 Sahara (Afrotropical region), South African Institute for Medical Research.

843 Gillies MT, De Meillon B. 1968. The Anophelinae of Africa South of the Sahara.
844 Johannesburg: South African Institute for Medical Research.

845 Gimnig JE, Walker ED, Otieno P, Kosgei J, Olang G, Ombok M, Williamson J,
846 Marwanga D, Abong'o D, Desai M. 2013. Incidence of malaria among mosquito
847 collectors conducting human landing catches in western Kenya. American
848 Journal of Tropical Medicine and Hygiene 88:301-308.

849 Hereford J. 2009. A quantitative survey of local adaptation and fitness trade‐offs.
850 The American Naturalist 173:579-588.

851 Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincker
852 P, Clark AG, Ribeiro JM, Wides R, et al. 2002. The genome sequence of the
853 malaria mosquito *Anopheles gambiae*. Science 298:129-149.

854 Huang DW, Sherman BT, Lempicki RA. 2008. Systematic and integrative analysis of
855 large gene lists using DAVID bioinformatics resources. Nature protocols 4:44-
856 57.

857 Jakobsson M, Rosenberg NA. 2007. CLUMPP: a cluster matching and permutation
858 program for dealing with label switching and multimodality in analysis of
859 population structure. Bioinformatics 23:1801-1806.

860 Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic
861 markers. Bioinformatics 24:1403-1405.

862 Jones CM, Liyanapathirana M, Agossa FR, Weetman D, Ranson H, Donnelly MJ,
863 Wilding CS. 2012. Footprints of positive selection associated with a mutation
864 (N1575Y) in the voltage-gated sodium channel of Anopheles gambiae.
865 Proceedings of the National Academy of Sciences 109:6614-6619.

866 Jones CM, Toé HK, Sanou A, Namountougou M, Hughes A, Diabaté A, Dabiré R,
867 Simard F, Ranson H. 2012. Additional selection for insecticide resistance in
868 urban malaria vectors: DDT resistance in Anopheles arabiensis from Bobo-
869 Dioulasso, Burkina Faso. PLoS One 7:e45995.

870 Kamdem C, Fossog BT, Simard F, Etouna J, Ndo C, Kengne P, Boussès P, Etoa F-X,
871 Awono-Ambene P, Fontenille D, et al. 2012. Anthropogenic habitat disturbance

872 and ecological divergence between incipient species of the malaria mosquito
873 Anopheles gambiae. PLoS One 7:e39453.

874 Kawecki TJ, Ebert D. 2004. Conceptual issues in local adaptation. Ecology Letters
875 7:1225-1241.

876 Keiser J, Utzinger J, De Castro MC, Smith TA, Tanner M, Singer BH. 2004.
877 Urbanization in sub-saharan Africa and implication for malaria control. The
878 American Journal of Tropical Medicine and Hygiene 71:118-127.

879 Korneliussen TS, Albrechtsen A, Nielsen R. 2014. ANGSD: analysis of next generation
880 sequencing data. BMC bioinformatics 15:356.

881 Lawniczak MKN, Emrich SJ, Holloway a K, Regier a P, Olson M, White B, Redmond S,
882 Fulton L, Appelbaum E, Godfrey J, et al. 2010. Widespread divergence between
883 incipient Anopheles gambiae species revealed by whole genome sequences.
884 Science 330:512–514.

885 Le Corre V, Kremer A. 2012. The genetic differentiation at quantitative trait loci
886 under local adaptation. Molecular Ecology 21:1548-1566.

887 Lee Y, Marsden CD, Norris LC, Collier TC, Main BJ, Fofana A, Cornel AJ, Lanzaro GC.
888 2013a. Spatiotemporal dynamics of gene flow and hybrid fitness between the M
889 and S forms of the malaria mosquito, Anopheles gambiae. Proceedings of the
890 National Academy of Sciences of the United States of America 110:19854-
891 19859.

892 Lee Y, Marsden CD, Norris LC, Collier TC, Main BJ, Fofana A, Cornel AJ, Lanzaro GC.
893 2013b. Spatiotemporal dynamics of gene flow and hybrid fitness between the M

894 and S forms of the malaria mosquito, Anopheles gambiae. Proceedings of the
895 National Academy of Sciences 110:19854-19859.

896 Liu C, Pitts RJ, Bohbot JD, Jones PL, Wang G, Zwiebel LJ. 2010. Distinct olfactory
897 signaling mechanisms in the malaria vector mosquito Anopheles gambiae. PLoS
898 Biology 8:27–28.

899 Livadas G, Mouchet J, Gariou J, Chastang R. 1958. Can one foresee the eradication of
900 malaria in wooded areas of South Cameroun. Rivista di Malariologia 37:229.

901 Lynd A, Weetman D, Barbosa S, Yawson AE, Mitchell S, Pinto J, Hastings I, Donnelly
902 MJ. 2010. Field, genetic, and modeling approaches show strong positive
903 selection acting upon an insecticide resistance mutation in Anopheles gambiae
904 ss. Molecular Biology and Evolution 27:1117-1125.

905 Maynard Smith J, Haigh j. 1974. The hitch-hiking effect of a favorable gene. Genet
906 Res 23:22-35.

907 Meirmans PG, Van Tienderen PH. 2004. GENOTYPE and GENODIVE: two programs
908 for the analysis of genetic diversity of asexual organisms. Molecular Ecology
909 Notes 4:792-794.

910 Moiroux N, Gomez MB, Pennetier C, Elanga E, Djènontin A, Chandre F, Djègbé I, Guis
911 H, Corbel V. 2012. Changes in anopheles funestus biting behavior following
912 universal coverage of long-lasting insecticidal nets in benin. Journal of
913 Infectious Diseases 206:1622–1629.

914 Mueller P, Chouaibou M, Pignatelli P, Etang J, Walker ED, Donnelly MJ, Simard F,
915 Ranson H. 2008. Pyrethroid tolerance is associated with elevated expression of

916 antioxidants and agricultural practice in Anopheles arabiensis sampled from an
917 area of cotton fields in Northern Cameroon. Molecular Ecology 17:1145-1155.

918 Mwangangi JM, Mbogo CM, Orindi BO, Muturi EJ, Midega JT, Nzovu J, Gatakaa H,
919 Githure J, Borgemeister C, Keating J, et al. 2013. Shifts in malaria vector species
920 composition and transmission dynamics along the Kenyan coast over the past
921 20 years. Malaria Journal 12:13.

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

926 Nkya TE, Akhouayri I, Kisinza W, David J-P. 2013. Impact of environment on
927 mosquito response to pyrethroid insecticides: facts, evidences and prospects.
928 Insect Biochemistry and Molecular Biology 43:407-416.

929 Norris LC, Main BJ, Lee Y, Collier TC, Fofana A, Cornel AJ, Lanzaro GC. 2015. Adaptive
930 introgression in an African malaria mosquito coincident with the increased
931 usage of insecticide-treated bed nets. Proceedings of the National Academy of
932 Sciences 112:815-820.

933 Nosil P, Vines TH, Funk DJ. 2005. Reproductive isolation caused by natural selection
934 against immigrants from divergent habitats. Evolution 59:705-719.

935 Nosil P, Feder JL. 2012. Widespread yet heterogeneous genomic divergence.
936 Molecular Ecology 21:2829–2832.

937 Nwane P, Etang J, Chouaibou M, Toto JC, Kerah-Hinzoumbé C, Mimpfoundi R,
938 Awono-Ambene HP, Simard F. 2009. Trends in DDT and pyrethroid resistance
939 in Anopheles gambiae ss populations from urban and agro-industrial settings in
940 southern Cameroon. BMC infectious diseases 9:163.

941 Nwane P, Etang J, Chouaibou M, Toto JC, Koffi A, Mimpfoundi R, Simard F. 2013.
942 Multiple insecticide resistance mechanisms in Anopheles gambiae sl
943 populations from Cameroon, Central Africa. Parasite and Vectors 6:41.

944 O'Loughlin SM, Magesa S, Mbogo C, Mosha F, Midega J, Lomas S, Burt A. 2014.
945 Genomic analyses of three malaria vectors reveals extensive shared
946 polymorphism but contrasting population histories. Molecular Biology and
947 Evolution 31:889-902.

948 Paradis E, Claude J, Strimmer K. 2004. APE: analyses of phylogenetics and evolution
949 in R language. Bioinformatics 20:289-290.

950 Pelz H-J, Rost S, Hünerberg M, Fregin A, Heiberg A-C, Baert K, MacNicoll AD, Prescott
951 CV, Walker A-S, Oldenburg J. 2005. The genetic basis of resistance to
952 anticoagulants in rodents. Genetics 170:1839-1847.

953 Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest
954 RADseq: an inexpensive method for de novo SNP discovery and genotyping in
955 model and non-model species. PLoS One 7:e37135.

956 Platt N, Kwiatkowska R, Irving H, Diabaté A, Dabire R, Wondji C. 2015. Target-site
957 resistance mutations (kdr and RDL), but not metabolic resistance, negatively

impact male mating competiveness in the malaria vector Anopheles gambiae. Heredity.

Plotree D, Plotgram D. 1989. PHYLIP-phylogeny inference package (version 3.2). cladistics 5:163-166.

Poupardin R, Riaz MA, Jones CM, Chandor-Proust A, Reynaud S, David J-P. 2012. Do pollutants affect insecticide-driven gene selection in mosquitoes? Experimental evidence from transcriptomics. Aquatic Toxicology 114:49-57.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics 81:559-575.

R Development Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria

Raj A, Stephens M, Pritchard JK. 2014. fastSTRUCTURE: variational inference of population structure in large SNP data sets. Genetics 197:573-589.

Riehle MM, Guelbeogo WM, Gneme A, Eiglmeier K, Holm I, Bischoff E, Garnier T, Snyder GM, Li X, Markianos K, et al. 2011. A cryptic subgroup of Anopheles gambiae is highly susceptible to human malaria parasites. Science 331:596–598.

Rinker DC, Zhou X, Pitts RJ, Rokas A, Zwiebel LJ. 2013. Antennal transcriptome profiles of anopheline mosquitoes reveal human host olfactory specialization in Anopheles gambiae. BMC Genomics 14:749.

980 Robert V, Macintyre K, Keating J, Trape J-F, Duchemin J-B, Warren M, Beier JC. 2003.
981 Malaria transmission in urban sub-Saharan Africa. American Journal of Tropical
982 Medicine and Hygiene 68:169-176.

983 Rosenberg NA. 2004. DISTRUCT: a program for the graphical display of population
984 structure. Molecular Ecology Notes 4:137-138.

985 Rowland M. 1991a. Activity and mating competitiveness of gamma HCH/dieldrin
986 resistant and susceptible male and virgin female Anopheles gambiae and An.
987 stephensi mosquitoes, with assessment of an insecticide-rotation strategy.
988 Medical and Veterinary Entomology 5:207.

989 Rowland M. 1991b. Behaviour and fitness of gamma HCH/dieldrin resistant and
990 susceptible female Anopheles gambiae and An. stephensi mosquitoes in the
991 absence of insecticide. Medical and Veterinary Entomology 5:193.

992 Santolamazza F, Della Torre A, Caccone A. 2004. Short report: A new polymerase
993 chain reaction-restriction fragment length polymorphism method to identify
994 *Anopheles arabiensis* from *An. gambiae* and its two molecular forms from
995 degraded DNA templates or museum samples. American Journal of Tropical
996 Medicine and Hygiene 70:604-606.

997 Savolainen O, Lascoux M, Merilä J. 2013. Ecological genomics of local adaptation.
998 Nature Review Genetics 14:807–820.

999 Scott JA, Brogdon WG, Collins FH. 1993. Identification of single specimens of the
1000 Anopheles gambiae complex by the polymerase chain reaction. American
1001 Journal of Tropical Medicine and Hygiene 49:520-529.

1002 Service MW. 1993. Mosquito Ecology. Field Sampling Methods, 2nd ed. London:
1003 Elsevier Applied Science Publishers, Ltd.

1004 Simard F, Ayala D, Kamdem GC, Etouna J, Ose K, Fotsing J-M, Fontenille D, Besansky
1005 NJ, Costantini C. 2009. Ecological niche partitioning between the M and S
1006 molecular forms of *Anopheles gambiae* in Cameroon: the ecological side of
1007 speciation. BMC Ecology 9:17.

1008 Slotman MA, Tripet F, Cornel AJ, Meneses CR, Lee Y, Reimer LJ, Thiemann TC, Fondjo
1009 E, Fofana A, Traore SF, et al. 2007. Evidence for subdivision within the M
1010 molecular form of Anopheles gambiae. Molecular Ecology 16:639-649.

1011 Sokhna C, Ndiath MO, Rogier C. 2013. The changes in mosquito vector behaviour and
1012 the emerging resistance to insecticides will challenge the decline of malaria.
1013 Clinical Microbiology and Infection 19:902–907.

1014 Song Y, Endepols S, Klemann N, Richter D, Matuschka F-R, Shih C-H, Nachman MW,
1015 Kohn MH. 2011. Adaptive introgression of anticoagulant rodent poison
1016 resistance by hybridization between old world mice. Current Biology 21:1296-
1017 1301.

1018 Stephens M and Scheet P. 2005. Accounting for decay of linkage disequilibrium in
1019 haplotype inference and missing-data imputation. Am J Hum Genet 76:449-462.

1020 Stephens M, Smith N J, Donnelly P. 2001. A new statistical method for haplotype
1021 reconstruction from population data. Am J Hum Genet 68:978-989.

1022 Stevenson J, St. Laurent B, Lobo NF, Cooke MK, Kahindi SC, Oriango RM, Harbach RE,
1023 Cox J, Drakeley C. 2012. Novel vectors of malaria parasites in the western
1024 highlands of Kenya. Emerging Infectious Diseases 18:1547–1549.

1025 Storz JF. 2005. Using genome scans of DNA polymorphism to infer adaptive
1026 population divergence. Molecular Ecology 14:671-688.

1027 Suwanchaichinda C, Brattsten L. 2001. Effects of exposure to pesticides on carbaryl
1028 toxicity and cytochrome P450 activities in Aedes albopictus larvae (Diptera:
1029 Culicidae). Pesticide Biochemistry and Physiology 70:63-73.

1030 Tajima F. 1989. Statistical method for testing the neutral mutation hypothesis by
1031 DNA polymorphism. Genetics 123:585-595.

1032 Takken W, Verhulst NO. 2013. Host Preferences of Blood-Feeding Mosquitoes.
1033 Annual Review of Entomology :433–453.

1034 Tantely ML, Tortosa P, Alout H, Berticat C, Berthomieu A, Rutee A, Dehecq J-S,
1035 Makoundou P, Labbé P, Pasteur N. 2010. Insecticide resistance in Culex pipiens
1036 quinquefasciatus and Aedes albopictus mosquitoes from La Reunion Island.
1037 Insect Biochemistry and Molecular Biology 40:317-324.

1038 Tene BF, Poupardin R, Costantini C, Awono-Ambene P, Wondji CS, Ranson H,
1039 Antonio-Nkondjio C. 2013. Resistance to DDT in an urban setting: common
1040 mechanisms implicated in both M and S forms of Anopheles gambiae in the city
1041 of Yaoundé Cameroon. PLoS One 8:e61408.

1042 Tene Fossog B, Antonio-Nkondjio C, Kengne P, Njiokou F, Besansky NJ, Costantini C.
1043 2013. Physiological correlates of ecological divergence along an urbanization

1044 gradient: differential tolerance to ammonia among molecular forms of the
1045 malaria mosquito Anopheles gambiae. BMC Ecology 13:1.

1046 Tene Fossog B, Ayala D, Acevedo P, Kengne P, Ngomo Abeso Mebuy I, Makanga B,
1047 Magnus J, Awono-Ambene P, Njiokou F, Pombi M, et al. 2015. Habitat
1048 segregation and ecological character displacement in cryptic African malaria
1049 mosquitoes. Evolutionary Applications 8:326–345.

1050 Tiffin P, Ross-Ibarra J. 2014. Advances and limits of using population genetics to
1051 understand local adaptation. Trends in Ecology and Evolution 29:673-680.

1052 Turissini DA, Gamez S, White BJ. 2014. Genome-wide patterns of polymorphism in
1053 an inbred line of the African malaria mosquito Anopheles gambiae. Genome
1054 biology and evolution 6:3094-3104.

1055 United Nations DoEaSA, Population Division. 2014. World Urbanization Prospects:
1056 The 2014 Revision, Highlights. (ST/ESA/SER.A/352).

1057 Vannini L, Reed TW, Willis JH. 2014. Temporal and spatial expression of cuticular
1058 proteins of Anopheles gambiae implicated in insecticide resistance or
1059 differentiation of M/S incipient species. Parasites and Vectors 7:24.

1060 Wang-Sattler R, Blandin S, Ning Y, Blass C, Dolo G, Toure YT, della Torre A, Lanzaro
1061 GC, Steinmetz LM, Kafatos FC, et al. 2007. Mosaic Genome Architecture of the
1062 Anopheles gambiae Species Complex. PLoS One 2(11): e1249

1063 Weill M, Chandre F, Brengues C, Manguin S, Akogbeto M, Pasteur N, Guillet P,
1064 Raymond M. 2000. The kdr mutation occurs in the Mopti form of Anopheles
1065 gambiae s.s. through introgression. Insect Molecular Biology 9:451-455.

1066 White BJ, Collins FH, Besansky NJ. 2011. Evolution of Anopheles gambiae in Relation
1067 to Humans and Malaria. Annual Review of Ecology, Evolution, and Systematics,
1068 Vol 42 42:111-132.

1069 White B, Lawniczak M, Cheng C, Coulibaly MB, Wilson MD, Sagnon N, Costantini C,
1070 Simard F, Christophides GK, Besansky NJ. 2011. Adaptive divergence between
1071 incipient species of Anopheles gambiae increases resistance to Plasmodium.
1072 Proceedings of the National Academy of Sciences 108:244–249.

1073 Windbichler N, Menichelli M, Papathanos PA, Thyme SB, Li H, Ulge UY, Hovde BT,
1074 Baker D, Monnat RJ, Jr., Burt A, et al. 2011. A synthetic homing endonuclease-
1075 based gene drive system in the human malaria mosquito. Nature 473:212-215.

1076 Wondji C, Frederic S, Petrarca V, Etang J, Santolamazza F, Della Torre A, Fontenille D.
1077 2005. Species and populations of the *Anopheles gambiae* complex in Cameroon
1078 with special emphasis on chromosomal and molecular forms of *Anopheles*
1079 *gambiae s.s.* Journal of Medical Entomology 42:998-1005.

1080 Wood O, Hanrahan S, Coetzee M, Koekemoer L, Brooke B. 2010. Cuticle thickening
1081 associated with pyrethroid resistance in the major malaria vector Anopheles
1082 funestus. Parasite and Vectors 3:67.

1083

1084

1085 FIGURE LEGENDS

1086 **Figure 1. *Anopheles gambiae* complex sibling species are genetically**
1087 **distinct.** A) PCA B) NJ Tree, and C) fastSTRUCTURE analyses clearly separate
1088 the four *An. gambiae* complex species that occur in Cameroon into discrete
1089 genetic clusters. Additional subdivision below the species level is apparent within
1090 *An. coluzzii* and *An. gambiae.* D) Species composition varies strongly between
1091 eco-geographic regions. Sampling sites are denoted by black dots.

1092

1093 **Figure 2. *Anopheles gambiae* is divided into three cryptic subpopulations.**
1094 A) PCA, B) NJ Tree, and C) fastSTRUCTURE analyses reveal subdivisions
1095 within *An. gambiae.* We term the most abundant group *GAM1*, while a second
1096 small, but widely distributed group, is termed *GAM2*. Finally, most individuals
1097 from the village of Nkolondom are genetically distinct from other *An. gambiae*
1098 suggestive of local adaptation.

1099

1100 **Figure 3. *Anopheles coluzzii* is divided into four subgroups.** A) PCA, NJ
1101 Tree, and C) fastSTRUCTURE reveal major population structuring between *An.*
1102 *coluzzii* from the northern Savannah eco-geographic region and *An. coluzzii* from
1103 the southern three forested regions. Within the south, D) PCA, E) NJ Tree, and
1104 F) fastSTRUCTURE analyses separate mosquitoes based on geographic origin,
1105 although clustering is not fully discrete indicating a dynamic interplay between
1106 local adaptation and migration.

1107

Figure 4. Phylogenetic relationships between populations show recent radiation within *An. gambiae* and *An. coluzzii* clades. In an unrooted, F_{ST}-based NJ tree, *An. melas* is most distant from all other species, while *An. gambiae* and *An. coluzzii* are sister species. Southern populations of *An. coluzzii* are more closely related to each other than to the northern savannah population. In contrast to geographic distance, the Douala subpopulation is genetically closer to Yaoundé rather than Coastal mosquitoes. Within *An. gambiae*, a relatively deep split is present between *GAM2* and *GAM1*, while *Nkolondom* appears to have recently diverged from *GAM1*.

Figure 5. Genome scans reveal footprints of global and local adaptation in *An. gambiae* subpopulations. A-C) Diversity and Tajima's *D* are plotted for each of the three subpopulations. Brown asterisks denote windows above the 99th percentile or below the 1st percentile of empirical distribution of Tajima's *D*. D-E) Both absolute (d_{xy}) and relative (F_{ST}) divergence between populations are plotted across 150-kb windows. Red asterisks indicate windows above the 99th percentile of empirical distribution of F_{ST}. In all populations, concordant dips in diversity and Tajima's *D* are evident near the pericentromeric region of 2L where the *para* sodium channel gene is located. Three other selective sweeps located on the X and 3R chromosomes are highlighted (grey boxes).

Figure 6. Strong positive selection acts on xenobiotic resistance loci in subpopulations of *An. coluzzii.* A-H) Grey boxes highlight five selective

sweeps characterized by extreme values of F_{ST} and Tajima's *D*. As in *An. gambiae*, sharp declines in diversity and allele frequency spectrum at the *para* sodium channel gene are present in all populations. A sweep encompassing a cluster of detoxification genes on 3R is limited to urban mosquitoes. The region centered on the resistance to dieldrin (*rdl*) gene also shows signs of local selection in southern subpopulations. 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 Yaoundé 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.

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Detailed map of sampling sites in Cameroon.

Figure S2. Bayesian information criterion was used to determine the most likely number of clusters/populations for A) all 941 samples, B) all *An. coluzzii*, C) *An. coluzzii* from the northern savannah, D) *An. arabiensis*, E) the 309 individuals used for genome scans F) all *An. gambiae*, G) southern *An. coluzzii*, and H) *An. melas.* BIC scores for 1 to 50 clusters are plotted. The lower the BIC score the better the model fits the observed genetic diversity.

Figure S3. Southern (pale pink) and northern populations (deep pink) of *An. coluzzii* are readily separated in PCA (top) and NJ trees (bottom) using SNPs exclusively from any of the five chromosomal arms.

Figure S4. No population substructure is detectable within northern *An. coluzzii* using PCA and NJ tree analysis.

Figure S5. Mean sequencing coverage per individual is plotted in 300kb non-overlapping windows across the genome. Only individuals used in the genome scans are included in the coverage calculation.

Figure 1.

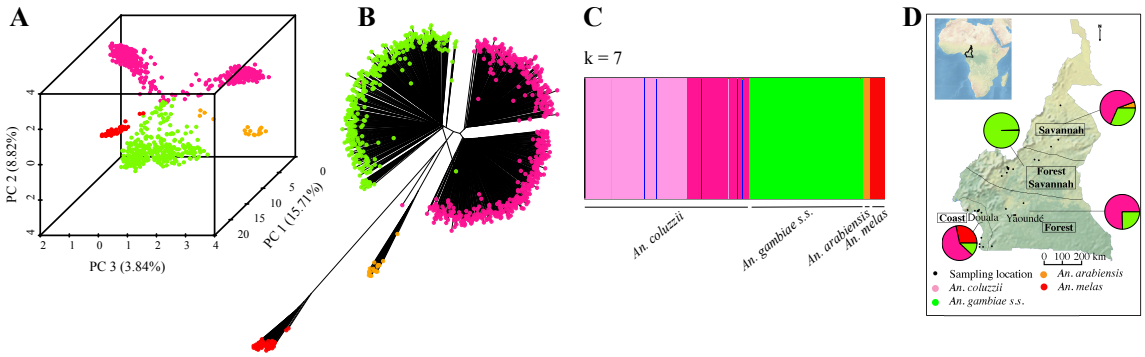


Figure 2.

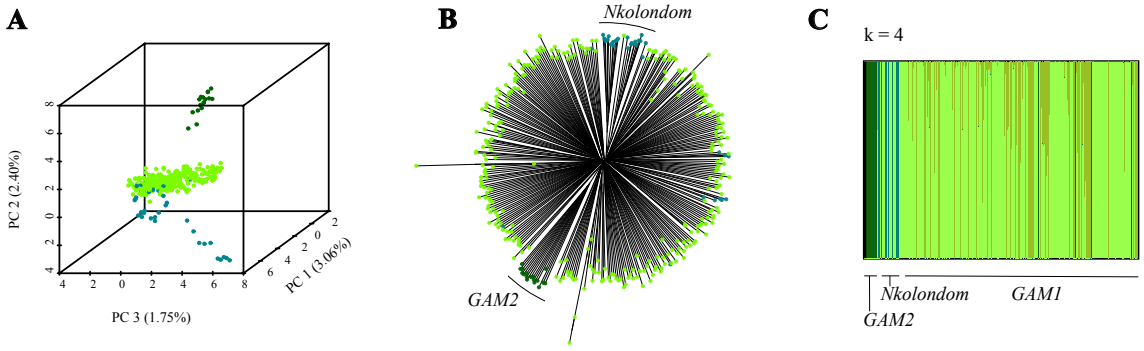


Figure 3.

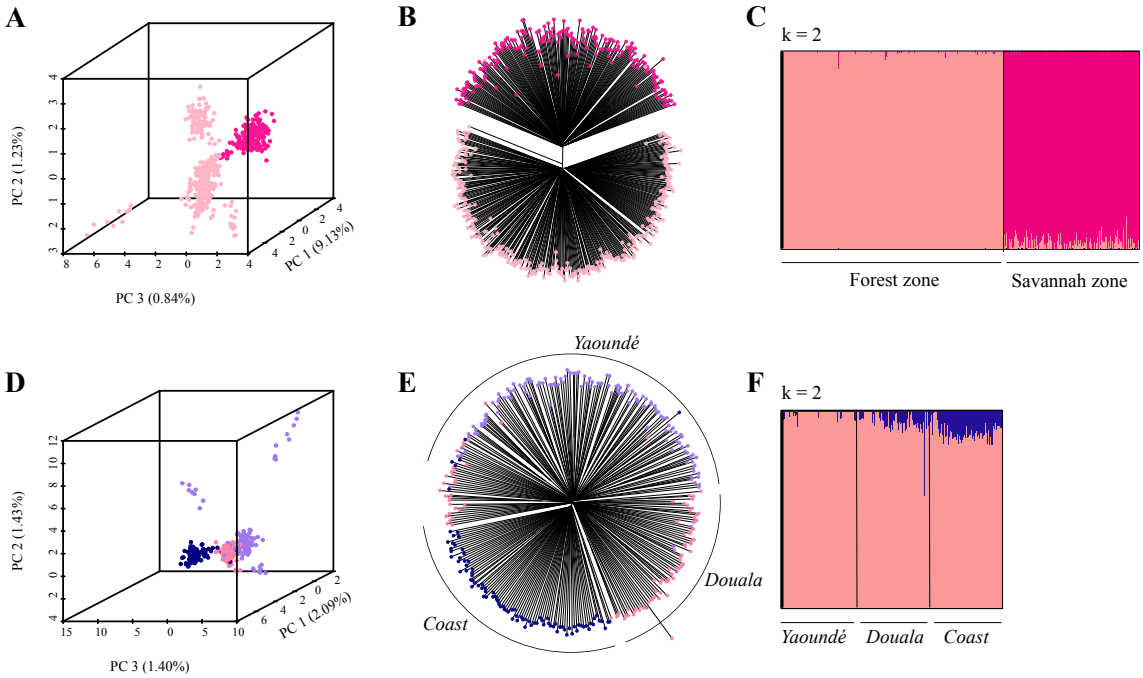


Figure 4.

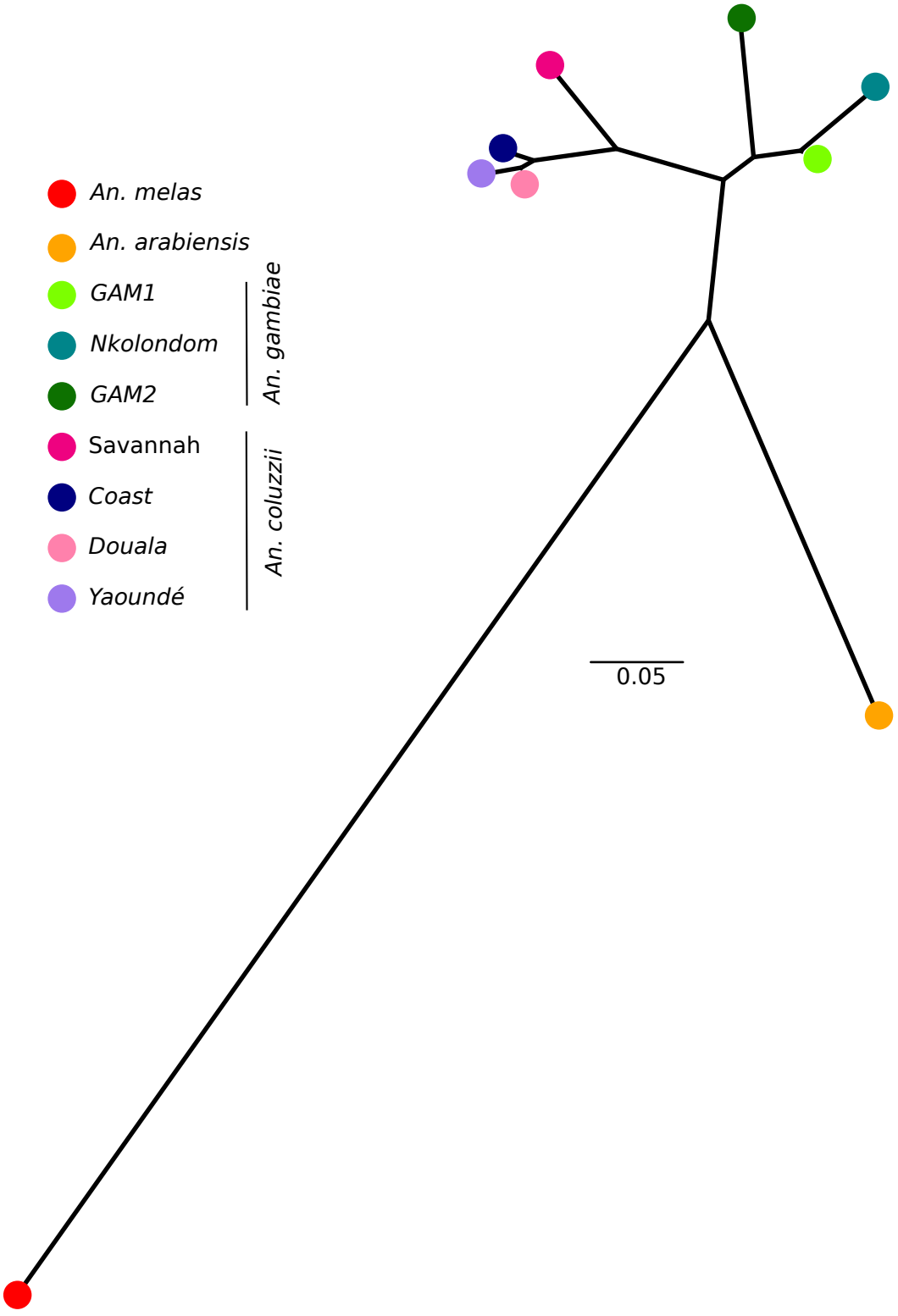


Figure 5.

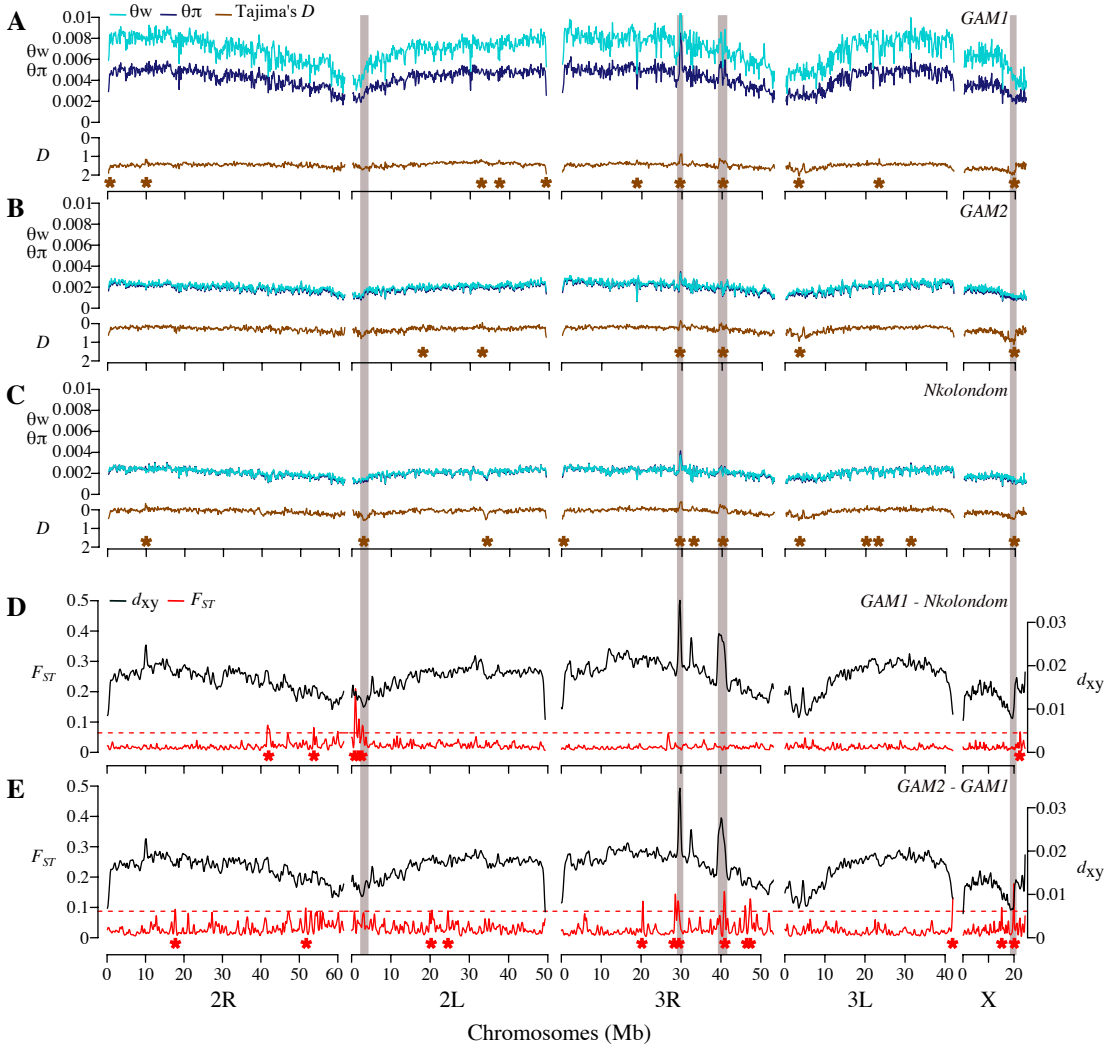


Figure 6.

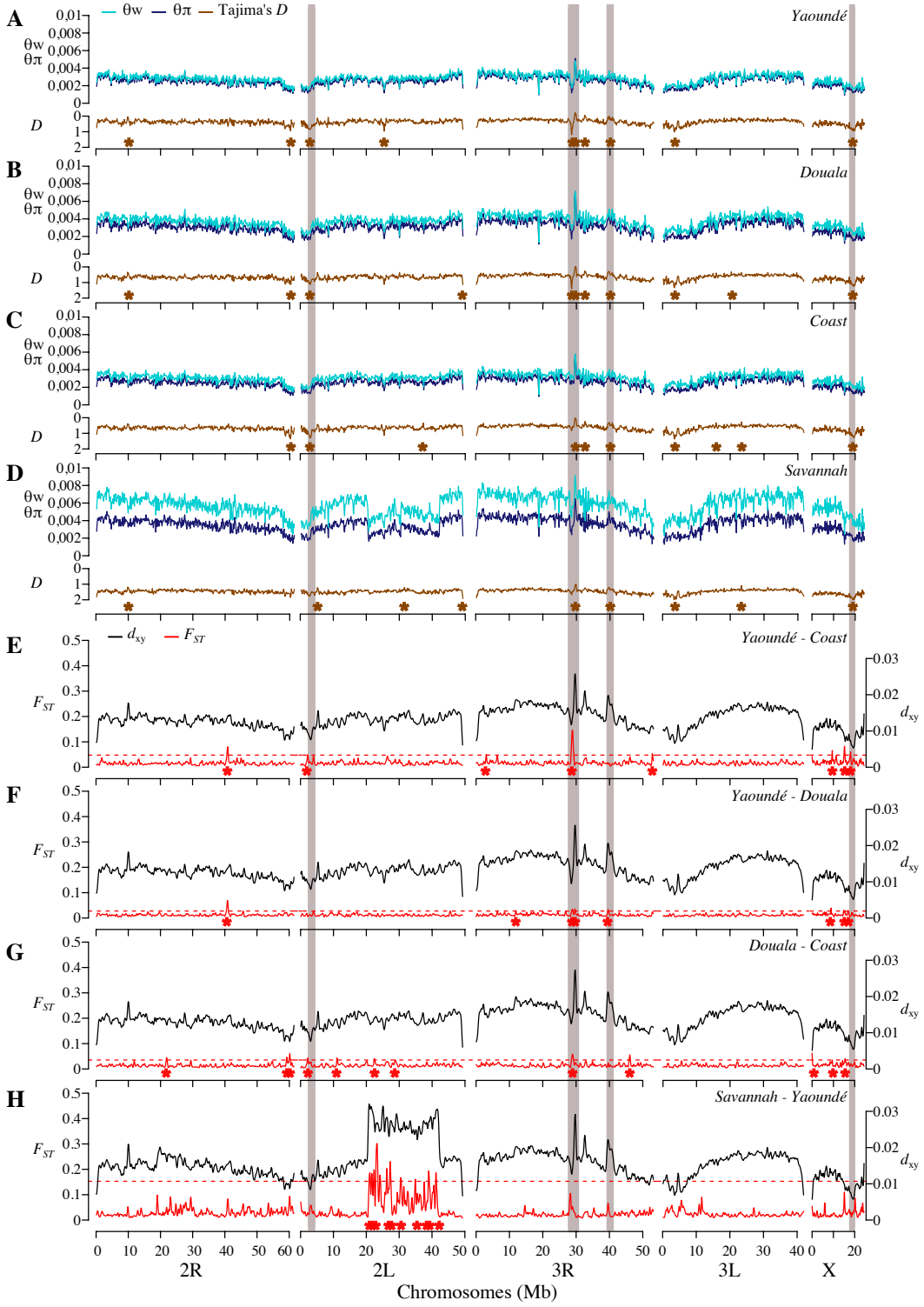


Figure 7.

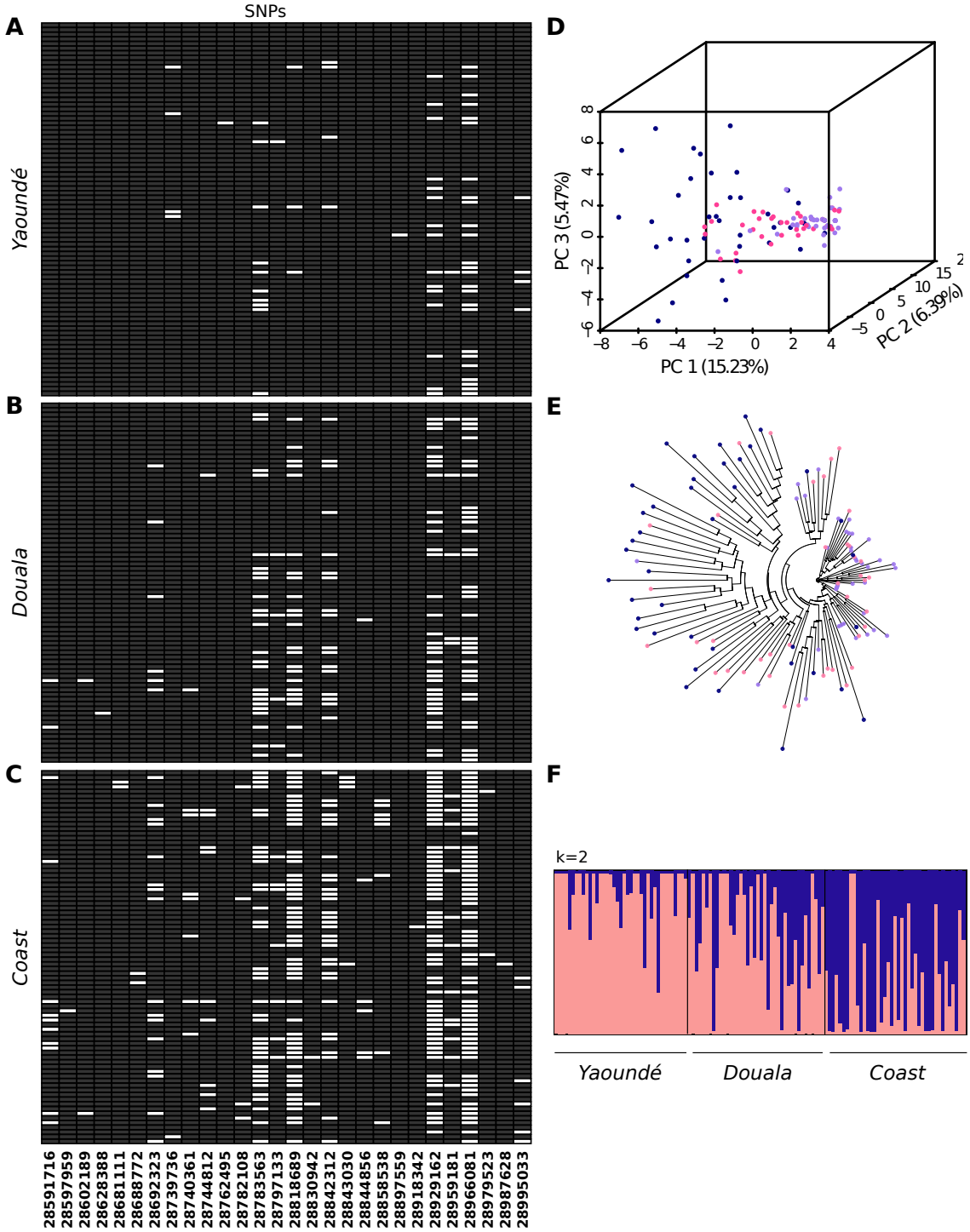


Figure S1.

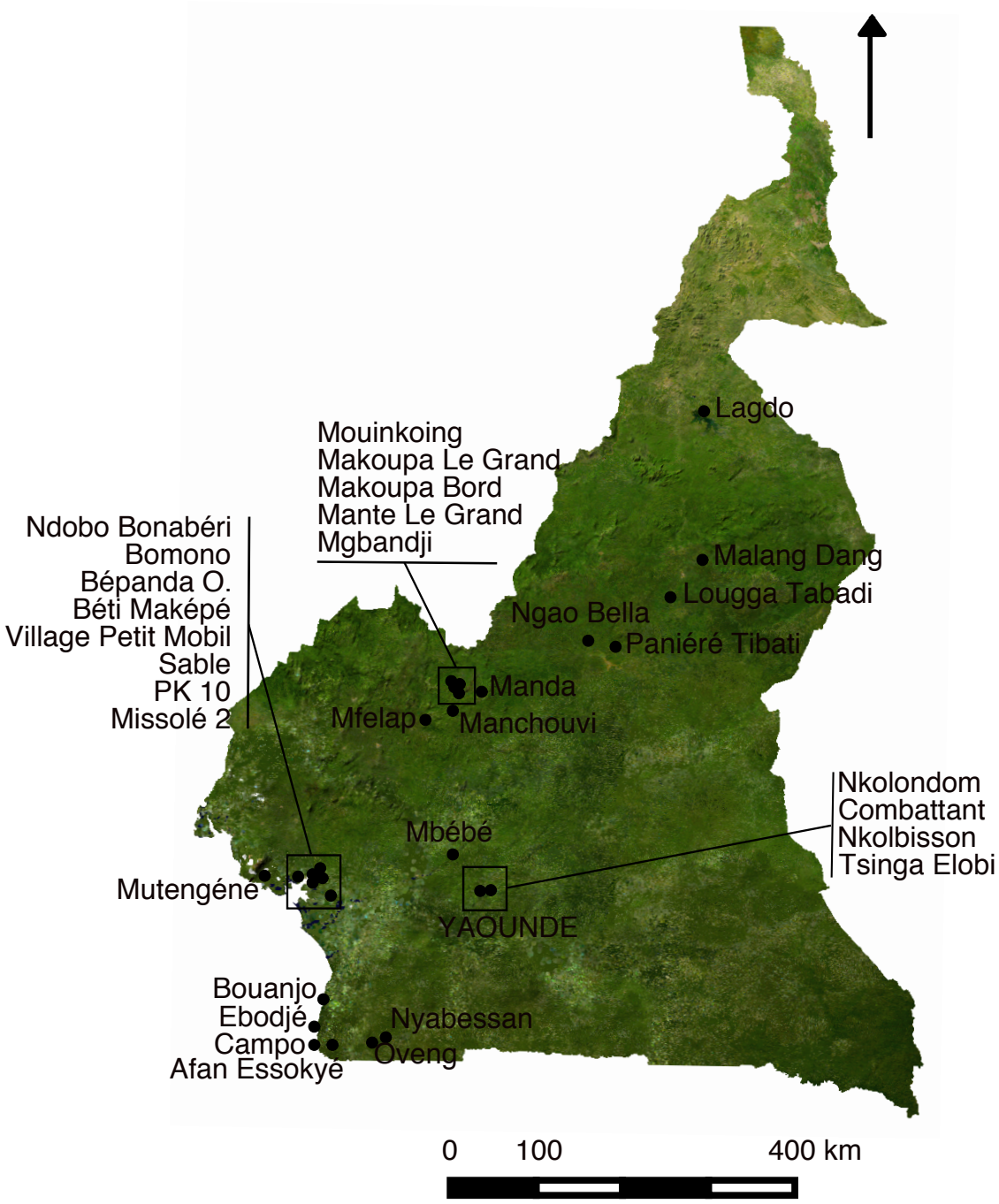


Figure S2.

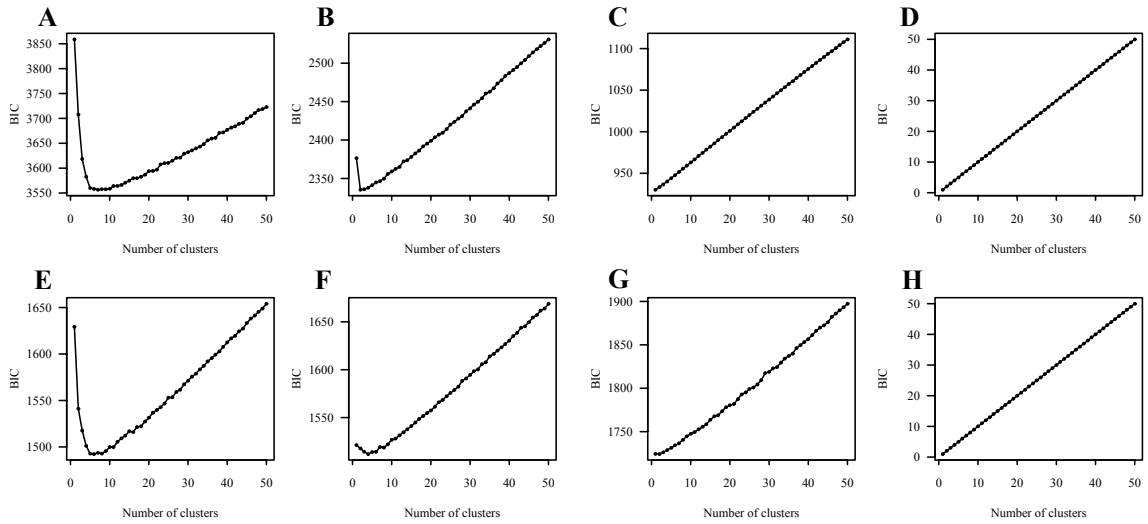


Figure S3.

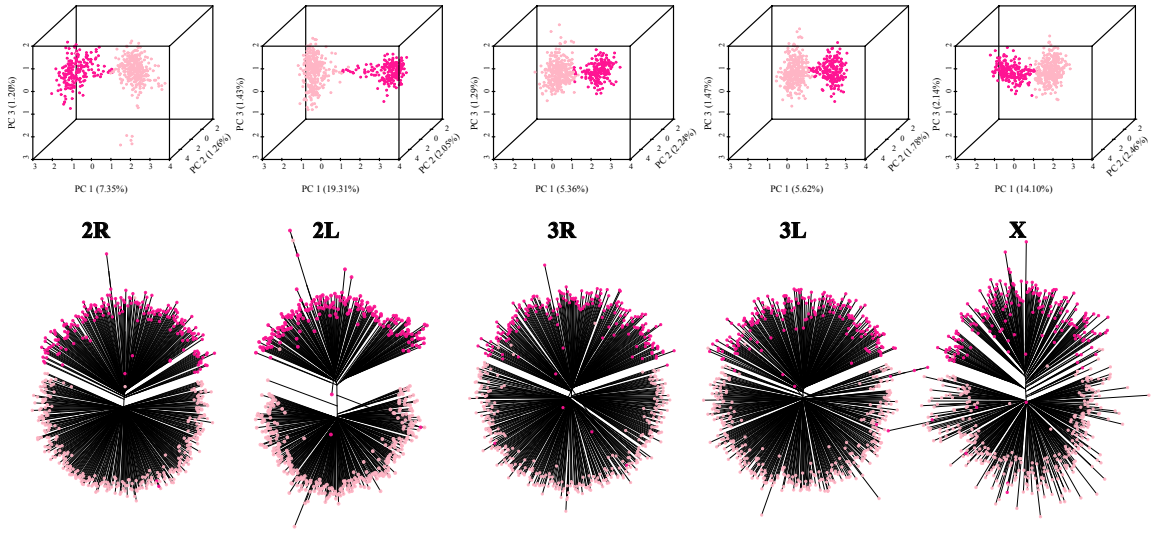


Figure S4.

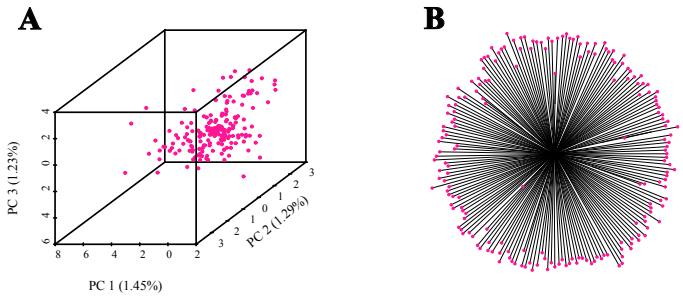
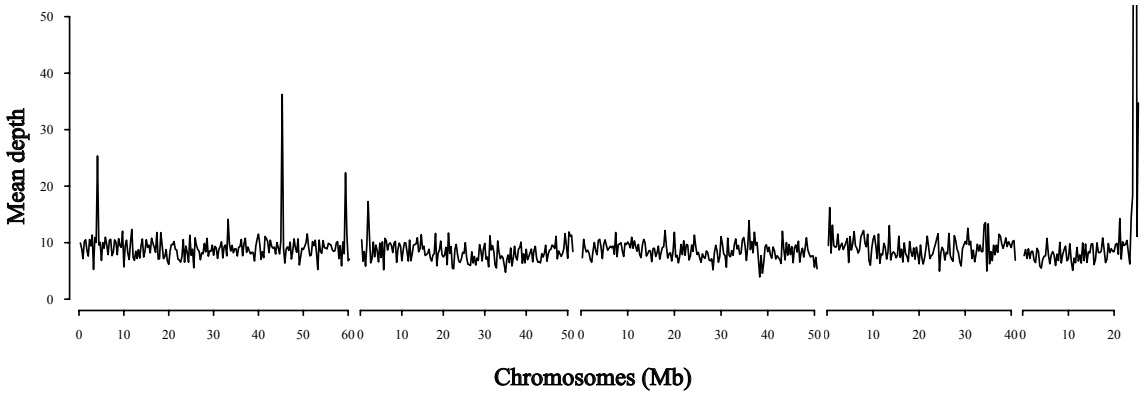


Figure S5.



1174

1175 SUPPLEMENTAL INFORMATION

1176 Consistently negative Tajima's *D* across all subgroups may reflect recent
1177 population expansions. To further address this hypothesis we modeled the
1178 demographic history of each population using a diffusion-based approach
1179 implemented in the software package ∂a∂i v 1.6.3 (Gutenkunst et al. 2009). We
1180 fit four alternative demographic models (*neutral, growth, two-epoch, bottle-*
1181 *growth*), without migration or recombination, to the folded allele frequency
1182 spectrum of each cryptic subgroup of *An. gambiae s.s.* and *An. coluzzii*. The best
1183 model was selected based on the highest composite log likelihood, the lowest
1184 Akaike Information Criterion (AIC), and visual inspection of residuals. As the
1185 choice of model can be challenging in recently diverged populations, we
1186 prioritized the simplest model when we found it difficult to discriminate between
1187 conflicting models. To obtain uncertainty estimates for the demographic
1188 parameters we used the built-in bootstrap function implemented in ∂a∂i to derive
1189 95% bootstrap confidence intervals.

1190 Results indicate that GAM1, GAM2, and Savannah populations have
1191 experienced recent size increases. However, for the southern populations of
1192 *Yaoundé, Coast, Douala, and Nkolondom* the best demographic model is a
1193 *bottle-growth* (Table S4). While most classical studies report *An. gambiae s.l.*
1194 populations that are in expansion (Donnelly et al. 2001), a more recent study
1195 employing RAD markers revealed that some East African populations have more
1196 complex demographic histories, often involving several changes in effective
1197 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 (Athrey et al. 2012). 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.

SUPPLEMENTAL REFERENCES

Athrey G, Hodges TK, Reddy MR, Overgaard HJ, Matias A, Ridl FC, Kleinschmidt I, Caccone A, Slotman M a. 2012. The effective population size of malaria mosquitoes: large impact of vector control. PLoS Genet. 8:e1003097.

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

Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5:e1000695.

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

Ecogeographic regions	Sampling locations	Geographic coordinates	Sampling methods				Total	*Anopheles gambiae sensu lato* populations identified								
			HLC-OUT	HLC-IN	LC	SPRAY		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 (Doula)	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.

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).

Table S5. Functional analysis of gene ontology terms in candidate regions.
Supplementary online information