POLLUTANTS AND INSECTICIDES DRIVE LOCAL ADAPTATION IN AFRICAN MALARIA MOSQUITOES

Colince Kamdem\textsuperscript{1}\textsuperscript{*}, Caroline Fouet\textsuperscript{1}, Stephanie Gamez\textsuperscript{1}, Bradley J. White\textsuperscript{1,2,*}

\textsuperscript{1}Department of Entomology, University of California, Riverside, CA 92521

\textsuperscript{2}Center for Disease Vector Research, Institute for Integrative Genome Biology, University of California, Riverside, CA 92521

\textsuperscript{*}Corresponding authors: guy.kamdemdjoko@ucr.edu; bwhite@ucr.edu
ABSTRACT

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

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

The Afrotropical \textit{Anopheles gambiae} complex is a group of at least nine isomorphic mosquito species exhibiting varying degrees of geographic and reproductive isolation \cite{12-14}. Owing to the critical role its members play in sustaining malaria transmission, a wealth of genomic data exists on the complex including whole genome assemblies for most of the species \cite{15,16}. Because radiation of the complex began \(~1.85\) mya, interspecific genetic comparisons will yield little insight into the establishment of divergence with gene flow. However, both ecological and genetic evidence suggests that contemporary local
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 [17-20]. Up to now, shallow sampling and/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. Worrisomely, ongoing local adaptation to urban sites with high prevalence of xenobiotics is both increasing malaria transmission spatially and further compromising the effectiveness of chemical control tools. Moreover, the extensive population structure within both species will present a formidable challenge to novel vector control strategies that rely on population replacement.
RESULTS

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

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

After aligning ddRADseq reads to the *An. gambiae* reference genome, we used STACKS to remove loci present in <80% of individuals, leaving 8,476 SNPs (~1 SNP every 30kb across the genome) for population structure inference [22, 23]. First, we performed principal component analysis (PCA) on genetic diversity across all 941 individuals (Figure 1A). The top three components explain 28.4% of the total variance and group individuals into five main clusters. Likewise, a neighbor-joining (NJ) tree, based on Euclidian distance of allele frequencies, shows five distinct clades of mosquitoes (Figure 1B). We hypothesized that these groups at least partially correspond to the four sibling species – *An. gambiae*, *An. coluzzii*, *An. arabiensis*, and *An. melas* – known to occur in Cameroon. To confirm, we typed a subset of 288 specimens using validated species ID PCRs and found that each cluster comprised a single species [24, 25]. In agreement
with previous surveys [26, 27], our collections indicate that the brackish water breeding An. melas is limited to coastal regions, while the arid-adapted An. arabiensis is restricted to the savannah. In contrast, An. gambiae and An. coluzzii are distributed across the four eco-geographic zones of Cameroon (Figure 1D). Lee and colleagues [28] recently reported frequent bouts of hybridization between An. gambiae and An. coluzzii in Cameroon. While both the PCA and NJ trees clearly separate the two species, the PCA does show intermixing of some rare individuals consistent with semi-permeable species boundaries.

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

Cryptic Population Structure within An. gambiae s.s.

To further resolve the population structure within 357 An. gambiae specimens, we performed population genetic analysis with a set of 9,345 filtered SNPs. Using a combination of PCA, NJ trees, and ancestry assignment, we
consistently identify three distinct subgroups within *An. gambiae* (Figure 2A-C).

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. 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 [30, 31], 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 robustly and consistently separate northern and southern
mosquitoes, suggesting incipient reproductive isolation between the two populations (Figure S2).

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 data suggest a dynamic and ongoing process of local adaptation within southern *An. coluzzii*. In contrast, no similar geographic clustering is observed in northern populations (Figure S1, S3).

**Relationships Between Species and Subgroups**

Population genomic analysis identified four different *An. gambiae* s.l. species present within our samples. Within *An. gambiae* and *An. coluzzii* we identified seven potential subgroups with apparently varying levels of isolation. To further explore the relationships between different populations, we built an unrooted NJ tree using pairwise levels of genomic divergence (*F*<sub>ST</sub>) between all species and subgroups (Figure 4, Table S2). As previously observed in a phylogeny based on whole genome sequencing [16], we find that *An. melas* is highly divergent from all other species (*F*<sub>ST</sub> ~ 0.8), while *An. arabiensis* shows
intermediate levels of divergence ($F_{ST} \sim 0.4$) from *An. gambiae* and *An. coluzzii*.

As expected, the sister species *An. gambiae* and *An. coluzzii* are more closely related to each other ($F_{ST} \sim 0.2$) 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 ($F_{ST} \sim 0.1$) 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 ($F_{ST} \sim 0.05$), 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.

**Using genome scans to identify selective sweeps**

To find potential targets of selection within subgroups we performed scans of nucleotide diversity ($\theta_w$, $\theta_n$) and allele frequency spectrum (*Tajima’s D*) using non-overlapping 150-kb windows across the genome. Scans of $\theta_w$, $\theta_n$, and *Tajima’s D* were conducted by importing aligned, but otherwise unfiltered, reads directly into ANGSD, which uses genotype likelihoods to calculate summary statistics [32]. Natural selection can increase the frequency of an adaptive variant within a population, leading to localized reductions in genetic diversity (and a skew towards rare polymorphisms) as the haplotype containing the adaptive variant(s) sweeps towards fixation [33, 34]. Thus, genomic regions harboring
targets of recent selection should exhibit extreme reductions in these metrics
relative to genome-wide averages [35].

We also performed genome scans using both a relative ($F_{ST}$) and absolute
($d_{xy}$) measure of divergence calculated with STACKS and ANGSD, respectively
[36]. If positive selection is acting on alternative haplotypes of the same locus in
two populations, values of $F_{ST}$ and $d_{xy}$ should increase at the target of selection.
Whereas spatially varying selection that acts on one population, but not the
other, should produce a spike in $F_{ST}$ between populations and no change in $d_{xy}$.
Finally, parallel selection on the same haplotype in two populations should lead
to a decrease in both metrics. For both diversity and divergence scans we used a
maximum of 40 mosquitoes per population, prioritizing individuals with the
highest coverage in populations where sample size exceeded 40.

Targets of Selection within An. gambiae subgroups

Our calculations of genome-wide diversity levels (Table S3) within An.
gambiae subgroups are comparable to previous estimates based on RAD
sequencing of Kenyan An. gambiae s.l. populations [37]. 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) [34]. Indeed, formal
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). An apparent selective sweep on the left arm of chromosome 2 near the centromere is found in all populations. Notably, the para sodium channel gene is embedded within this sweep. 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 [38-42].

Other sweeps are population-specific. For example, a major sweep unique to the Nkolondom population is found on chromosome 2L from ~33-35 Mb. Nkolondom is suburban region of Yaoundé where larvae can be readily collected from irrigated garden plots that likely contain elevated levels of pesticides directed at agricultural pests [30, 31]. Intriguingly, the region contains six epidermal growth factor (\( EGF \)) genes, which could facilitate larval development of this subgroup in pesticide-laced agricultural water. It should be noted that due to the reduced representation sequencing approach we used, our analysis is necessarily conservative, highlighting only clear instances of positive selection, which are likely both recent and strong [10]. Additional smaller sweeps outside our power of detection are likely prevalent in all subgroups.

We next performed divergence based genome scans. At the putative \( kdr \) 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 at the
$kdr$ sweep, while local values of $F_{ST}$ actually increase. While not definitive, the
dramatic 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. At the Nkolondom specific 2L $EGF$ sweep, we
observe an increase in $F_{ST}$ between GAM1-Nkolondom and no clear change in
$d_{xy}$, indicative of ongoing local adaptation in Nkolondom.

Targets of Selection within An. coluzzii subpopulations

As above, we first used diversity 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). Examination of genome-wide plots
reveals that all subgroups exhibit dips in diversity and Tajima’s $D$ at the $kdr$ locus
(Figure 6A-D). A large sweep at ~25 Mb on 2L centered on the resistance to
dieldrin ($rdl$) locus is clearly present in all southern groups. In the northern
savannah population, a pronounced dip in diversity occurs at $rdl$, but Tajima’s $D$
stays constant. While we can confidently infer that a resistant $rdl$ allele has
recently swept through all southern populations of An. coluzzii, evidence for a
sweep in the northern population is inconclusive. Regional and population-
specific sweeps are also evident including a sweep on 3R is nearly exclusive to
mosquitoes collected in the urban areas of Yaoundé and Douala. A sharp,
dramatic drop in both diversity and Tajima’s D occurs on 3R from ~28.5-29.0 Mb
with the decline being more dramatic in Yaoundé than Douala. Geographical
limitation of the sweep to urban mosquitoes strongly suggests it may contain
variant(s) that confer adaptation to extreme levels of anthropogenic disturbance.
Compellingly, the selective sweep contains a cluster of both Glutathione S-
transferase (GSTE1-GSTE7) and cytochrome P450 (CYP4C27, CYP4C35,
CYP4C36) genes. Both gene families can confer metabolic resistance to
insecticides and pollutants in mosquitoes [43-45]. 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 [46, 47].

Genome scans of both relative and absolute divergence between
populations reveal intriguing patterns at sites of putative selective sweeps. 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). 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. Differentiation between southern and
northern populations at rdl is obscured by the high divergence between
alternative arrangements of the 2La 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. Finally, when comparing mosquitoes from the two urban areas (Douala and Yaoundé), we observe only a slight bump in $F_{ST}$ and a dramatic dip in $d_{xy}$ indicative of an ongoing, shared selective sweep.

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

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

Anthropogenic Mediated Selection

Human activity has altered the evolutionary trajectory of diverse taxa. In insects, spatially varying intensity of insecticide application can drive divergence between populations, potentially leading to reproductive isolation. While a plausible scenario, scant empirical evidence supports the hypothesis [48]. Previous studies have documented dramatic reductions in the population size of An. gambiae s.l. after introduction of long lasting insecticide treated nets (LLINs), but did not determine the influence of exposure on population structuring [49, 50]. Among Cameroonian populations of both An. gambiae and An. coluzzii, we find a pervasive signature of selection at the para sodium channel gene. We infer that this sweep confers globally beneficial resistance to LLINs, which are ubiquitous
in Cameroon and treated with pyrethroids that target para [51]. In contrast, selective sweeps centered on other insecticide resistance genes are restricted to specific geographic locations/populations. For example, a sweep at the rdl locus is limited to southern populations of An. coluzzii. Initial selection for dieldrin resistance likely occurred during massive indoor residual spraying campaigns conducted by the WHO in southern Cameroon during the 1950s. Indeed, the spraying was so intense that it temporarily eliminated An. gambiae s.l. from Yaoundé (and likely other locations in the forest region) [52]. However, due to high human toxicity, dieldrin has been banned for use in mosquito control since the mid-1970s. In the absence of insecticide exposure, resistant rdl mosquitoes are significantly less fit than wild type mosquitoes [53-55], making the continued persistence of resistant alleles in southern An. coluzzii populations puzzling. One plausible explanation is that other cyclodiienes targeting rdl, such as fipronil and lindane, are still commonly used in agriculture and may frequently runoff into An. coluzzii larval habitats, imposing strong selection for resistant mosquitoes. A similar phenomenon was recently proposed to explain the maintenance of resistance rdl alleles in both Culex and Aedes mosquitoes [56].

Mosquitoes inhabiting Cameroon’s two major cities, Yaoundé and Douala, provide a clearer example of how xenobiotic exposure can directly influence population structure. Both cities have seen exponential human population growth over the past 50 years, creating a high concentration of hosts for anthropophilic mosquitoes. Despite elevated levels of organic pollutants and insecticides in urban relative to rural larval sites, surveys show substantial year-round
populations of *An. gambiae* and *An. coluzzii* in both cities [57-59]. Bioassays of insecticide resistance demonstrate that urban mosquitoes have significantly higher levels of resistance to multiple insecticides compared to rural mosquitoes [30, 31, 57, 60, 61]. In support of human mediated local adaptation, we find a selective sweep in urban *An. coluzzii* mosquitoes centered on a cluster of GSTE/CYP450 detoxification genes. While the specific ecological driver of the selective sweep is unknown, GSTE and P450 enzymes detoxify both organic pollutants and insecticides [62-64]. Indeed, the synergistic effects of the two types of xenobiotics could be exerting intense selection pressure for pleiotropic resistance in urban mosquitoes [44, 45, 65]. Regardless of the underlying targets of selection, it is clear that mosquitoes inhabiting highly disturbed urban and suburban landscapes are genetically differentiated from rural populations. Further analysis of specific sweeps using a combination of whole genome resequencing and emerging functional genetics approaches (e.g. CRISPR/Cas9) should help resolve the specific targets of local adaptation in urban mosquitoes, while also shedding light on the evolutionary history of the enigmatic subgroup *GAM2*.

*Impacts on Vector Control*

Just five decades ago, there was not a single city in Sub-Saharan African with a population over 1 million; today there are more than 40. Population shifts to urban areas will only continue to increase with the United Nations estimating that 60% of Africans will live in large cities by 2050 [66]. When urbanization commenced, it was widely assumed that malaria transmission would be minimal
because rural *Anopheles* vectors would not be able to complete development in the polluted larval habitats present in cities [67]. However, increasingly common reports of endemic malaria transmission in urban areas across Sub-Saharan Africa unequivocally demonstrate that anophelines are exploiting the urban niche [68-70]. Specifically, our study shows that *An. gambiae* s.l. from the urban and suburban centers of southern Cameroon form genetically distinct subgroups relative to rural populations. Local adaptation to urban environments is accompanied by strong selective sweeps centered on putative xenobiotic resistance genes, which are likely driven by a combination of exposure to organic pollutants and insecticides in larval habitats. The rapid adaptation of *Anopheles* to the urban landscape poses a growing health risk to Africans as levels of resistance in these populations negate the effectiveness of almost all commonly used insecticides. Moreover, repeated instances of beneficial alleles introgressing between *An. gambiae* s.l. species make the emergence of highly resistant subgroups even more troubling [14, 16, 41, 42, 71]. In essence, urban populations can serve as a reservoir for resistance alleles, which have the potential to rapidly move between species/populations as needed. Clearly, sustainable malaria vector control, urban or otherwise, requires not only more judicious use of insecticides, but also novel strategies not reliant on chemicals. Towards this goal, various vector control methods that aim to replace or suppress wild mosquito populations using genetic drive are currently under development (e.g. [72]). While promising, the complexities of ongoing cryptic
diversification within African *Anopheles* must be explicitly planned for prior to the release of transgenic mosquitoes.
MATERIALS AND METHODS

Mosquito collections

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

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 [76]. Preparation of ddRAD libraries largely followed [77]. 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 contained 288 individuals and was subjected to single end, 100 bp sequencing across one or
two flow cells lanes run on an Illumina HiSeq2500. A detailed library preparation protocol is available at mosquitogenomics.org/protocols.

Raw sequence reads were demultiplexed and quality filtered using the STACKS v 1.29 process_radtags pipeline [22, 23]. After removal of reads with ambiguous barcodes, incorrect restriction sites, and low sequencing quality (mean Phred < 33), GSNAP was used to align reads to the An. gambiae PEST reference genome (AgamP4.2) allowing up to five mismatches per read. After discarding reads that perfectly aligned to more than one genomic position, we used STACKS to identify unique RAD tags and construct consensus assemblies for each. Individual SNP genotypes were called using default setting in the maximum-likelihood statistical model implemented in the STACKS genotypes pipeline.

Population Genomic Analysis

Population genetic structure was assessed using the SNP dataset output by the populations program of STACKS. We used PLINK v 1.19 to retrieve subsets of genome-wide SNPs as needed [78]. PCA, neighbor-joining tree analyses, and Bayesian information criterion (BIC) were implemented using the packages adegenet and ape in R [79, 80]. Ancestry analyses were conducted in fastSTRUCTURE v 1.0 [29] using the logistic method. The choosek.py script was used to find the appropriate number of populations (k); in cases where a range of k was suggested, the BIC-inferred number of clusters was chosen [81]. Ancestry assignment of individual mosquitoes was then visualized with DISTRUCT v 1.1.
We input pairwise $F_{ST}$ values into the program Fitch from the Phylip [83] suite to create the population-level NJ tree.

**Genome Scans for Selection**

We used ANGSD v 0.612 [32] to calculate nucleotide diversity ($\theta_w$ and $\theta_\pi$) and Tajima’s $D$ in 150-kb non-overlapping windows. Unlike most genotyping algorithms, ANGSD does not perform hard SNP calls, instead taking genotyping uncertainty into account when calculating summary statistics. Similarly, absolute divergence ($d_{xy}$) was calculated using ngsTools [84] based on genotype likelihoods generated by ANGSD. Kernel smoothed values for 150-kb windows for all four metrics ($\theta_w$, $\theta_\pi$, $D$, $d_{xy}$) 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. Selective sweeps were identified by locating regions where Tajima’s $D$ was in the top percentile of empirical values in at least one population. Regions with low Tajima’s $D$, but unusually high or low read depth (Figure S4) were deemed unreliable due to the likelihood of repeats and local misassembly. We used a subset of 1,000 randomly chosen SNPs to calculate average pairwise $F_{ST}$ between populations in GENODIVE using up to 40 individuals – prioritized by coverage – per population [85]. To determine if selective sweeps were enriched for specific functional annotation classes, we used the program DAVID 6.7 with
default settings [86]. Haplotypes across the GSTE/CYP450 sweep were
reconstructed by PHASE v 2.1.1 using the default recombination model (87).
ACKNOWLEDGEMENTS

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

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


16. Fontaine, M.C., Pease, J.B., Steele, A., Waterhouse, R.M., Neafsey, D.E.,


18. Wang-Sattler, R., Blandin, S., Ning, Y., Blass, C., Dolo, G., Toure, Y.T.,

19. Lee, Y., Marsden, C.D., Norris, L.C., Collier, T.C., Main, B.J., Fofana, A.,

20. Caputo, B., Nwakanma, D., Caputo, F., Jawara, M., Oriero, E.,
Prominent intraspecific genetic divergence within Anopheles gambiae sibling species triggered by habitat discontinuities across a riverine landscape. Molecular Ecology 23, 4574-4589.


Anopheles gambiae complex in Cameroon with special emphasis on chromosomal and molecular forms of Anopheles gambiae s.s. J Med Entomol 42, 998-1005.


resistance mutation in Anopheles gambiae ss. Molecular Biology and Evolution 27, 1117-1125.


Bayoh, M.N., Mathias, D.K., Odiere, M.R., Mutuku, F.M., Kamau, L.,
(2010). Anopheles gambiae: historical population decline associated with
regional distribution of insecticide-treated bed nets in western Nyanza
Province, Kenya.

Bowen, H.L. (2013). Impact of a mass media campaign on bed net use in
Cameroon. Malar J 12, 10.1186.

foresee the eradication of malaria in wooded areas of South Cameroun.

HCH/dieldrin resistant and susceptible male and virgin female Anopheles
gambiae and An. stephensi mosquitoes, with assessment of an
insecticide-rotation strategy. Medical and Veterinary Entomology 5, 207.

resistant and susceptible female Anopheles gambiae and An. stephensi
mosquitoes in the absence of insecticide. Medical and Veterinary
Entomology 5, 193.

Platt, N., Kwiatkowska, R., Irving, H., Diabaté, A., Dabire, R., and Wondji,
C. (2015). Target-site resistance mutations (kdr and RDL), but not
metabolic resistance, negatively impact male mating competitiveness in the
malaria vector Anopheles gambiae. Heredity.


practice in Anopheles arabiensis sampled from an area of cotton fields in
Northern Cameroon. Molecular Ecology 17, 1145-1155.

Urbanization Prospects: The 2014 Revision, Highlights.

(ST/ESA/SER.A/352).

Barnish, G., Konradsen, F., Klinkenberg, E., Townson, H., and Trape, J.-

68. Robert, V., Macintyre, K., Keating, J., Trape, J.-F., Duchemin, J.-B.,
Warren, M., and Beier, J.C. (2003). Malaria transmission in urban sub-
Saharan Africa. The American Journal of Tropical Medicine and Hygiene
68, 169-176.

69. Keiser, J., Utzinger, J., De Castro, M.C., Smith, T.A., Tanner, M., and
malaria control. The American Journal of Tropical Medicine and Hygiene
71, 118-127.

malaria transmission in sub-Saharan Africa: a systematic review. Journal
of tropical medicine 2012.

71. Weill, M., Chandre, F., Brengues, C., Manguin, S., Akogbeto, M., Pasteur,
Mopti form of Anopheles gambiae s.s. through introgression. Insect Mol
Biol 9, 451-455.


PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics 81, 559-575.


FIGURE LEGENDS

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

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

**Figure 3.** *Anopheles coluzzii* is divided into four subgroups. A) PCA, NJ Tree, and C) fastSTRUCTURE reveal major population structuring between *An. coluzzii* from the northern Savannah eco-geographic region and *An. coluzzii* from the southern three forested regions. Within the south, D) PCA, E) NJ Tree, and F) fastSTRUCTURE analyses separate mosquitoes based on geographic origin, although clustering is not fully discrete indicating a dynamic interplay between local adaptation and migration.
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 three subpopulations. A dashed line marks the 1st percentile empirical distribution of Tajima's $D$. 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. A population-specific sweep in Nkolondom is present at ~34 Mb on 2L. D-E) Both absolute ($d_{xy}$) and relative ($F_{ST}$) divergence between populations are plotted with a dashed line marking the 99th percentile distribution of $F_{ST}$. An increase in $F_{ST}$ between GAM1-Nkolondom occurs at the Nkolondom-specific 2L sweep.

Figure 6. Strong positive selection acts on xenobiotic resistance loci in subpopulations of *An. coluzzii*. A-D) Sharp declines in diversity and allele
frequency spectrum at the *para* sodium channel gene are present in all
populations. A sweep centered on the resistance to dieldrin (*rdl*) gene is present
in southern subpopulations, while a dramatic sweep encompassing a cluster of
detoxification genes on 3R is limited to urban mosquitoes. E-G) No evidence for
locally elevated divergence is observed at the *para* or *rdl* loci suggesting a
shared sweep amongst populations. In contrast, urban-rural mosquitoes show
extreme levels of divergence at the detoxification-enriched sweep on 3R.

**Figure 7. Spatially varying selection between urban and coastal**

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

Figure S1. 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 S2. 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 S3. No population substructure is detectable within northern An. coluzzii using PCA and NJ tree analysis.

Figure S4. 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.

A  
B  
C  
D  
E  
F  
G  
H  

Chromosomes (Mb)
Figure 7.
Figure S1.
Figure S2.
Figure S3.
Figure S4.
Consistently negative Tajima’s $D$ across all subgroups may reflect recent population expansions. To further address this hypothesis we modeled the demographic history of each population using a diffusion-based approach implemented in the software package $\partial a\partial i$ v 1.6.3 (1). We fit four alternative demographic models (neutral, growth, two-epoch, bottle-growth), without migration or recombination, to the folded allele frequency spectrum of each cryptic subgroup of An. gambiae s.s. and An. coluzzii. The best model was selected based on the highest composite log likelihood, the lowest Akaike Information Criterion (AIC), and visual inspection of residuals. As the choice of model can be challenging in recently diverged populations, we prioritized the simplest model when we found it difficult to discriminate between conflicting models. To obtain uncertainty estimates for the demographic parameters we used the built-in bootstrap function implemented in $\partial a\partial i$ to derive 95% bootstrap confidence intervals.

Results indicate that GAM1, GAM2, and Savannah populations have experienced recent size increases. However, for the southern populations of Yaoundé, Coast, Douala, and Nkolondom the best demographic model is a bottle-growth (Table S4). While most classical studies report An. gambiae s.l. populations that are in expansion (e.g. 2), a more recent study employing RAD markers revealed that some East African populations have more complex demographic histories, often involving several changes in effective population size ($N_e$) as we observed in southern forest populations of both An. coluzzii and
An. gambiae. It has also been shown that Anopheles mosquitoes can experience drastic declines in $N_e$ due to insecticidal campaigns (3). Such events affect demographic parameters and could be a plausible explanation for the difficulty we encountered in distinguishing between bottle-growth and two-epoch models in some populations.


<table>
<thead>
<tr>
<th>Ecogeographic regions</th>
<th>Village Name</th>
<th>Geographic coordinates</th>
<th>Sampling methods</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Savanna</td>
<td>Lagdo</td>
<td>9°02'56&quot;N, 13°39'22&quot;E</td>
<td>52</td>
<td>143</td>
</tr>
<tr>
<td></td>
<td>Lougga Tabadi</td>
<td>7°06'36&quot;N, 13°12'36&quot;E</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Malang Dang</td>
<td>7°26'06&quot;N, 13°33'14&quot;E</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Ngao Bella</td>
<td>6°30'04&quot;N, 12°25'34&quot;E</td>
<td>28</td>
<td>114</td>
</tr>
<tr>
<td></td>
<td>Paniéré Tibati</td>
<td>6°28'08&quot;N, 12°37'44&quot;E</td>
<td>5</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td>87</td>
<td>313</td>
</tr>
<tr>
<td>F-S Transition</td>
<td>Makoupa Bord</td>
<td>6°06'06&quot;N, 11°11'29&quot;E</td>
<td>2</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>Makoupa Le Grand</td>
<td>6°02'28&quot;N, 11°10'39&quot;E</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>Manchoutvi</td>
<td>5°52'48&quot;N, 11°06'36&quot;E</td>
<td>0</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>Manda</td>
<td>5°43'32&quot;N, 10°52'06&quot;E</td>
<td>1</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Mante Le Grand</td>
<td>6°03'44&quot;N, 11°12'18&quot;E</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Mfelap</td>
<td>5°43'32&quot;N, 10°52'06&quot;E</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Mgbandji</td>
<td>6°05'49&quot;N, 11°08'26&quot;E</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Mouinkoing</td>
<td>6°02'56&quot;N, 11°24'36&quot;E</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td>3</td>
<td>165</td>
</tr>
<tr>
<td>Forest (urban)</td>
<td>Bepanda Omnisport (Douala)</td>
<td>4°03'18&quot;N, 9°43'16&quot;E</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Beti Maképé (Douala)</td>
<td>4°03'54&quot;N, 9°45'40&quot;E</td>
<td>0</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>Bomono Gare (Douala)</td>
<td>4°04'55&quot;N, 9°35'35&quot;E</td>
<td>0</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>PK10 (Douala)</td>
<td>4°02'49&quot;N, 9°46'47&quot;E</td>
<td>0</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Missolé 2 (Douala)</td>
<td>3°59'17&quot;N, 9°54'22&quot;E</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Ndobo Bonabéri (Douala)</td>
<td>4°04'39&quot;N, 9°40'12&quot;E</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Sable (Douala)</td>
<td>4°04'52&quot;N, 9°43'34&quot;E</td>
<td>28</td>
<td>33</td>
</tr>
<tr>
<td></td>
<td>Village Petit Mobil (Douala)</td>
<td>4°00'16&quot;N, 9°45'18&quot;E</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>Nkobisson (Yaoundé)</td>
<td>3°52'29&quot;N, 11°26'58&quot;E</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>Nkolondom (Yaoundé)</td>
<td>3°58'20&quot;N, 11°30'56&quot;E</td>
<td>0</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Tsinga Elobi (Yaoundé)</td>
<td>3°52'49&quot;N, 11°30'23&quot;E</td>
<td>0</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>Combattant (Yaoundé)</td>
<td>3°52'36&quot;N, 11°30'46&quot;E</td>
<td>18</td>
<td>66</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td>47</td>
<td>293</td>
</tr>
<tr>
<td>Forest (rural)</td>
<td>Mbébé</td>
<td>4°10'00&quot;N, 11°04'00&quot;E</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Nyabessan Centre</td>
<td>2°24'00&quot;N, 10°24'00&quot;E</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>Oveng</td>
<td>2°44'00&quot;N, 11°27'00&quot;E</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td></td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td>Location</td>
<td>Latitude, Longitude</td>
<td>HLC-OUT</td>
<td>HLC-IN</td>
<td>LC</td>
</tr>
<tr>
<td>------------</td>
<td>----------------------</td>
<td>---------</td>
<td>--------</td>
<td>----</td>
</tr>
<tr>
<td>Afan Essokyé</td>
<td>2°22'01&quot;N, 9°58'59&quot;E</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Bouanjo</td>
<td>2°48'00&quot;N, 9°54'00&quot;E</td>
<td>0</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Campo</td>
<td>2°22'01&quot;N, 9°49'01&quot;E</td>
<td>48</td>
<td>16</td>
<td>27</td>
</tr>
<tr>
<td>Ebodjé</td>
<td>2°30'00&quot;N, 9°49'05&quot;E</td>
<td>35</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Mutengéné</td>
<td>4°06'53&quot;N, 9°14'51&quot;E</td>
<td>0</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>84</td>
<td>4</td>
<td>28</td>
</tr>
</tbody>
</table>

**Total** 226      51     394     270   941

HLC-OUT, human landing catches performed outdoor; HLC-IN, human landing catches performed indoor; LC, larval collection; SPRAY, spray catches.
Table S2. Pairwise comparison of genetic distance ($F_{ST}$) among cryptic subgroups and sibling species of *An. gambiae s.l.*

<table>
<thead>
<tr>
<th></th>
<th>Yaoundé</th>
<th>Douala</th>
<th>Coast</th>
<th>Savannah</th>
<th>GAM2</th>
<th>Nkolondom</th>
<th>GAM1</th>
<th>An. arabiensis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yaoundé</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Douala</td>
<td>0.016</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coast</td>
<td>0.035</td>
<td>0.023</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Savannah</td>
<td>0.127</td>
<td>0.109</td>
<td>0.103</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAM2</td>
<td>0.244</td>
<td>0.199</td>
<td>0.230</td>
<td>0.168</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nkolondom</td>
<td>0.197</td>
<td>0.201</td>
<td>0.200</td>
<td>0.247</td>
<td>0.161</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAM1</td>
<td>0.183</td>
<td>0.153</td>
<td>0.178</td>
<td>0.188</td>
<td>0.090</td>
<td>0.050</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>An. arabiensis</td>
<td>0.451</td>
<td>0.406</td>
<td>0.498</td>
<td>0.383</td>
<td>0.478</td>
<td>0.372</td>
<td>0.327</td>
<td>-</td>
</tr>
<tr>
<td>An. melas</td>
<td>0.851</td>
<td>0.836</td>
<td>0.830</td>
<td>0.792</td>
<td>0.841</td>
<td>0.818</td>
<td>0.781</td>
<td>0.872</td>
</tr>
</tbody>
</table>
Table S3. Average nucleotide diversity in seven cryptic subgroups of *An. coluzzii* and *An. gambiae s.s.*

<table>
<thead>
<tr>
<th>Population</th>
<th>θπ (bp⁻¹)</th>
<th>θw (bp⁻¹)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yaoundé</td>
<td>0.0025</td>
<td>0.0028</td>
</tr>
<tr>
<td>Douala</td>
<td>0.0031</td>
<td>0.0037</td>
</tr>
<tr>
<td>Coast</td>
<td>0.0025</td>
<td>0.0031</td>
</tr>
<tr>
<td>Savannah</td>
<td>0.0035</td>
<td>0.0057</td>
</tr>
<tr>
<td>GAM2</td>
<td>0.0019</td>
<td>0.0020</td>
</tr>
<tr>
<td>Nkolondom</td>
<td>0.0020</td>
<td>0.0021</td>
</tr>
<tr>
<td>GAM1</td>
<td>0.0042</td>
<td>0.0069</td>
</tr>
</tbody>
</table>
**Table S4.** Parameters of demographic models inferred from folded Site Frequency Spectrum (SFS) of autosomal SNPs in seven cryptic subpopulations of *An. gambiae*

<table>
<thead>
<tr>
<th>Species</th>
<th>Population</th>
<th>Best Model</th>
<th>Log Likelihood</th>
<th>Final Pop Size&lt;sup&gt;a&lt;/sup&gt; (95% CI)</th>
<th>Bottleneck Size&lt;sup&gt;b&lt;/sup&gt; (95% CI)</th>
<th>Time&lt;sup&gt;c&lt;/sup&gt; (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anopheles coluzzii</td>
<td>Yaoundé</td>
<td>Bottle-growth</td>
<td>-83.45</td>
<td>1.18 (0.92 - 1.69)</td>
<td>24.85 (11.94 - 125.77)</td>
<td>0.58 (0.43 - 1.01)</td>
</tr>
<tr>
<td></td>
<td>Douala</td>
<td>Bottle-growth</td>
<td>-84.50</td>
<td>2.20 (1.52 - 3.77)</td>
<td>5.66 (2.80 - 142.42)</td>
<td>0.59 (0.41 - 1.77)</td>
</tr>
<tr>
<td></td>
<td>Coast</td>
<td>Bottle-growth</td>
<td>-82.95</td>
<td>1.72 (1.28 - 2.33)</td>
<td>23.83 (10.95 - 128.52)</td>
<td>0.67 (0.50 - 1.10)</td>
</tr>
<tr>
<td>Anopheles gambiae s.s.</td>
<td>Savannah</td>
<td>Two-epoch</td>
<td>-102.65</td>
<td>6.73 (6.35 - 7.21)</td>
<td></td>
<td>0.62 (0.54 - 0.72)</td>
</tr>
<tr>
<td></td>
<td>GAM2</td>
<td>Two-epoch</td>
<td>-37.48</td>
<td>3.60 (1.74 - 7.70)</td>
<td>3.60 (0.56 - 9.30)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nkolodom</td>
<td>Bottle-growth</td>
<td>-74.92</td>
<td>1.08 (0.82 - 1.12)</td>
<td>24.95 (47.91 - 99.84)</td>
<td>0.63 (0.47 - 0.91)</td>
</tr>
<tr>
<td></td>
<td>GAM1</td>
<td>Two-epoch</td>
<td>-103.33</td>
<td>6.75 (6.33 - 7.24)</td>
<td></td>
<td>0.68 (0.60 - 0.78)</td>
</tr>
</tbody>
</table>

<sup>a</sup> Ratio of contemporary to ancient population size.

<sup>b</sup> Ratio of population size after instantaneous change to ancient population size.

<sup>c</sup> Time in the past at which instantaneous change happened and growth began (in units of 2*Na generations).