Genomic insights into adaptive divergence and speciation between malaria vectors of the *Anopheles nili* group

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

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

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

*Corresponding authors: colincek@ucr.edu; bwhite@ucr.edu

Abstract

Ongoing speciation in most African malaria vectors gives rise to cryptic species that differ remarkably in their behaviour, ecology and capacity to vector malaria parasites. Any vector control measure can be undermined if cryptic populations exist and are mistargeted within vector species. To examine population structure and the potential impacts of recent large-scale control programs, we have investigated genomic patterns of differentiation in mosquitoes belonging to a large taxonomic group that diverged ~3-Myr ago. Using 4343 single nucleotide polymorphisms (SNPs), we detected strong population structure characterized by high F_{ST} values between multiple divergent populations. Delineating cryptic species within the *Anopheles nili* group is challenging due to incongruence between morphology, ribosomal DNA and SNP markers consistent with incomplete lineage sorting and/or gene flow. Throughout the genome, a very high proportion of loci are fixed (F_{ST} = 1) within putative species, which suggests that ecological and/or reproductive barriers are maintained by strong selection on a substantial number of genes.

Key words: *Anopheles nili*, divergent selection, high- *F*ST regions, speciation.

1. Introduction

One of the principal goals of population genetics is to summarize genetic similarities and differences between populations (Wright 1984). This task can be relatively straightforward for some taxa, but the genetic relationship among populations can also be difficult to summarize, especially for species whose evolutionary history is complex and reticulate. Most *Anopheles* mosquitoes have very complex range-wide population structure due to cryptic speciation and ongoing gene flow, which complicate species and population delimitations (Krzywinski & Besansky 2003; Harbach 2013). For example, almost all major African malaria vectors belong to large taxonomic groups encompassing multiple incipient species relatively isolated reproductively and geographically from one another (Reviewed by Sinka *et al.* 2010; Antonio-nkondjio & Simard 2013; Dia *et al.* 2013; Lanzaro & Lee 2013). These characteristics make them promising model systems to study speciation and the processes which contribute to reproductive barriers (e.g. Turner *et al.* 2005; Lawniczak *et al.* 2010; Neafsey *et al.* 2010; Fontaine *et al.* 2015), but can also have far-reaching practical consequences.

When the cryptic genetic structure underlying variations in behavior, distribution and vectorial capacity within mosquito species is not taken into account, vector populations can be mistargeted and major efforts at malaria control undermined (e.g. Molineaux & Gramiccia 1980; Van Bortel *et al.* 2001). The genetic structure of *Anopheles* species is most often based on macrogeographic or regional subdivisions (e.g. Touré *et al.* 1998; Wondji *et al.* 2005; Antonio-Nkondjio *et al.* 2007; Slotman *et al.* 2007; Ayala *et al.* 2011; Ndo *et al.* 2013; Pinto *et al.* 2013; Caputo *et al.* 2014; Kamdem *et al.* 2017), but can also involve more subtle differentiation between larvae and adults, or between adult populations found in or

around human dwellings (e.g. Riehle *et al.* 2011). This type of microgeographic genetic structure can have potentially significant impacts on malaria epidemiology and insecticide resistance because the main vector control measures target exclusively indoor mosquito populations (WHO 2016).

The recent scaling up of insecticide-treated nets usage and indoor insecticide spraying to a lesser extent have led to a dramatic reduction of malaria morbidity across the continent, but have also inevitably increased insecticide resistance among vectors (Hemingway *et al.* 2016; Ranson & Lissenden 2016). Although a direct causal link is not established, populations from urban areas and regions where pesticides are intensely used are strongly differentiated genetically implying a certain correlation between insecticide exposure and local adaptation (Kamdem *et al.* 2017). It has also been shown that this increased usage of xenobiotics can drive selective sweeps across genomes of the two major vectors *An. gambiae* and *An. funestus* (Clarkson *et al.* 2014; Barnes *et al.* 2017; Kamdem *et al.* 2017). The success of current malaria control measures relies among others on accurate knowledge of population structure and the environmental drivers that contribute or have contributed to genetic differentiation between mosquito populations. Extensive sampling combined with recent genomic tools and analytical approaches have helped resolve fine-scale population structure and genomic targets of selection in *An. gambiae* and *An. funestus* (White *et al.* 2011; Reidenbach *et al.* 2012; O'Loughlin *et al.* 2014; Barnes *et al.* 2017; Fouet *et al.* 2017; Kamdem *et al.* 2017), but important taxa remain understudied.

The present work focused on a set of African malaria vectors representing a large taxonomic unit named "*An. nili* group". Species that compose this group are characterized by reticulate evolution and complex phylogenies that have been challenging to resolve

72 (Kengne *et al.* 2003; Awono-Ambene *et al.* 2004, 2006, Ndo *et al.* 2010, 2013; Peery *et al.*
73 2011; Sharakhova *et al.* 2013). To date, four species distinguishable by slight morphological
74 variations are known within the group and occur in forested areas of Central and West
75 Africa: *An. nili sensu stricto* (hereafter *An. nili*), *An. ovengensis*, *An. carnevalei*, and *An.*
76 *somalicus*. Populations of *An. nili* and *An. ovengensis* are very anthropophilic and efficient
77 vectors of *Plasmodium* in rural areas where malaria prevalence is particularly high
78 (Antonio-Nkondjio *et al.* 2006).

79 To delineate genomic patterns of differentiation, we sampled mosquito populations
80 throughout the range of species of the *An. nili* group in Cameroon and used reduced
81 representation sequencing to develop genome-wide SNP markers that we genotyped in 145
82 individuals. We discovered new strongly differentiated subpopulations within *An.*
83 *ovengensis* and *An. nili*. We analyzed the genetic differentiation at 4343 SNPs and revealed
84 the presence of a very high number of outlier loci that are targets of selection. Further
85 studies using a more complete reference genome will help answer important questions
86 concerning the functional and phenotypic characteristics of differentiated loci as well as the
87 contribution of recent selective events.

88 **2. Materials and methods**

89 **(a) Mosquito species**

90 We surveyed 28 locations within the geographic ranges of species of the *An. nili*
91 group previously described in Cameroon (Figure 1) (Awono-Ambene *et al.* 2004, 2006;
92 Antonio-Nkondjio *et al.* 2009; Ndo *et al.* 2010, 2013). We collected larvae and adult
93 mosquitoes in and around human dwellings using several sampling techniques (Service
94 1993) (Table S1). The four currently known members of the *An. nili.* group were identified
95 using morphological keys and a diagnostic PCR, which discriminates species on the basis of
96 point mutations of the ribosomal DNA (Gillies & De Meillon 1968; Gillies & Coetzee 1987;
97 Kengne *et al.* 2003; Awono-Ambene *et al.* 2004).

98

99 **(b) Library preparation, sequencing and SNP discovery**

100 We created double-digest RAD libraries using a modified version of the protocol
101 described by Peterson *et al.* 2012. Genomic DNA of mosquitoes was extracted using the
102 DNeasy Blood and Tissue kit (Qiagen) and the Zymo Research MinPrep kit for larvae and
103 adult samples respectively. Approximately 50 ng (10µl) of DNA of each mosquito was
104 digested simultaneously with *MluC1* and *NlaIII* restriction enzymes. Digested products were
105 ligated to adapter and barcode sequences enabling identification of individuals. Samples
106 were pooled, purified, and 400-bp fragments selected. The resulting libraries were
107 amplified via PCR and purified, and fragment size distribution was checked on BioAnalyzer
108 (Agilent Technologies, Inc., USA). PCR products were quantified, diluted and sequenced on
109 Illumina HiSeq2000 (Illumina Inc., USA).

110

111 **(b) SNP discovery and genotyping**

112 The *process_radtags* program of the Stacks v 1.35 pipeline (Catchen *et al.* 2011,
113 2013) was used to demultiplex and clean raw reads. Reads that passed quality filters were
114 aligned to the *An. nili* Dinderesso draft genome assembly made up of 51048 short contigs
115 (~200-30512bp long) using Gsnap (Wu & Nacu 2010). The *ref_map.pl* program of Stacks
116 was used to call SNPs within consensus RAD loci. We set the minimum number of reads
117 required to form a stack to three and allowed two mismatches during catalogue creation.
118 We generated SNP files in different formats for further downstream analyses using the
119 *populations* program of Stacks and Plink v1.09 (Purcell *et al.* 2007).

120

121 **(d) Population genomics analyses**

122 We analyzed the genetic structure of *An. nili* sensu lato (s.l.) populations using
123 Principal Component Analysis (PCA) and Neighbor-Joining trees (NJ). We also examined
124 ancestry proportions and admixtures between populations in Admixture v1.23 (Alexander
125 *et al.* 2009) and Structure v2.3.4 (Pritchard *et al.* 2000). We used the R package *adegenet*
126 (Jombart 2008) to implement the PCA. Neighbor-Joining trees were generated from SNP
127 allele frequencies via a matrix of Euclidian distance using the R package *ape* (Paradis *et al.*
128 2004). We ran Admixture with 10-fold cross-validation for values of k from 1 through 8. We
129 analyzed patterns of ancestry from k ancestral populations in Structure, testing five
130 replicates of k = 1-8. We used 200000 iterations and discarded the first 50000 iterations as
131 burn-in for each Structure run. Clumpp v1.1.2 (Jakobsson & Rosenberg 2007) was used to
132 summarize assignment results across independent runs. To identify the optimal number of
133 genetic clusters in our sample, we applied simultaneously the lowest cross-validation error

134 in Admixture, the ad-hoc statistic deltaK (Evanno *et al.* 2005; Earl & VonHoldt 2012) and
135 the Discriminant Analysis of Principal Component (DAPC) method implemented in
136 *adegenet.* To examine the extent of genome divergence, we estimated locus-specific genetic
137 differentiation (F_{ST}) using the *populations* program of the Stacks pipeline. Mean F_{ST} values
138 were also used to assess pairwise divergence among populations. To infer the demographic
139 history of populations, we used the diffusion approximation method implemented in the
140 ∂a∂i package (Gutenkunst *et al.* 2009). Single-population models were fitted to allele
141 frequency spectra and the best model was selected with the lowest likelihood and Akaike
142 Information Criterion as well as visual inspections of residuals.
143

144 **3. Results**

145 **(a) SNP genotyping**

146 We collected mosquitoes from four locations out of 28 sampling sites (Figure 1,
147 Table S1) and sequenced 145 individuals belonging, according to morphological
148 identifications and diagnostic PCRs, to two species (*An. nili* (n = 24) and *An. ovengensis* (n =
149 121)). We assembled 197724 RAD loci that mapped to unique positions throughout the
150 reference genome. We applied stringent filtering rules and retained 408 loci present in all
151 populations and in at least 50% of individuals in each population. Within these loci, we
152 identified 4343 high-quality biallelic markers that were used to analyze population
153 structure and genetic differentiation.
154

155 **(b) Morphologically-defined species do not correspond to genetic clusters**

156 PCA and NJ trees show that the genetic variation across 4343 SNPs is best explained
157 by more than two clusters, implying subdivisions within *An. nili* and *An. ovengensis* (Figure
158 2). Three subgroups are apparent within *An. nili* while two distinct clusters segregate in *An.*
159 *ovengensis*. These five subpopulations are associated with the different sampling sites
160 suggesting local adaptation of divergent populations. Importantly, Structure and Admixture
161 analyses reveals that, at k = 2, one population identified by morphology and diagnostic
162 PCRs as *An. nili* has almost the same ancestry pattern as the largest *An. ovengensis* cluster
163 (Figure 3). Such discrepancies between morphology-based and molecular taxonomies can
164 be due to a variety of processes including phenotypic plasticity, introgressive hybrydization
165 or incomplete lineage sorting (i.e., when independent loci have different genealogies by
166 chance) (Arnold 1997; Combosch & Vollmer 2015; Fontaine *et al.* 2015; Weng *et al.* 2016).
167 At k = 2 and k = 3, some populations also depict half ancestry from each morphological
168 species suggestive of gene flow. We found a conflicting number of genetic clusters in our
169 samples reflecting the complex history of subdivisions and admixtures among populations
170 (Figure 4). The Evanno et al. method, which highlights early divergence between *An. nili* and
171 *An. ovengensis* indicates two probable ancestors. DAPC and the Admixture cross-validation
172 error, which are more sensitive to recent hierarchical population subdivisions, show five or
173 more distinct clusters as revealed by PCA and NJ trees (Figure 4).

174 As suggested by the long internal branches connecting subpopulations on NJ trees,
175 there is strong differentiation between and within morphological species characterized by
176 globally high F_{ST} values (Table 1). Relatively lower F_{ST} observed between certain clusters
177 can be due to greater inter-population migration and intermixing or more recent
178 divergence. Overall, both genetic structure and differentiation reveal a set of populations

179 whose ancestries and species status are likely confounded by hybridization and/or
180 incomplete lineage sorting. The current taxonomy based on morphology and ribosomal
181 DNA does not effectively describe the optimal reproductive units among populations of this
182 group of mosquitoes.
183
184 **(c) Genomic signatures of divergent selection and demographic history**
185 Estimates of genetic differentiation varied substantially across 4343 SNP loci (Figure
186 5). The distribution of locus-specific F_{ST} between the five subpopulations revealed a bimodal
187 shape characterized by two peaks around 0 and 1. The great majority of SNPs have low to
188 moderate divergence, but a substantial number of variants are extremely differentiated
189 between populations. The maximum F_{ST} among SNPs is 1 and the number of loci with $F_{ST} = 1$
190 varies from 406 between the populations we termed *An. nili* goup 1 and *An. nili* group 2 to
191 838 between the subgroups called *An. nili* group 3 and *An. ovengensis* group 2. This pattern
192 of genome-wide divergence suggests that a very high number of sites with abrupt
193 differentiation, which likely contain genes that contribute to divergent selection and/or
194 reproductive isolation, coexist with regions of weak divergence that can be freely
195 exchanged between species.
196 Models of population demography indicate that all subgroups have experienced an
197 increase in effective size in more or less recent past (Table 2). Nevertheless, confidence
198 intervals of population parameters are high in some populations, and our results should be
199 interpreted with the necessary precautions. The population growth is less significant in *An.*
200 *nili* group 1.
201

202 **4. Discussion**

203 **(a) Genetic differentiation**

204 Advances in sequencing and analytical approaches have opened new avenues for the
205 study of genomes of disease vectors. We have focused on malaria mosquitoes of the *An. nili*
206 group, whose taxonomy and population structure have been challenging to resolve with
207 low-resolution markers. We analysed genetic structure using genome-wide SNPs and found
208 strong differentiation and local adaption among populations belonging to the two
209 morphological species *An. nili* and *An. ovengensis*. The exact number of subpopulations
210 remains contentious, with the suggested number of divergent clusters varying from two to
211 five within both species. Significant population structure at eight microsatellite loci has
212 been described among *An. nili* populations from Cameroon, with F_{ST} values as high as 0.48
213 between samples from the forest area (Ndo *et al.* 2013). By contrast, *An. ovengensis* was
214 discovered recently and the population structure of this vector remains understudied. This
215 species was initially considered as a sibling of *An. nili* (Kengne *et al.* 2003; Awono-Ambene
216 *et al.* 2004, 2006), but more recent studies have started to challenge the assumed
217 relatedness between the two species due to the high divergence of their polytene
218 chromosomes (Sharakhova *et al.* 2013). Our findings call for careful review of the current
219 taxonomy within this group of species, which is a necessary first step for accurate
220 delineation of the role played by the different subpopulations in malaria transmission.

221 Our samples were collected from locations characterized by more or less degraded
222 forest in the equatorial region of Cameroon. In those habitats, larvae of *An. nili* s.l. exploit
223 relatively similar breeding sites consisting of slow-moving rivers (Antonio-Nkondjio *et al.*
224 2009). Given the similarity of habitats among populations, it is hard to pinpoint the

225 environmental variables that may be correlated with adaptive divergence. One of the most
226 expected outcomes of current large-scale malaria control measures that are underway in
227 Sub-Saharan African countries concerns the effects of increased insecticide exposure on
228 genetic diversity and population structure of vectors. The inferred demographic history of
229 the different subpopulations of the *An. nili* group suggests that there are currently no
230 bottlenecks that can be potentially linked with increased usage of insecticides and
231 insecticide-treated nets (e.g. Athrey *et al.* 2012). Nevertheless, further study is needed to
232 understand whether recent selection associated with human intervention is among the
233 drivers of local adaptation as observed in *An. gambiae* (Kamdem *et al.* 2017).
234
235 **(b) Genomic architecture of reproductive isolation**
236 The genomic architecture of reproductive isolation can reveal crucial information on
237 the mode of speciation and has been the focus of much research over the last few decades
238 (e.g. Turner *et al.* 2005; Harr 2006; Nadeau *et al.* 2012; Ellegren *et al.* 2012; Carneiro *et al.*
239 2014; Burri *et al.* 2015). One influential concept of speciation coined the "genic view of
240 species" proposes that boundaries between species are properties of individual genes or
241 genome regions and not of whole organisms or lineages (Key 1968; Barton & Hewitt 1985;
242 Harrison 1990; Rieseberg *et al.* 1999; Wu 2001; Nosil & Feder 2012; Harrison & Larson
243 2014). We have found that a high number of SNPs are strongly differentiated between
244 populations and often fixed within subgroups. These variants can presumably be
245 assimilated with regions of the genome bearing signals of reproductive isolation. However,
246 recent studies analyzing interactions between pairs of related taxa at their hybrid zone
247 indicate a complex relationship between genetic differentiation and gene flow at the

248 genome level (e.g. Gompert *et al.* 2012; Hamilton *et al.* 2013a; b, Larson *et al.* 2013, 2014;
249 Parchman *et al.* 2013; Taylor *et al.* 2014). Highly divergent genomic regions do not
250 necessarily coincide with regions of reduced gene flow. Several alternative interpretations
251 exist for the numerous high-F_{ST} regions we detected in all pairwise comparisons (Noor &
252 Bennett 2009; Nachman & Payseur 2012; Cruickshank & Hahn 2014; Delmore *et al.* 2015),
253 but careful examination of these outlier of differentiation will reveal significant insights into
254 the markers of ecological divergence and/or reproductive isolation. More complete genome
255 assemblies will be necessary to better delineate specific regions of the genome under
256 natural selection, and thereby clarify the genomic basis of phenotypic fitness differences
257 between divergent populations.

258 Signals consistent with gene flow between *An. nili* and *An. ovengensis* are apparent
259 in our data despite significant time since divergence (~3M-yr) (Ndo *et al.* 2013). Some
260 individuals depict almost half ancestry from each morphological species. The disagreement
261 observed between morphology/PCR and molecular taxonomies also suggests that
262 incongruent genealogies may be widespread along chromosomes due to hybridization
263 and/or incomplete lineage sorting. However, hybridization can be difficult to detect
264 because other factors such as incomplete lineage sorting or technical artefacts can leave
265 signatures that are similar to those of interspecific gene flow (Patterson *et al.* 2012; Liu *et*
266 *al.* 2014). A complete reference genome is also necessary to analyze the detailed
267 distribution of genealogies across genomic windows and to disentangle the relative
268 contribution of processes that generate the putative admixtures and species confusion
269 observed among divergent populations (Fontaine *et al.* 2015; Weng *et al.* 2016).
270

5. Conclusions and implications

Delineating the fine-scale population structure of mosquito populations is crucial to understand their epidemiological significance and their potential response to vector control measures. Moreover, recent malaria control efforts are affecting interspecific gene flow, genetic differentiation, population demography and natural selection in mosquitoes (Athrey *et al.* 2012; Clarkson *et al.* 2014; Norris *et al.* 2015; Barnes *et al.* 2017; Kamdem *et al.* 2017). Deciphering the signatures of all these processes across mosquito genome is important to minimize their negative impacts on vector control. Our findings shed some light on the complex evolutionary history of members of the *An. nili* group and provide a framework for future investigations into the genetic basis of ecological and reproductive barriers.

Acknowledgements

Funding for this project was provided by the University of California Riverside and NIH grants 1R01AI113248 and 1R21AI115271 to BJW. We thank populations and administrative authorities of the sampling sites included in this study for their collaboration.

Data Archiving Statement

Data for this study are available at: to be completed after manuscript is accepted for publication

References

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, **19**, 1655–1664.

Antonio-Nkondjio C, Kerah CH, Simard F *et al.* (2006) Complexity of the malaria vectorial system in Cameroon: contribution of secondary vectors to malaria transmission. *Journal of medical entomology*, **43**, 1215–1221.

Antonio-Nkondjio C, Ndo C, Awono-Ambene P *et al.* (2007) Population genetic structure of the malaria vector Anopheles moucheti in south Cameroon forest region. *Acta Tropica*, **101**, 61–68.

Antonio-Nkondjio C, Ndo C, Costantini C *et al.* (2009) Distribution and larval habitat characterization of Anopheles moucheti, Anopheles nili, and other malaria vectors in

305 river networks of southern Cameroon. *Acta Tropica*, **112**, 270–276.
306 Antonio-nkondjio C, Simard F (2013) Highlights on Anopheles nili and Anopheles moucheti,
307 Malaria Vectors in Africa. In: *Anopheles mosquitoes - New insights into malaria vectors*
308 (ed Manguin S)
309 Arnold ML (1997) *Natural Hybridization and Evolution*.
310 Athrey G, Hodges TK, Reddy MR *et al.* (2012) The effective population size of malaria
311 mosquitoes: large impact of vector control. *PLoS genetics*, **8**, e1003097.
312 Awono-Ambene HP, Kengne P, Simard F, Antonio-Nkondjio C, Fontenille D (2004)
313 Description and bionomics of Anopheles (Cellia) ovengensis (Diptera: Culicidae), a new
314 malaria vector species of the Anopheles nili group from south Cameroon. *Journal of*
315 *medical entomology*, **41**, 561–568.
316 Awono-Ambene HP, Simard F, Antonio-Nkondjio C *et al.* (2006) Multilocus enzyme
317 electrophoresis supports speciation within the Anopheles nili group of malaria vectors
318 in Cameroon. *The American journal of tropical medicine and hygiene*, **75**, 656–658.
319 Ayala D, Fontaine MC, Cohuet A *et al.* (2011) Chromosomal inversions, natural selection and
320 adaptation in the malaria vector Anopheles funestus. *Molecular biology and evolution*,
321 **28**, 745–758.
322 Barnes KG, Weedall GD, Ndula M *et al.* (2017) Genomic Footprints of Selective Sweeps from
323 Metabolic Resistance to Pyrethroids in African Malaria Vectors Are Driven by Scale up
324 of Insecticide-Based Vector Control. *PLoS Genetics*, **13(2): e10**, 1–22.
325 Barton N, Hewitt G. (1985) Analysis of hybrid zones. *Annual Review of Ecology and*
326 *Systematics*, **16**, 113–148.
327 Van Bortel W, Harbach RE, Trung HD *et al.* (2001) Confirmation of Anopheles varuna in
328 vietnam, previously misidentified and mistargeted as the malaria vector Anopheles
329 minimus. , **65**, 729–732.
330 Burri R, Nater A, Kawakami T *et al.* (2015) Linked selection and recombination rate
331 variation drive the evolution of the genomic landscape of differentiation across the
332 speciation continuum of Ficedula flycatchers. *Genome Research*, **25**, 1656–1665.
333 Caputo B, Nwakanma D, Caputo FP *et al.* (2014) Prominent intraspecific genetic divergence
334 within Anopheles gambiae sibling species triggered by habitat discontinuities across a
335 riverine landscape. *Molecular ecology*, **23**, 4574–89.
336 Carneiro M, Albert FW, Afonso S *et al.* (2014) The genomic architecture of population
337 divergence between subspecies of the European Rabbit. *PLoS Genetics*, **10**.
338 Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and
339 genotyping Loci de novo from short-read sequences. *G3 (Bethesda, Md.)*, **1**, 171–82.
340 Catchen J, Hohenlohe P a., Bassham S, Amores A, Cresko W. (2013) Stacks: An analysis tool
341 set for population genomics. *Molecular Ecology*, **22**, 3124–3140.
342 Clarkson CS, Weetman D, Essandoh J *et al.* (2014) Adaptive introgression between
343 Anopheles sibling species eliminates a major genomic island but not reproductive
344 isolation. *Nature communications*, **5**, 4248.
345 Combosch DJ, Vollmer S V. (2015) Trans-Pacific RAD-Seq population genomics confirms
346 introgressive hybridization in Eastern Pacific Pocillopora corals. *Molecular*
347 *Phylogenetics and Evolution*, **88**, 154–162.
348 Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation
349 are due to reduced diversity, not reduced gene flow. *Molecular ecology*, **23**, 3133–57.
350 Delmore KE, Hübner S, Kane NC *et al.* (2015) Genomic analysis of a migratory divide reveals

351 candidate genes for migration and implicates selective sweeps in generating islands of
352 differentiation. *Molecular Ecology*, 1873–1888.
353 Dia I, Guelbeogo M, Ayala D (2013) Advances and Perspectives in the Study of the Malaria
354 Mosquito Anopheles funestus. In: *Anopheles mosquitoes - New insights into malaria*
355 *vectors* (ed Manguin S)
356 Earl DA, VonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for
357 visualizing STRUCTURE output and implementing the Evanno method. *Conservation*
358 *Genetics Resources*, **4**.
359 Ellegren H, Smeds L, Burri R *et al.* (2012) The genomic landscape of species divergence in
360 Ficedula flycatchers. *Nature*, **491**, 756–60.
361 Evanno G, Goudet J, Regnaut S (2005) Detecting the number of clusters of individuals using
362 the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.
363 Fontaine MC, Pease JB, Steele a. *et al.* (2015) Extensive introgression in a malaria vector
364 species complex revealed by phylogenomics. *Science*, **347**, 1258524–1258524.
365 Fouet C, Kamdem C, Gamez S, White BJ (2017) Extensive genetic diversity among
366 populations of the malaria mosquito Anopheles moucheti revealed by population
367 genomics. *Infection, Genetics and Evolution*, **48**, 27–33.
368 Gillies MT, Coetzee M (1987) *A supplement to the Anophelinae of Africa south of the Sahara*.
369 The South African Institute for Medical Research, Johannesburg.
370 Gillies MT, De Meillon B (1968) *The Anophelinae of Africa South of the Sahara*. Publications
371 of the South African Institute for Medical Research, Johannesburg.
372 Gompert Z, Lucas LK, Nice CC *et al.* (2012) Genomic regions with a history of divergent
373 selection affect fitness of hybrids between two butterfly species. *Evolution*, **66**, 2167–
374 2181.
375 Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD (2009) Inferring the joint
376 demographic history of multiple populations from multidimensional SNP frequency
377 data. *PLoS genetics*, **5**, e1000695.
378 Hamilton JA, Lexer C, Aitken SN (2013a) Genomic and phenotypic architecture of a spruce
379 hybrid zone (Picea sitchensis × P. glauca). *Molecular Ecology*, **22**, 827–841.
380 Hamilton J a., Lexer C, Aitken SN (2013b) Differential introgression reveals candidate genes
381 for selection across a spruce (Picea sitchensis × P. glauca) hybrid zone. *New Phytologist*,
382 **197**, 927–938.
383 Harbach RE (2013) The Phylogeny and Classification of Anopheles. In: *Anopheles*
384 *mosquitoes - New insights into malaria vectors* (ed Manguin S)
385 Harr B (2006) Genomic islands of differentiation between house mouse subspecies. , 730–
386 737.
387 Harrison RG (1990) Hybrid zones: windows on evolutionary process. In: (eds Futuyma D, J.
388 Antonovics), pp. 69–128. New York.
389 Harrison RG, Larson EL (2014) Hybridization, introgression, and the nature of species
390 boundaries. *Journal of Heredity*, **105**, 795–809.
391 Hemingway J, Ranson H, Magill A *et al.* (2016) Averting a malaria disaster: Will insecticide
392 resistance derail malaria control? *The Lancet*, **387**, 1785–1788.
393 Jakobsson M, Rosenberg N (2007) CLUMPP: a cluster matching and permutation program
394 for dealing with multimodality in analysis of population structure. *Bioinformatics*, **23**,
395 1801– 1806.
396 Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers.

397 *Bioinformatics*, **24**, 1403–1405.
398 Kamdem C, Fouet C, Gamez S, White BJ (2017) Pollutants and insecticides drive local
399 adaptation in African malaria mosquitoes. *Molecular Biology and Evolution*, **Accepted**.
400 Kengne P, Awono-Ambene, H. P. Antonio-Nkondjio C, Simard F, Fontenille D (2003)
401 Molecular identification of the Anopheles nili group African malaria vectors. *Med Vet*
402 *Entomol*, **17**, 67–74.
403 Key KHL (1968) The concept of stasipatric speciation. *Syst. Zool.*, **17**, 14–22.
404 Krzywinski J, Besansky NJ (2003) Molecular systematics of Anopheles : from subgenera to
405 subpopulations. *Annual Review of Entomology*, **48**, 111–139.
406 Lanzaro GC, Lee Y (2013) Speciation in Anopheles gambiae — The Distribution of Genetic
407 Polymorphism and Patterns of Reproductive Isolation Among Natural Populations. In:
408 *Anopheles mosquitoes - New insights into malaria vectors* (ed Manguin S)
409 Larson EL, Andrés JA, Bogdanowicz SM, Harrison RG (2013) Differential introgression in a
410 mosaic hybrid zone reveals candidate barrier genes. *Evolution*, **67**, 3653–3661.
411 Larson EL, White TA, Ross CL, Harrison RG (2014) Gene flow and the maintenance of
412 species boundaries. *Molecular Ecology*, **23**, 1668–1678.
413 Lawniczak MKN, Emrich SJ, Holloway a K *et al.* (2010) Widespread divergence between
414 incipient Anopheles gambiae species revealed by whole genome sequences. *Science*
415 *(New York, N.Y.)*, **330**, 512–4.
416 Liu KJ, Dai J, Truong K *et al.* (2014) An HMM-Based Comparative Genomic Framework for
417 Detecting Introgression in Eukaryotes. *PLoS Computational Biology*, **10**.
418 Molineaux L, Gramiccia G (1980) *The Garki Project. Research on the epidemiology and*
419 *control of malaria in the sudan savanna of West Africa*. World Health Organization.
420 Nachman MW, Payseur BA (2012) Recombination rate variation and speciation : theoretical
421 predictions and empirical results from rabbits and mice. , 409–421.
422 Nadeau NJ, Whibley A, Jones RT *et al.* (2012) Genomic islands of divergence in hybridizing
423 Heliconius butterflies identified by large-scale targeted sequencing. , 343–353.
424 Ndo C, Antonio-Nkondjio C, Cohuet A *et al.* (2010) Population genetic structure of the
425 malaria vector Anopheles nili in sub-Saharan Africa. *Malaria journal*, **9**, 161.
426 Ndo C, Simard F, Kengne P *et al.* (2013) Cryptic Genetic Diversity within the Anopheles nili
427 group of Malaria Vectors in the Equatorial Forest Area of Cameroon (Central Africa).
428 *PLoS ONE*, **8**, 1–12.
429 Neafsey DE, Lawniczak MKN, Park DJ *et al.* (2010) SNP genotyping defines complex gene-
430 flow boundaries among African malaria vector mosquitoes. *Science (New York, N.Y.)*,
431 **330**, 514–7.
432 Noor M a F, Bennett SM (2009) Islands of speciation or mirages in the desert? Examining
433 the role of restricted recombination in maintaining species. *Heredity*, **103**, 439–444.
434 Norris LC, Main BJ, Lee Y *et al.* (2015) Adaptive introgression in an African malaria
435 mosquito coincident with the increased usage of insecticide-treated bed nets.
436 *Proceedings of the National Academy of Sciences*, 201418892.
437 Nosil P, Feder JL (2012) Widespread yet heterogeneous genomic divergence. *Molecular*
438 *Ecology*, **21**, 2829–2832.
439 O'Loughlin SM, Magesa S, Mbogo C *et al.* (2014) Genomic Analyses of Three Malaria Vectors
440 Reveals Extensive Shared Polymorphism but Contrasting Population Histories.
441 *Molecular biology and evolution*, 1–14.
442 Paradis E, Claude J, Strimmer K (2004) Analyses of Phylogenetics and Evolution in R

443 language. *Bioinformatics*, **20**, 289–290.
444 Parchman TL, Gompert Z, Braun MJ *et al.* (2013) The genomic consequences of adaptive
445 divergence and reproductive isolation between species of manakins. *Molecular ecology*,
446 **22**, 3304–17.
447 Patterson N, Moorjani P, Luo Y *et al.* (2012) Ancient admixture in human history. *Genetics*,
448 **192**, 1065–1093.
449 Peery A, Sharakhova M V, Antonio-Nkondjio C *et al.* (2011) Improving the population
450 genetics toolbox for the study of the African malaria vector Anopheles nili:
451 microsatellite mapping to chromosomes. *Parasites & Vectors*, **4**, 202.
452 Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: An
453 Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-
454 Model Species. *PLoS ONE*, **7**, e37135.
455 Pinto J, Egyir-Yawson A, Vicente J *et al.* (2013) Geographic population structure of the
456 African malaria vector Anopheles gambiae suggests a role for the forest-savannah
457 biome transition as a barrier to gene flow. *Evolutionary applications*, **6**, 910–24.
458 Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using
459 multilocus genotype data. *Genetics*, **155**, 945–959.
460 Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: a toolset for whole-genome
461 association and population-based linkage analysis. *American Journal of Human*
462 *Genetics*, **81**.
463 Ranson H, Lissenden N (2016) Insecticide Resistance in African Anopheles Mosquitoes: A
464 Worsening Situation that Needs Urgent Action to Maintain Malaria Control. *Trends in*
465 *Parasitology*, **32**, 187–196.
466 Reidenbach KR, Neafsey DE, Costantini C *et al.* (2012) Patterns of genomic differentiation
467 between ecologically differentiated M and S forms of Anopheles gambiae in West and
468 Central Africa. *Genome biology and evolution*, **4**, 1202–12.
469 Riehle MM, Guelbeogo WM, Gneme A *et al.* (2011) A cryptic subgroup of Anopheles gambiae
470 is highly susceptible to human malaria parasites. *Science (New York, N.Y.)*, **331**, 596–8.
471 Rieseberg LH, Whitton J, Gardner K (1999) Hybrid Zones and the Genetic Architecture of a
472 Barrier to Gene Flow Between Two Sunflower Species.
473 Service MW (1993) *Mosquito ecology: field sampling methods* (UK Elsevier Applied Science,
474 London, Ed,).
475 Sharakhova M V., Peery A, Antonio-Nkondjio C *et al.* (2013) Cytogenetic analysis of
476 Anopheles ovengensis revealed high structural divergence of chromosomes in the
477 Anopheles nili group. *Infection, Genetics and Evolution*, **16**, 341–348.
478 Sinka ME, Bangs MJ, Manguin S *et al.* (2010) The dominant Anopheles vectors of human
479 malaria in Africa, Europe and the Middle East: occurrence data, distribution maps and
480 bionomic précis. *Parasites & vectors*, **3**, 117.
481 Slotman M a, Tripet F, Cornel a J *et al.* (2007) Evidence for subdivision within the M
482 molecular form of Anopheles gambiae. *Molecular ecology*, **16**, 639–49.
483 Taylor SA, Curry RL, White TA, Ferretti V, Lovette I (2014) Spatiotemporally consistent
484 genomic signatures of reproductive isolation in a moving hybrid zone. *Evolution*, **68**,
485 3066–3081.
486 Touré Y, Petrarca V, Traoré S *et al.* (1998) The distribution and inversion polymorphism of
487 chromosomally recognized taxa of the Anopheles gambiae complex in Mali, West
488 Africa. *Parassitologia*, **40**, 477–511.

489 Turner TL, Hahn MW, Nuzhdin S V. (2005) Genomic islands of speciation in Anopheles
490 gambiae. *PLoS Biology*, **3**, 1572–1578.
491 Weng D, Yu Y, Hahn MW, Nakhleh L (2016) Reticulate evolutionary history and extensive
492 introgression in mosquito species revealed by phylogenetic network analysis.
493 *Molecular Ecology*, **Special is**.
494 White B, Lawniczak M, Cheng C *et al.* (2011) Adaptive divergence between incipient species
495 of Anopheles gambiae increases resistance to Plasmodium. *Proceedings of the National*
496 *Academy of Sciences*, **108**, 244–249.
497 WHO (2016) *World malaria report 2016*.
498 Wondji C, Simard F, Petrarca V (2005) Species and populations of the Anopheles gambiae
499 complex in Cameroon with special emphasis on chromosomal and molecular forms of
500 Anopheles gambiae ss. *Journal of medical entomology*.
501 Wright S (1984) *Evolution and the Genetics of Populations: Genetics and Biometric*
502 *Foundations v. 1 (Genetic & Biometric Foundations); New Edition*. University of Chicago
503 press.
504 Wu C (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**,
505 851–865.
506 Wu TD, Nacu S (2010) Fast and SNP-tolerant detection of complex variants and splicing in
507 short reads. *Bioinformatics*, **26**, 873–881.
508
509 **Author contributions**
510 Conceived and designed the experiments: CK CF BJW. Performed the experiments: CK CF SG
511 BJW. Analyzed the data: CK CF. Wrote the paper: CK CF BJW.
512
513
514

515 **Tables**

516

517 **Table 1**: Pairwise F_{ST} between divergent subpopulations.

F_{ST}	*An. nili* group 1	*An. nili* group 2	*An. nili* group 3	*An. ovengensis* group 1	*An. ovengensis* group 2
***An. nili* group 1**	-				
***An. nili* group 2**	0.374	-			
***An. nili* group 3**	0.506	0.552	-		
***An. ovengensis* group 1**	0.135	0.275	0.364	-	
***An. ovengensis* group 2**	0.432	0.458	0.492	0.349	-

518

519 **Table 2:** Demographic models of different subgroups.

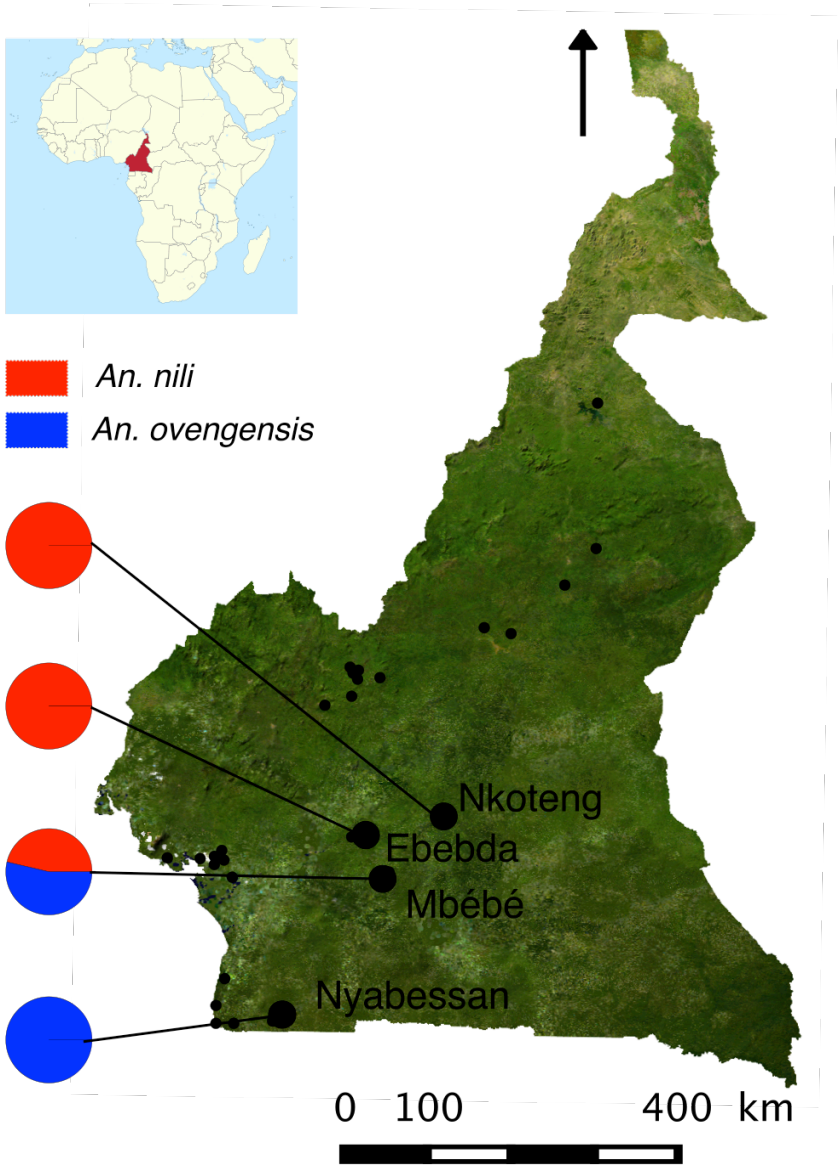
Population	Best Model	Log Likelihood	Final Population Size[a] (95% CI)	Time[b] (95% CI)
An. nili group 1	*Growth*	-18.42	6.41 (5.326 - 20.71)	3.70 (1.11 - 13.31)
An. nili group 2	*Two-epoch*	-19.97	17.87 (9.33 - 35.50)	11.27 (4.93 - 19.64)
An. ovengensis group 1	*Growth*	-112.18	13.04 (12.15 - 17.26)	0.70 (0.58 - 1.08)
An. Ovengensis group 2	*Growth*	-22.98	19.95 (14.45 - 45.70)	5.11 (2.33 - 15.13)

[a] Relative to ancestral population size.

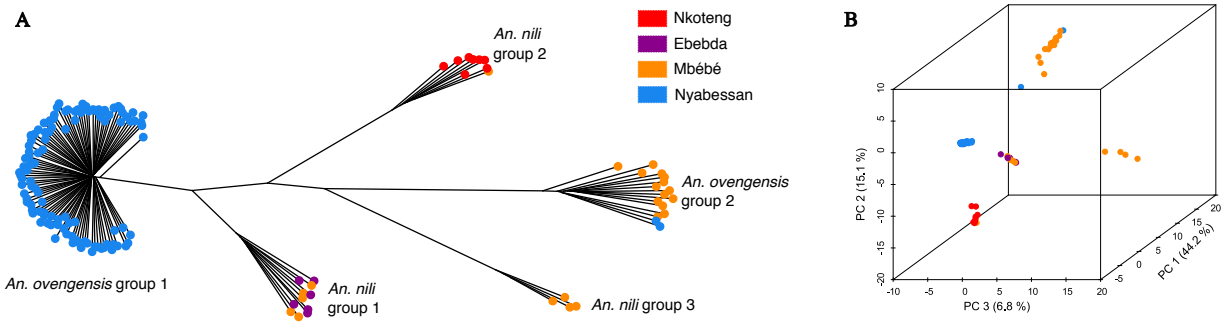
[b] Expressed in units 2Ne generations from start of growth to present.

Figures

Figure 1: Map showing the sampling locations and the relative frequencies of *An. nili* and *An. ovengensis*. Small and large black dots indicate respectively the 28 locations surveyed and the four sampling sites where mosquitoes were collected.

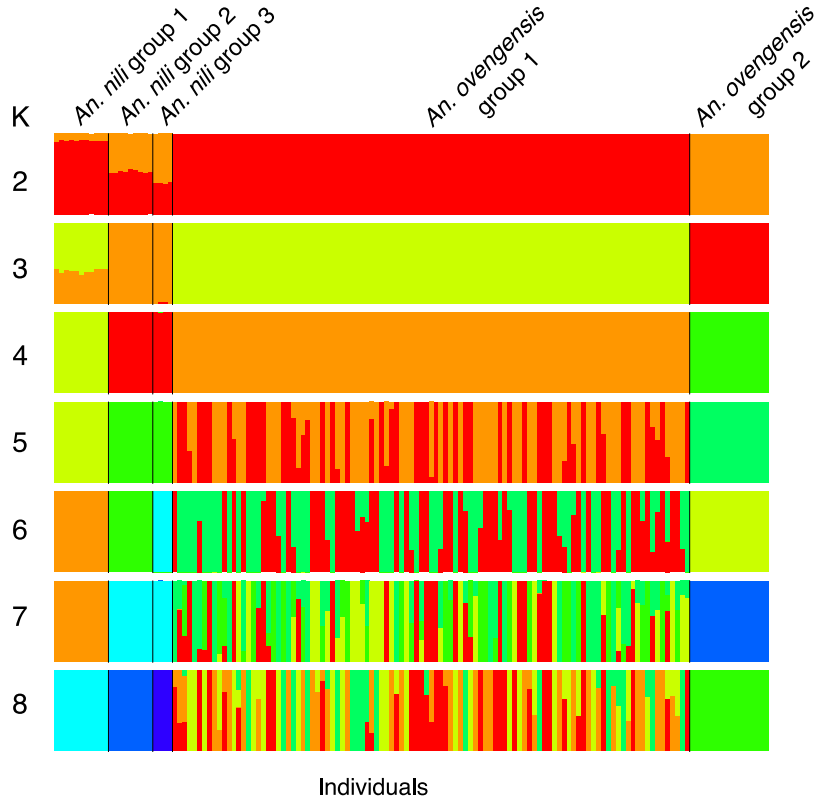


527 **Figure 2**: Population genetic structure inferred from 4343 SNPs using PCA (A) and a
528 neighbor-joining tree (B). The percentage of variance explained is indicated on each PCA
529 axis.
530



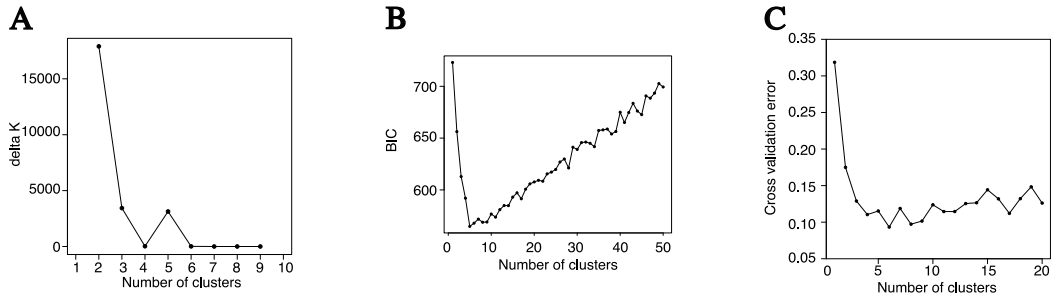
531
532
533

534 **Figure 3:** Ancestry proportions inferred in ADMIXTURE with k = 2 – 8.
535

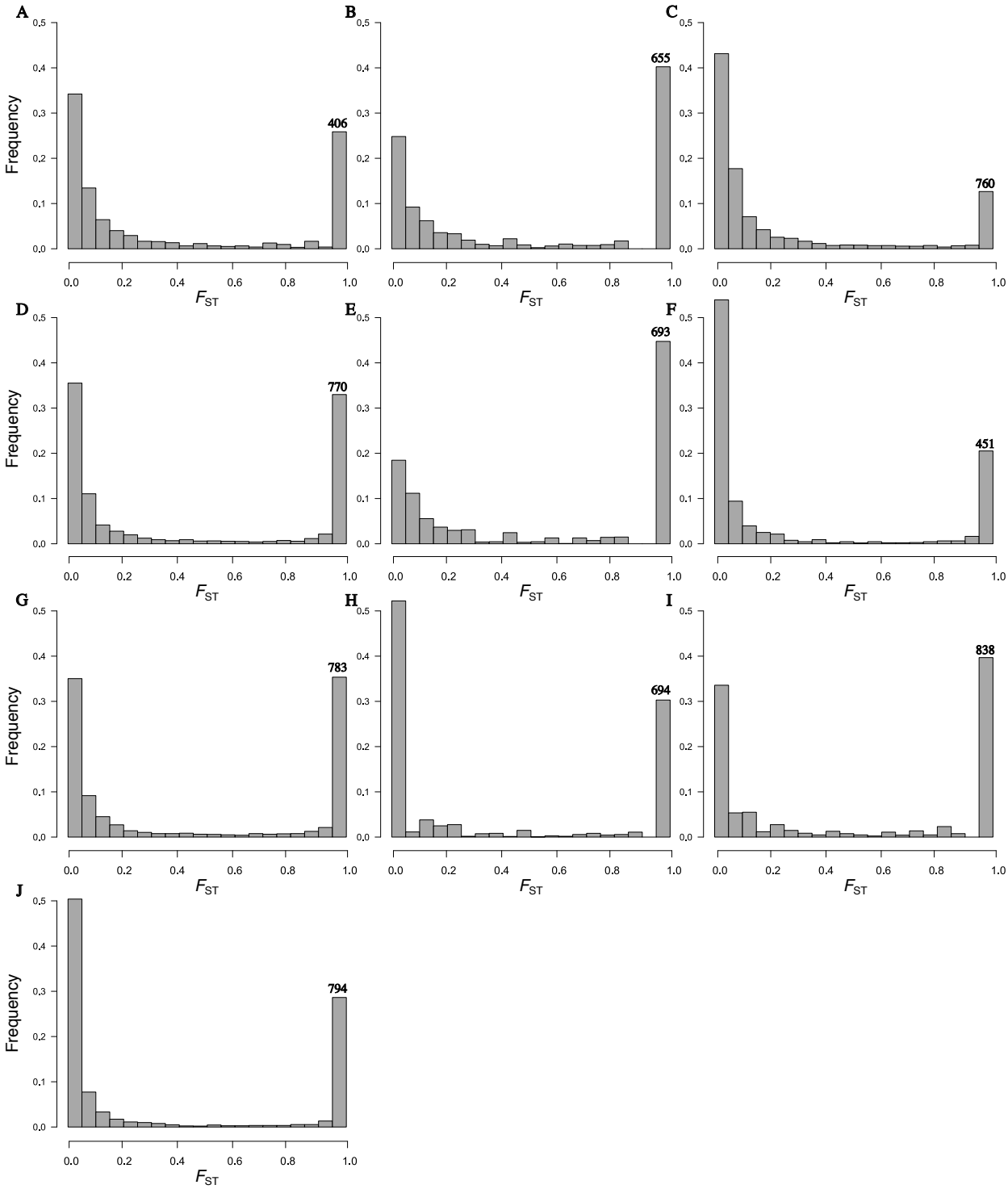


536
537
538

Figure 4: Identification of the optimal number of genetic clusters using the delta k method of Evanno et al (A), DAPC (B) and 10-fold cross-validation in ADMIXTURE (C). The lowest BIC and CV error and the highest delta k indicate the most probable number of clusters.



547 **Figure 5:** Distribution of F_{ST} values across 4343 SNPs between *An. nili* group 1 and *An. nili*
548 group 2 (A); *An. nili* group 1 and *An. nili* group 3 (B); *An. nili* group 1 and *An. ovengensis*
549 group 1 (C); *An. nili* group 1 and *An. ovengensis* group 2 (D); *An. nili* group 2 and *An. nili*
550 group 3 (E); *An. nili* group 2 and *An. ovengensis* group 1 (F); *An. nili* group 2 and *An.*
551 *ovengensis* group 2 (G); *An. nili* group 3 and *An. ovengensis* group 1 (H); *An. nili* group 3 and
552 *An. ovengensis* group 2 (I); *An. ovengensis* group 1 and *An. ovengensis* group 2 (J). The
553 number of SNPs with F_{ST} = 1 is indicated in each pairwise comparison.
554


A
B
C
D
E
F
G
H
I
J
Frequency
F_{ST}
406
655
760
770
693
451
783
694
838
794

555
556

557 **Supplemental information**

558

559 **Table S1:** Information on mosquito samples included in this study.

Sampling locations	Geographic coordinates	Sampling methods			Total
		HLC-OUT	HLC-IN	LC	
Ebebda	4°20'00"N, 11°17'00"E			6	6
Nkoteng	4°31'00"N, 12°02'00"E			8	8
Nyabessan	2°24'00"N, 10°24'00"E	63	44		107
Mbébé	4°10'00"N, 11°04'00"E	13	3	8	24
Total		76	47	22	145

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

560