bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Extensive genetic diversity among populations of the malaria mosquito *Anopheles moucheti* revealed by population genomics

Caroline Fouet[1]*, Colince Kamdem[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: caroline.fouet@ucr.edu; bwhite@ucr.edu

Key words: *Anopheles moucheti*, population genomics, RADseq, *de novo* assembly

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

11 **Abstract**

12 **Background**: Recent intensive efforts to control malaria in African countries expose
13 vector populations to additional adaptive challenges. The main malaria mosquitoes
14 of the continent display an array of adaptive strategies to cope with such challenges.
15 The development of genomic resources will empower genetic studies that are
16 crucial to understand the evolutionary history and adaptive potential of these
17 vectors.

18 **Methodology/Principal findings**: Here we constructed double-digest Restriction
19 Associated DNA (ddRAD) libraries and generated 6461 Single Nucleotide
20 Polymorphisms (SNPs) that we used to explore the population structure and
21 demographic history of wild-caught *Anopheles moucheti* from Cameroon. The
22 genome-wide distribution of allelic frequencies among sampled populations best
23 fitted that of an old population at equilibrium, characterized by a weak genetic
24 structure and extensive genetic diversity, presumably due to a large long term
25 effective population size. In contrast to other important African malaria vectors,
26 polymorphic chromosomal inversions play little role in the genome architecture and
27 evolutionary adaptation of *An. moucheti*.

28 **Conclusions/Significance**: Our study provides the first investigation of the genetic
29 structure and diversity in *An. moucheti* at the genomic scale. Despite a weak genetic
30 structure and absence of adaptive divergence, the adaptive potential of this
31 mosquito remains significant owing to a great diversity and standing genetic
32 variation that can be used to face current vector control measures and other rapid
33 anthropogenic and environmental changes.

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Background

An. moucheti sensu lato is a group of three related mosquito species (*An. moucheti moucheti*, *An. moucheti nigeriensis*, and *An. moucheti bervoetsi*) distributed across forested areas in Sub-Saharan Africa and distinguishable from each other by slight morphological differences [1]. The nominal species *An. moucheti moucheti* (hereafter *An. moucheti*) is the most widespread and one of the leading malaria vectors in the equatorial region [2–4]. The species breeds in slow moving streams and rivers close to villages where it contributes to high malaria transmission and often outcompete other main malaria vectors species. Despite this epidemiological significance, the evolutionary history and the evolvability of this mosquito remain understudied. Early investigations of the genetic structure based on allozymes and microsatellites detected no significant differentiation among *An. moucheti* populations over substantial geographic areas. This finding was particularly intriguing given the widespread distribution of *An. moucheti* in the evergreen forest areas of Central Africa. Moreover, African anopheline mosquito populations are increasingly exposed to intense selective pressure associated with the use of insecticides and insecticide-impregnated materials associated with recent expansive malaria control campaigns [5]. Such pressures are strong driving forces that often contribute to rapid diversification of vector populations at the scales of a few decades [6–8]. As a result, a more detailed characterization of the genomic architecture of *An. moucheti* will be crucial to: (1) understand the genetic bases allowing populations of this species to adapt to such a wide geographic area and (2)

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

predict the evolutionary potential of this vector in face of selective constraints imposed by rapidly changing environments.

Thanks to recent progresses in sequencing technology, high-resolution sequence information can be generated for virtually any living organism. However, while significant progresses have been made in genomic studies of other African anopheline mosquitoes [9], *An. moucheti* has lagged behind mainly because of the absence of genomic resources. So far, this species crucially lacks a reference genome assembly, a physical or linkage map and laboratory strains that are essential to generate high-quality sequencing information and to enable robust interpretations of natural polymorphisms.

To start filling this gap, we have conducted the first screening of genetic variation across the genome of wild populations of *An. moucheti*. We have performed a high-throughput sequencing of reduced representation libraries in 87 wild-caught individuals from Cameroon and used a *de novo* assembly to identify thousands of RAD loci scattered throughout the genome. Using high-quality Single Nucleotide Polymorphisms (SNPs) identified within these loci, we have investigated the genetic relatedness and population history of our samples. We found that populations of *An. moucheti* are characterized by extensive gene flow and a great genetic diversity. This vector appears particularly adapted to challenge the selective pressures imposed by vector controls and rapid human-induced environmental modifications.

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Methods

Mosquito sampling and sequencing

This study included two *An. moucheti* populations from the Cameroonian equatorial rainforest. A total of 98 mosquitoes (97 adults and 1 larva) were collected in August and November 2013 Olama and Nyabessan, respectively (Table 1). The two locations are separated by ~200 km (Fig. 1A) and are crossed respectively by the Nyong and the Ntem rivers that provide the breeding sites for *An. moucheti* larvae. Specimens were identified as *An. moucheti moucheti* using morphological identification keys and a diagnostic PCR [1,10,11]. We extracted genomic DNA of specimens using either the DNeasy Blood and Tissue kit (Qiagen) or the Zymo Research MinPrep kit. We used 10ul (~50ng) of genomic DNA to prepare double-digest Restriction-site Associated DNA libraries following a modified protocol of Peterson et al.,2012 [12]. *MluC1* and *NlaIII* restriction enzymes were used to digest DNA of individual mosquitoes yielding RAD-tags of different sizes to which short unique DNA sequences (barcodes) were ligated to enable the identification of reads belonging to each specimen. The digestion products were purified and pooled. DNA fragments with size ~400bp were selected and amplified via PCR. The distribution of fragment sizes was checked on a BioAnalyzer (Agilent Technologies, Inc., USA) before sequencing. The sequencing was performed on an Illumina HiSeq2000 platform (Illumina Inc., USA) (Genomic Core Facility, University of California, Riverside) to yield single-end reads of 101 bp.

De novo assembly and SNP discovery

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

We used the bioinformatics pipeline Stacks v1.35 [13] to process Illumina short reads. The program *process_radtags* was first used to sort the reads according to the barcodes and to trim all reads to 96-bp in length by removing index and barcode sequences from the ends of the reads. Reads with ambiguous barcodes, those that did not contain the *NlaIII* recognition site and those with low-quality scores (average Phred score < 33) were excluded. The program *ustacks* was then utilized for a *de novo* assembly enabling the identification of consensus RAD loci in each individual in our populations. We allowed a maximum of 2 nucleotide mismatches between stacks (M parameter in *ustacks*) and we required a minimum of three reads to create a stack (m parameter in *ustacks*). Using the *cstacks* program, a catalogue of loci was built to synchronize variations across all individuals in our populations. Finally, we utilized the *populations* program to calculate population genetic parameters and output SNPs in different formats. To avoid bias associated with less informative SNPs or possible false positive SNPs (due to sequencing or pipeline errors), only RAD loci scored in at least 70-80% of individuals were retained for further analyses.

Population genomic analyses

SNP files outputted by the *populations* program were used to assess the population genetic structure with a Principal Component Analysis (PCA) and a Neighbor-joining tree analysis using respectively the R packages *adegenet* and *ape* [14,15]. We also explored patterns of ancestry and admixture in *An. moucheti* individuals in ADMIXTURE v1.23 [16] with 10-fold cross-validation for k assumed ancestral populations (k= 1 through 6). The optimal number of clusters was confirmed using

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

the Discriminant Analysis of Principal Component (DAPC) method which explores the number of genetically distinct groups by running a k-means clustering sequentially with increasing numbers and by comparing different clustering solutions using Bayesian Information Criterion (BIC) [14]. We examined the population genetic diversity, conformity to Hardy-Weinberg equilibrium and demographic background using several statistics calculated with the *populations* program. Precisely, to assess the global genetic diversity per population, we calculated the overall nucleotide diversity (π) and the frequency of polymorphic sites within population. To make inferences on the demographic history and to test for departures from Hardy–Weinberg equilibrium, we utilized the Allele frequency spectrum and the Wright's inbreeding coefficient (F_{IS}). To quantify the geographic and genetic differentiation between allopatric populations, we estimated the genome-wide average F_{ST} [17] on 2000 randomly selected SNPs in Genodive v1.06 [18]. We also conducted an hierarchical Analysis of Molecular Variance (AMOVA) [19] on the same SNP set to quantify the effects of the geographic origin on the genetic variance among individuals. The statistical significance of F_{ST} and AMOVA was assessed with 10000 permutations. Finally, to examine the genomic architecture divergence, we inspected the genome-wide distribution of locus-specific estimates of F_{ST}.

Identification of polymorphic chromosomal inversions

The neutral recombination rate is notoriously reduced in genomic regions bearing chromosomal inversions, which results in elevated linkage disequilibrium (LD) in those regions relative to the rest of the genome. Thus, assessing genome-wide

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

147 patterns of LD can reveal clusters of highly correlated SNPs (LD blocks)
148 corresponding potentially to chromosomal inversions. The R package LDna [20]
149 allow the examination of the distinct LD network clusters within the genome of
150 nonmodel species without the need of a linkage map or reference genome. We
151 calculated LD, estimated as the r^2 correlation coefficient between all pairs of SNPs, in
152 PLINK v1.09 [21]. To avoid spurious LD due to the strong correlation between SNPs
153 located on the same RAD locus, we randomly selected only one SNP within each RAD
154 locus resulting in a dataset of 1056 variants containing less than 10% missing data.
155 LDna was then used to identify clusters of highly correlated SNPs that were
156 interpreted in light of the population demographic history, the population genetic
157 structure and previous cytological studies of *An. moucheti*.
158

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

159 **Results**

160 **De novo assembly**

161 In total, 518,218 unique 96-bp RAD loci were identified from *de novo* assembly of
162 reads in 98 individuals. We retained 946 loci that were present in all sampled
163 populations and in at least 75% of individuals in every population, and we identified
164 3027 high-quality biallelic SNPs from these loci.

165 **Population genetic structure**

166 First, we tested for the presence of cryptic genetic subdivision within *An. moucheti*
167 with PCA, NJ trees and the ADMIXTURE ancestry model. A NJ tree constructed from
168 a matrix of Euclidian distance using allele frequencies at 3027 genome-wide SNPs
169 showed a putative subdivision of *An. moucheti* populations in two genetic clusters
170 (Fig. S1). The first three axes of PCA also revealed a number of outlier individuals
171 separated from a main cluster (Fig. S1). However, when we ranked our sequenced
172 individuals based on the number of sequencing reads, we noticed that one of the
173 putative genetic clusters corresponded to a group of individuals having the lowest
174 sequencing coverage. We excluded all these individuals and reduced our dataset to
175 78 individuals. We conducted a *de novo* assembly and analyzed the relationship
176 between the 78 remaining individuals at 6461 genome-wide SNPs using PCA, NJ
177 trees and ADMIXTURE. Both the k-means clustering (DAPC) and the variation of the
178 cross-validation error as a function of the number of ancestral populations in
179 ADMIXTURE revealed that the polymorphism of *An. moucheti* resulted from only one
180 ancestral population (k = 1) (Fig. 1E,F). PCA and NJ depicted a homogeneous cluster
181 comprising all 78 individuals providing additional evidence of the lack of genetic or

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

geographic structuring among populations. Unsurprisingly, the overall F_{ST} was remarkably low between populations from the two sampling locations Olama and Nyabessan (F_{ST} = 0.008, p < 0.005). Similarly, the distribution of F_{ST} values across the 6461 SNPs showed a large dominance of very low F_{ST} values (Fig. 2). The highest per locus F_{ST} was only 0.126, while 5006 of the 6461 loci revealed F_{ST} near zero. The modest geographic differentiation was also well illustrated by a hierarchical AMOVA, which showed that the genetic variance was explained essentially by within-individual variations (99.7%). Finally, we found very low overall Wright's inbreeding coefficient (F_{IS}= 0.0014 in Nyabessan and F_{IS} = 0.0025 in Olama) (Table 2) suggesting that allelic frequencies within both populations were in accordance with proportions expected under the Hardy-Weinberg equilibrium.

Genetic diversity and demographic history

The estimates of the overall nucleotide diversity (π = 0.0020 and π = 0.0016, respectively, in Olama and Nyabessan) (Table 2) were within the range of average values found in other African *Anopheles* species using RAD-seq approaches [8,22–24]. Notorious demographic expansions have been described in natural populations of this insect clade [25], and the values of π observed in *An. moucheti* likely reflect the level genetic diversity of a population with large effective size. The great genetic diversity of *An. moucheti* was also illustrated by the percentage of polymorphic sites. Of the 6461 variant sites, 89.60% were polymorphic in Olama and 34.82% in Nyabessan (Table 2). The difference observed between the two locations can be related to the sample size (n = 19 in Nyabessan and n = 59 in Olama) or to demographic particularities that persists between the two geographic sites despite a

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

massive gene flow. To infer the demographic history of *An. moucheti*, we examined the Allele Frequency Spectrum (AFS), summarized as the distribution of the major allele in one population. This approach represented a substitute to model-based methods that provide powerful examinations of the history of genetic diversity by modeling the AFS at genome-wide SNP variants, but that couldn't be implemented here due to the lack of reference genome assembly. The frequency distribution of the major allele *p* (Fig. 3) indicates that the majority of loci that are polymorphic between Olama and Nyabessan are fixed within each population as shown by the predominance of SNPs at frequencies equal to 1. Ranges of allele frequencies are similar in Olama and Nyabessan (between 0.47 and 1 in Olama and between 0.34 and 1 in Nyabessan). These frequency ranges are expected for old populations at equilibrium capable of accumulating high amount of genetic diversity.

Polymorphic chromosomal inversions

LD analyses at 1056 highly filtered SNPs revealed a globally low LD in the *An. moucheti* genome (average genome-wide r^2 = 0.0149) as expected in highly polymorphic populations with large effective size. We used LDna to cluster the LD values and to identify Single Outlier Clusters (SOC) that can be associated with distinct or multiple evolutionary phenomena in the *An. moucheti* history. Results showed the presence of two independent LD blocks in our samples (Fig. 4). Cytogenetic analyses have identified three polymorphic chromosomal inversions in Cameroon samples [26] and the two LD clusters could well correspond to linked SNPs within inversions. Studies on *Anopheles baimaii* [20] have shown that when SOCs are generated by polymorphic inversions, SNPs within the SOCs can clearly

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

228 separate the three expected karyotypes (inverted homozygotes, heterozygotes and
229 uninverted homozygotes). We conducted downstream analyses with a PCA using
230 respectively 30 and 34 SNPs identified within the two SOCs. As shown on Fig,
231 although individuals were spread along three PCA axes, no distinct cluster could be
232 identified for the first SOC. likewise, except three outlier individuals, the second SOC
233 formed a single genetic cluster (Fig. 4B,C). The two LD clusters were apparently not
234 associated with polymorphic inversions, but caution must be used when
235 interpreting this result as evidence of absence of inversions in our samples because
236 other LDna analyses conducted on *An. funestus* and *An. gambiae* which contain tens
237 of polymorphic inversions did not systematically identified the segregating
238 karyotypes of key adaptive inversions (results not shown). However, based on our
239 findings and previous cytogenetic studies, we can reasonably think that inversion
240 polymorphisms are less important in the genome architecture *An. moucheti*
241 compared with that of other important African malaria vectors such as *An. gambiae*
242 and *An. funestus* whose populations often segregate along clines of multiple
243 polymorphic inversions in nature.
244

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussion

We have analyzed genome-wide polymorphism and characterized some of the baseline population genomic parameters in *An. moucheti*, an important malaria vector in the African rainforest. We found very little differentiation among our samples, with most of the genetic variation distributed within individuals. Although a more substantial sampling will be necessary to fully appreciate the population genetic structure of this species, our sampling likely reflects the current dynamic of *An. moucheti* populations in Cameroon. It is worth mentioning that we have surveyed a total of 28 locations across the country, some of which were known from several past surveys to harbor *An. moucheti* populations [1–4,27,28], but we confirmed the presence of the species in only 2 villages. Extant populations of *An. moucheti* are distributed in patches of favorable habitats along river networks where larval populations breed. Our results indicate that despite this apparent fragmentation, connectivity and gene flow are high among population aggregates. The weak population genetic structure of *An. moucheti* observed with genome-wide markers corroborated results obtained with microsatellites and allozymes [2,3]. A survey of eight microsatellite loci revealed that the highest F_{ST} among Cameroonian populations was as low as 0.003. Nevertheless, a substantial differentiation was found between samples from different countries consistent with an isolation-by-distance model [3]. It is clear that a deep sequencing of continental populations is necessary to further clarify the status of these putative subpopulations that seem to be endemic to different countries. But, samples collected at lower spatial scales like ours are also relevant as they can allow robust inferences about ongoing selective

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

268 processes. We have found that signatures of selection are very rare in the genome of
269 *An. moucheti* populations from the Cameroonian rainforest. Populations remain
270 largely undifferentiated throughout the genome, with F_{ST} values near zero across the
271 vast majority of segregating sites. Interestingly, polymorphic chromosomal
272 inversions that are the prime source of adaptive polymorphisms in many other
273 important malaria vectors seem to be less vital in the evolution of *An. moucheti*. The
274 characterization of chromosomal inversions with cytogenetic methods can be
275 laborious, particularly with regard to the identification of new inversions [26,29].
276 We have implemented a recently designed method that used Next Generation
277 Sequencing and LD estimates to identify paracentric inversions. Although the
278 method holds great promise in identifying inversions in genomes of nonmodel
279 species like *An. moucheti*, care should be taken when interpreting the results. So far,
280 only three paracentric polymorphic inversions have been discovered in *An. moucheti*
281 thanks to cytogenetics [26]. The ecological, behavioral or functional roles of these
282 inversion polymorphisms remain unknown. In keeping with cytogenetic
283 observations, LD analyses revealed the presence of only a few LD clusters that could
284 not be unambiguously associated with inversions. On the other hand, the low overall
285 LD observed across the genome reflected the significant genetic polymorphism that
286 seems to prevail in *An. moucheti* populations. This polymorphism translates into
287 exceptional levels of overall genetic diversity and very high percentage of
288 polymorphic sites that are in the range of values found in other mosquito species
289 undergoing a significant demographic expansion [8,24,25]. One of our main goals
290 was to know if the genetic diversity of *An. moucheti* can provide cues about its

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

291 extensive spatial distribution and can help to predict its environmental resilience. In
292 principle, evolutionary responses of species to human-induced or natural changes
293 rely largely on available heritable variation, which reflects the evolutionary
294 potential and adaptability to novel environments [30]. Therefore, the screening of
295 genome-wide variation is supposed to be a sensible approach that may provide a
296 generalized measure of evolutionary potential in species like *An. moucheti* for which
297 direct ecological, evolutionary or functional tests are impossible [31]. Our
298 population genomic analyses have depicted *An. moucheti* as a species with a great
299 genetic diversity and consequently a sustainable long-term adaptive resilience.
300 Implications of our findings in malaria epidemiology and control can be very
301 significant. First, *An. moucheti* is essentially endophilic and is particularly sensitive
302 to the principal measures currently employed to control malaria in Sub-Saharan
303 Africa such as the massive use of Insecticide Treated Nets (ITNs) and Indoor
304 Residual insecticide Spraying (IRS). For example, estimates of population effective
305 size in one village in Equatorial Guinea indicated that both mass distribution of ITNs
306 and IRS campaigns resulted in a decline of approximately 55% of *An. moucheti* [32].
307 However, the great genetic diversity and the massive gene flow we observed within
308 populations could easily enable *An. moucheti* to challenge population declines and
309 recover from shallow bottlenecks. Moreover, most insecticide resistance
310 mechanisms found in insects exploit standing genetic variation that allow to rapidly
311 respond to the evolutionary challenge by increasing the frequency of existing
312 variations rather than that of *de novo* mutations [33]. As a result, despite the current
313 sensitivity of *An. moucheti* to common insecticides, the significant amount of

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

314 standing genetic variation provides the species with a great potential to challenge
315 insecticides and other types of human-induced stress.
316
317 Recent advances in sequencing allow sensitive genomic data to be generated for
318 virtually any species [30]. However, the most important information we can obtain
319 from population resequencing approaches often depends on the availability and the
320 quality of genomic resources such as a well-annotated reference genome.
321 Nevertheless, the reduced genome sequencing strategy (RAD-seq) offers a cost-
322 effective strategy that can been used to effectively study the genetic variation in a
323 broad range of species from yeast to plants, insects, etc., in the absence of reference
324 genome. We have extended this approach to the study of the genetic structure of an
325 understudied epidemiologically important mosquito species. We have provided
326 both significant baseline population genomic data and the methodological validation
327 of one approach that should motivate further studies on this species and other
328 understudied anopheline mosquitoes lacking genomic resources.
329

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

330 **Acknowledgements**

331 Funding for this project was provided by the University of California Riverside and

332 NIH grants 1R01AI113248 and 1R21AI115271 to BJW. We thank populations and

333 authorities of the locations surveyed for their kind collaboration.

334

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References

1. Kengne P, Antonio-Nkondjio C, Awono-Ambene HP, Simard F, Awolola TS, Fontenille D. Molecular differentiation of three closely related members of the mosquito species complex, Anopheles moucheti, by mitochondrial and ribosomal DNA polymorphism. Med. Vet. Entomol. 2007;21:177–82.

2. Antonio-Nkondjio C, Simard F, Cohuet A, Fontenille D. Morphological variability in the malaria vector, Anopheles moucheti, is not indicative of speciation: evidences from sympatric south Cameroon populations. Infect. Genet. Evol. 2002;2:69–72.

3. Antonio-Nkondjio C, Ndo C, Kengne P, Mukwaya L, Awono-Ambene P, Fontenille D, et al. Population structure of the malaria vector Anopheles moucheti in the equatorial forest region of Africa. Malar. J. 2008;7:120.

4. Antonio-Nkondjio C, Ndo C, Costantini C, Awono-Ambene P, Fontenille D, Simard F. Distribution and larval habitat characterization of Anopheles moucheti, Anopheles nili, and other malaria vectors in river networks of southern Cameroon. Acta Trop. 2009;112:270–6.

5. Bowen HL. Impact of a mass media campaign on bed net use in Cameroon. Malar. J. Malaria Journal; 2013;12:36.

6. Norris LC, Main BJ, Lee Y, Collier TC, Fofana A, Cornel AJ, et al. Adaptive introgression in an African malaria mosquito coincident with the increased usage of insecticide-treated bed nets. Proc. Natl. Acad. Sci. 2015;201418892.

7. Clarkson CS, Weetman D, Essandoh J, Yawson AE, Maslen G, Manske M, et al. Adaptive introgression between Anopheles sibling species eliminates a major genomic island but not reproductive isolation. Nat. Commun. 2014;5:4248.

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

358 8. Kamdem C, Fouet C, Gamez S, White BJ. Pollutants and insecticides drive local
359 adaptation in African malaria mosquitoes. bioRxiv. 2016;
360 9. Neafsey DE, Waterhouse RM, Abai MR, Aganezov SS, Alekseyev MA, Allen JE, et al.
361 Highly evolvable malaria vectors: The genomes of 16 Anopheles mosquitoes. Science
362 (80-.). 2015;347:1258522–1258522.
363 10. Gillies MT, De Meillon B. The Anophelinae of Africa South of the Sahara. Second
364 Edi. Johannesburg: Publications of the South African Institute for Medical Research;
365 1968.
366 11. Gillies MT, Coetzee M. A supplement to the Anophelinae of Africa south of the
367 Sahara. Johannesburg: The South African Institute for Medical Research; 1987.
368 12. Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double Digest RADseq:
369 An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and
370 Non-Model Species. PLoS One. 2012;7:e37135.
371 13. Catchen J, Hohenlohe P a., Bassham S, Amores A, Cresko W. Stacks: An analysis
372 tool set for population genomics. Mol. Ecol. 2013;22:3124–40.
373 14. Jombart T. adegenet: a R package for the multivariate analysis of genetic
374 markers. Bioinformatics. 2008;24:1403–5.
375 15. Paradis E, Claude J, Strimmer K. Analyses of Phylogenetics and Evolution in R
376 language. Bioinformatics. 2004;20:289–90.
377 16. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in
378 unrelated individuals. Genome Res. 2009;19:1655–64.
379 17. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population
380 structure. Evolution (N. Y). 1984;38:1358–70.

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

381 18. Meirmans P, Van Tienderen P. GENOTYPE and GENODIVE: two programs for the
382 analysis of genetic diversity of asexual organisms. Mol. Ecol. Notes. 2004;4:792–4.
383 19. Excoffier L, Smouse PE, Quattro JM. Analysis of molecular variance inferred from
384 metric distances among DNA haplotypes: Application to human mitochondrial DNA
385 restriction data. Genetics. 1992;131:479–91.
386 20. Kemppainen P, Knight CG, Sarma DK, Hlaing T, Prakash A, Maung Maung YN, et
387 al. Linkage disequilibrium network analysis (LDna) gives a global view of
388 chromosomal inversions, local adaptation and geographic structure. Mol. Ecol.
389 Resour. 2015;n/a – n/a.
390 21. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, et al. PLINK:
391 a toolset for whole-genome association and population-based linkage analysis. Am. J.
392 Hum. Genet. 2007;81.
393 22. O'Loughlin SM, Magesa S, Mbogo C, Mosha F, Midega J, Lomas S, et al. Genomic
394 Analyses of Three Malaria Vectors Reveals Extensive Shared Polymorphism but
395 Contrasting Population Histories. Mol. Biol. Evol. 2014;1–14.
396 23. Kamdem C, Fouet C, Gamez S, White BJ. Genomic signatures of introgression at
397 late stages of speciation in a malaria mosquito. bioRxiv. 2016;
398 24. Fouet C, Kamdem C, White BJ. Chromosomal inversions facilitate chromosome-
399 scale evolution in Anopheles funestus. bioRxiv. 2016;
400 25. Donnelly MJ, Licht MC, Lehmann T. Evidence for recent population expansion in
401 the evolutionary history of the malaria vectors Anopheles arabiensis and Anopheles
402 gambiae. Mol. Biol. Evol. 2001;18:1353–64.
403 26. Sharakhova M V., Antonio-Nkondjio C, Xia a., Ndo C, Awono-Ambene P, Simard F,

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

et al. Polymorphic chromosomal inversions in Anopheles moucheti, a major malaria vector in Central Africa. Med. Vet. Entomol. 2014;28:337–40.

27. Antonio-Nkondjio C, Kerah CH, Simard F, Awono-Ambene P, Chouaibou M, Tchuinkam T, et al. Complexity of the malaria vectorial system in Cameroon: contribution of secondary vectors to malaria transmission. J. Med. Entomol. 2006;43:1215–21.

28. Antonio-Nkondjio C, Demanou M, Etang J, Bouchite B. Impact of cyfluthrin (Solfac EW050) impregnated bed nets on malaria transmission in the city of Mbandjock : lessons for the nationwide distribution of long-lasting insecticidal nets (LLINs) in Cameroon. Parasit. Vectors. Parasites & Vectors; 2013;6:10.

29. Kirkpatrick M. How and why chromosome inversions evolve. PLoS Biol. 2010;8.

30. Ellegren H. Genome sequencing and population genomics in non-model organisms. Trends Ecol. Evol. Elsevier Ltd; 2014;29:51–63.

31. Harrisson K a., Pavlova A, Telonis-Scott M, Sunnucks P. Using genomics to characterize evolutionary potential for conservation of wild populations. Evol. Appl. 2014;n/a – n/a.

32. Athrey G, Hodges TK, Reddy MR, Overgaard HJ, Matias A, Ridl FC, et al. The effective population size of malaria mosquitoes: large impact of vector control. PLoS Genet. 2012;8:e1003097.

33. Messer PW, Petrov D. Population genomics of rapid adaptation by soft selective sweeps. Trends Ecol. Evol. Elsevier Ltd; 2013;28:659–69.

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Author contributions

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

Competing interests

The authors declare that they have no competing interests.

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Tables

Table 1: Information on *An. moucheti* samples included in this study.

Sampling locations	Geographic coordinates	Sampling methods			Total
		HLC-OUT	HLC-IN	LC	
Nyabessan	2°24'00"N, 10°24'00"E	21	15	1	37
Olama	3°26'00"N, 11°17'00"E	30	31	0	61
Total					98

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

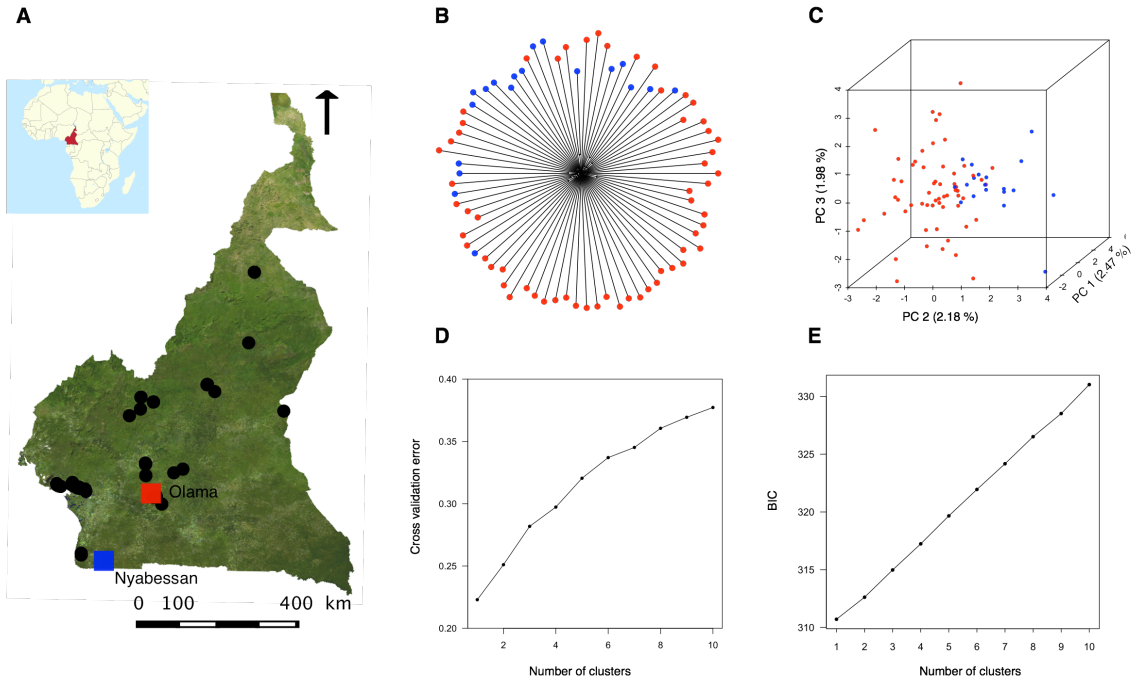
Table 2: Population genomic parameters based on 6461 variant sites reflecting the genetic diversity and conformity to Hardy-Weingberg equilibrium.

	n	Sites	Observed heterozygosity	F_{IS}	π	% polymorphic sites
Variant positions						
Olama	54.12	6 461	0.0445	0.0631	0.0505	89.60
Nyabessan	16.78	6 461	0.0334	0.0372	0.0402	34.82
All positions						
Olama	54.47	165 975	0.0017	0.0025	0.0020	3.49
Nyabessan	16.86	165 975	0.0013	0.0014	0.0016	1.36

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figures

Figure 1: Relationship between *An. moucheti* individuals from Olama and Nyabessan. (A) Map of the study site showing both the locations surveyed (small black dots) and the two villages (large red and blue squares) where *An. moucheti* samples were collected. (B) and (C) Absence of genetic structure within populations illustrated by neighbor-joining and PCA. The percentage of variance explained by each PCA axis is indicated. (D) and (E) Plots of the ADMIXTURE cross-validation error and the Bayesian Information Criterion (BIC) (DAPC) as a function of the number of genetic clusters indicating that k = 1. The lowest BIC and CV error indicate the suggested number of clusters.

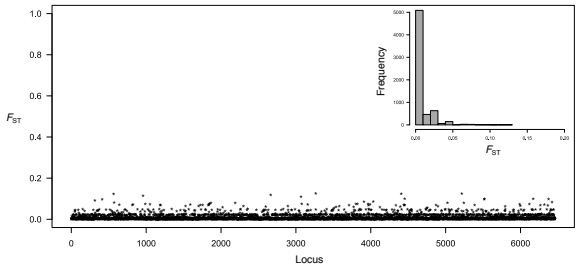


bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

450 **Figure 2**: Frequency distribution of F_{ST} between Olama and Nyabessan across 6461

451 SNP loci and plot of these F_{ST} values along arbitrary positions in the genome.

452

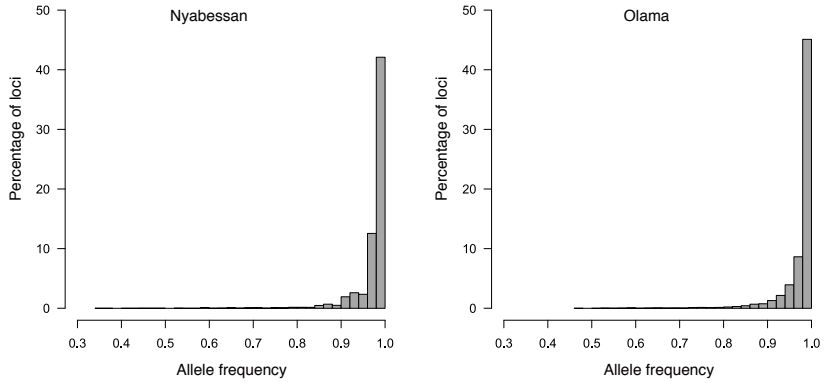


453

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

454 **Figure 3**: Allele Frequency Spectrum for 6461 SNP loci in Nyabessan and Olama
455 populations. The x-axis presents the frequency of the major allele and the y-axis the
456 frequency distribution of loci in each class of the major allele frequency.

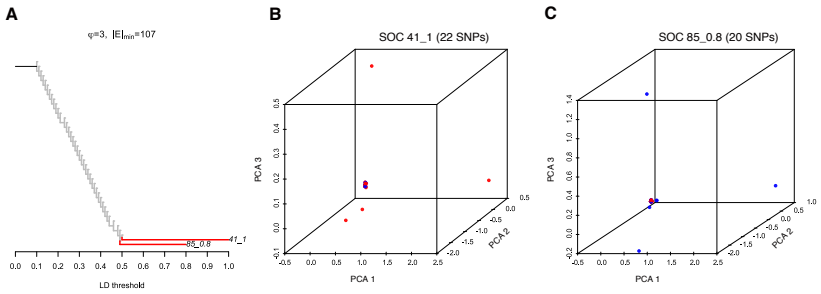
457



458

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

459 **Figure 4**: LDna analyses on 1056 SNPs showing the presence of two Single Outlier
460 Clusters (SOCs) of linkage disequilibrium in *An. moucheti*. (A) Graph presenting the
461 results obtained with values of the two parameters: φ (which controls when clusters
462 are defined as outliers) and $|E|_{\min}$, the minimum number of edges required for a LD
463 cluster to be considered as an outlier, indicated on top. LD thresholds are shown on
464 the x-axis. (B) and (C) PCA indicating the population genetic structure inferred from
465 SNPs within the two SOCs (red: Olama; blue: Nyabessan).
466



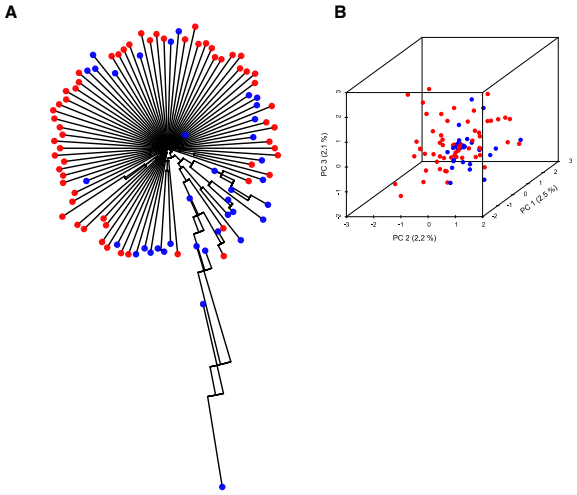
467

bioRxiv preprint doi: https://doi.org/10.1101/068247; this version posted August 6, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

468 **Supplemental Material**

469 **Figure S1**: Selection of individuals included in final analyses based on the average
470 per individual sequencing coverage. (A) and (B) Neighbor-joining tree and PCA
471 indicating spurious population structure due to individuals with low sequencing
472 coverage in Olama (red) and Nyabessan (blue).

473



474

475