

## Research paper

Genetic analysis and population structure of the *Anopheles gambiae* complex from different ecological zones of Burkina Faso

Abdou Azaque Zouré<sup>a,b,\*</sup>, Grégoire Noël<sup>a</sup>, Aboubacar Sombié<sup>c</sup>, Zéphirin Somda<sup>c</sup>, Athanase Badolo<sup>c</sup>, Frédéric Francis<sup>a</sup>

<sup>a</sup> Functional and Evolutionary Entomology, TERRA, Gembloux Agro-Bio Tech, University of Liège, Passage des Déportés 2, 5030 Gembloux, Belgium

<sup>b</sup> Institute of Health Sciences Research, (IRSS/CNRST)/Department of Biomedical and Public Health, 03 BP 7192 Ouagadougou, Burkina Faso

<sup>c</sup> Laboratoire d'Entomologie Fondamentale et Appliquée, UFR/SVT, Université Joseph KI-ZERBO, BP 7021, Ouagadougou 03, Burkina Faso

## ARTICLE INFO

## Keywords:

*Anopheles gambiae* s.l.

rDNA

COI

genetic structure

Burkina Faso

## ABSTRACT

The *Anopheles gambiae* complex (Diptera: Culicidae) is the most important vector for malaria in Sub-Saharan Africa, besides other vectors such as *Anopheles funestus*. Malaria vector control should encompass specific identification, genetic diversity and population structure of *An. gambiae* to design vector control strategies.

The aim of this study was to determine the distribution of sibling species of the *An. gambiae* complex according to climatic regions related to cotton-growing or cotton-free areas by using polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP). Then, variation in mitochondrial cytochrome c oxidase I (COI) was used to assess the genetic structure within and between populations from our selected ecological zones.

At the sibling species level, the following proportions were found across all samples ( $n = 180$ ): *An. coluzzii* 65.56%, *An. gambiae* stricto sensu (s.s.) 21.11%, and *An. arabiensis* 3.33%. Hybrids between *An. gambiae* s.s. and *An. coluzzii* (7.78%) and hybrids between *An. coluzzii* and *An. arabiensis* (2.22%) were found. The phylogenetic tree and Integer Neighbour-Joining (IntNJ) haplotype network did not reveal any distinct genetic structure pattern related to climatic or agricultural conditions in Burkina Faso. The  $F_{st}$  (Wright's F-statistic) values close to zero showed a free gene flow and no differentiation in *An. gambiae* complex populations. Furthermore, neutrality indices calculated by Tajima's D, Fu and Li's D\*, Fu and Li's F\*, Fu's Fs tests suggested an excess of rare mutations in the populations.

Overall, variation in the proportions of *An. gambiae* s.s., *An. coluzzii* and *An. arabiensis* was found according to climatic regions, but COI analysis did not evidence any population structuring of the *An. gambiae* complex. These scientific contributions can be used as a basis for further in-depth study of the genetic diversity of the *An. gambiae* complex for epidemiological risk assessment of malaria in Burkina Faso.

## 1. Introduction

In Sub-Saharan Africa, malaria is still a public health challenge (WHO, 2018). Early diagnosis of human infection by the parasite and medical therapy (prevention and treatment) constitute the basis of malaria control, while indoor residual spraying (IRS) and long-lasting insecticide nets (LLINs) are used for vector control (Diallo et al., 2015; WHO, 2018). Moreover, malaria control should encompass identification of the vector species or sub-species as well as their population structure (Besansky et al., 2003). To date, 485 described species belong to the *Anopheles* genus (Diptera: Culicidae) (Harbach and Kitching,

2016). An estimated 70 species are regularly associated with transmission of *Plasmodium* sp. to humans worldwide (Stevenson and Norris, 2017). Members of *Anopheles gambiae* sensu lato (s.l.) (named the *Anopheles gambiae* complex) are among the most important vectors for malaria transmission in Sub-Saharan Africa (Benson and Falade, 2013). The *An. gambiae* complex is also involved in the transmission of Bancroft filaria (*Wuchereria bancrofti*) and several arboviruses such as O'Nyong-Nyong virus, Tataguin virus, Chikungunya virus and Ilesha virus (Carnevale and Robert, 2009; Hegde et al., 2015). Before 1962, *An. gambiae* s.l. was considered to be a single species (Coetzee et al., 2000). But the evolution of molecular biology techniques caused *An.*

\* Corresponding author at: Functional and Evolutionary Entomology, TERRA, Gembloux Agro-Bio Tech, University of Liège, Passage des Déportés 2, 5030 Gembloux, Belgium.

E-mail addresses: [abdouzaque@gmail.com](mailto:abdouzaque@gmail.com), [entomologie.gembloux@ulg.ac.be](mailto:entomologie.gembloux@ulg.ac.be) (A.A. Zouré).

<https://doi.org/10.1016/j.meegid.2020.104261>

Received 13 October 2019; Received in revised form 20 February 2020; Accepted 21 February 2020

Available online 22 February 2020

1567-1348/ © 2020 Published by Elsevier B.V.

*gambiae* s.l. to break into seven sibling species, all morphologically indistinguishable (Borja-Cacho and Matthews, 2008). Among them, *An. arabiensis* Patton 1904, *An. gambiae* sensu stricto (s.s.) Giles 1902 (S-form or Savannah or Bamako form), and *An. coluzzii* Coetzee and Wilkerson 2013 (M-form or Mopti form) are the most anthropophilic vectors of malaria in Sub-Saharan Africa (Coetzee et al., 2013; Della Torre et al., 2001; Pates et al., 2014; Wiebe et al., 2017). Although hybridisation is possible in many sympatric areas, natural hybrids between these sibling species are very rare (Yawson et al., 2007; Pates et al., 2014). Speciation processes (Miles et al., 2017) and chromosomal inversion indicate genetic heterogeneity between *An. arabiensis*, *An. gambiae* s.s. and *An. coluzzii* in West Africa (Coetzee et al., 2000; Borja-Cacho and Matthews, 2008). Therefore, each of these sibling species (*An. arabiensis*, *An. gambiae* s.s. and *An. coluzzii*) has been clearly protected by effective reproductive isolation (Della Torre et al., 2001). This speciation process could also be explained by a combination of different factors, such as the intensity of rainfall or the type of agricultural system in any given area, as *Anopheles* mosquitoes depend on standing water for a breeding habitat (Bomblies, 2012). In Burkina Faso (Africa), the differences in epidemiological facies of malaria follow the different rainfall levels in the three climatic regions, which are the Sahelian (short malaria transmission period of two to three months per year), Sudano-Sahelian (seasonal transmission four to six months per year) and Sudanian (perennial transmission) regions (Carnevale and Robert, 2009; Abiodun et al., 2016). One-third of the gross domestic product of Burkina Faso comes from the agricultural economic sector, primarily food grains (sorghum, millet, maize, and rice). Because of the use of insecticides, farmland may be considered as a source of selection pressure on *Anopheles* mosquitoes, especially in cotton-growing areas (Hien et al., 2017) where cotton is the main cash crop (IFPRI, 2010). During the agricultural seasons of the last decade (2000), Burkina Faso was the leading cotton producer in West Africa with about 600,000 tons of seed cotton/year (Hauchart, 2007). Research suggests that certain fertilizer contents may influence mosquito egg-laying behaviour. The use of these fertilizers is combined with insecticide treatments, which have attractive properties for mosquitoes and are found in the waters where mosquitoes lay their eggs (Darriet and Corbel, 2008; Darriet et al., 2012). In sum, temperature, rainfall and agriculture could influence the species density and structure of the *An. gambiae* species complex (Minakawa and Sonye, 2002).

An accurate and predictive understanding of the geographic distribution of these *Anopheles* species should permit the identification of areas in which particular co-habituating species are potentially involved in further disease transmission (Levine et al., 2004). In Burkina Faso, most research studies have been carried out on *An. gambiae* s.l. (Dabiré et al., 2012) but not specifically according to ecological conditions. Therefore, the aim of this study was firstly to use molecular techniques to determine the distribution of the sibling species of the *An. gambiae* complex (*An. arabiensis*, *An. gambiae* s.s. and *An. coluzzii* in the present case) within and between populations according to their localisation in the Sahelian, Sudano-Sahelian and Sudanian climatic regions in cotton-growing or cotton-free areas. Secondly, a fragment of the COI gene was sequenced to distinguish the population structure according to agricultural and climatic features. The distribution of *An. gambiae* s.l. sibling species and its population structure could be an important monitoring tool for malaria control in Burkina Faso.

## 2. Materials and methods

### 2.1. Sample collection

Adult females of the *An. gambiae* s.l. complex were collected from human residential areas in different localities of Burkina Faso (Fig. 1) using a Prokopack aspirator (a semi-craft production) between August and October 2017. Malaria transmission generally reaches its seasonal peak at that time of year and corresponds to the population peak of *An.*

*gambiae* s.l. in the three climate regions (INSD, 2015). The Sudanian climatic region (annual rainfall > 900 mm; average annual temperature of 31 °C) contributes to malaria transmission all year round. In the Sudano-Sahelian region (annual rainfall between 600 mm and 900 mm; average annual temperature of 33 °C) and the Sahelian region (annual rainfall < 600 mm; average annual temperature of 35 °C), malaria transmission is seasonal and lasts four to six months and two to three months, respectively (Kabore et al., 2017). The collection times ranged between 6:00–8:00 a.m. and 5:00–7:00 p.m. Mosquitoes were collected indoors and outdoors in six localities (two in each climatic region): Kongoussi (KG); Ouahigouya (OH); Sanguié (SG); Kombissiri (KB); Houndé (HD), and Bobo (BB) (three villages in each collection site), including three sites in cotton-growing areas and three in cotton-free areas. Thirty individuals from the *An. gambiae* complex were collected per locality.

### 2.2. Morphological identification of mosquitoes

All samples ( $n = 180$ ) were brought back to the Laboratoire d'Entomologie Fondamentale et Appliquée (Ouagadougou, Burkina Faso). All specimens were prepared and identified morphologically according to the identification keys of Gillies and De Meillon (1968) and Gillies and Coetzee (1987). These morphological keys can distinguish differences among groups of African *Anopheles* species but do not distinguish the sibling species. Specimens were washed in sterile phosphate-buffered saline (PBS 1 ×, pH 7.8), and dissected into three parts (wings/legs, head/thorax, and midgut) preserved in ethanol 70%.

### 2.3. Molecular identification

All our dissected samples were brought to Gembloux Agro-Bio Tech, Functional & Evolutionary Entomology (Gembloux, Belgium) and stored at  $-80^{\circ}\text{C}$  until DNA extraction. First, the samples were washed twice in sterile distilled water to remove traces of ethanol and to avoid contamination. DNA was extracted from the mosquito samples (wings/legs) using a QIAamp DNA Micro Kit (QIAGEN, Valencia, CA) according to the manufacturer's instructions. Briefly, the samples were crushed individually with plastic pestles in 1.5-ml microtubes with 180  $\mu\text{l}$  of buffer ATL, and 20  $\mu\text{l}$  of proteinase K were added. Mixtures were vortexed thoroughly for 15 s and incubated in a thermomixer (56 °C, 450 rpm) overnight. After processing and washing with the adequate product provided in the kit, the final DNA product was dissolved in 20  $\mu\text{l}$  of elution buffer AE. DNA purity and quality were assessed using a spectrophotometer (Nanodrop®), and DNA was stored at  $-20^{\circ}\text{C}$  until PCR analysis.

A total of 180 mosquitoes of the *An. gambiae* complex was identified as being *An. gambiae* s.s., *An. coluzzii* or *An. arabiensis* by the PCR-RFLP method using the intergenic spacer (IGS) of ribosomal DNA (rDNA) as the target gene (Fanello et al., 2002). We combined the protocols established by Scott et al. (1993) and Favia et al. (2001) with the enzyme *Hha* I and restriction site GCG°C. Briefly, 25 ng of DNA were amplified in 25  $\mu\text{l}$  of master mix containing 1 × Buffer-MgCl<sub>2</sub>, 2.5 mM MgCl<sub>2</sub>, 0.2 mM of each dNTP, 12.5 ng of UN primer, 6.25 ng of GA primer, 18.75 ng of AR primer, 25 ng of QD primer, 12.5 ng of ME primer and 4 U of Platinum™ Taq DNA Polymerase (ThermoFisher Scientific™) (Table 1).

The PCR started with an initial step of ten minutes at 94 °C to activate the DNA polymerase, followed by 30 cycles each consisting of denaturation for 30 s at 94 °C, hybridisation for 30 s at 50 °C, extension for 30 s at 72 °C, and a final cycle for seven minutes at 72 °C. After amplification, 10  $\mu\text{l}$  of PCR products were added directly to the 10- $\mu\text{l}$  mix containing 5 U of *Hha* I (10 U/ $\mu\text{l}$ . ThermoFisher Scientific™), 2 ng of BSA (10 ng/ $\mu\text{l}$ ), and 1 × of RE Buffer C (10 ×); digestion was performed at 37 °C for a minimum of three hours, following the manufacturer's instructions. Amplified fragments were revealed and confirmed by electrophoresis on a 2% agarose gel and visualised under UV light. The

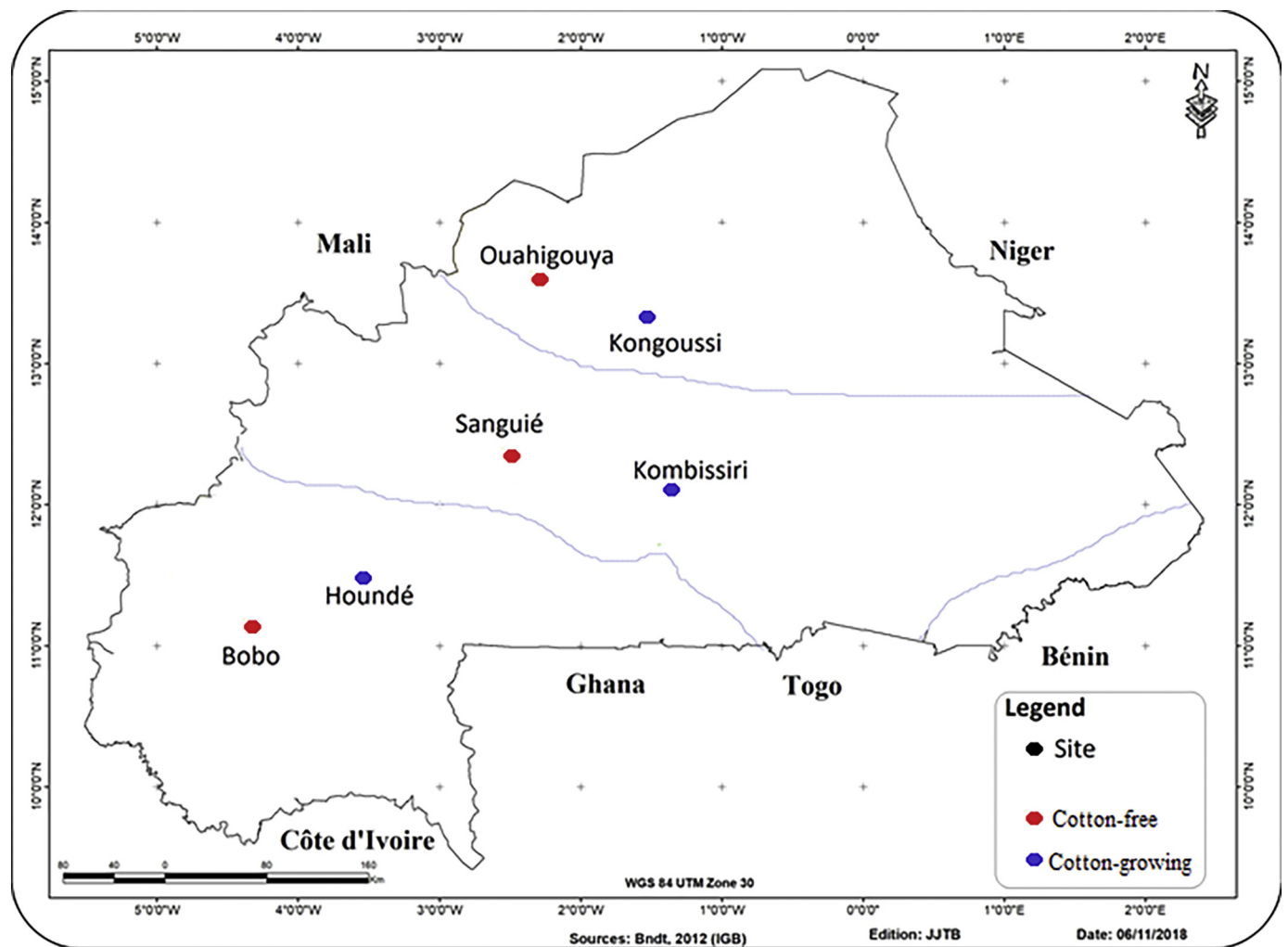


Fig. 1. Map of the collection sites showing climatic regions and cotton-growing or cotton-free areas: Sahelian (Kongoussi; Ouahigouya), Sudano-Sahelian (Kombissiri; Sanguié) and Sudanian (Houndé; Bobo).

product sizes following the PCR and the fragment lengths of the sibling species following *Hha* I digestion are presented in Table 1.

To study the population structure and to build a phylogenetic tree of our dataset related to the *An. gambiae* species complex, the COI gene or barcode was also amplified from our DNA extracts ( $n = 180$ ) using

primers COI (Forward) and Cul-Rev (Reverse) (Table 1). The PCR reaction mixtures were as follows: a final volume of 25  $\mu$ L containing 10 mM Tris-HCl (pH = 8.4), 150 mM KCl, 0.10% Triton X – 100, 0.2  $\mu$ g/ $\mu$ L of bovine serum albumin (BSA), 1.5 mM MgCl<sub>2</sub>, 0.2 mM of each dNTP, 30 ng of each primer, 2 U of Platinum™ *Taq* DNA

**Table 1**  
List of primers used for molecular identification of *Anopheles gambiae* complex members and amplification of the cytochrome oxidase subunit COI.

Sequences of primers/Names of sibling species	Length of the fragments after <i>Hha</i> I digestion
PCR primers for the intergenic space (IGS) of ribosomal DNA (rDNA) *	
UN primer [5'-GTG TGC CCC TTC CTC GAT GT-3']	
GA primer [5'-CTG GTT TGG TCG GCA CGT TT-3']	
<i>An. gambiae</i> s.s. (S-form)	390 bp and 257 bp <sub>110 pb</sub> , 23 bp
<i>An. coluzzii</i> (M-form)	390 bp and 367 bp <sub>23 bp</sub>
AR primer [5'-AAG TGT CCT TCT CCA TCC TA-3']	
<i>An. arabiensis</i> Patton	315 bp and 292 bp <sub>23 bp</sub>
QD primer [5'-CAG ACC AAG ATG GIT AGT AT-3']	
<i>An. quadriannulatus</i> (Theobald)	153 bp and 124 bp <sub>6 pb</sub> , 23 bp
ME primer [5'-TGA CCA ACC CAC TCC CTT GA-3']	
<i>An. melas</i> Theobald	464 bp and 435 bp <sub>6 pb</sub> , 23 bp
<i>An. merus</i> Dönitz	466 bp and 437 bp <sub>29 pb</sub>
PCR primers for the cytochrome oxidase subunit COI **	
Forward-COI (5'-TTG ATT TTT TGG TCA TCC AGA AGT-3')	
Reverse-Cul-Rev (5'-TAG AGC TTA AAT TCA TTG CAC TAA TC-3')	877 bp

bp: base pair; \* and \*\* according to Fanello et al. (2002) and Oshaghi et al. (2006) respectively.

Polymerase (ThermoFisher Scientific™) and 25 ng of DNA sample (Oshaghi et al., 2006). The PCR profile was as follows: four minutes of initial denaturation at 94 °C, 32 cycles of one minute at 94 °C, one minute at 55 °C, two minutes at 72 °C, and a final extension step of seven minutes at 72 °C (Norris and Norris, 2015). A total of 10 µl of each PCR product was loaded for electrophoresis on a 1.5% agarose gel, stained with SYBR™ Safe™ DNA, and visualised under UV light. The remainder of each successful PCR reaction was purified using a NucleoSpin® gel and PCR clean-up kit (Macherey-Nagel GmbH & Co. KG, Germany). After the purification step, COI amplicons were sequenced in both directions using Sanger sequencing at GATC Biotech (A Eurfins Genomics Company, Germany).

#### 2.4. Molecular data processing and phylogenetic analysis

Barcode sequencing (Ratnasingham and Hebert, 2007) is commonly used for identifying *Anopheles* (Culicidae) species (Mohanty et al., 2009). The sequences obtained from Sanger sequencing were corrected manually using BioEdit v.7.2.6 software (Hall, 1999) and edited to acquire the consensus sequences. After BLASTn examination to control the quality of all sequences, each barcode sequence ( $n = 151$ ) included a maximum of 765 bp. All COI sequences are available at GenBank under accession numbers MK681509-MK681659.

For phylogenetic reconstruction, we followed the guidelines of Harrison and Langdale (2006). The first step in the phylogenetic methodology relies on sequence alignment to highlight homologies in nucleotide positions. All COI sequences were aligned using ClustalX v.2.1 (Larkin et al., 2007) with default parameters and pairwise deletion for gap treatments.

Since nucleotide positions within a sequence do not evolve independently (e.g., unequal base frequencies or substitution rates) (Posada, 2003), the second step consisted in selecting the best evolutionary model that explained the empirical evolution of sequences (i.e., our aligned sequence dataset). Therefore, we searched for the best substitution model using PartitionFinder v.2.1.1. (Lanfear et al., 2016) with a linked branch length parameter, the AICc model selection metrics and the greedy search algorithm (Lanfear et al., 2012). The transversal (TVM) model (Posada, 2003) was selected. The third step of phylogenetic reconstruction consisted in generating and analysing the phylogenetic tree with support values. We then conducted Bayesian Inference (BI) for the *An. gambiae* complex using BEAST2 v.2.5.2. environment (Bouckaert et al., 2014) to reconstruct an ultrametric tree. The yule model was set up as 'Prior Tree' with a Markov Chain Monte Carlo (MCMC) length of five million. We then summarised posterior probabilities using TreeAnnotator v.2.5.2. (included in BEAST environment) with a burn-in of 10%. Phylogenetic trees were built using FigTree v.1.4.3. (Rambaut, 2017). For tree topologies, we considered posterior probabilities above 0.95 as the threshold for well-supported nodes (Wilcox et al., 2002).

#### 2.5. Population structure visualisation and statistical analysis

Regardless of the subspecies, we defined six populations of the *An. gambiae* complex corresponding to sampling localities based on the combination of agricultural pressure and climatic regions. All our population structure analyses used COI as a molecular marker. Population diversity indices such as number of polymorphic (segregating) sites (S), haplotype number (h), haplotype diversity (Hd), nucleotide diversity ( $\pi$ ) and average number of pairwise nucleotide differences within a population (K) were estimated using DnaSP v.6.12.01 software (Librado and Rozas, 2009). The haplotype number (h) refers to a nucleotide sequence that may be common to several individuals but differs from other haplotypes in at least one nucleotide substitution. Haplotype diversity (Hd) defines the probability for two randomly selected genes in a sample to be identical (Nei, 1987). Nucleotide diversity ( $\pi$ ) measures the average nucleotide divergence between all sequence pairs in a given

sample with the probability that two randomly selected genes are different by a nucleotide (Tajima, 1983). The neutrality indices of Tajima's D, Fu and Li's D\* and F\*, and Fu's Fs were also calculated in each population with DnaSP v.6.12.01. Tajima's D was estimated by comparing the differences between the number of segregating sites (S) and the average number of nucleotide differences (K). Fu and Li's D\* and F\* statistics compare estimates of theta based on mutations in the internal and external branches of a genealogy. Fu's Fs were used to assess the structure of the haplotype frequency distribution. In addition, the average number of pairwise nucleotide differences (Kxy) and nucleotide substitutions per site (Dxy) between populations were also calculated using DnaSP v.6.12.01. Pairwise genetic differences were estimated for all populations by calculating Wright's F-statistic (Fst) and carrying out analyses of molecular variance (AMOVA) and gene flow (Nm) in Arlequin 3.5.1.2 software (Excoffier et al., 2005). According to Wright (1969), the Fst metric estimates the extent of gene flow between various genotypes (Raymond and Rousset, 2013). AMOVA was tested to detect population (i.e. locality) segregation with a Euclidian distance matrix derived from the combined climatic region and agricultural pressure data. We implemented the non-parametric permutation approach ( $n = 999$ ) to test if our populations were significantly structured.

Isolation by distance was calculated by examining the correlation between genetic distance Fst/(1-Fst) and geographical distance in log (km) of the samples. Null hypotheses assumed that the regression slope of genetic distance over geographical distance was null.

For the analysis and visualisation of *An. gambiae* s.l. haplotypes, we used the representation of reticulate networks. This method encompasses reticulation events such as evolutionary histories of the different haplotypes. The edges provide the evolutionary history of the lineages such as speciation events, and the nodes correspond to supposed ancestors (Huson and Bryant, 2006). Therefore, our final aligned barcode dataset (in FASTA format) was transformed in NEXUS format (Maddison et al., 1997), and then POPART v.1.7. (Leigh and Bryant, 2015) was used to build haplotype networks for climatic region and agricultural areas with the IntNJ method and a reticulation tolerance parameter ( $\alpha$ ) of 0.5. This approach of haplotype network construction is more suitable for a dataset with low genetic divergence, as expected in a species complex.

### 3. Results

#### 3.1. Identification and distribution of the *Anopheles gambiae* complex

We identified 180 *Anopheles* specimens using the PCR-RLFP method. *An. coluzzii* was the dominant species with 65.56% occurrence among all samples, followed by *An. gambiae* s.s. (21.11%) and *An. arabiensis* (3.33%). These three species of the *An. gambiae* complex were found in all climatic regions as well as in cotton-growing and cotton-free areas. Hybrids between *An. gambiae* s.s. and *An. coluzzii* (7.78%) and between *An. coluzzii* and *An. arabiensis* (2.22%) were also found (Table 2). *An. coluzzii*, *An. arabiensis* and their hybrid sequences were present in all climatic regions. Across the three climatic regions, chi-square goodness of fit tests highlighted a significant dominance of *An. coluzzii* ( $P < .006$ ; ddl = 2;  $\chi^2 = 10.403$ ) in the Sahelian region, followed by the Sudano-Sahelian and Sudanian regions. *An. gambiae* s.s. ( $P < .001$ ; ddl = 2;  $\chi^2 = 16.769$ ) was dominant in the Sudano-Sahelian region, followed by the Sudanian region, and hybrid specimens between *An. gambiae* s.s. and *An. coluzzii* ( $P < .015$ ; ddl = 2;  $\chi^2 = 8.400$ ) were dominant in the Sudanian region, followed by the Sudano-Sahelian region. However, this difference in species distribution among climatic regions was not observed for the *An. arabiensis* population. No significant difference was observed in the distribution of *An. gambiae* populations between cotton-free and cotton-growing areas.



**Table 2**Distribution of the *Anopheles gambiae* species complex ( $n = 180$ ) depending on the climatic regions and related to areas of insecticide exposure.

Climate	Agriculture pressures	Sites	Latitude	Longitude	Hybrids between				
					<i>An. coluzzii</i>	<i>An. gambiae</i> s.s.	<i>An. arabiensis</i>	<i>An. arabiensis</i> and <i>An. coluzzii</i>	<i>An. gambiae</i> s.s. and <i>An. coluzzii</i>
Sahelian	Cotton-growing	Kongoussi	13°19'31.063"N	-1°31'32.535"W	26	0	3	0	1
Sahelian	Cotton-free	Ouahigouya	13°34'0.573"N	-2°24'39.566"W	30	0	0	0	0
Sudano-Sahelian	Cotton-growing	Kombissiri	12°4'27.57"N	-1°20'31.571"W	14	13	0	3	0
Sudano-Sahelian	Cotton-free	Sanguié	12°9'0.67"N	-2°41'54.192"W	19	7	2	1	1
Sudanian	Cotton-growing	Houndé	11°29'21.926"N	-3°31'23.131"W	13	11	0	4	2
Sudanian	Cotton-free	Bobo	11°9'53.719"N	-4°18'18.555"W	16	7	1	6	0

### 3.2. Phylogenetic reconstruction

Our phylogenetic analyses based on COI (Fig. 2) were poorly supported (except for some nodes) and revealed no significant clusters according to population origin. The topology of the unrooted COI phylogeny tree was not consistent with classical molecular taxonomy for *An. gambiae* s.s., *An. coluzzii* or *An. arabiensis*. Therefore, no genetic structure was observed in these three species' populations based on COI as a single locus. Moreover, no population cluster was related to any climatic region or agricultural use in Burkina Faso. In supplementary materials A1, neighbour-joining (NJ) trees are used as a standard phylogenetic approach in DNA barcoding (Saitou and Nei, 1987) to visualise genetic distances and pairwise p-distances (i.e. proportions of variable sites) (Tamura et al., 2004).

### 3.3. Diversity indices and nucleotide polymorphism

Among 151 sequences, haplotype and nucleotide diversity indices were highest in the Houndé population and lowest in the Bobo population (Sudanian region). However, the *An. gambiae* complex had a high haplotypic diversity (Hd) (0.95252 for all 151 sequences) combined with low nucleotide diversity ( $\pi$ ) at 0.00604 (Table 3). The average number of nucleotide differences (K) was 4.13310 (Table 3). Among the 765 nucleotide sites, 684 sites were found (excluding fixed gaps and missing data). A total number of 114 mutations was noted, three of which were parsimony-informative sites.

Neutrality indices calculated from Tajima's D, Fu and Li's D\*, Fu and Li's F\*, and Fu's Fs tests were negatively significant in all populations except in Ouahigouya, suggesting an excess of rare mutations in this population (Table 4). The inter-population nucleotide differences (Kxy) varied from 3.4331 (Ouahigouya and Sanguié) to 4.91244 (Kombissiri and Sanguié); while the average number of nucleotide substitutions per site between all these populations (Dxy) varied from 0.00504 (Bobo and Sanguié) to 0.00737 (Kombissiri and Kongoussi) (Table 5). Pairwise genetic distance (Fst) between populations varied from -0.01366 (between Ouahigouya and Bobo) to 0.0263 (between Sanguié and Ouahigouya) (Table 6). The Fst values between Kongoussi and Ouahigouya, Kombissiri and Kongoussi, Bobo and Kongoussi, Ouahigouya and Bobo, and between Sanguié and Houndé were negative, indicating no differentiation in the COI sequence (Lehmann et al., 1996). For all the other populations taken pairwise (e.g. Sanguié and Ouahigouya, Sanguié and Koudougou), the Fst value ranged from 0.00052 to 0.0263, indicating that these populations were differentiated by low gene flow (Table 6). In addition, isolation by distance was not significantly correlated with the geographical distance between populations (Pearson correlation  $R = 0.376$ ,  $P < .167$ ) (Supplementary materials A2). Furthermore, both IntNJ haplotype networks (Fig. 3) showed no distinct genetic structure pattern according to climatic or agricultural conditions. Haplotype 1 and haplotype 2 clustered 22 (14.6%) and 18 (11.9%) barcode sequences from all three climatic zones and the two agricultural pressures, respectively. Most *An. gambiae* complex haplotypes encompassed only one barcode sequence.

### 3.4. Analysis of molecular variance (AMOVA)

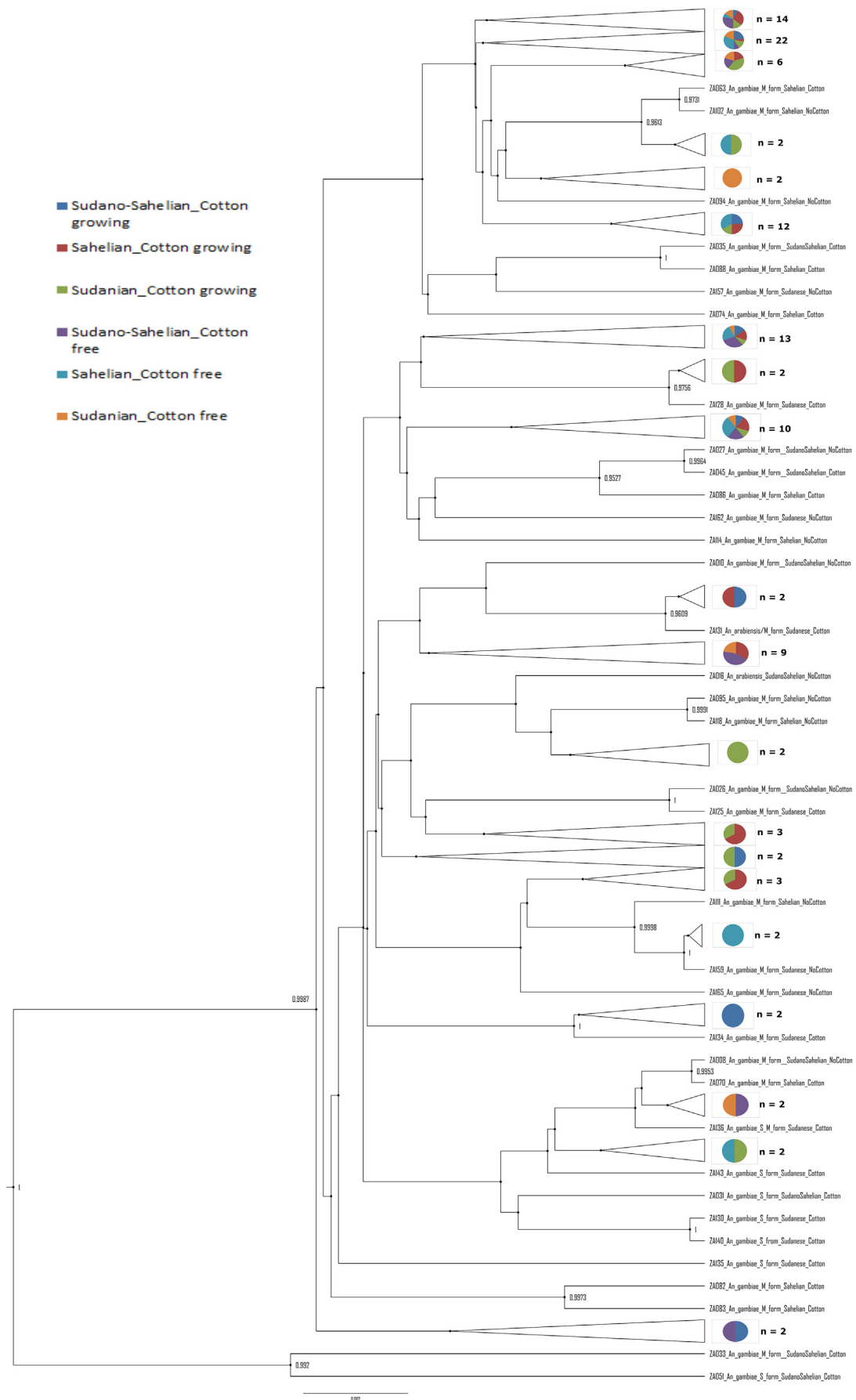
AMOVA results showed that 99.27% of the diversity was distributed within populations and 0.59% among populations, but the value was non-significant (Table 7). No significant difference in variance ( $F$  statistics) was observed, pointing to the absence of a genetic structure in the population. The low  $F_{ST}$  value (0.00723,  $P = .26$ ) of variance within populations indicated non-significant genetic variability, suggesting that the sampled populations were unstructured (Table 7).

## 4. Discussion

### 4.1. Distribution of *An. arabiensis*, *An. gambiae* s.s. and *An. coluzzii*

In Sub-Saharan Africa, *An. arabiensis*, *An. coluzzii* and *An. gambiae* s.s. are sympatric, with different occurrences in Burkina Faso (Costantini et al., 2009) according to climatic conditions (Coetzee et al., 2000). Among the seven species of the *An. gambiae* complex, our PCR-RFLP analysis from indoor and outdoor collections revealed three species: *An. gambiae* s.s. (21.11%), *An. coluzzii* (65.55%), and *An. arabiensis* (3.33%). In contrast, another study in Burkina Faso found *An. gambiae* in a higher proportion (51.7%) than *An. coluzzii* (21.6%) and *An. arabiensis* (26.7%) (Costantini et al., 2009). We found a smaller population of—largely exophilic—*An. Arabiensis* as compared to *An. gambiae* s.s. and *An. coluzzii*. Our results corroborate the findings of a previous study suggesting that *An. gambiae* s.s. dominates in the humid Sudanian region, while *An. arabiensis* is better represented in the Sahel steppes (Coetzee et al., 2000). But urbanisation in West Africa and Burkina Faso is linked to the distribution of *An. arabiensis*, so its distribution does not necessarily follow ecoclimatic regions (Markianos et al., 2016). In general, *An. gambiae* s.s. and *An. coluzzii* occupy wetlands (forest, wet savannah) but can be found in arid Sahelian regions bordering the Sahara, from Mauritania to Chad, Senegal, Mali, Niger and Nigeria. These two species are absent in the horn of Africa, southern Arabia and most of the Indian Ocean islands except in Madagascar, Comoros and Mauritius (Coetzee et al., 2000; Simard et al., 2009). Previous studies on the *An. gambiae* complex found *An. gambiae* s.s. and *An. coluzzii* to be more common in West Africa than in Central and Eastern Africa, while *An. arabiensis* had a wider range including the arid areas south of the Sahara and the horn of Africa (Kyalo et al., 2017). Furthermore, *An. gambiae* s.s. was found mostly dominant in savannah (Sudanian) regions, while *An. coluzzii* was typical of the Niger River delta (Simard et al., 2009) and predominant in shrublands (Sahelian region) (Lee et al., 2009), as they are in Burkina Faso and Mali. *An. coluzzii* has a decisive adaptive advantage in the dry season, while *An. gambiae* s.s. distribution is linked to its advantage in the rainy season (Bomblies, 2012).

*An. coluzzii* is dominant in southwestern Burkina Faso, while *An. gambiae* s.s. occurs in rice-growing areas in a relatively high proportion, with a maximum peak at 51% in October (end of the rainy season) (Dabiré et al., 2008). *An. coluzzii* is mainly found in the Sudano-Sahelian region and is highly predominant (93%) in the irrigated rice fields



(caption on next page)

**Fig. 2.** Collapsed BI tree constructed using 151 barcode sequences of *Anopheles gambiae* complex from Burkina Faso. For collapsed branches, pie charts (n = number of sequences) correspond to sequence proportion of the combination of the climatic locations and agricultural condition. Labels corresponded to a sample name followed by its PCR-RFLP species identification, its climatic locations and agricultural condition. Only posterior probabilities > 0.95 were displayed on nodes of the phylogeny.

area, while *An. gambiae* s.s. is found predominantly in the Sudanese zone (Namountougou et al., 2012). Migration models of polymorphism suggest that expansions of *An. gambiae* s.s. and *An. coluzzii* are related to agricultural pressure (Crawford and Lazzaro, 2010) while *An. arabiensis* has a high potential by responding to selective pressures from environmental changes and vector control efforts (Ishimura et al., 2015).

We found new hybrids between *An. gambiae* s.s. and *An. coluzzii* (7.78%) and between *An. arabiensis* and *An. coluzzii* (2.22%). Although hybridisation has been shown to be possible in many sympatric areas, natural hybrids between these three species of the *An. gambiae* complex are very rare (Pates et al., 2014; Pombi et al., 2017). A study in Uganda reported occasional hybridisation between *An. arabiensis*, *An. gambiae* s.s. and *An. coluzzii* in nature, showing that hybrid forms between these three species may occur (Weetman et al., 2014). In West Africa, hybridisation between *An. gambiae* s.s. and *An. coluzzii* is extremely low for the most part, with hybrid frequencies below 1%, comparable to hybrids between *An. gambiae* and *An. arabiensis* (Della Torre et al., 2005). Hybrids of all sympatric members of the *An. gambiae* complex have been encountered in nature with frequencies ranging from 0.02% to 0.76% (Besansky et al., 2003). All hybrid females are fertile, and some even show higher fertility than the female parent (Besansky et al., 2003). However, in our study, the frequency of hybrids between *An. gambiae* s.s. and *An. coluzzii* was 7.7%, and the frequency of hybrids between *An. arabiensis* and *An. coluzzii* was 2.22%. An earlier study in south-eastern Burkina Faso found no hybrid between *An. arabiensis* and *An. gambiae* s.s., but a 6% hybrid frequency between *An. coluzzii* and *An. gambiae* s.s. (Robert et al., 1989). In contrast, GOUNDRY, a recent cryptic subgroup of *An. gambiae* s.l. found in Burkina Faso (Crawford et al., 2016), displays a high rate of hybridisation (36%) (Riehle et al., 2011). GOUNDRY is a hybrid between *An. coluzzii* and *An. gambiae* s.s. and is more closely related to *An. coluzzii* (Crawford et al., 2016).

#### 4.2. Phylogenetic analysis

Phylogenetic analysis as a tool for taxonomic studies is useful in mosquito systematics when the sequences of the genome and molecular markers of the *An. gambiae* complex are available (Mohanty et al., 2009). The *Anopheles* genus has an impact on human evolution (i.e. sickle cell anaemia as a mode of resistance to malarial protozoa), but little work has been done to understand the evolution and phylogenetic relationships of these mosquitoes (Harbach, 2013). The topologies of the phylogenetic trees we obtained from the COI gene did not allow for a reasonable classification of *An. gambiae* s.s., *An. coluzzii* and *An. arabiensis*, as compared to other molecular techniques using the IGS of rDNA. Phylogenetic relationships between these sub-species are not fully delineated, especially when based on morphology or the

**Table 3**  
COI haplotype diversity and nucleotide diversity in different agro-climatic populations of *An. gambiae* complexes

Population code (Number of DNA Sequences)	Number of polymorphic (segregating) sites (S)	Number of haplotypes (h)	Haplotype diversity (Hd)	Average number of differences (K)	Nucleotide diversity (Pi (π))
KG <sup>a</sup> (29)	34	21	0.96305	4.24384	0.00620
OH <sup>b</sup> (29)	21	18	0.95567	3.51232	0.00513
SG <sup>b</sup> (25)	28	16	0.92667	3.17333	0.00464
KB <sup>a</sup> (31)	58	19	0.94194	5.88602	0.00861
HD <sup>a</sup> (28)	28	24	0.98942	3.82804	0.00560
BB <sup>b</sup> (18)	22	13	0.90196	3.61438	0.00528
Total (151)	102	74	0.95252	4.13310	0.00604

Each sample collection site had a unique population code: Kongoussi (KG); Ouahigouya (OH); Sanguié (SG); Kombissiri (KB); Houndé (HD) and Bobo (BB).

<sup>a</sup> : collection site cotton-growing and <sup>b</sup>: collection site cotton-free.

**Table 4**

Summary of neutrality tests in different geographic populations of the *An. gambiae* complex.

Population code	Fu and Li's D <sup>a</sup>	Fu and Li's F <sup>a</sup>	Fu's Fs	Tajima's D
KG <sup>a</sup>	-2.73374*	-2.95469*	-14.297	-2.18264**
OH <sup>b</sup>	-0.80365	-1.13087	-9.725	-1.30076
SG <sup>b</sup>	-2.89057*	-3.04071*	-10.123	-1.93802*
KB <sup>a</sup>	-3.02537*	-3.27671**	-8.765	-2.26197**
HD <sup>a</sup>	-2.44943	-2.63967*	-22.564	-1.80675
BB <sup>b</sup>	-2.86949**	-3.03067**	-7.697	-1.98661*
Total	-4.42854**	-4.27649**	-91.382	-2.52597***

Statistical significance: \*,  $P < .05$ ; \*\*,  $P < .02$ ; \*\*\*,  $P < .01$ . <sup>a</sup>: collection site cotton-growing and <sup>b</sup>: collection site cotton-free.

**Table 5**

Pairwise average number of nucleotide differences (Kxy) (above the diagonal) and nucleotide divergence (Dxy) (below the diagonal) between geographic populations of the *An. gambiae* complex.

	KG	OH	SG	KB	HD	BB
KG <sup>a</sup>		3.86801	3.73379	5.04227	4.04187	3.90996
OH <sup>b</sup>	0.00565		3.4331	4.71302	3.7032	3.51533
SG <sup>b</sup>	0.00546	0.00502		4.59097	3.47857	3.44889
KB <sup>a</sup>	0.00737	0.00689	0.00671		4.91244	4.75269
HD <sup>a</sup>	0.00591	0.00541	0.00509	0.00718		3.75000
BB <sup>b</sup>	0.00572	0.00514	0.00504	0.00695	0.00548	

<sup>a</sup>: collection site cotton-growing and <sup>b</sup>: collection site cotton-free.

mitogenome (Harbach, 2004; Hanemaaijer et al., 2019). The considerable sequence similarity, common genetic elements and common ancestral molecular polymorphisms of the *An. gambiae* complex make it difficult to determine the direction of its evolution (Besansky et al., 1994; Mohanty et al., 2009). The mitochondrial gene COI is commonly sequenced for *Anopheles* phylogeny (Ratnasingham and Hebert, 2007; Mohanty et al., 2009). In Nigeria, Matthews et al. (2007) used the COI gene to assess *Anopheles* population genetic structures and found a trend towards decreasing genetic diversity in *An. arabiensis* from the northern savannah to the southern rainforest (Matthews et al., 2007). Phylogenetic analysis is used as a tool for dynamic population studies because the genome sequences and molecular markers of the *An. gambiae* complex are now available (Mohanty et al., 2009). Therefore, gene flows and population structures are built based on allelic distribution and dispersal of genes of interest, such as the mitochondrial gene COI and the internal transcribed spacer 2 (ITS2) rDNA region (Sum et al., 2014; Norris and Norris, 2015). Their spatial structure analysis is therefore necessary to ensure the success of selective vector

**Table 6**

Pairwise *F*<sub>st</sub> values (above the diagonal) and the natural logarithm of geographical distance (km) (below the diagonal) between populations of the *An. gambiae* complex.

	KG	OH	SG	KB	HD	BB
KG <sup>a</sup>						
OH <sup>b</sup>	4.6001576	−0.0026	0.00675	−0.00449	0.00147	−0.0049
SG <sup>b</sup>	5.0205856	4.92870191	0.0263	0.00294	0.00892	−0.01366
KB <sup>a</sup>	5.2034571	5.30727711	4.8291134	0.01335	−0.00636	0.01596
HD <sup>a</sup>	5.6947417	5.55914198	4.9917922	5.50655031	0.01128	0.00052
BB <sup>b</sup>	5.9542818	5.81740873	5.4705886	5.82422863	4.5282891	0.00768

<sup>a</sup>: collection site cotton-growing and <sup>b</sup>: collection site cotton-free.

control campaigns and to predict epidemiological risks (Stevenson and Norris, 2017).

The COI is a mitochondrial protein-coding gene that was used to infer the phylogeny of *Anophelinae* (Sallum et al., 2002). This gene encompasses the barcode region that was used for identifying sibling and cryptic species of the genus *Anopheles* (Wang et al., 2017; Mohanty et al., 2009). However, the phylogenetic relationships among *An. gambiae* s.s., *An. coluzzii* and *An. arabiensis* are not fully delineated although their mitochondrial DNA (mtDNA) genome sequences are available (Hanemaaijer et al., 2019). In fact, *An. gambiae* s.s., *An. coluzzii* and *An. arabiensis* are a monophyletic taxon (Benson and Falade, 2013; Kamali et al., 2014), even though the subdivision of *An. coluzzii* into distinct West, Central, and Southern African genetic clusters suggests restrictions to gene flow (Pinto et al., 2013).

#### 4.3. Population genetic structure

Geographical distance between populations may lead to genetic differentiation; therefore, we considered each climatic region or cotton-growing area as a basic reproductive unit. We observed non-significant intraspecific nucleotide variation and nucleotide diversity ( $\pi$ ) in *An. gambiae* complex populations for the mitochondrial (COI) gene sequences. Our *F*<sub>st</sub> values varied from −0.01366 to 0.0263 across all populations; *F*<sub>st</sub> values close to zero show a free gene flow and the absence of differentiation in the *An. gambiae* complex populations. Furthermore, our neutrality indices calculated using Tajima's D, Fu and Li's D\*, Fu and Li's F\*, Fu's Fs tests were significantly negative in all populations except Ouahigouya. A significantly negative value ( $P < .05$ ) of D and Fs is expected in cases of population expansion, purifying selection or selective scanning (Tajima, 1983).

AMOVA indicated that a high proportion of the total genetic variance was due to variation within populations (99.27%), but the value was non-significant. These results indicate that the distribution of genetic variance between the sampling sites of each climate region did not take significant genetic differentiation between *Anopheles gambiae* populations into account (Sum et al., 2014). These observations suggest interspecies gene flow among *An. coluzzii*, *An. gambiae* s.s. and *An. arabiensis* populations. Further gene analysis by Fontaine et al. (2016) showed that *An. gambiae* s.s. was more closely related to *An. arabiensis*.

There was no isolation by distance, and the haplotype network showed no distinct genetic structure. Although the locations sampled in each climatic region were separated by a maximum of 90 km, these results reveal that there are no barriers to gene flow for mosquitoes from Burkina Faso. This high level of gene flow can be explained either by the wide flight range of the mosquitoes (up to seven km in a lifetime) or by the extensive human-associated migration throughout Burkina Faso (Donnelly et al., 2002; Chen et al., 2004). Our results are similar to those found in Mali, where interspecies gene flow was low among villages (Taylor et al., 2001), and they are consistent with the majority of previous studies on *An. gambiae* in similar regions of Africa (Lehmann et al., 2003; Prugnolle et al., 2008). In western Kenya, low genetic differentiation was found between island and mainland populations. The authors explained that wind or human-assisted dispersal probably

caused a large extent of the gene flow between the populations (Chen et al., 2004).

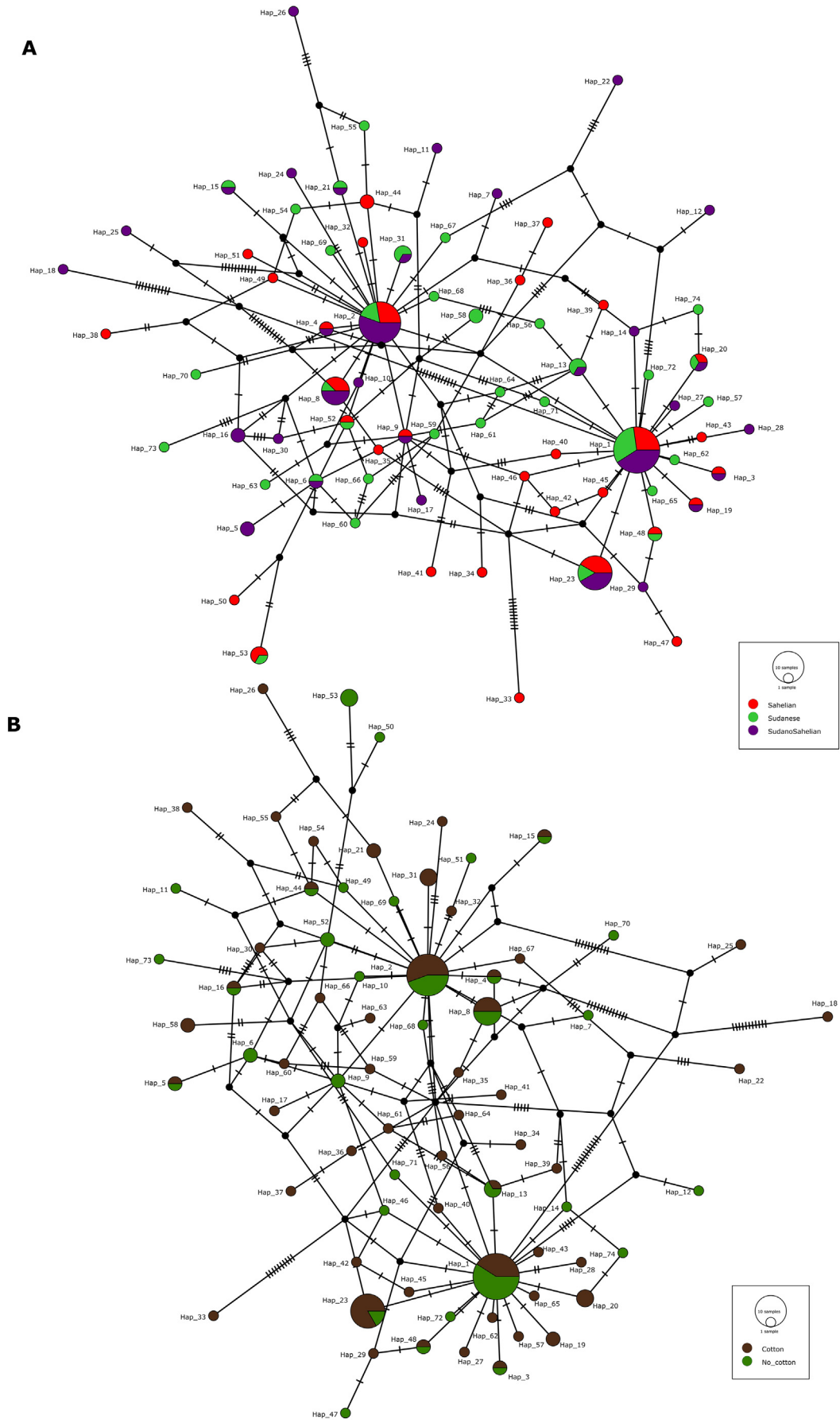
In the Nyanza province of Kenya, *An. gambiae* populations have been found only very slightly differentiated genetically across collection sites (two populations with a maximum distance of about 68 km) (Prugnolle et al., 2008). In a malaria endemic region of southern Tanzania, high levels of genetic differentiation and genetic divergence among *An. arabiensis* populations have been observed, as opposed to *An. gambiae* s.s. which displayed no differentiation. However, it appeared that genetic differentiation was not attributed to physical barriers or distance but possibly to ecological diversification (Ng'Habi et al., 2011). In Cameroon (Central Africa), both *An. coluzzii* and *An. gambiae* s.s. have been found to occur in sympatry, with low geographic differentiation despite a minimum geographical distance of 35 km between collection sites (Wondji et al., 2002). All these authors worked on microsatellites, while we worked on the COI gene sequence (Wondji et al., 2002; Prugnolle et al., 2008; Borja-Cacho and Matthews, 2010; Ng'Habi et al., 2011). Furthermore, these previous studies showed low genetic differentiation between some high-transmission sites in Western and Eastern Africa although they are separated by thousands of kilometres with a large local population of *An. gambiae* in these areas (Prugnolle et al., 2008). Reidenbach et al. (2012) concluded that *An. coluzzii*, *An. gambiae* s.s. and *An. arabiensis* are evolving collectively on independent evolutionary trajectories through the analysis of 400,000 single nucleotide polymorphisms (SNPs) across the genomes of paired population samples from Mali, Burkina Faso and Cameroon (Coetzee et al., 2013). Based on population genomic evidence from whole genome sequencing and SNP genotyping, *An. coluzzii* and *An. gambiae* s.s. are cohesive and exclusive taxonomic groups across their shared range. This may contribute substantially to the lack of gene flow (Coetzee et al., 2013). Lastly, many studies on gene flow across Africa suggest that the *An. gambiae* complex exchanges individuals at a rate sufficient in magnitude to prevent them from diverging genetically. There is no 'isolation by distance', hence, no relationship between the levels of genetic divergence (*F*<sub>st</sub>) and geographic distance (Lanzaro and Tripet, 2009). However, Miles et al. (2017) recently sequenced the genomes of 765 specimens of *An. gambiae* and *An. coluzzii* sampled from 15 locations across Africa, and identified over 50 million SNPs. Contrary to our study, their data revealed high levels of genetic diversity, a complex population structure, and patterns of gene flow, with evidence of ancient expansions, recent bottlenecks, and local variation in the effective population (Miles et al., 2017).

The limitations of this study are that nuclear markers of phylogeny such as nuclear ribosomal internal transcribed spacer 2 (ITS2) could be used instead of the COI marker to determine the extent of population structure and gene flow. Additionally, the present study concerns only one country (Burkina Faso). It will gain in robustness by extending sampling to several malaria-affected countries or localities in Sub-Saharan Africa.

## 5. Conclusions

This study was carried out in Burkina Faso to assess the taxonomic





(caption on next page)

**Fig. 3.** Haplotype networks ( $n = 74$ ) of COI from the *Anopheles gambiae* complex in Burkina Faso. The haplotype networks show genetic distance between two sequences. Each coloured circle corresponds to haplotype; the circle size displays the number of the sequence, and the colour displays the origin where the sequence was generated: (A) according to climatic region of Burkina Faso (i.e. Sahelian, Sudanian, Sudano-Sahelian) and (B) according to agriculture (i.e. cotton-growing and cotton-free areas). Each black circle corresponds to missing haplotypes. Each edge between two haplotypes corresponds to single changes between sequences, which could be increased according to the number of straps on the edge.

**Table 7**

Analysis of molecular variance (AMOVA) of spatial variation of mitochondrial DNA COI gene sequence in 6 populations of *An. gambiae* complex.

Source of variation	Df	Sum of squares	Variance components	Percentage of variation	F-statistics	P-value
Among groups	2	8.092	0.004 <i>V<sub>a</sub></i>	0.140	$F_{CT} = 0.001$	0.34
Among populations within groups	3	11.408	0.019 <i>V<sub>b</sub></i>	0.590	$F_{SC} = 0.006$	0.12
Within populations	145	481.096	3.317 <i>V<sub>c</sub></i>	99.270	$F_{ST} = 0.007$	0.26
Total	150	500.596	3.342	100	–	–

*F<sub>CT</sub>*, *F<sub>SC</sub>* and *F<sub>ST</sub>* are the F-statistics. *F<sub>CT</sub>* denotes genetic differentiation among groups; *F<sub>SC</sub>* denotes genetic differentiation among populations within groups; *F<sub>ST</sub>* denotes genetic differentiation between populations among groups. Groups (three) refer to collected localities: Kongoussi (KG) and Ouahigouya (OH); Sanguié (SG) and Kombissiri (KB); Houndé (HD) and Bobo (BB). *V<sub>a</sub>*, *V<sub>b</sub>*, *V<sub>c</sub>* are the associate variance components.

and genetic structure of the *Anopheles gambiae* complex, a major vector of malaria. Our genomic approaches revealed the distribution of *An. gambiae* s.s., *An. coluzzii* and *An. arabiensis* populations according to selected agro-climatic areas. Using the COI gene, no genetic differentiation was evidenced in the *An. gambiae* complex in Burkina Faso. However, this scientific contribution can not only be used as a basis for in-depth genetic diversity studies of the *An. gambiae* complex in Burkina Faso, but also extended to neighbouring countries to further refine the model and control of malaria transmission.

#### Availability of data and materials

Data generated or analysed during this study is included in this published article and its supplementary information files. All datasets are available from the corresponding author on reasonable request.

#### Authors' contributions

AAZ, AB and FF conceived and designed the study. AAZ, AS, ZS and AB supervised the data and sample collection. AAZ performed practical work and wrote the manuscript. AAZ, GN, AB and FF analysed and interpreted the results. GN performed haplotype networks analysis and phylogenetic reconstruction. AB, GN and FF revised the manuscript. All authors were major contributors in writing the manuscript. All authors read and approved the final manuscript.

#### Ethics statement and consent to participate

This study was approved by the National Health Ethic Committee in Burkina Faso (reference number 2017–9-143 on 12 September 2017). For mosquito collection in residential areas, written and informed consent was obtained from homeowners in each location.

#### Declaration of Competing Interest

The authors declare that they have no competing interests.

#### Acknowledgements

We are sincerely grateful to all the inhabitants of the mosquito collection villages for participating in this study. We acknowledge Nicolas Poncelet who helped in PCR analysis. We also thank the mosquito collectors (Martin Tientiga et al.) and Mr. Dramé, who dissected the mosquitoes, for their valuable contribution to the successful completion of the study. We also acknowledge Dr. Sylvie Cornélius from MiVEGEC (Maladies infectieuses et Vecteurs, Ecologie, Génétique, Evolution et Contrôle) and Institut de Recherche pour le

Développement (IRD) for graciously giving us the KIMUSU strain (insecticide-sensitive) of *Anopheles gambiae* as a positive control in PCR in our study. Finally, we sincerely thank Prof. Kevin J. Emerson for his helpful suggestions and also the second reviewer.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.meegid.2020.104261>.

#### References

- Abiodun, G.J., Maharaj, R., Witbooi, P., Okosun, K.O., 2016. Modelling the influence of temperature and rainfall on the population dynamics of *Anopheles arabiensis*. *Malar. J.* 15, 1–15. <https://doi.org/10.1186/s12936-016-1411-6>.
- Benson, O., Falade, M.O., 2013. Molecular evolution and phylogenomics of the *Anopheles gambiae* complex. *J. Mosq. Res.* 3, 65–70. <https://doi.org/10.5376/jmr.2013.03.0009>.
- Besansky, N.J., Powell, J.R., Caccone, A., Hamm, D.M., Scott, J.A., Collins, F.H., 1994. Molecular phylogeny of the *Anopheles gambiae* complex suggests genetic introgression between principal malaria vectors. *Proc. Natl. Acad. Sci. U. S. A.* 91, 6885–6888. <https://doi.org/10.1073/pnas.91.15.6885>.
- Besansky, N.J., Krzywinski, J., Lehmann, T., Simard, F., Kern, M., Mukabayire, O., Fontenille, D., Touré, Y., Sagnon, N., 2003. Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: evidence from multilocus DNA sequence variation. *Proc. Natl. Acad. Sci. U. S. A.* 100, 10818–10823. <https://doi.org/10.1073/pnas.1434337100>.
- Bombliès, A., 2012. Modeling the role of rainfall patterns in seasonal malaria transmission. *Clim. Chang.* 112, 673–685. <https://doi.org/10.1007/s10584-011-0230-6>.
- Borja-Cacho, D., Matthews, J., 2008. The molecular forms of *Anopheles gambiae*: a phenotypic perspective. *Infect. Genet. Evol.* 6, 737–746. <https://doi.org/10.1016/j.meegid.2008.06.003>.
- Borja-Cacho, D., Matthews, J., 2010. Population structure of *Anopheles gambiae* along the Kenyan coast. *Acta Trop.* 114, 103–108. <https://doi.org/10.1016/j.actatropica.2010.01.011>.
- Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* 10, 1–6. <https://doi.org/10.1371/journal.pcbi.1003537>.
- Carnevale, P., Robert, V., 2009. Les anophèles Biologie, transmission du *Plasmodium* et lutte antivectorielle. In: IRD éditions, <https://doi.org/10.4000/books.irdeditions.10374>.
- Chen, H., Minakawa, N., Beier, J., Yan, G., 2004. Population genetic structure of *Anopheles gambiae* mosquitoes on Lake Victoria islands, west Kenya. *Malar. J.* 3, 1–8. <https://doi.org/10.1186/1475-2875-3-48>.
- Coetzee, M., Craig, M., Le Sueur, D., 2000. Distribution of African malaria mosquitoes belonging to the *Anopheles gambiae* complex. *Parasitol. Today* 16, 74–77. [https://doi.org/10.1016/S0169-4758\(99\)01563-X](https://doi.org/10.1016/S0169-4758(99)01563-X).
- Coetzee, M., Hunt, R.H., Wilkerson, R., Torre, A. Della, Coulibaly, M.B., Besansky, N.J., 2013. *Anopheles coluzzii* and *Anopheles amharicus*, new members of the *Anopheles gambiae* complex. *Zootaxa* 3619, 246–274. <https://doi.org/10.11646/zootaxa.3619.3.2>.
- Costantini, C., Ayala, D., Guelbeogo, W.M., Pombi, M., Some, C.Y., Bassole, I.H.N., Ose, K., Fotsing, M., Sagnon, N.F., Fontenille, D., Besansky, N.J., Simard, F., 2009. Living at the edge: biogeographic patterns of habitat segregation conform to speciation by niche expansion in *Anopheles gambiae*. *BMC Ecol.* 9, 1–27. <https://doi.org/10.1186/1472-6785-9-16>.
- Crawford, J.E., Lazzaro, B.P., 2010. The demographic histories of the M and S molecular

- forms of *Anopheles gambiae* s.s. *Mol. Biol. Evol.* 27, 1739–1744. <https://doi.org/10.1093/molbev/msq070>.
- Crawford, J.E., Riehle, M.M., Markianos, K., Bischoff, E., Guelbeogo, W.M., Gneme, A., Sagnon, N., Vernick, K.D., Nielsen, R., Lazzaro, B.P., 2016. Evolution of GOUNDRY, a cryptic subgroup of *Anopheles gambiae* s.l., and its impact on susceptibility to *Plasmodium* infection. *Mol. Ecol.* 25, 1494–1510. <https://doi.org/10.1111/mec.13572>.
- Dabiré, K.R., Diabaté, A., Djogbenou, L., Ouari, A., N'Guessan, R., Ouédraogo, J.B., Hougaard, J.M., Chandre, F., Baldet, T., 2008. Dynamics of multiple insecticide resistance in the malaria vector *Anopheles gambiae* in a rice growing area in South-Western Burkina Faso. *Malar. J.* 7, 1–9. <https://doi.org/10.1186/1475-2875-7-188>.
- Dabiré, R.K., Namountougou, M., Sawadogo, S.P., Yaro, L.B., Toé, H.K., Ouari, A., Gouagna, L.C., Simard, F., Chandre, F., Baldet, T., Bass, C., Diabaté, A., 2012. Population dynamics of *Anopheles gambiae* s.l. in Bobo-Dioulasso city: Bionomics, infection rate and susceptibility to insecticides. *Parasit. Vectors* 5, 1–9. <https://doi.org/10.1186/1756-3305-5-127>.
- Darriet, F., Corbel, V., 2008. Influence des engrais de type NPK sur l'oviposition d'*Aedes aegypti*. *Parasite* 15, 89–92.
- Darriet, F., Rossignol, M., Chandre, F., 2012. The combination of NPK fertilizer and deltamethrin insecticide favors the proliferation of pyrethroid-resistant *Anopheles gambiae* (Diptera: Culicidae). *Parasite* 19, 159–164. <https://doi.org/10.1051/parasite/2012192159>.
- Della Torre, A., Fanello, C., Akogbeto, M., Dossou-yovo, J., Favia, G., Petrarca, V., Coluzzi, M., 2001. Molecular evidence of incipient speciation within *Anopheles gambiae* s.s. in West Africa. *Insect Mol. Biol.* 10, 9–18. <https://doi.org/10.1046/j.1365-2583.2001.00235.x>.
- Della Torre, A., Tu, Z., Petrarca, V., 2005. On the distribution and genetic differentiation of *Anopheles gambiae* s.s. molecular forms. *Insect Biochem. Mol. Biol.* 35, 755–769. <https://doi.org/10.1016/j.ibmb.2005.02.006>.
- Diallo, Mouctar, Sangaré, D., Traoré, A., Mariko, M., Diarra, S., Diallo, Moussa, Coulibaly, B., Dao, A., Doumbo, O., Traoré, C.F., 2015. Etude de l'impact des gîtes larvaires sur l'inféctivité des gamétocytes de *Plasmodium falciparum* chez *Anopheles gambiae* s.l. en zone d'endémie palustre de Nanguilabougou-Mali. *Mali Med.* 30, 28–33.
- Donnelly, M.J., Simard, F., Lehmann, T., 2002. Evolutionary studies of malaria vectors. *Trends Parasitol.* 18, 75–80. [https://doi.org/10.1016/S1471-4922\(01\)02198-5](https://doi.org/10.1016/S1471-4922(01)02198-5).
- Excoffier, L., Laval, G., Schneider, S., 2005. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol. Bioinformatics Online* 1, 47–50.
- Fanello, C., Santolamazza, F., Della Torre, A., 2002. Simultaneous identification of species and molecular forms of the *Anopheles gambiae* complex by PCR-RFLP. *Med. Vet. Entomol.* 16, 461–464. <https://doi.org/10.1046/j.1365-2915.2002.00393.x>.
- Favia, G., Lanfrancotti, A., Spanos, L., Sidén-Kiamos, I., Louis, C., 2001. Molecular characterization of ribosomal DNA polymorphisms discriminating among chromosomal forms of *Anopheles gambiae* s.s. *Insect Mol. Biol.* 10, 19–23. <https://doi.org/10.1046/j.1365-2583.2001.00236.x>.
- Fontaine, M.C., Waterhouse, R.M., Steele, A., Pease, J.B., Neafsey, D.E., Sharakhov, I.V., Jiang, X., Hall, A.B., Kakani, E., Mitchell, S.N., Wu, Y., Smith, H.A., Hahn, M.W., Besansky, N.J., 2016. Extensive introgression in a malaria vector species complex revealed by phylogenomics. *Science* (80-) 347, 1–20. <https://doi.org/10.1126/science.1258524>. Extensive.
- Gillies, M.T., Coetzee, M., 1987. A Supplement to the Anophelinae of Africa South of the Sahara. *Publ. South African Inst. Med. Res.* 55, 63. <https://doi.org/10.1046/j.1365-294X.1997.00177.x>.
- Gillies, M.T., De Meillon, B., 1968. The Anophelinae of Africa South of the Sahara (Ethiopian Zoogeographical Region). *South African Institute for Medical Research, P.O. Box 1038, S. Africa, Johannesburg.*
- Hall, T.A., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* 41, 95–99.
- Hanemaaijer, M.J., Houston, P.D., Collier, T.C., Norris, L.C., Fofana, A., Lanzaro, G.C., Cornel, A.J., Lee, Y., 2019. Mitochondrial genomes of *Anopheles arabiensis*, *An. gambiae* and *An. coluzzii* show no clear species division. *F1000Research* 7, 1–14. <https://doi.org/10.12688/f1000research.13807.2>.
- Harbach, R.E., 2004. The classification of genus *Anopheles* (Diptera: Culicidae): a working hypothesis of phylogenetic relationships. *Bull. Entomol. Res.* 94, 537–553. <https://doi.org/10.1079/BER2004321>.
- Harbach, R.E., 2013. The Phylogeny and Classification of *Anopheles*, *Anopheles* Mosquitoes—New Insights into Malaria Vectors. <https://doi.org/10.5772/3392>.
- Harbach, R.E., Kitching, I.J., 2016. The phylogeny of anophelinae revisited: inferences about the origin and classification of *Anopheles* (Diptera: Culicidae). *Zool. Scr.* 45, 34–47. <https://doi.org/10.1111/zsc.12137>.
- Harrison, J.C., Langdale, J.A., 2006. A step by step guide to phylogeny reconstruction. *Plant J.* 45, 561–572. <https://doi.org/10.1111/j.1365-313X.2005.02611.x>.
- Hauchart, V., 2007. Le Burkina Faso, un producteur de coton face à la mondialisation et à la dépendance économique. *Regard sur un Sud*. *Cybergeo Eur. J. Geogr.* 0–10. <https://doi.org/10.4000/cybergeo.2665>.
- Hegde, S., Rasgon, J.L., Hughes, G.L., 2015. The microbiome modulates arbovirus transmission in mosquitoes. *Curr. Opin. Virol.* 15, 97–102. <https://doi.org/10.1016/j.coviro.2015.08.011>.
- Hien, A.S., Soma, D.D., Hema, O., Bayili, B., Namountougou, M., Gnankiné, O., Baldet, T., Diabaté, A., Dabiré, K.R., 2017. Evidence that agricultural use of pesticides selects pyrethroid resistance within *Anopheles gambiae* s.l. populations from cotton growing areas in Burkina Faso, West Africa. *PLoS One* 12, 1–15. <https://doi.org/10.1371/journal.pone.0173098>.
- Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* 23, 254–267. <https://doi.org/10.1093/molbev/msj030>.
- IFPRI, R., 2010. Diagnostic du secteur agricole de pays: dépenses publiques. pp. 2010.
- INSD, 2015. Annuaire statistiques 2014. Institut national de la statistique et de la démographie. Ouagadougou, Burkina Faso.
- Ishimura, R., Nagy, G., Dotu, I., Zhou, H., Yang, X., Senju, S., Nishimura, Y., Chuang, J.H., Susan, L., 2015. Population Genetic Structure of *Anopheles Arabiensis* (Diptera: Culicidae) in a Rice Growing Area of Central Kenya. *J. Med. Entomol.* 345, 455–459. <https://doi.org/10.1126/science.1249749>. Ribosome.
- Kabore, B., Kam, S., Ouédraogo, G.W.P., Bathiebo, D.J., 2017. Etude de l'évolution climatique au Burkina Faso de 1983 à 2012 : cas des villes de Bobo dioulasso. *Ouagadougou et Dori. Arab. J. Earth Sci.* 4, 50–59.
- Kamali, M., Marek, P.E., Peery, A., Antonio-Nkondjio, C., Ndo, C., Tu, Z., Simard, F., Sharakhov, I.V., 2014. Multigene phylogenetics reveals temporal diversification of major African malaria vectors. *PLoS One* 9, 1–9. <https://doi.org/10.1371/journal.pone.0093580>.
- Kyalo, D., Amratia, P., Mundia, C.W., Mbogo, C.M., Coetzee, M., Snow, R.W., 2017. A geo-coded inventory of anophelines in the Afrotropical Region south of the Sahara: 1898–2016. *Wellcome Open Res.* 2, 57. <https://doi.org/10.12688/wellcomeopenres.12187.1>.
- Lanfear, R., Calcott, B., Ho, S.Y.W., Guindon, S., 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses research article. *Mol. Biol. Evol.* 29, 1695–1701. <https://doi.org/10.1093/molbev/ms020>.
- Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol.* 34, 772–773. <https://doi.org/10.1093/molbev/msw260>.
- Lanzaro, G.C., 2009. Gene flow among populations of *Anopheles gambiae*: a critical review. *Crit. Rev.* 61, 713–722. <https://doi.org/10.1002/iub.196>.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Willm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. <https://doi.org/10.1093/bioinformatics/btm404>.
- Lee, Y., Cornel, A.J., Meneses, C.R., Fofana, A., Andrianarivo, A.G., Mcabee, R.D., Fondjo, E., Traoré, S.F., Lanzaro, G.C., 2009. Ecological and genetic relationships of the Forest-M form among chromosomal and molecular forms of the malaria vector *Anopheles gambiae* sensu stricto. *Malar. J.* 8, 1–15. <https://doi.org/10.1186/1475-2875-8-75>.
- Lehmann, T., Hawley, W.A., Kamau, L., Fontenille, D., Simard, F., Collins, F.H., 1996. Genetic differentiation of *Anopheles gambiae* populations from East and West Africa: Comparison of microsatellite and allozyme loci. *Heredity* (Edinb). 77, 192–200. <https://doi.org/10.1038/hdy.1996.124>.
- Lehmann, T., Licht, M., Elissa, N., Maega, B.T.A., Chimumbwa, J.M., Watsenga, F.T., Wondji, C.S., Simard, F., Hawley, W.A., 2003. Population structure of *Anopheles gambiae* in Africa. *J. Hered.* 94, 133–147. <https://doi.org/10.1093/jhered/esh024>.
- Leigh, J.W., Bryant, D., 2015. POPART: full-feature software for haplotype network construction. *Methods Ecol. Evol.* 6, 1110–1116. <https://doi.org/10.1111/2041-210X.12410>.
- Levine, R.S., Peterson, A.T., Benedict, M.Q., 2004. Geographic and ecologic distributions of the *Anopheles gambiae* complex predicted using a genetic algorithm. *Am. J. Trop. Med. Hyg.* 70, 105–109. <https://doi.org/10.1093/ajtmph/70.2.105>.
- Librado, P., Rozas, J., 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 25, 1451–1452.
- Maddison, D., Swofford, D.L., Maddison, W., 1997. NEXUS: an Extensible File Format for Systematic Information NEXUS: an extensible file format for systematic information. *Syst. Biol.* 46, 590–621. <https://doi.org/10.1093/sysbio/46.4.590>.
- Markianos, K., Bischoff, E., Mitri, C., Guelbeogo, W.M., Gneme, A., Eiglmeier, K., Holm, I., Sagnon, N., Vernick, K.D., Riehle, M.M., 2016. Genetic structure of a local population of the *Anopheles gambiae* complex in Burkina Faso. *PLoS One* 11, 1–16. <https://doi.org/10.1371/journal.pone.0145308>.
- Matthews, S.D., Meehan, L.J., Onyabe, D.Y., Vaineis, J., Nock, I., Ndams, I., Conn, J.E., 2007. Evidence for late Pleistocene population expansion of the malarial mosquitoes, *Anopheles arabiensis* and *Anopheles gambiae* in Nigeria. *Med. Vet. Entomol.* 21, 358–369. <https://doi.org/10.1111/j.1365-2915.2007.00703.x>.
- Miles, A., Harding, N.J., Bottà, G., Clarkson, C.S., Antão, T., Kozak, K., Schrider, D.R., Kern, A.D., Redmond, S., Sharakhov, I., Pearson, R.D., Bergey, C., Fontaine, M.C., Donnelly, M.J., Lawniczak, M.K.N., Ayala, D., Besansky, N.J., Burt, A., Caputo, B., Torre, A., Della, Godfray, H.C.J., Hahn, M.W., Midega, J., Neafsey, D.E., O'Loughlin, S., Pinto, J., Riehle, M.M., Vernick, K.D., Weetman, D., Wilding, C.S., White, B.J., Troco, A.D., Diabaté, A., Costantini, C., Rohatgi, K.R., Elissa, N., Coulibaly, B., Dinis, J., Mbogo, C., Bejon, P., Mawejje, H.D., Stalker, J., Rockett, K., Drury, E., Mead, D., Jeffreys, A., Hubbard, C., Rowlands, K., Isaacs, A.T., Jyothi, D., Malangone, C., Vauterin, P., Jeffery, B., Wright, I., Hart, L., Kluczyński, K., Cornelius, V., Macinnis, B., Henrichs, C., Giacomantonio, R., Kwiatkowski, D.P., 2017. Genetic diversity of the African malaria vector *Anopheles gambiae*. *Nature* 552, 96–100. <https://doi.org/10.1038/nature24995>.
- Minakawa, N., Sonye, G., 2002. The effects of climatic factors on the distribution and abundance of malaria vectors in Kenya. *J. Med. Entomol.* 39, 833–841. <https://doi.org/10.1603/0022-2585-39.6.833>.
- Mohanty, A., Swain, S., Kar, S.K., Hazra, R.K., 2009. Analysis of the phylogenetic relationship of *Anopheles* species, subgenus *Celia* (Diptera: Culicidae) and using it to define the relationship of morphologically similar species. *Infect. Genet. Evol.* 9, 1204–1224. <https://doi.org/10.1016/j.meegid.2009.06.021>.
- Namountougou, M., Simard, F., Baldet, T., Diabaté, A., Ouédraogo, J.B., Martin, T., Dabiré, R.K., 2012. Multiple insecticide resistance in *Anopheles gambiae* s.l. Populations from Burkina Faso, West Africa. *PLoS One* 7. <https://doi.org/10.1371/journal.pone.0048412>.
- Nei, M., 1987. *Molecular Evolutionary Genetics*.

- NgHabi, K.R., Knols, B.G., Lee, Y., Ferguson, H.M., Lanzaro, G.C., 2011. Population genetic structure of *Anopheles arabiensis* and *Anopheles gambiae* in a malaria endemic region of southern Tanzania. *Malar. J.* 10, 289. <https://doi.org/10.1186/1475-2875-10-289>.
- Norris, L.C., Norris, D.E., 2015. Phylogeny of anopheline (Diptera: Culicidae) species in southern Africa, based on nuclear and mitochondrial genes. *J. Vector Ecol.* 40, 16–27. <https://doi.org/10.1111/jvec.12128>. Phylogeny.
- Oshaghi, M.A., Yaaghoobi, F., Abaie, M.R., 2006. Pattern of mitochondrial DNA variation between and within *Anopheles stephensi* (Diptera: Culicidae) biological forms suggests extensive gene flow. *Acta Trop.* 99, 226–233. <https://doi.org/10.1016/j.actatropica.2006.08.005>.
- Pates, H.V., Curtis, C.F., Takken, W., 2014. Hybridization studies to modify the host preference of *Anopheles gambiae*. *Med. Vet. Entomol.* 28, 68–74. <https://doi.org/10.1111/mve.12070>.
- Pinto, J., Egyir-Yawson, A., Vicente, J., Gomes, B., Santolamazza, F., Moreno, M., Charlwood, J., Simard, F., Elissa, N., Weetman, D., Donnelly, M., Caccone, A., Della Torre, A., 2013. Geographic population structure of the African malaria vector *Anopheles gambiae* suggests a role for the forest-savannah biome transition as a barrier to gene flow. *Evol. Appl.* 6, 910–924. <https://doi.org/10.1111/eva.12075>.
- Pombi, M., Kengne, P., Gimonneau, G., Tene-Fossog, B., Ayala, D., Kamdem, C., Santolamazza, F., Guelbeogo, W.M., Sagnon, N., Petrarca, V., Fontenille, D., Besansky, N.J., Antonio-Nkondjio, C., Dabiré, R.K., Della Torre, A., Simard, F., Costantini, C., 2017. Dissecting functional components of reproductive isolation among closely related sympatric species of the *Anopheles gambiae* complex. *Evol. Appl.* 10, 1102–1120. <https://doi.org/10.1111/eva.12517>.
- Posada, D., 2003. Using MODELTEST and PAUP \* to select a model of nucleotide substitution. *Curr. Protoc. Bioinformatics* 00, 6.5.1–6.5.14. <https://doi.org/10.1002/0471250953.bi0605s00>.
- Prugnolle, F., Durand, P., Jacob, K., Razakandrainibe, F., Arnathau, C., Villarreal, D., Rousset, F., de Meeûs, T., Renaud, F., 2008. A comparison of *Anopheles gambiae* and *Plasmodium falciparum* genetic structure over space and time. *Microbes Infect.* 10, 269–275. <https://doi.org/10.1016/j.micinf.2007.12.021>.
- Rambaut, A., 2017. Figtree 1.4.3. Available from <http://tree.bio.ed.ac.uk/software/figtree/>. (accessed 18 Aug. 2017).
- Ratnasingham, S., Hebert, P.D.N., 2007. BARCODING, BOLD: The Barcode of Life Data System ([www.barcodinglife.org](http://www.barcodinglife.org)). *Mol. Ecol. Notes* 7, 355–364. <https://doi.org/10.1111/j.1471-8286.2006.01678.x>.
- Raymond, M., Rousset, F., 2013. An exact test for population differentiation. *Evolution* (N. Y.) 49, 1280–1283.
- Riehle, M.M., Guelbeogo, W.M., Gneme, A., Eiglmeier, K., Bischoff, E., Garnier, T., Snyder, G.M., Li, X., Markianos, K., Sagnon, N.F., Vernick, K.D., 2011. A cryptic subgroup of *Anopheles gambiae* is highly susceptible to human malaria parasites. *Science* (80-) 331, 596–598. <https://doi.org/10.1126/science.1196759.A>.
- Robert, V., Petrarca, V., Carnevale, P., Zoulani, A., Coluzzi, M., 1989. Analyse Cytogénétique Du Complexe *Anopheles gambiae* Dans La Région De Bobo-Dioulasso (Burkina Faso). *Ann. Parasitol. Hum. Comp.* 64, 290–311. <https://doi.org/10.1051/parasite/1989644290>.
- Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.
- Sallum, M.A., Schultz, T.R., Foster, P.G., Aronstein, K., Wirtz, R.A., Wilkerson, R.C., 2002. Phylogeny of *Anophelinae* (Diptera: Culicidae) based on nuclear ribosomal and mitochondrial DNA sequences. *Syst. Entomol.* 27, 361–382. <https://doi.org/10.1111/jvec.12128>.
- Scott, J.A., Brogdon, W.G., Collins, F.H., 1993. Identification of single specimens of the *Anopheles gambiae* complex by the polymerase chain reaction. *Am. J. Trop. Med. Hyg.* 49, 520–529. <https://doi.org/10.4269/ajtmh.1993.49.520>.
- Simard, F., Ayala, D., Kamdem, G.C., Pombi, M., Etouana, J., Ose, K., Fotsing, J.M., Fontenille, D., Besansky, N.J., Costantini, C., 2009. Ecological niche partitioning between *Anopheles gambiae* molecular forms in Cameroon: the ecological side of speciation. *BMC Ecol.* 9. <https://doi.org/10.1186/1472-6785-9-17>.
- Stevenson, J.C., Norris, D.E., 2017. Implicating cryptic and novel anophelines as malaria vectors in Africa. *Insects* 8, 1–18. <https://doi.org/10.3390/insects8010001>.
- Sum, J.-S., Lee, W.-C., Amir, A., Braima, K.A., Jeffery, J., Abdul-Aziz, N.M., Fong, M.-Y., Lau, Y.-L., 2014. Phylogenetic study of six species of *Anopheles* mosquitoes in Peninsular Malaysia based on inter-transcribed spacer region 2 (ITS2) of ribosomal DNA. *Parasit. Vectors* 7, 309. <https://doi.org/10.1186/1756-3305-7-309>.
- Tajima, F., 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105, 437–460.
- Tamura, K., Nei, M., Kumar, S., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc. Natl. Acad. Sci.* 101, 11030–11035.
- Taylor, C., Touré, Y.T., Carnahan, J., Norris, D.E., Dolo, G., Traoré, S.F., Edillo, F.E., Lanzaro, G.C., 2001. Gene flow among populations of the malaria vector, *Anopheles gambiae*, in Mali, West Africa. *Genetics* 157, 743–750.
- Wang, G., Li, C., Zheng, W., Song, F., Guo, X., Wu, Z., Luo, P., Yang, Y., He, L., Zhao, T., 2017. An evaluation of the suitability of COI and COII gene variation for reconstructing the phylogeny of, and identifying cryptic species in, anopheline mosquitoes (Diptera Culicidae). *Mitochondrial DNA Part A DNA Mapping. Seq. Anal.* 28, 769–777. <https://doi.org/10.1080/24701394.2016.1186665>.
- Weetman, D., Steen, K., Rippon, E.J., Mawejje, H.D., Donnelly, M.J., Wilding, C.S., 2014. Contemporary gene flow between wild *An. gambiae* s.s. and *An. arabiensis*. *Parasit. Vectors* 7, 1–9. <https://doi.org/10.1186/1756-3305-7-345>.
- WHO, 2018. World Malaria Report 2018. World Health Organization, Geneva, Switzerland.
- Wiebe, A., Longbottom, J., Gleave, K., Shearer, F.M., Sinka, M.E., Massey, N.C., Cameron, E., Bhatt, S., Gething, P.W., Hemingway, J., Smith, D.L., Coleman, M., Moyes, C.L., 2017. Geographical distributions of African malaria vector sibling species and evidence for insecticide resistance. *Malar. J.* 16, 1–10. <https://doi.org/10.1186/s12936-017-1734-y>.
- Wilcox, T.P., Zwickl, D.J., Heath, T.A., Hillis, D.M., 2002. Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Mol. Phylogenet. Evol.* 25, 361–371. [https://doi.org/10.1016/S1055-7903\(02\)00244-0](https://doi.org/10.1016/S1055-7903(02)00244-0).
- Wondji, C., Simard, F., Fontenille, D., 2002. Evidence for genetic differentiation between the molecular forms M and S within the Forest chromosomal form of *Anopheles gambiae* in an area of sympatry. *Insect Mol. Biol.* 11, 11–19. <https://doi.org/10.1046/j.0962-1075.2001.00306.x>.
- Yawson, A.E., Weetman, D., Wilson, M.D., Donnelly, M.J., 2007. Ecological zones rather than molecular forms predict genetic differentiation in the malaria vector *Anopheles gambiae* s.s. in Ghana. *Genetics* 175, 751–761. <https://doi.org/10.1534/genetics.106.065888>.