BioPublisher®

Journal of Mosquito Research 2014, Vol.4, No.13, 1-10
http://jmr.biopublisher.ca

Research Article **Open Access**

Molecular Characterization of Mosquitoes of *Anopheles gambiae* Species Complex (Diptera: Culicidae) from Sudan and Republic of Southern Sudan

Asma Mahmoud Hamza[1], El Amin El Rayah[2], Sumaia Mohamed Ahmed Abukashawa[2]

1. Department of Biology, Faculty of Education, University of Kassala, Kassala State, Sudan.
2. Department of Zoology, Faculty of Science, University of Khartoum, Khartoum, Sudan.

Corresponding author email: Asmakassala@gmail.com

Journal of Mosquito Research, 2014, Vol.4, No.13 doi: 10.5376/jmr.2014.04.0013
Received: 14 Jun., 2014
Accepted: 13 Jul., 2014
Published: 30 Aug., 2014

Preferred citation for this article:
Hamza et al., 2014, Molecular Characterization of Mosquitoes of *Anopheles gambiae* Species Complex (Diptera: Culicidae) from Sudan and Republic of Southern Sudan, Journal of Mosquito Research, Vol.4, No.13 1−10 (doi: 10.5376/jmr.2014.04.0013)

Abstract Mosquitoes of the *Anopheles gambiae* complex, namely *Anopheles arabiensis* (Patton, 1905) and *Anopheles gambiae* (Giles, 1902) are the major vectors of human malaria in the African continent. This study was conducted mainly to investigate the molecular biology of members of the *An. gambiae* complex in Sudan and Republic of Southern Sudan. The molecular investigation involved identification of members of the *An. gambiae* complex using polymerase chain reaction (PCR) techniques based on DNA specific nucleotide differences in the intergenic spacer region (IGS) of the ribosomal RNA gene claster (rRNA) and partial sequencing and analysis of IGS regions. *Adult Anopheles* mosquitoes were collected from four sites in Kassala State, Sudan and from one site in Western Bahr El Ghazal State, Republic of Southern Sudan. In addition, *An. arabiensis* specimens, obtained from Sennar town laboratory colony (Sudan) was used in the study. Collection of *Anopheles* mosquitoes was done by hand capture using sucking tube (aspirator) during the rainy seasons of 2008, 2009 and 2010. The molecular investigation predicts the existence of two species within the *An. gambiae* complex, namely *An. arabiensis* and *An. gambiae*. *An. arabiensis* was found as the predominant *Anopheles* mosquitoes in all the collection sites while *An. gambiae* was found sympatrically with *An. arabiensis* in Republic of Southern Sudan. The analysis of the IGS fragments revealed moderate level of genetic variations within and between the *An. arabiensis* populations. *An. gambiae* individuals showed high genetic similarity. The genetic analysis revealed little population differentiation (Fst=0.067) and high migration rate (Nm=3.51) which indicated high gene flow between *An. arabiensis* populations collected from Kassala State localities. The phylogenetic relationships between the different populations of *An. arabiensis* and *An. gambiae* were investigated. The IGS regions of rRNA gene have been shown to be powerful markers for species identification and studying the genetic structure of members of *An. gambiae* complex.

Keywords *An. arabiensis*; *An. gambiae*; Ribosomal RNA gene (rRNA); Intergenic spacer region (IGS); Sudan; Republic of Southern Sudan

Background

Malaria is a major health problem in Africa. Members of *An. gambiae* complex have been identified as major vectors of human malaria parasites in the African continent. The *An. gambiae* complex is comprised of seven genetically and behaviourally distinct species that are morphologically undistinguishable (Davidson et al., 1967; Service, 1985; Hunt et al., 1998). Within this species complex the most important vectors of human malaria are: *An. gambiae* (Giles, 1902) and *An. arabiensis* (Patton, 1905) which are distributed over 70% of the Sub-Saharan Africa with *An. arabiensis* being distributed over the dry savanna and semi-arid parts of Africa (Service, 1980; Bryan, 1983; Lindsay et al. 1998).The two species are more adapted to the human environment; they are sympatric and synchronic over most of their geographical distribution range (Petrarca et al., 1998).

Of the seven recognized species of the An. gambiae complex, *An. arabiensis* and *An. gambiae* are the most abundant and most important vector of human malaria in Sudan and Republic of Southern Sudan. *An. arabiensis* has been regarded from many localities in Sudan, from the extreme south up to the northern borders with Egypt. It was reported from many localities in Kassala State, eastern Sudan by (Haridi, 1972; Petrarca et al., 1986; Himeidan, 2004). *An. gambiae* is restricted to Republic of Southern Sudan

and found sympatrically with *An. arabiensis* (Zahar, 1985; Petrarca et al., 1986).

Identification of species within the *An. gambiae* group is essential for the correct evaluation of malaria vector ecology studies and control programs (Gale, 1987). Understanding the genetic structure of mosquito populations is important for addressing important biological and public health issues such as evolution, spread of insecticide resistance alleles and epidemiology of vector-borne diseases (Cramption et al., 1994; Tripet et al., 2001; Fanello et al., 2003).

Several methods for identifying species of mosquito complexes have been developed such as polymerase chain reaction (PCR) techniques. This has become the standard method to species identification and studying the genetic structure (Scott et al., 1993; Wilkins et al., 2006). Ribosomal RNA gene claster (rRNA) is one of the most widely used regions of the genome to infer genetic variations and phylogenetic relationships. rRNA gene is a gene family consisting of many copies (100-500) of genes and encodes the ribosomal RNA. In eukaryotes, the rRNA gene is composed of tandem repeated units separated from each other by intergenic non transcribed spacers (IGS). Each repeat contains the coding genes for 18S, 5.8S and 28S in a respective order separated from each other by an external transcribed spacer (ETS) at the 5´end of the 18S gene, and two internal transcribed spacers (ITS). The first internal transcribed spacer (ITS1) is located between 18S and 5.8S genes, whereas the ITS2 separates 5.8S and 28S rRNA genes (Hillis and Dixon, 1991). The coding regions of 18S, 5.8S and 28S are highly conserved; whereas the non-coding spacers (ITS1, ITS2 and IGS) are highly variable and evolve at a faster rate than the coding regions. They can be highly variable in length and nucleotides sequence between closely related species (Olsen and Woese, 1993; Aransay, 2000). Highly conserved region repeats can be used for studying relationships across phyla (Gerib, 1985) while more variable regions can be used for lower taxonomic levels. The IGS region contains species-specific nucleotide sequences and has facilitated discrimination of species in *An. gambiae* group (Collins et al., 1987; Paskewiz et al.; 1993; Scott et al., 1993).

The ability to amplify DNA using PCR techniques has greatly facilitated DNA sequence comparisons (Innis et al., 1990) and resulted in the development and use of species-specific diagnostic PCR primer pairs (Paskewitz and Collins, 1990). Nucleotide sequencing of PCR products has made it suitable both for conformation and characterization of mutants detected by one of the described screening methods such as single nucleotide polymorphism (SNP). SNP occurs when a single nucleotide in the genome sequence is changed. SNPs are the commonest type of nucleotide sequence variations in genome but they have only recently been used to investigate the evolutionary and demographic history of populations and speciation (Brumfield et al., 2003).

Several molecular markers have been used in studying population genetic structure and gene flow in anopheline mosquitoes. These markers range from classical genetic markers (e. g. mtDNA or rRNA gene) to methods used to detect and identify SNPs and finally to highly polymorphic markers (e. g. RAPDs, microsatellite DNAs) (Norris, 2002). Classical genetic markers are characterized by targeting a defined gene or genetic fragment for analysis using different techniques to evaluate genetic variability down to resolution of SNPs and sequencing.

Gene flow can be defined as the movement of genes within and between the different populations (Ferris et al., 1983) and it cannot be measured, but estimated using either a direct or an indirect method (Slatkin, 1995). Direct estimates of gene flow are based on the observation of organism dispersal within a defined range of time and space (Taylor et al., 2001). The indirect method is based on the estimation of allele frequencies, obtained by electrophoretic survey of proteins or DNA sequence using molecular markers (Donnelly and Townson, 2000).

The IGS regions have previously been used as a powerful tool in studying the genetic and phylogenetic divergence between closely related species. Phylogenetic trees are powerful means for summarizing the evolutionary relationships. Many different criteria can be used to construct phylogenetic trees from morphological or molecular data. In integrated mosquito control programmes, taxonomic and phylogenetic studies had been quite useful in understanding the

vectorial capacity and insecticide resistance in malaria vectors (Sharma and Chaudhry, 2010).

In the present study we conducted molecular investigation of members of the *An. gambiae* complex in Sudan and Republic of Southern Sudan. The investigation involved molecular identification of the members of the complex and analysis of partial sequence of IGS regions of rRNA gene.

1 Materials and methods

1.1 Collection sites and mosquitoes used in the study

Sudan and Republic of Southern Sudan are situated in the eastern part of the African continent, between 22 ° and 38 ° degrees East (E) longitude and between 4 ° and 22 ° degrees North (N) latitude (Figure 1). The area is crossed by the River Nile and its tributaries. Sudan has different climatic regions, ranging from desert (dry hot-arid) in the North, to semi-desert and savanna in the South. Republic of Southern Sudan has equatorial (humid-tropical) climate.

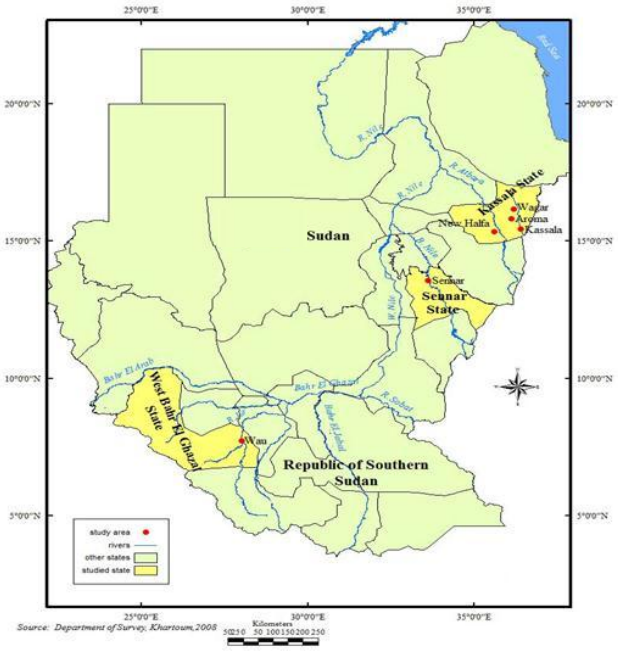

Figure 1Map of Sudan and Republic of Southern Sudan showing the collection sites

Field samples of females *An. gambiae* complex were collected from four different sites in Kassala State, eastern Sudan (dry area) and from one site in Western Bahr El Ghazal State, northern Republic of Southern Sudan (humid area). In addition samples of *An. arabiensis* from Sennar laboratory colony were used in the study. The laboratory colony materials originating from Sennar State (central Sudan) and maintained at Sennar town since 2007 (Figure 1). The major consideration in the selection of the study sites is that they represent different ecological regions with different environmental conditions that may have an effect on the distribution and the population genetics of the *An. gambiae* species complex. The selection is also based on the easy accessibility of the collection sites. The distance between each two sites in Kassala State exceeds 30 Km. This distance is more than the expected dispersal distance of *Anopheles* mosquitoes from the favorable breeding sites. Table 1 shows the collection sites and the species and number of mosquitoes analyzed.

Indoor resting wild adult *Anopheles* mosquitoes were caught from rooms by hand capture using sucking tube (aspirator) during the rainy seasons 2008, 2009 and 2010. *Anopheles* mosquitoes collected were fixed and preserved individually in 70% ethanol and stored at -20 ℃ for subsequent processing. Members of *An. gambiae* complex were morphologically separated from other anopheline mosquitoes using the morphological identification keys of Gillies and De-Mellion (1968) and Gillies and Coetzee (1987).The processing of the materials for this study was carried out at the Genetics and Molecular Biology laboratory at the Department of Zoology, Faculty of Science, University of Khartoum, Sudan.

1.2 DNA extraction, amplification and sequencing

Genomic DNA was extracted from thorax and abdomen tissues of individual mosquitoes according to the method of Collins et al. (1987) with minor modifications described by Proft et al. (1999). The quantity and quality of the extracted DNA were estimated using the Nanodrop spectrophotometer (ND-1000) for absorbance.

PCR and partial sequencing of the IGS region was based on the diagnostic method of species-specific nucleotide sequences in the IGS regions of rRNA gene. The ribosomal set of primers developed by Scott et al. (1993) was used. Three primers, of which 2 were specific to *An. arabiensis* and *An. gambiae* and one was a common (universal) primer to both species were used.

Table 1 Collection sites, the species and number of females *An. gambiae* species complex used in the study

Country	State	Collection site	Species	No. of mosquitoes analyzed
Sudan	Kassala	Kassala	*An. arabiensis*	59
		New Halfa	*An. arabiensis*	61
		Aroma	*An. arabiensis*	59
		Wagar	*An. arabiensis*	56
	Sennar	Sennar	*An. arabiensis*	56
Republic of Southern Sudan	Western Bahr El Ghazal	Wau	*An. arabiensis*	90
			An. gambiae	35
				Total=416

Species-specific *An. arabiensis* primer:
3 ´AR- AAG TGT CCT TCT CCA TCC TA

Species-specific *An. gambiae* primer:
3 ´GA- CTG GTT TGG TCG GCA CGT TT

Universal 5 ´primer sequence:
Un-5 ´GTG TGC CCC TTC CTC GAT GT

Amplification reaction was performed following a slightly modified version (in the master mix and the times of the program of amplification) of the protocol described by Scott et al. (1993). PCR reagents were obtained from Vivantitis. PCR was performed in a total volume of 25 μL using a thermocycler (Techne, Touchgene Gradient). 5 μL of DNA template (10ng / μL) were used with 20 μL PCR mix containing 2.5 μL of 10X PCR buffer Mg^{++} free, 1 μL of dNTPs mix, each at 10 mM, 1.2 μL $MgCl_2$, 3 units of Taq polymerase and 2 μL of each forward and reverse primers (20 pmol /25 μL). 11 μL of sterile deionized water were added to make the final volume to 25 μL.

The PCR reaction was carried out with a program of 30 cycles of denaturation at 94 ℃ for 1 min, annealing at 50 ℃ for 1 min and extension at 72 ℃ for 1 min. After the PCR was completed, 5 μL from the PCR product were mixed with 5 μL of loading dye, electrophoresed through a 2.5% agarose gel and stained with ethidium bromide stain following the standard method described by Sambrook, et al. (1989). The amplified fragments were visualized by illumination with short wave ultraviolet light and photodocumented.

Samples of purified PCR products were sequenced at Macrogen (www. macogen. com).

1.3 Sequence analysis

Chromatograms of sequence results and texts of the seven populations of *An. arabiensis* and *An. gambiae* were analyzed using different computer software programs. Mismatching alignments were checked by eye for sequence reading errors. The forward and reverse sequence strands of each specimen were matched. The consensus sequences of *An. arabiensis* and *An. gambiae* populations were aligned along with the published reference sequences of the GenBank sequences of Scott et al. (1993). Sequencing alignments were done using the software BioEdit and CLC Main Work Bench-version 6.5. The polymorphisms in the analyzed segments were exported using software Mega 5.05 (Tamura, 2011). Then the clustered sequences were directed for further analysis.

Using the software DnaSp-version 5.10.0, the frequency of each haplotype, haplotype diversity (Nei**,** 1987) and nucleotide diversity (Tajima, 1983) were calculated. Population genetic differentiation using Wrights F-statistics (F*st*) and levels of gene flow were determined through the effective number of migrants (Nm) between locations using DnaSp-version 5.10.0 (Hudson et al., 1992).

Mega 5.05 software was used to construct trees of individuals and haplotypes of *An. arabiensis* and *An. gambiae*. A neighbor-joining tree (Saitou et al., 1987) using Kimura 2-parameter model (Kimura, 1980) with 1000 bootstrapping replicates was constructed based on the aligned sequences to identify possible phylogenetic lineages.

2 Results

2.1 Molecular identification of species of *An. gambiae* complex

The quantity and quality of template DNA was found to be suitable for PCR amplification. The mean DNA quantity was 10-22ng /μL and DNA quality range was 1.7-2.2 for single female mosquitoes.

Female *An. gambiae* species complex were identified as *An. arabiensis* and *An. gambiae* by the results of

the PCR identification. A total of 315bp and 390bp segments of the IGS region of rRNA gene sequences of *An. arabiensis* and *An. gambiae*, respectively, were amplified (Figure 2). *An. arabiensis* was found as the predominant *Anopheles* mosquitoes in all the collection sites. All samples from Kassala State were identified as *An. arabiensis*. In Southern Sudan *An. arabiensis* was found sympatrically with *An. gambiae* and represented 72% of the *An. gambiae* complex in the area.

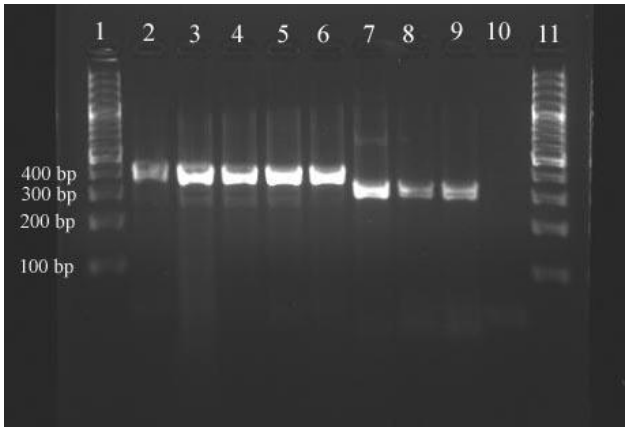

Figure 2 Agarose gel showing DNA fragments of PCR amplification using *An. gambiae* and *An. arabiensis* DNA from Sudan and Republic of Southern Sudan.

Note: Lanes 1 and 11, a100 bp DNA ladder of standard size; Lane 2, positive control of *An. gambiae*; Lanes 3- 6 represent amplified fragments of 390 bp of *An. gambiae*; Lane 7, *An. arabiensis* from Sennar laboratory colony materials; Lanes 8 and 9, amplified fragments of 315 bp of field females *An. arabiensis*; Lane 10, negative control.

2.2 Sequence alignment and characterization of IGS regions of rRNA gene

The amplified regions were found to be corresponding to the rRNA gene segments characterized by Scott et al. (1993).The sequences of the IGS region of *An. arabiensis* and *An. gambiae* populations were aligned along with the published sequences of the IGS region of *An. arabiensis* [GenBank: accession number, U10138] and *An. gambiae* [GenBank: accession number, U10135] (Scott et al., 1993) (Figure 3). Sequence analysis was carried out on 255bp and 334bp of *An. arabiensis* and *An. gambiae*, respectively. The analyzed regions were equivalent to position (506-760 for *An. arabiensis* and 456-789 for *An. gambiae*) in the *An. gambiae* reference sequences of Scott et al. (1993). The sequences of the analyzed regions were published in the GenBank with accession numbers [KC491792- KC491797] and [KC491806-KC491834] for *An. arabiensis* and [KC491798-KC491805] for *An. gambiae*.

2.3 Single nucleotide polymorphisms

Mismatching alignments of *An. arabiensis* sequences indicated that variations detected by sequencing are substitutions. Three polymorphic sites were identified within *An. arabiensis* populations and no gaps were present. The polymorphic sites were located in positions of 590bp, 693bp and 713bp in the *An. arabiensis* reference sequences (Figure 3). Position 590bp (T-A) was found only in colony specimens [accession number, KC491828] while position 693bp (C-T) found only in New Halfa and Aroma populations [accession number, KC491814]. Position 713bp (A- T) was found in all the populations and characterized most individuals of *An. arabiensis* collected from Wau town [accession number, KC491795].

Out of 35 *An. arabiensis* sequenced, 23 % yielded sequences that were identical to the *An. arabiensis* IGS reference sequence. The direct sequences revealed no substitution, insertion or deletion events within any *An. gambiae* sequences, i.e all *An. gambiae* individuals yielded sequences identical to the published GenBank sequence (Scott et al., 1993).

2.4 Haplotypes estimation and genetic diversity

Comparison of IGS regions of rRNA gene of field *An. arabiensis* individuals and colony specimens' revealed 4 different haplotypes with 3 polymorphic sites (Figure 3). The 4 haplotypes are shared between all *An. arabiensis* populations and their average guanine-cytosine (GC) content was observed to be 0.519%. There are two major groups of haplotypes within *An. arabiensis* populations, one being identical to the *An. arabiensis* IGS reference sequence. The most frequent haplotype is (*An. arabiensis* II) with (57%) frequency.

The haplotype number in each population varies from 1-3 and haplotype diversity values (Hd) differ from 0.286 in Aroma population to 0.600 in Kassala and New Halfa populations. Nucleotide diversity (π) values range from 0.00131 in Wau *An. arabiensis* population to 0.00341 in New Halfa population. *An. gambiae* population showed the lowest haplotype diversity and nucleotide diversity values (0.000) (Table 2).

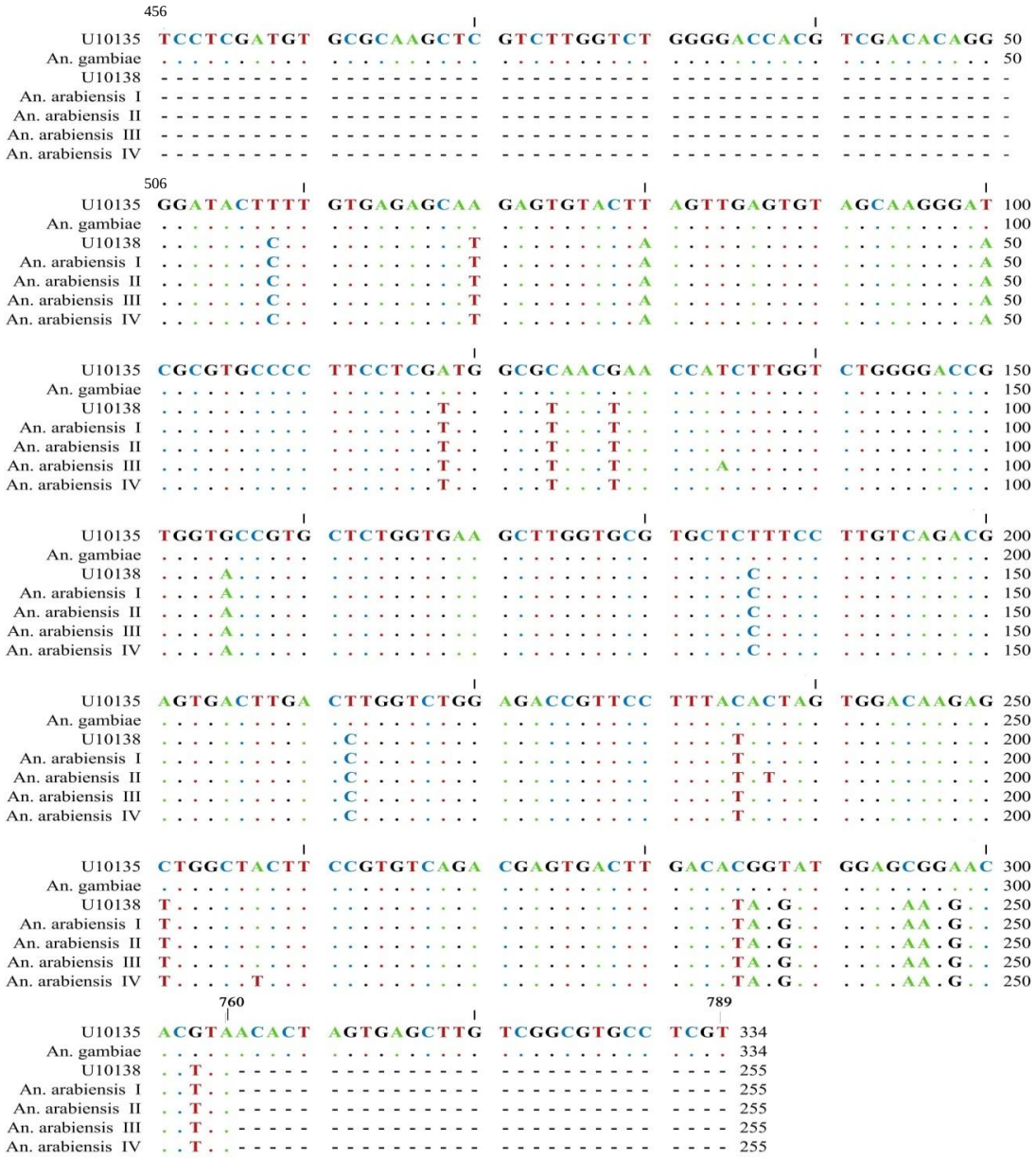

Figure 3 Partial alignments of IGS regions of rRNA gene of *An. gambiae* and *An. arabiensis* from Sudan and Republic of Southern Sudan. The sequences numbered with reference to the published IGS sequence of *An. arabiensis* in GenBank [accession number: U10138]. A dot in the alignment indicates that the sequence is identical with that of the consensus sequence. (*An. arabiensis* I, II, III and IV), the different haplotypes of *An. arabiensis*

Table 2 Statistical data of 255 bp of IGS regions of rRNA gene polymorphism within *An. arabiensis* and *An. gambiae* populations collected from Sudan and Republic of Southern Sudan

No.	Population	N	S	H	Hd	π
1	Kassala	5	1	2	0.6000	0.0025
2	New Halfa	6	2	3	0.6000	0.0034
3	Aroma	7	2	2	0.2857	0.0023
4	Wager	4	1	2	0.5000	0.0020
5	Sennar	7	2	3	0.5238	0.0030
6	Wau *An. arabiensis*	6	1	2	0.3333	0.0013
7	Wau *An. gambiae*	8	0	1	0.0000	0.0000
8	Total	43	22	5	0.7276	0.0264

Note: N, sample size; S, number of polymorphic sites; H, number of haplotypes; Hd, Haplotype diversity; π, nucleotide diversity

2.5 Phylogenetic relationships

All the individuals of *An. arabiensis* and *An. gambiae* were found to be separated into two main clades: (Figure 4) one consisting of two groups, the first group consisted of *An. gambiae* sequences revealed from the present study in addition to the reference sequence of IGS of *An. gambiae* [accession number: U10135]. A fairy *An. melas* sequence from GenBank [accession number: U10139] (Scott et al., 1993) was used to root the phylogeny. This comprised the second group (out group). The other clade consisting of all *An. arabiensis* populations plus the reference sequence of IGS of *An. arabiensis* [accession number: U10138].

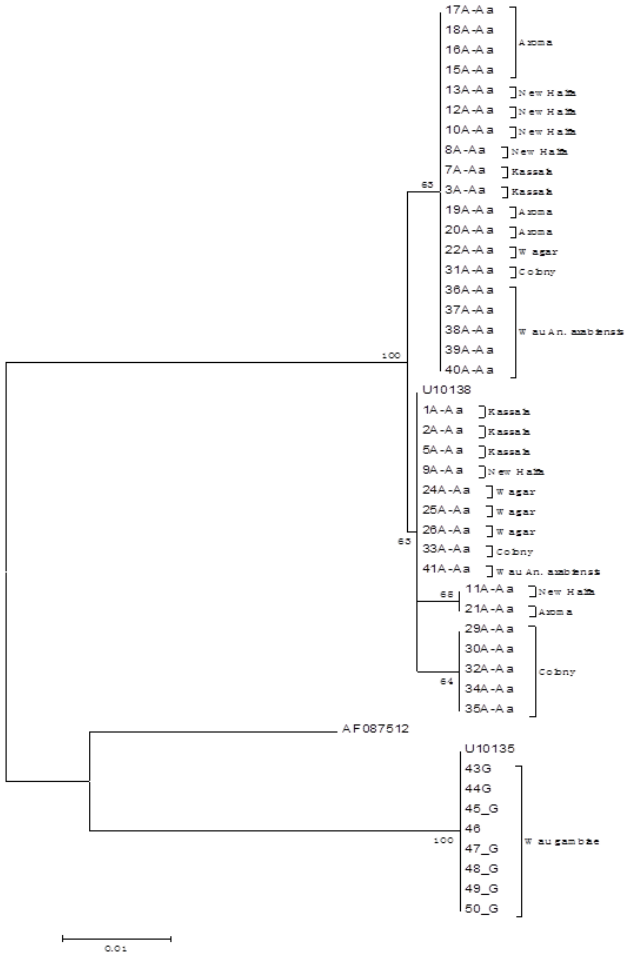

Figure 4 Neighbor joining phylogenetic tree of IGS regions of individuals of *An. gambiae* complex conducted in Mega 5.05 soft ware program. The sequences of species with accession numbers U10138, U10135 and AF087512 were obtained from the GenBank. Values at the nodes indicate bootstrap support of 1000 replicates

The phylogenetic tree shows obvious interactions between the colony population and the 5 field populations collected from different collection sites. Colony specimens showed a greater tendency to group with *An. arabiensis* obtained from GenBank and lie close to *An. gambiae* while *An. arabiensis* collected from Wau town showed opposed tendency.

The haplotypes of members of *An. gambiae* complex were found to be separated into two main clades: (Figure 5) one consisting of *An. arabiensis* haplotypes (*An. arabiensis* I, II, III and IV) generated in the present study and the reference sequence of *An. arabiensis*. This group confirms that the haplotype *An. arabiensis* I is identical to that one of the GenBank, whereas haplotype (*An. arabiensis* IV) is in a separate group. The second clade consisting of two separate groups. One of them consisting of *An. melas* reference sequence (out group) and the other consisting of *An. gambiae* sequences revealed from the present study and the reference sequence of *An. gambiae*. This group confirms that *An. gambiae* collected from Wau town was identical to that one of the GenBank.

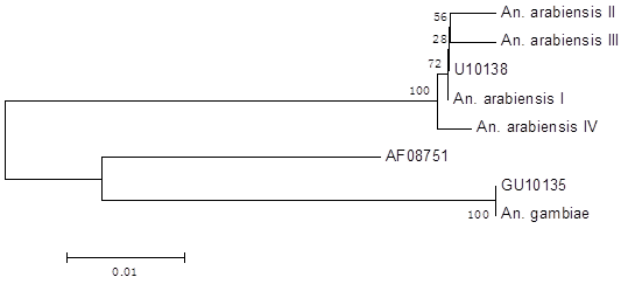

Figure 5 Neighbor joining phylogenetic tree of IGS regions of *An. gambiae* complex conducted in Mega 5.05 software program. The sequences of *An. arabiensis* I, II, III and IV and *An. gambiae* were generated in the present study. The sequences of species with accession numbers U10138, U10135 and AF087512 were obtained from GenBank. Values at the nodes indicate bootstrap support of 1000 replicates

3 Discussion

Understanding of the epidemiology of vector-borne diseases depends on the accurate identification of proven and suspected vectors of disease. Species assessment was performed using a PCR targeting a sequence in the IGS region of the rRNA gene. The molecular method predicts the existence of two species within *An. gambiae* complex, namely *An. arabiensis* and *An. gambiae*. *An. arabiensis* is the predominant malaria vector species in all of the collection sites and found sympatrically with *An.*

gambiae in Republic of Southern Sudan. It showed predominance of (100%) in Kassala State. This result confirms the previous findings revealed by other researchers in eastern Sudan, who showed that *An. arabiensis* is considered as the main malaria vector in Ksssala State e.g. Haridi (1972), Petrarca et al. (1986) and Himeidan (2004). In Republic of Southern Sudan, *An. arabiensis* showed predominance (72%) of *An. gambiae* complex found in the area. Many studies showed that the predominance of *An. arabiensis* is quite widespread in similar climatic conditions in the African continent, e. g. Coosemans et al. (1989) and Collins et al. (1988). Climatic factors, such as precipitation and temperature are important determinants of the range and relative abundance of member species of *An. gambiae* group. *An. gambiae* is usually predominant in humid environment but *An. arabiensis* is more common in arid dry areas (Lindsay et al., 1998).

For identification of members of *An. gambiae* complex, the protocol developed by Scott et al. (1993) became a reference method that was applied to the all developmental stages and both sexes (Fanello et al., 2002; Fettene and Temu, 2003). In our study we used this protocol to identify members of the *An. gambiae* complex found in the collection sites. The protocol was also used by many workers to identify members of the *An. gambiae* complex from different part of the African continent. e. g. Berzosa et al. (2002), Nyanjom et al. (2003), Pock Tsy et al. (2003) and Oyewolea and Awololab (2006).

The present study described an analysis of sequences of a 255bp and 334bp IGS regions of *An. arabiensis* and *An. gambiae*, respectively. The sequences were published in the GenBank with accession numbers [KC491792- KC491797] and [KC491806- KC491834] for *An. arabiensis* and [KC491798- KC491805] for *An. gambiae*. The sequencing survey indicated lower genetic variations within and between the populations of *An. arabiensis* and high genetic similarity within *An. gambiae* individuals from Republic of Southern Sudan.

The study examined the population genetic structure of *An. arabiensis* and *An. gambiae* mosquitoes. The population genetic differentiation index (*F*st) of four populations of *An. arabiensis* found in Kassala State was estimated as (*F*st=0.067) indicating low amount of genetic differentiation between *An. arabiensis* populations in the area. *F*st values reported in the present study from Kassala State for *An. arabiensis* populations are consistent with the general population genetic structure pattern of *An. arabiensis* based on microsatellite and other molecular markers. Donnelly and Townson (2000) reported *F*st value of 0.035 for nine *An. arabiensis* populations from north Sudan to south Mozambique. Besansky et al. (1997) reported that *An. arabiensis* populations from Kenya, South Africa and west Senegal had F*st* value of 0.038 by allozyme and 0.44 by mDNA sequence variation. In contrast high *F*st value (0.96-0.258) was reported by Onyabe and Conn (2001) for populations of *An. arabiensis* from Nigeria using microsatellite data. The level of genetic differentiation in *An. arabiensis* populations may vary significantly among different regions of the genome or may be due to variations in the immature stage habitats.

In this molecular investigation the migration rate (Nm), which indicate the rate of gene flow was estimated for the four *An. arabiensis* populations from Kassala State by an indirect method using IGS regions of rRNA gene markers. A high value of migration rate (Nm=3.51) was reported indicating a high gene flow and a high gene exchange among the different four populations of Kassala State. Therefore, active dispersal of this mosquito due to the wind dispersal or human transportation may be the reason behind the continuous gene flow between these four localities.

The two sympatric populations of *An. arabiensis* and *An. gambiae* found in Republic of Southern Sudan showed highest genetic difference (*F*st=0.991) and gene flow estimate (Nm=0.000) among the two populations. This highest *F*st value provides strong evidence of subdivision of the sympatric populations into two species and no gene flow or gene exchange occurred between them.

The IGS regions was previously been used as a useful tool in studying the level of genetic and phylogenetic divergence between closely related species. The phylogenetic dendrograms generated by Neighbor Joining method revealed that *An. gambiae* closely paired with *An. melas* (from GenBank) forming a single clade while *An. arabiensis* individuals (or

haplotypes) clustered together formed another clade closest to *An. melas*. The inclusion of *An. gambiae* with *An. melas* in a single clade was due to their close genetic homology and phylogenetic relationship. The trees also showed that maximum closeness was present within the individuals of *An. gambiae* and *An. arabiensis* (a bootstrap value of 100).

The IGS regions of rRNA gene sequences have been shown to be powerful markers for species identification and studying the genetic structure of *An. arabiensis* and *An. gambiae* populations. Studies on the genetic structure of the sibling species of the *An. gambiae* complex are still required, and the PCR techniques targeting IGS regions can be an exceedingly useful instrument in these studies.

Authors' contributions

Hamza A.M. participated in field sampling, performed the technical work, analyzed the results and drafted the manuscript. El Rayah El A. contributed to the reading and revising the manuscript. Abukashawa S.M. supervised the research group and participated in revising the draft manuscript. All authors read and approved the final manuscript.

Acknowledgements

We thank Department of Zoology, Faculty of Science, University of Khartoum for the institutional support and use of the facilities of the Genetics and Molecular Biology Laboratory. We thank Mr Abd El Gader Morwak, Ministry of Health, Western Bahr El Ghazal State, Republic of Southern Sudan for his assistance in the collection of mosquito samples from Republic of Southern Sudan and Mr Abd El Monem Faza, Director of Dr El Gadal Center, Sennar, Federal Ministry of Health, Sudan for providing the laboratory colony materials of *An. arabiensis*. We thank Dr Rania Baleela and Dr Mai Masri, Department of Zoology, Faculty of Science, University of Khartoum for their collaboration and assistance in analysis of the results. This research was supported by Kassala University.

References

Aransay A.M., Scoulica E. and Tselentis Y., 2000, Phylogenetic relationships of Phlebotominae sand flies inferred from small subunit nuclear ribosomal DNA, Insect Biochemistry and Molecular Biology, 9: 157-168
http://dx.doi.org/10.1046/j.1365-2583.2000.00168.x

Berzoa P.J., Cano J., Roche J., Rubio J.M., Garcia L., Moyaro E., Guerra A., Mateos J.C., Petrarca V., Do Rosario V. and Benito A., 2002, Malaria vectors in Bioko Island (Equatorial Guinea: PCR determination of the members of *Anopheles gambiae* Giles complex (Diptera: Culicidae) and pyrethroid knock down resistance (Kdr) in *Anopheles gambiae* sensu strict, Journal of Vector Ecology, 27 (1): 102-126

Besansky N.J., Lehmann T., Fahey G.T., Fontenille D., Braack L.E.O., Hawley W.A. and Collins F.H., 1997, Patterns of mitochondrial variation within and between the African malaria vectors, *Anopheles gambiae* and *Anopheles arabiensis*, suggest extensive gene flow, Genetics, 147: 1817-1828

Brumfield R.T., Beerli P., Nickerson D.A. and Edwards S.V., 2003, The utility of single nucleotide polymorphisms in inferences of population history, Trends in Ecology and Evolution, 18: 249-256
http://dx.doi.org/10.1016/S0169-5347(03)00018-1

Bryan J.H., 1983, *Anopheles gambiae* and *Anopheles melas* at Brefet: The *gambiae* and their role in malaria transmission, Ann Trop Med Parasitol, 77: 1-12

Collins F.H., Mendez M.A., Rasmussen M.O., Mehaffey P.C., Besansky N.J. and Finnerty V.A., 1987, Ribosomal RNA gene probe differentiated member species of the *Anopheles gambiae* complex, American Journal of Tropical Medicine and Hygiene, 36: 37-41

Collins F.H., Petrarca V., Mpofu S., Brandung-Bennett A.D., Were J.B.O., Rasmussen M.O. and Finnertyti V., 1988, Comparison of DNA probe and cytogenetic methods for identifying field collected *Anopheles gambiae* complex mosquitoes, American Journal of Tropical Medicine and Hygiene,39 (6): 545-550

Coosemans M., Petrarca V., Barutwanayo M. and Coluzzi M., 1989, Species of the *Anopheles gambiae* complex and chromosomal polymorphism in a rice-growing area of the Rusizi Valley (Republic of Burundi), Parassitologia, 31: 113-122

Crampton J.M., Warren A. and Lycett G.J., 1994, Genetic manipulation of insect vectors as a strategy for the control of vector-borne disease, Ann. Trop. Med. Parasitol, 88: 3-12

Davidson G., Paterson H.E., Coluzzi M., Mason G.F. and Micks D.W., 1967, The *Anopheles gambiae* complex, In Genetics of Insect Vectors of Disease edited by: Wright J.W. Pal R. Elsevier, 1967:211-250

Donnelly M.J. and Townson H., 2000, Evidence for extensive genetic differentiation among populations of the malaria vector *Anopheles arabiensis* in eastern Africa, Insect Molecular Biology, 9 (4): 357-367
http://dx.doi.org/10.1046/j.1365-2583.2000.00197.x

Fanello C., Petrarca V., della Torre A., Santolamazza F., Dolo G., Coulibaly M., Alloueche A., Courtis C.F., Toure Y.T. and Coluzzi M., 2003, The pyrethroid knock-down resistance gene in the *Anopheles gambiae* complex in Mali and further indication of incipient speciation within *An. gambiae* s. s., Insect Molecular Biology, 12 (3): 241-245
http://dx.doi.org/10.1046/j.1365-2583.2003.00407.x

Fanello C., Santolamazza F. and Della Torre A., 2002, Simultaneous identification of species and molecular forms of *Anopheles gambiae* complex by PCR-RFLP, Medical and Veterinary Entomology, 16:461-464
http://dx.doi.org/10.1046/j.1365-2915.2002.00393.x

Ferris S.D., Sage R.D., Huang C.M., Nielsen J.T., Ritte U. and Wilson A.C., 1983, Flow of mitochondrial DNA across a species boundary, Proceedings of the National Academy of Science, USA, 91: 2290-2294
http://dx.doi.org/10.1073/pnas.80.8.2290

Fettene M. and Temu E.A., 2003, Species-specific primer for identification of *Anopheles quadriannulatus* sp B (Diptera: Culicidae) from Ethiopia using a multiplex polymerase chain reaction assay, Journal of Medical Entomology, 40 (1): 112-115
http://dx.doi.org/10.1603/0022-2585-40.1.112

Gale K.R. and Crampton J.M., 1987, DNA probes for species identification of mosquitoes in the *Anopheles gambiae* complex, Medical and Veterinary Entomology, 1(2): 127-136
http://dx.doi.org/10.1111/j.1365-2915.1987.tb00333.x

Gerib S.A., 1985, Evolution of ribosomal DNA: In Molecular Evolutionary Genetics, R. J. Maclntyre (ed) Plenum Press, New York, 205-243
http://dx.doi.org/10.1007/978-1-4684-4988-4_7

Giles G.M., 1902, A handbook of the Gnats or Mosquitoes, 2nd edition, London: John Bale, Sons and Danielsson

Gillies M.T. and Coetzee M., 1987, A supplement to the Anophelinae of Africa South of the Sahara (Afrotropical region), South African Institute for Medical Research, Johannesburg, South Africa

Gillies M.T. and De-Mellion B., 1968, The Anophelinae of Africa South of the Sahara, South African Institute for Medical Research, Johannesburg, South Africa

Haridi A.M., 1972, Partial exophily of *Anopheles gambiae* species B in Khashm El-Girba area in eastern Sudan, Bull World Health Organization, 46: 39-46

Hillis D.M. and Dixon M.T., 1991, Ribosomal DNA: Molecular evolution and phylogenetic inference, The Quarterly Review of Biology, 66: 411 http://dx.doi.org/10.1086/417338

Himeidan Y.El S., Dukeen M.Y., El Rayah El A. and Adam I., 2004, *Anopheles arabiensis*: abundance and insecticide resistance in an irrigated area of eastern Sudan, Eastern Mediterranean Health Journal, 10:167-174

Hudson R.R., Slatkin M. and Maddison W.P., 1992, Estimation of levels of gene flow from DNA sequence data, Genetics 132:583-589

Hunt R.H., Coetzee M. and Fettene M., 1998, The *Anopheles gambiae* complex: a new species from Ethiopia, Transactions of the Royal Society of Tropical Medicine and Hygiene, 92: 231-235 http://dx.doi.org/10.1016/S0035-9203(98)90761-1

Innis M.A., Geifand D.H, Sninsky J.J. and White T. J., 1990, PCR Protocols, Academic Press, San Diego

Kimura M., 1980, A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution, 16:111-120 http://dx.doi.org/10.1007/BF01731581

Lindsay S.W. and Mortens W.J., 1998, Medical Biology Climate and Malaria, Bull World Health Organization, 76: 33-45

Nei M., 1987, Molecular Evolutionary Genetics. New York: Columbia University Press

Norris D.E., 2002, Genetic markers for study of the anopheline vectors of human malaria, International Journal for Parasitology, 32: 1607-1615 http://dx.doi.org/10.1016/S0020-7519(02)00189-3

Nyanjoum S.R., Chen H., Gebre-Micaael T., Bekele E., Sbililu J., Githure J., Beier J.C. and Yan G., 2003, Population genetic structure of *Anopheles arabiensis* mosquitoes in Ethiopia and Eritrea, Journal of Heredity, 94 (6), The American Genetic Association, 94: 457-463

Olsen G.J. and Woese C.R., 1993, Ribosomal RNA: a key to phylogeny, FASEBJ, 7: 113-123

Onyabe D.Y. and Conn J.E., 2001, Population genetic structure of the malaria mosquito *Anopheles arabiensis* across Nigeria suggests range expansion, Molecular Ecology, 10 (11): 2577-2591 http://dx.doi.org/10.1046/j.0962-1083.2001.01387.x

Oyewolea I.O. and Awololab T.S., 2006, Impact of urbanization on bionomics and distribution of malaria vectors in Lagos, southwestern Nigeria, J Vect Borne Dis, 43: 173-178

Paskewitz S.M., Coetzee M. and Hunt R.H., 1993, Evolution of the polymerase chain reaction method for identifying members of the *Anopheles gambiae* complex (Diptera: Culicidae) in South Africa, Journal of Medical Entomology, 30 (5): 953-957

Paskewitz S.M. and Collins F.M., 1990, Use of the polymerase chain reaction to identify mosquito species of the *Anopheles gambiae* complex, Journal of Medical and Veterinary Entomology, 4: 367-373 http://dx.doi.org/10.1111/j.1365-2915.1990.tb00453.x

Patton W.S., 1905, The culicid fauna of the Aden hinterland, their haunts and habits, J Bombay Nat Hist Sot, 16: 623-637

Petrarca V., Nugud A.D., Ahmed M.A.H., Haridi A.M., Abd El Nur O.M. and Coluzzi M., 1986, Dati Preliminari sul complesso *Anopheles gambiae* in Sudan, Parasitologia, 28: 304-306

Petrarca V., Sabatinelli G., Toure Y.T. and Di Deco M.A., 1998, Morphometric multivariate analysis of field samples of adult *Anopheles arabiensis* and *Anopheles gambiae* s. s (Diptera: Culicidae), Journal of Medical Entomology, 35(1): 16-25

Pock Tsy J.M., Duchemin J.B., Marramo L., Rabarison P., Le Goff G., Rajaonarivelo V. and Robert V., 2003, Distribution of the species of the *Anopheles gambiae* complex and first evidence of *Anopheles merus* as a malaria vector in Madagascar, Malaria Journal, 2: 33 http://dx.doi.org/10.1186/1475-2875-2-33

Proft J., Maier W.A. and Kampen H., 1999, Identification of six sibling species of the *Anopheles maculipennis* complex (Diptera: Culicidae) by a polymerase chain reaction assay, Parasitol Res, 85: 837-843 http://dx.doi.org/10.1007/s004360050642

Saitou N. and Nei M., 1987, The neighbor-joining method: A new method for reconstructing phylogenetic trees, Molecular Biology and Evolution, 4: 406-425

Sambrook J., Fritsch E.F. and Maniatis T., 1989, Molecular Cloning: a laboratory manual, 2nd edition, Cold Spring Harbor, N.Y., Cold Spring, Harbor Laboratory Press

Scott J.A., Borgdon W.C. and Collins F.H., 1993, Identification of single specimen of *Anopheles gambiae* complex by the polymerase chain reaction, American Journal for Tropical Medicine and Hygiene, 49: 520-529

Service M.W., 1980, A guide to Medical Entomology, Macmillan International College Editions, London

Service M.W., 1985, *Anopheles gambiae*: Africa's principal malaria vector (1902-1984), Bulletin of the Entomological Society of America, 31: 8-12

Sharma M. and Chaudhry S., 2010, Molecular Cytogenetics of some *Anopheles* Mosquitoes (Culicidae: Diptera), Electronic Journal of Biology, 6(1): 13-18

Slatkin M., 1995, A measure of population subdivision based on microsatellite allele frequencies, Genetics, 139: 457-462

Tajima F., 1983, Evolutionary relationship of DNA sequences infinite populations, Genetics, 105: 437-460

Tamura K., Peterson D., Peterson N., Stecher G., Nei M. and Kumar S., 2011, MEGA 5: Molecular Evolutionary Genetics Analysis

Taylor C., Toure Y.T., Carnahan J., Norris D.E., Dolo G., Traore S.F., Edillo F.E. and Lanzaro G.C., 2001, Gene flow among populations of the malaria vector *Anopheles gambiae* in Mali, West Africa, Genetics,157 (2): 743-750

Tripet F., Toure Y.T., Taylor C.E., Norris D.E. and Dolo G., 2001, DNA analysis of transferred sperm reveals significant levels of gene flow between molecular forms of *Anopheles gambiae*, Molecular Ecology, 10 (7): 1725-1732 http://dx.doi.org/10.1046/j.0962-1083.2001.01301.x

Wilkins E.E., Howell P.I. and Benedict M.Q., 2006, IMP PCR primers detect single nucleotide polymorphisms for *Anopheles gambiae* species identification, Mopti and Savanna rDNA types, and resistance to dieldrin in *Anopheles arabiensis*, Malaria Journal, 5:125 http://dx.doi.org/10.1186/1475-2875-5-125

Zahar A.R., 1985, Vectors binomics in the epidemiology and control of malaria, Part 1, The WHO African Region and the Southern WHO Eastern Mediterranean Region