Population Genetic of Anopheles Arabiensis Patton(Diptera: Culicidae) the Malaria Vector in the Republic of Sudan.

Mashair Mustafa (mashairm@hotmail.com)
University of Al Butana
Zairi Jaal
University Sains Malaysia
Sumia Abu Kashawa
University of Khartoum
Siti Azizah Mohd Nor
University Malaysia Terengganu

Research

Keywords: Anopheles arabiensis, microsatellite loci, population genetic, gene flow, bottleneck, Sudan.

Posted Date: December 22nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-131539/v1

Population genetic of *Anopheles arabiensis* Patton(Diptera: Culicidae) the malaria vector in the Republic of Sudan.

MashairSirElKhatimMustafa[1], Zairi Jaal[2], Sumia Abu Kashawa[3], Siti Azizah Mohd Nor[4]

[1]Department of Biology & Chemistry, Al Butana University, 200, Rufaa, Sudan

[2]School of Biological Sciences, Universiti Sains Malaysia, 11800 Penang, Malaysia.

[3]Faculty of Science, Department of Zoology, University of Khartoum, Sudan.

[4]Institute of Marine Biotechnology, Universiti Malaysia Terengganu, 21030 Kuala Terengganu, Malaysia.

mashairm@hotmail.com, zairijaal@gmail.com, abukashawa@hotmil.com, sazizah@usm.my.

Background

Anopheles arabiensis is a member of *An. gambiae* complex and a main malaria vector in Sudan. There is no sufficient *An. arabiensis* population genetic data available an understanding of vector population structure and genetics are important to the malaria vector control programs. The objective of this investigation is to study the population structure, gene flow and isolation by distance among *An. arabiensis* for developing control strategies

Methods

Mosquitoes were collected from six sites in Sudan using pyrethrum spray catch of indoor resting mosquitoes. Anopheline mosquitos were identified morphologically and based on species specific nucleotide sequences in the ribosomal DNA intergenic spacers (IGS). Seven microsatellite loci published *An. gambiae* primers were used to amplify the DNA of *An. arabiensis* samples.

Results

PCR confirmed that *An. arabiensis* was the main malaria vector found in the six localities. Of the seven microsatellite loci utilized, six were found to be highly polymorphic across populations, with high allelic richness and heterozygosity with the remaining one being monomorphic. Deviation from Hardy-Weinberg expectations were found in 21 out of 42 tests in the six populations due to heterozygotes deficiency. Bayesian clustering analysis revealed two gene pools, grouping samples into two population clusters; one includes four and the other includes two populations. The genetic distances between pairs of populations ranged from 0.06 to 0.24. Significant F_{ST} was observed between all *An. arabiensis* populations. Kr population indicated high genetic differentiation (F_{ST} ranged from 0.17 to 0.24). High gene flow (Nm= 1.6–8.2) was detected between clusters. There was evidence of a bottleneck event in the Hj population. No isolation by distance pattern was detected among populations.

Conclusions

This study revealed low levels of population differentiation with high gene flow among six *An. arabiensis* populations in Sudan.

Keywords

Anopheles arabiensis, microsatellite loci, population genetic, gene flow, bottleneck, Sudan.

Background

In Sudan, malaria remains as one of the most important public health problems plaguing its population. Thirty one *Anopheles* species have been recognized globally but only a few of them are malaria vectors [1].With the exception of Southern Sudan, the main malaria vector in Sudan is *An. arabiensis*[3-9]. However, in Southern Sudan *An. gambiae* and *An. funestus*, are the major vectors of malaria and their vectorial capacity maybe similar to that of *An. arabiensis*[2].

55 *Anopheles arabiensis* is a highly adaptable species with the capability to feed on multiple host
56 species, both indoors and outdoors and acclimatize to a wide range of larval habitats [10-11].At
57 present, indoor residual spraying (IRS) and insecticide treated bednets (ITNs) are the main
58 methods extensively used for vector control worldwide. These methods have proven valuable in
59 reducing malaria burden[12-13], but their usefulness have been threatened by increasing
60 prevalence of insecticide resistance in the most important malaria vectors as well as the
61 phenotypic flexibility of *An. arabiensis*. Thus, there is an urgent need for effective and
62 sustainable alternatives to these traditional vector control strategies.

63 Microsatellites are genetic markers of short tracts of tandemly repeated DNA sequences. These
64 markers have become the genetic marker of choice for studying the population genetics of many
65 eukaryotic species including mosquitoes. For instance, they have been widely utilised in such
66 studies in the *An. gambiae* complex [14-20]. Furthermore, they could be developed into PCR-
67 based molecular markers that are very useful for small organisms with limited extractable DNA
68 [21].

69 To date, one hundred and fifty polymorphic microsatellite loci have been characterized in
70 *An. gambiae*. s.l.[14, 22] which have been widely used to explain the population structure and
71 gene flow within and between members of the *An. gambiae* complex [23-25]. The majority of
72 these studies have mostly been conducted on *An. gambiae* with some limited data on *An.*
73 *arabiensis*. Present literature on *An. arabiensis* has revealed the lack of subpopulation
74 differentiation in relation to larval habitat utilization [24]. Lack of annual bottlenecks in response
75 to changes in the environment has also been documented [17, 26]. Large effective population
76 size and/or recent range expansion as opposed to group migration [26, 27] have been attributed

to the widespread gene flow. This is based on several experimental studies which have reported a short flight range for this malaria vector species [28] among villages in Gambia. On the other hand, there is evidence in support of population structuring[29] from West Africa and eastern outer islands[27] of Eastern Africa. Furthermore, limited gene flow has been observed between the west and south east of the Rift Valley and in Southern Zambia,[17, 23] respectively. Geographic distance and habitat alterations have been suggested as the main contributors of genetic isolation.

Anopheles arabiensis has changeable deme sizes ranging from as low as 25 km [27] to a few 1000 kilometres [17]. It was observed that in the Mwea Rice Scheme of Central Kenya, *An. arabiensis* mosquito densities decrease with increasing distance from the Scheme [30-31]. On the contrary, the human blood index [10] in addition to malarial transmission [30] by this species was significantly lower inside than in the outer areas of the rice scheme. With respect to their study, agricultural practices can alter local environment [32] as well as affect the number and diversity of larval habitats. All these factors influence mosquito reproductive fitness, survivorship and fertility [33]. Such alteration may change malarial transmission indices [34] and can lead to subpopulation differentiation [35-36] as was observed in this agricultural scheme. But on the other hand the lack of evident geographical barriers that could have restricted gene flow between mosquito populations in the surrounding areas had led to the generation of a single panmictic population. A number of studies, for example one that was conducted by Dolo et al,[37]in the irrigated area of Sahel in Mali had suggested that the existence of mosquito colonies in an adjoining non-irrigated area during the dry season is maintained through migration of a few individuals from the irrigated areas.

99 Several comparative population studies between *An. arabiensis* and *An. gambiae* have
100 shown a higher level of genetic differentiation in the latter species. Significant genetic
101 differentiation, F_{ST} = 0.072- 0.100 were observed for *An. gambiae* populations between western
102 Kenya and coastal Kenya using microsatellite markers. Lehmann et al,[38-39] suggested the
103 Great Rift Valley as a major gene flow barrier for this species. However, nonsignificant genetic
104 differentiation was identified for *An. arabiensis* population from the two areas using the same
105 loci [23]. Similarly, Donnelly and Townson [27] noted nonsignificant genetic differentiation of
106 *An. arabiensis* populations in Malawi and Sudan. It thus appeared that different mechanisms of
107 gene movement were in operation between the two species. Considering *An. gambiae*, these
108 studies were discordant with another study on *An. gambiae* s.s. (mean F_{ST} = 0.006) which was
109 genetically undifferentiated across the 6,650 km2 of the Kilombero valley landscape southern
110 Tanzania. This suggested that the genetic differentiation in other populations was not due to
111 physical barriers or distance. One plausible explanation is that there was environmental
112 diversification even within the Kilombero valley [19]. Thus, the differentiated populations of *An.*
113 *gambiae* could have been maintained by some degree of reproductive isolation.

114 With respect to *An. arabiensis,* several studies have reported varying levels of genetic
115 differentiation. Nyanjom et al,[15] detected low F_{ST} but statistically significant genetic structure
116 for *An. arabiensis* populations in Ethiopia and Eritrea. Similarly, [40] also documented low but
117 significant difference between populations of *An. arabiensis* in Northern Sudan using
118 microsatellite markers. On the other hand Simard et al, [29] reported high levels of genetic
119 differentiation in two island populations of *An. arabiensis* populations that were 240 km apart in
120 the Indian Ocean (F_{ST} 0.080–0.215). High levels of genetic differentiation were also detected

among *An. arabiensis* populations (mean F_{ST} = 0.066) in Kilombero valley southern Tanzania [19].

Therefore, the objective of this investigation was to study the population structure and gene flow among *An. arabiensis* populations in Sudan based on microsatellite markers which may assist in developing control strategies.

Materials and Methods

Study Areas

A total of 200 specimens of *An. arabiensis* were collected from June 2010 to May 2011from six different localities in Sudan representing different ecological zones separated by the River Nile and its tributaries. Three localities were located in Khartoum State, namely 1. Mygoma (My) 2. Al Haj Yousif (Hj) 3. El Gerif West (Gw). My and Hj are nearest to each other and located east of the Blue Nile. Al Haj Yousif (Hj) is North east of Helt Koko, where animals are bred for milk production in a rich green area on the west bank of the Blue Nile. The fourth and fifth sites are in Kassala State namely 4. Alhalang Shemal (H.sh), located on the east bank of the AlGash River area of non-agricultural land and 5. Alkrmota (Kr) located on the west bank of the Al Gash River in the centre of an agricultural area which is surrounded by groves of fruit and vegetables in all directions. The sixth population is in the Sennar state (Se), and was collected from 6. Abu Algoni (Se), a farming area which is located on the west bank of the Blue Nile River. The pairwise distances between the six localities ranged from 3.93- 569.25 Km.

Microsatellite PCR Amplification

Molecular classification of *An. gambiae* species complex in this study was conducted based on the ribosomal DNA intergenic spacers (IGS) [41]. DNA extraction from individual *Anopheles* was conducted using the DNeasy blood and Tissue E kit, (QIAGEN, Valencia, CA). Seven published *An. gambiae* microsatellite loci primers [14] were used to amplify the DNA samples. The PCR reactions were performed in a gradient thermal cycler (MJ Research PTC-200 Peltier Thermal Cycler) for 30 cycles. The PCR mix contained 1 uL of genomic DNA, 5X PCR buffer (Promega, Madison, WI), 15 pmol of each fluorescent labelled (NED, HEX or FAM) forward primer, 200 mMof each dATP, dCTP, dGTP and dTTP, 1.2 uL of 25mm $MgCl_2$ and 0.5 U *Taq* DNA polymerase (Promega, Madison, WI) in a 20 uL total reaction volume. Singleplex PCR amplification was conducted for loci AGXH678, AG2H290, AG2H603, AG2H143, AG3H29, AG3H45, AG3H158 (Table 1). The thermal cycling conditions were; an initial hold at 95ºC for 2min, followed by 30 cycles of 94˚C for 30s, 55˚C for 30s and 72˚C for 30s and a final extension at 72˚C for 5min. Satisfactory PCR products as detected in a 2% agarose gel were sent to the service provider (First BASE Laboratories Sdn. Bhd., Selangor, Malaysia) into two primer multiplex sets for fragment analysis. Set A contained a mixture of AGXH678, AG2H290 and AG3H45, while set B was a mixture of AG2H603, AG2H143, AG3H29 and AG3H158. Loci AGXH678, AG3H29, AG3H45 and AG3H158 are found outside the inversion regions of the chromosome, while AG2H603 and AG2H143 loci are found within fixed inversion of chromosome 2La, and AG2H 290 in the 2R polymorphic inversion.

Data Analysis

Allelic data scoring of alleles was carried out as described in Arif et al, [42]. Screening of all the genotypic data was executed using Micro-Checker v2.2.3 [43] to check for presence of

null alleles and stuttering or large allele dropouts. The Monte Carlo simulation (bootstrap) method was applied to generate expected homozygote and heterozygote frequencies of alleles. The HWE was used to calculate expected allele frequencies and the frequency of any null allele detected with significance level at $P< 0.05$ obtained through 1000 permutations. To ensure compatibility with different software analyses the raw data was converted into several specific data formats using CONVERT [44]. Significant relationship between alleles at any two loci was tested using the likelihood ratio test of linkage disequilibrium based on Expectation-Maximization (EM) algorithm [45]. This was applied to all pairwise comparisons of loci using Arlequin version 3.11 [46] with 10000 permutations followed by false discovery rate (FDR) adjustment [47] at 95% significant level.

Population genetic diversity was measured based on allelic richness (A_R), adjusted for different sample size and number of alleles (N_A). Inbreeding coefficient (F_{IS}) for each locus and population [48] was estimated in FSTAT v.2.9.3 [49]. To test for global deviation from HWE in a population, the $F_{IS}$was estimated. Mean genetic heterozygosity, observed (H_O) and expected (H_E) heterozygosities per locus and population were estimated over all loci. Testing of deviation from Hardy-Weinberg equilibrium (HWE) was conducted using the exact tests with 10000 steps in Markov chain and 10000 dememorization steps in Arlequin version 3.11 [46]. Multiple testing of HWE was adjusted using False discovery rate (FDR) corrections with a global significance level of 0.05

Estimates of population differentiation, using Wrights F_{ST} and Slatkin R_{ST} [50]over all loci were conducted. F_{ST} is based on the infinite allele model (IAM) which hypothesizes that each new allele is generated at a given rate, μ [51]. Slatkin R_{ST} is based on the stepwise mutational model

185 (SMM). The model assumes that new alleles are generated by adding or deleting a single repeat
186 unit of the microsatellite, with an equal probability μ/2. Consequently, alleles that vary most in
187 sizes is expected to be more distantly related than alleles of similar sizes

188 The program BOTTLENECK V 1.2.02 [52] was used to detect whether the populations
189 had experienced recent effective population size reduction. Two-phase (TPM) models, infinite
190 allele (IAM) and stepwise mutation (SMM), and deviation from HWE was estimated using a
191 two-tailed Wilcoxon sign-rank test followed by FDR adjustment. Qualitative descriptor of allele
192 frequency ("mode-shift" indicator) was also performed in BOTTLENECK to discriminate
193 "shifted mode" populations (bottleneck) from stable populations [53]. Mantel test was used to
194 investigate the correlation between geographical and genetic distances among populations using
195 Arlequin 3.11 [46].

196 Assignment of individuals to their respective source populations based on multilocus
197 genotypic data was determined in STRUCTURE version 2.3 [54]. An assumption of correlated
198 allele frequency among populations [55] and admix model was used with the burn in period and
199 MCMC length, each at 10000 and 10 iterations, respectively. The probabilities of genotype
200 assignment into each individual group were performed across replicates using CLUMPP version
201 1.1.2 [57] and the graphical presentation was carried out using Structure Harvester [58]. Finally,
202 based on genetic distance a Neighbour-joining tree was constructed to determine the
203 phylogenetic tree among the six populations using MEGA 5.0.5 [59].

204 **Results**

All 200 individuals from six populations of *An. arabiensis* in Sudan were successfully genotyped and scored for all seven microsatellite loci. No evidence for scoring error due to null alleles, large allele dropout or stuttering was detected after assessing with Microchecker.

Allelic Frequency Distribution and Linkage Disequilibrium

All microsatellite markers of *An. arabiensis* populations were found to be polymorphic in at least one population. However, locus AG3H29 was monomorphic in most populations, may be due to presence of null alleles. Loci AG2H143 and AG3H45 were moderately polymorphic with number of allele per locus ranging from 6-12 and 4-10 respectively. The total number of alleles per locus ranged from 2 to 12 with an average of 7.6. The mean A_R ranged from a minimum of 3.5 in Hj to a maximum of 6.8 in Se and the mean observed heterozygosity of alleles per locus ranged from 0.55 to 0.67, while means of expected heterozygosity ranged from 0.55 to 0.62 (Table 1 and Figure 1). Tests for linkage disequilibrium revealed that 14 pairwise comparisons (11.11% out of 126 pairwise comparisons) after Bonferroni correction were significantly deviated from the random association of alleles at two or more loci with the highest linkage disequilibrium detected from pairwise loci comparisons in Kr (5 pairs in loci AGXH678, AG2H290, AG2H143 and AG3H158), followed by My (4 in loci AGXH678, AG2H290, AG2H603 and AG2H143), H.sh (2 pairs in loci AG2H290, AG2H603), Hj (1 pair in AG3H158) and Se (2 pairs in loci AGXH678 and AG2H143). No linkage disequilibrium of loci was observed in Gw.

Hardy-Weinberg Equilibrium (HWE) and F_{IS}

Each locus was tested separately for significant departure from HWE. Observed heterozygosity varied from 0.12 to 0.94 while expected heterozygosity ranged from 0.11 to 0.76 (Table 2). Deviations from HWE were found in 21 out of 42 tests in the six populations. These were observed in locus AGXH678 (Gw and H.sh), locus AG2H290 (H.sh, Hj, My and Kr), locus AG2H603 (Gw, H.sh, Hj, My and Kr) locus AG2H143 (Gw, Hj, My and Se), locus AG3H45 (H.sh, Hj, My, Se and Kr), AG3H158 in Kr. There was no consistent pattern according to locus or population and fairly equal numbers of heterozygote deficiencies and heterozygote excess were observed. This could be due to population subdivision rather than the existence of null alleles based on the Microchecker results. Locus AG3H29 showed no deviations from Hardy-Weinberg equilibrium and this locus was monomorphic in all populations except in H.sh and Kr populations. Inbreeding coefficient (F_{IS}) over all loci showed that most of the populations had high excess of heterozygosity. F_{IS} negative values indicate no inbreeding.

Genetic Structure

A hierarchical AMOVA performed after defining into three groups (Khartoum, Sennar and Kassala) revealed that 2.75% of the total genetic variance (F_{CT}) was contributed by 'among groups' variation while 13.61% (F_{SC}) was 'among populations within group' variation (Table 3). 83.63% of the genetic variation was attributed between individuals within population i.e. intrapopulation variation. All hierarchical levels i.e. F_{SC} and F_{CT} and within population revealed significant variation.

246 Significant differentiation among populations, (F_{ST}) was observed between all *An.*
247 *arabiensis* pairwise comparisons for all loci (Table 4). F_{ST} ranged from 0.06 to 0.24. All showed
248 that each of the population was highly differentiated from at least one other population. But if Kr
249 is excluded and in most cases, Se as well, the comparisons will indicate moderate genetic
250 differentiation for all. Moderate genetic differentiation was observed between *An. arabiensis*
251 populations from My (Khartoum State) and Se (Sennar State) ($F_{ST=}$0.06) and Gw (Khartoum
252 State) and H.sh (Kassala state) ($F_{ST=}$0.08). High genetic differentiation was observed between
253 *An. arabiensis* populations from Kr and other populations F_{ST}= 0.17-0.24. Therefore, Kr
254 population was most genetically isolated from the rest. However, pairwise estimates of R_{ST}
255 ranged between 0.00 to 0.84. High R_{ST} value was found between Hj and Gw (0.84) and lowest
256 values between Kr and H.sh, My and H.sh and Se and H.sh. Gene flow (Nm) calculated from
257 mean F_{ST} statistics ranged from 1.5 – 9.05 suggesting high gene flow between populations (Table
258 5). High migration rate and therefore high gene flow was detected among Se (Sennar State) with
259 My and Gw (both Khartoum State); Hj and My (both Khartoum State). Very little gene flow was
260 observed between Kr (Kassala State) and other populations. When considering the loci outside
261 the inversion (AGXH678, AG3H29, AG3H45 and AG3H158), F_{ST} statistics ranged from 0.026
262 to 0.32 with mean $F_{ST=}$ 0.13. For loci inside inversion (AG2H603; AG2H290 and AG2H143),
263 F_{ST}ranged from 0.019 to 0.20 with mean F_{ST}= 0.16.

264 Gene flow (Nm) for loci outside the chromosomal inversions ranged from 1.04 to 43.52.
265 The highest gene flow was between My and H.sh with Nm= 43.52, while little gene flow was
266 found between Kr and Hj with Nm= 1.04. For loci inside the fixed chromosomal inversion, 2La,

a range of Nm= 1.38 to 19.99 was observed, with the highest gene flow between Se and My with Nm = 19.99, while little gene flow was detected between Se and Kr with Nm= 1.38.

The genetic distance ranged from 6% between My and Se (although are geographically distant) to 32% between Kr and Hj. The genetic distance was high between Kr and H.sh. Thus, although the detailed magnitudes of population differentiation vary among population comparisons in the various analyses, in summary Kr is most distant or differentiated from other populations and followed by Se to a certain extent for several pairwise comparisons.

Population bottleneck

The Wilcoxon test (Table 6) indicated that all populations excluding Se were significant for IAM mutation-drift equilibrium ($\alpha<0.05$), but with normal L-shaped distribution. This shows that these populations had not experienced population bottleneck. However, population Hj with shifted mode suggest recent population size reduction. TPM analysis showed insignificance for all populations except Hj while the SMM analysis showed insignificance for all populations.

Although there was significant differentiation, F_{ST}, between the different populations, Mantel tests showed no significant correlation ($r^2= 0.09$, $P> 0.05$) between genetic differentiation measured as linearized F_{ST} ($F_{ST/(1-F_{ST})}$ and geographic distance (km) (Figure 2). Thus genetic and geographic distances are not correlated.

Population Structure

Based on the programme STRUCTURE after calculations of the delta K and plotting its value against the assumed number of populations (K= 10), a peak at K =2 revealed two main

clusters (Figure 3). The dataset was further analyzed by assigning individuals between the two suggested clusters. Figure 4 explains the analysis for assignment of the most likely K (K = 2).This analysis is ingeneralagreement with the F_{ST} analyses where Kr and to a lower degree, Se are distant from the other populations although My is closely related to the latter. In the F_{ST}analysis,Kr and Se were also found to be genetically distant.However, the phylogenetic tree showed that only Kr is distant from the other populations. On the other hand, *An. arabiensis* populations from Hj, GW and H.sh;My and Se are highly related genetically (Figure 5).

Discussion

Allele Frequency and Linkage Disequilibrium

In this study, a set of seven microsatellite markers, specific for *An. gambiae* [14] was used to analyse the population genetics among six *An. arabiensis* populations in Sudan. Allele number per locus ranging from 2 to 12 with an average of 7.6 alleles per locus as well as heterozygosity levels are concordant with other studies in this species conducted in Central Kenyaand Madagascar[18] and Eastern Africa [27].But higher heterozygosity values were observed in the eastern Africa Islands of Reunion and Mauritius [26, 29].Similar values have also been obtained for other species for example in *An. gambiae s.s*. [39],*An. funestus* in east and southern Africa[60]and *An. atroparvus* [61]in southern Europe. Muturi et al, [18]suggested that the level of allelic polymorphism could provide powerful measures to identify population subdivision.

Analyses of linkage disequilibrium showed that all six *An. arabiensis* populations in Sudan suggest the existence of population subdivision. Similarly, using a different suite of

microsatellite markers for *An. arabiensis* in Southern Tanzania Ng'habi et al, [19] observed high linkage disequilibrium which they attributed to the presence of population subdivision. The significant linkage disequilibrium observed in this study could be attributed to heterozygote deficits due to several factors-departure from random mating, as a result of inbreeding or selection for certain genotypes following ecological and environmental changes.

Population differentiation and population structure

The population pairwise F_{ST} values ranged from 0.06 to 0.24. These values divided the six populations into two groups as defined by F_{ST} values of; Group 1 comprising of Kr and Se with intra and inter group F_{ST} values of 0.16- 0.24 (high genetic differentiation)and Group 2 comprising of H.sh, My, Hj and Gw with pairwise intra-group F_{ST} values of 0.08-0.13 (moderate genetic differentiation). This was in agreement with the Bayesian cluster analysis performed with STRUCTURE showing that the most likely *K* value identified was *K*=2.The K 1 group includes Hj, Gw, My and H.sh populations and K 2 include Se and Kr populations.

The F_{ST} from pooled loci reported in this study was inconcordance with the reported F_{ST} values for *An. arabiensis* from Ethiopia and Eritrea [15] and Northern Sudan [40]. Both studies detected low F_{ST} but statistically significant genetic structure in *An. arabiensis* populations. However, Donnelly and Townson [27] did not detect significant genetic structure for *An. arabiensis* populations within Malawi and Sudan. Chen et al, [64] also detected a low, but significant, genetic structure of *An. gambiae* in Lake Victoria islands (F_{ST} = 0.019) and among the six villages in the mainland (F_{ST}= 0.010).

A high level of genetic differentiation was found in *An. arabiensis* between certain populations in the current study in agreement with that observed between the Reunion and Mauritius Islands (F_{ST} ranged from 0.080 to 0.215), located 240 km apart in East Africa Simard *et al*.,[29] . Similarly, high levels of genetic differentiation was detected among *An. Arabiensis* populations (mean F_{ST} = 0.066) [19]. This was not in agreement with *An. gambiae s. s*.,(mean F_{ST} = 0.006)which was genetically undifferentiated across the 6,650 km2 of the Kilombero valley landscape in southern Tanzania.studied genetic structure of *An. gambiae* populations among islands in north-western Lake Victoria, Uganda with F_{ST} ranging from 0.014–0.105 and concluded that these populations were significantly genetically differentiated [65]. Similar to F_{ST} values of 0.20 and 0.30 as described byVicente et al, [61],Kamau et al,Kamau, [66] observed F_{ST} values of 0.25 while [67], who studied genetic population structure and introgression in *An. dirus* mosquitoes in South-east Asia, reported F_{ST} value of 0.21-0.39.

The high differentiation of Kr from the other populations may be due to its ecology which was far from agriculture areas compared to the restwhich were near or within agriculture areas. Kr population is presumably reproductively isolated due to nonrandom mating or ecologically isolated as an effect of the AlGash River which acts as a physical barrier separating Kr from other populations. Coluzzi,[68] hypothesised that inversions may play an important role in the isolation process among species in the *An. gambiae* complex and between the various forms of *An. gambiae* s.s. He postulated that inversions can group co-adapted gene complexes that confer adaptation in temporarily isolated peripheral populations with marginal ecological conditions. When secondary contact with the source population occur, these inversions protect the co-adapted gene complexes from recombination, resulting in stable inversion polymorphisms and/or

expansions of the population into new habitat, finally resulting in a more permanent isolation and differentiation.

There is evidence that R_{ST} may not be the best estimator of population substructure for microsatellites data.Forbes et al,[69] studied microsatellite evolution in congeneric mammals of the domestic and bighorn sheep and found that F_{ST} was more sensitive to differences between allopatric populations compared to R_{ST} Perez-Lezaun et al, [70] conducted studies on microsatellite variation and the differentiation of modern humans and observed that genetic distance methods such as F_{ST} which do not take into account mutational relationships among alleles and which are associated with occurrence of differentiation through drift, have been shown to be better estimators of patterns of human evolution than do R_{ST} which are calculated based on distance methods. Schuget al,[71] suggested that mutation rates at microsatellite loci are lower than had been previously assumed. By that reasoning, Donnelly et al, [72] concluded that if the East African *An. arabiensis* populations had experienced a lower mutation rate and only became isolated very recently, R_{ST} estimators may not be sufficiently sensitive to detect any differentiation. Constrains upon accumulation of mutations in each population and a lower rate of mutation at these loci would increase the influence of drift on allele frequencies. The present study suggests that F_{ST} is a more appropriate estimator than R_{ST} for recently diverged populations and therefore low mutations can be detected, using F_{ST}.Kr population was found to be differentiated from the other populations and reproductively isolated but this was not detected using R_{ST}.

High genetic similarity between My and Hj which are geographically close to each other was observed, suggesting that they may represent a single population or gene pool. According to

372 [51], neighbouring populations are expected to be genetically more related than distant
373 populations and high gene flow generally prevent local adaptation. Consequently human
374 transportation or wind dispersal may be the reason behind the continuous gene flow between
375 these two localities and to a lower degree, two others namely GW and H.sh.The high gene flow
376 found between these particular populations in this study is interesting with regards to the spread
377 of insecticide resistance in Sudan. It is concordant with Kent et al, [17] who studied spatial and
378 temporal genetic structure of *An. arabiensis* in Southern Zambia. They observed high gene flow
379 between Macha and Namwala populations, in Southern Zambia. This was also in agreement with
380 Muturi *et al*. [18]who found high gene flow among the three populations of *An. arabiensis* in
381 Central Kenya. [73] detected a significant correlation between gene flow and commercial traffic
382 by planes and/or boats between islands on *Aedes polynesiensis* populations from islands in
383 French Polynesia.

384 The Mantel test showed no evidence of isolation by distance in this study,similar with
385 that reported by Nyanjom et al, [15] in populations of *An. arabiensis* from Ethiopia and Eritrea.
386 Isolation by distance could occur due to the limited flight range of *An. Arabiensis* Adams, [74]
387 but this was not the case for this study. It is known that the the distributional range of any species
388 is largely shaped by historical and geographical events. The species will extend its range until it
389 reaches a physical (mountain ranges, deserts and major geographic feature) or other forms of
390 barriers (example climatic changes). However, there are no variable ecological zones, or great
391 physical barriers which could have led to population structuring. These results points to great
392 impact of transportation in the genetic structure of *An. arabiensis* along the River Nile. Thus,
393 presumably, any differentiation could largely be due to other factors than geographic distance. A

394 similar observation was seen within *An. atroparvus* Vicente et al, [61] where no correlation
395 between geographic or genetic distances was detected in a study conducted in southern Europe.
396 In agreement, Kamau et al, [76] using microsatellite loci revealed that there were no significant
397 relation between geographic and genetic distance in *An.arabiensis* and *An. gambiae* suggesting
398 that levels of genetic differentiation are not related to geographical distance and not associated to
399 the side on which populations were sampled in relation to the Rift Valley. This finding is in
400 contrast with Azrag, [40] who found that there was significant genetic differentiation related to
401 geographic distance in their study on *An. arabiensis* in northern Sudan. Their samples showed
402 increasing differentiation with increasing geographical separation. Chen et al, [64] who studied
403 population genetic structure of *An. gambiae* mosquitoes on Lake Victoria islands, west Kenya
404 revealed a significant correlation between geographic distance and pairwise distance. On the
405 other hand,Failloux et al,[73] found no significant effect of geographic distance on the
406 population genetic structure on *Aedes polynesiensis* populations from islands in French Polynesia
407 in contrast to the genetic structuring pattern of *Culex pipiensquinques fasciatus* from the same
408 islands. In the latter species, genetic differentiation increased considerably ($P < 0.01$) with
409 geographic distance [75]. These differences may be due to the variable biology of the two
410 species as well as their histories of colonization. This study conclude high population
411 differentiation and continuous gene flow among the studied populations, however without any
412 signs of isolation. The high migration rate and lack of interpopulation genetic variation among
413 the Sudanese population was attributed to the continuous human and domestic animals
414 movement among the studied localities that might help in the distribution of *An. arabiensis*.

415

Population size bottleneck

No severe bottleneck or reduction in population size was detected in the *An. arabiensis* populations of Sudan except in AlHajYousif (Hj). This was evident from the significance in Wilcoxon sign-rank test which was also supported by the "shifted mode" allele distribution. In comparison, the other populations had relatively higher rare alleles than common alleles, a sign that these populations are experiencing mutation-drift equilibrium.The situation at AlHajYousif (Hj) is likely due to the effective vector control programme in this area. This finding did not agree with the Muturi et al,[18] who studied the population genetic structure of *An. arabiensis* in central Kenya. They did not find any evidence of genetic bottlenecks in the area under different agricultural practices. Furthermore, there was no evidence of a genetic bottleneck in *An. arabiensis* despite a drastic reduction in mosquito numbers during the drought year in southern Zambia as reported byKent et al,[17]. This is similar to the present study, where there is a reduction in *An. arabiensis* during the dry season but no occurrence of genetic bottlenecks apart from Hj which is under vector control programmes. Hj,My and Gw in Khartoum state have very strong programmes of malarial control 'Khartoum Malaria Free Initiative' started from 2001 to 2009. The significant achievement in malarial control in Khartoum state is highly evident. For example, the percentage of malaria cases among the followers of health services decreased from 20 percentage in 2001 to just 3.3 percentage in 2008 and the parasitological incidence has gone down from 91 to just 4 per 10,000 population. Another programme was initiated in 2011 and is due to end in 2015. The objective of this initiative is to decrease the number of morbidity and mortality of malaria cases by 90% by 2015 in northern Sudan compared to the number of reported cases in 2009. However, results of this study in My and Gw did not show any

reduction. This could be explained due to the resistance to insecticide. Therefore, no reduction in these populations means they are expanding or developing up quickly.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All data generated or analysed during this study are included in the text.

Competing interests

The authors declare that they have no competing interests.

Funding

This work funded by TWAS-USM Postgraduate Fellowship FR number: 3240189425 for the first Authors PhD.

Authors' contributions

MSM: did the field and practical labwork, data analyses and prepared the manuscript. ZJ and SMN supervised the study and edited the manuscript. SA supervised the field work.

ACKNOWLEDGEMENTS:

This project was mainly supported by the Postgraduate Research Grant Scheme (USM-RU-PRGS) of the Universiti Sains Malaysia. Thanks also go to TWAS for providing the PhD fellowship, and to the Unversity of Albutana for granting me the study leave. Special thanks go to the Zoology Department, Faculty of Science, University of Khartoum and the Malaria

Research Center in Sennar for allowing me to use their laboratory and the technical assistance given during the period of the study.

REFERENCE

1. Nugud AD, Eltyeb RA, Abd-El-Nur OM. Vectors of malaria in Sudan. Joint Workshop on Scientific Cooperation, the Federal Ministry of Agricultural and Forestry, Sudan, ICIPE, Kenya, Khartoum, 6-7 Dec. 1997.

2. El Sayed BB Transmission and risk factors of Urban and Peri-Urban malaria in Khartoum state, Sudan, phD Thesis 1998, Dept. Of Zoology, University of Khartoum.
3. Abdel Elnur OM, Dukeen MYH. Vectors of malaria in Sudan, National conference for malaria, Malaria Administration1992, Ministry of Health, Khartoum, Sudan.

4. Petrarca V, Nugud AD, Ahamed MA, Haridi AM, Di Deco MA, et al., Cytogenetics of *An. arabiensis* relationships with East and West African populations. *Medical and Veterinary Entomology* 2000, 14(2):149- 164.

5. Himeidan YE, Dukeen MY, EL-Rayah EA, Adam I, *Anopheles arabiensis*: abundance and insecticide resistance in an irrigated area of eastern Sudan. *Eastern Mediterranean Health Journal* 2004, 10(1-2)**:**167-174.

6. Abdalla H, Matambo TS, Koekemoer LL, Mnzavae AP, Hunt RH, Coetzee M. Insecticide susceptibility and vector status of natural populations of *Anopheles arabiensis* from Sudan. *Royal Society of Tropical Medicine and Hygiene*2008, 102**:** 263-271.

7. Tasneem MAO. Species identification and infectivity rate of malaria vector in two endemic malaria areas in Sudan. *Egyptian Academic Journal of biological Science*2010, 2 (1): 1- 15.

8. Seidahmed OME, Abdelmajed MA, Mustafa MS, Mnzava AP. Insecticide susceptibility status of the malaria vector *Anopheles arabiensis* in Khartoum city, Sudan: differences between urban and periurban areas. *Eastern Mediterranean Health Journal* 2012, 8(7)**:** 769-776.

9. Yagoop JSH, Bashir N, OHM, Assad Y. Susceptibility of *Anopheles arabiensis* (Diptera: Culicidae) adults to some commonly used agricultural insecticides in El Rahad Agricultural Corporation, Central Sudan. Scholarly *Journal of Agricultural Science* 2013, 3(1)**:** 10-20.

10. Muturi J, Muriu S, Shililu J, Mwangangi J, Jacob B, Mbogo C, Githure, J, etal. Effect of rice cultivation on malaria transmission in central Kenya. *American Journal for Tropical Medicine and Hygiene*2008, 78:270–275.

11. Mwangangi J, Muturi J, Shililu J, Muriu S, Jacob B, Kabiru E, Mbogo C, etal. Contribution of different aquatic habitats to adult *Anopheles arabiensis* and *Culex quinquefasciatus* (Diptera: Culicidae) production in a rice agroecosystem in Mwea, Kenya. *Journal of Vector Ecology*2008, 33(1):129-138.

12. Lengeler C. Insecticide-treated bed nets and curtains for preventing malaria. *Cochrane Database of Systematic Reviews* 2004 CD000363.

13. Fegan GW, Noor AM, Akhwale WS, Cousens S, Snow RW. Effect of expanded insecticide-treated bednet coverage on child survival in rural Kenya: a longitudinal study. *Lancet*2007, 370:1035–1039.

14. Zheng L, Benedict MQ, Cornel AJ, Collins FH, Kafatos FC.An integrated genetic map of the African human malaria vector mosquito, *Anopheles gambiae***.** *Genetics*,1996,143**:**941-952.

15. Nyanjom SR, Chen H, Gebre-Michael T, Bekele E, Shililu J, Githure J, Beier JC, etal. Population genetic structure of *Anopheles arabiensis* mosquitoes in Ethiopia and Eritrea. *Journal of Heredity*2003, 94:457-463.

16. Temu EA, Yan G. microsatellite and mitochondrial genetic differentiation of *Anopheles arabiensis* (Diptera: Culicidae) from western Kenya, the Great Rift Valley, and coastal Kenya. *American Journal of Tropical Medicine and Hygiene*2005,73(4), 726–733.

17. Kent RT, Mharakurwa S, Norris DE. Spatial and temporal genetic structure of *Anopheles arabiensis* in Southern Zambia over consecutive wet and drought years. *American Journal of Tropical Medicine and Hygiene*2007,**77:**316-323**.**

18. Muturi J, Kim C, Baliraine F, Musani S, Jacob B, John Githure J, etal. Population Genetic Structure of *Anopheles Arabiensis* (Diptera: Culicidae) in a Rice Growing Area of Central Kenya. *Journal of Medical Entomology*2010, 47(2): 144–151.

19. Ng'habi KR, Knols BG, Lee Y, Ferguson HM, Lanzaro G C. Population genetic structure of *Anopheles arabiensis* and *Anopheles gambiae* in a malaria endemic region of southern Tanzania. *Malaria Journal*2011, 10:289.

20. Deitz KC, Athrey G, Reddy MR, Hans j, Overgaard HI, Matias A, Jawara M, etal. Genetic isolation within the malaria mosquito *Anopheles melas*. *Molecular Ecology* 2012, 21, 4498–4513.

543 21. Aransay AM, Ready PD, Morillas-Marquez F. Population differentiation of *Phlebotomus*
544 *perniciosus* in Spain following postglacial dispersal. *Heredity* 2003, 90; 316 – 25.
545
546 22. Zheng L, Collins FH, Kumar V, Kafatos FC. A detailed genetic map for the X
547 chromosome of the malaria vector, *Anopheles gambiae*. *Science*1993, 261:605–608.
548
549 23. KamauL, Mukabana WR, Hawley WA, Lehmann T, Irungu LW, Orago AA, etal.
550 Analysis of genetic variability in *Anopheles arabiensis* and *Anopheles gambiae* using
551 microsatellite loci. *Insect Molecular Biology*1999, 8:287–297.
552
553 24. Kamau L, Munyekenye GO, Vulule JM, Lehmann T. Evaluating genetic differentiation
554 of *Anopheles arabiensis* in relation to larval habitats in Kenya. *Infection*, *Genetics* and
555 *Evolution* 2007, 7**:**293-297.
556
557 25. Moreno M, Salgueiro P, Vicente JL, Cano J, Berzosa PJ, de Lucio A, Simard F, etal.
558 Genetic population structure of *Anopheles gambiae* in Equatorial Guinea. *Malaria.*
559 *Journal* 2007, 6:137.
560
561 26. Simard F, Lehmann T, Lemasson JJ, Diatta M, Fontenille D. Persistence of *Anopheles*
562 *arabiensis* during the severe dry season conditions in Senegal: an indirect approach using
563 microsatellite loci *Insect Molecular Biology*2000, 9:467–479.

27. Donnelly MJ Townson H. Evidence for extensive genetic differentiation among populations of the malaria vector *Anopheles arabiensis* in Eastern Africa. *Insect Molecular Biology*2000; 9:357–367.

28. Thomson MC, Connor SJ, Quinones ML, Jawara M, Todd J, Greenwood BM. Movement of *Anopheles gambiae s.l.* malaria vectors between villages in The Gambia. *Medical* and Veterinary *Entomology*1995, 9: 413–419.

29. Simard F, Fontenille D, Lehmann T, Girod R, Brutus L, Gopaul R, Dournon C, etal. High amounts of genetic differentiation between populations of the malaria vector *Anopheles arabiensis* from West Africa and eastern outer islands. *American Journal of Tropical Medicine and Hygiene*1999, *60:* 1000–1009.

30. Mutero C, Kabutha C, Kimani V, Kabuage L, Gitau G, Ssennyonga J, Githure J, etal. A transdisciplinary perspective on the links between malaria and agroecosystems in Kenya. *Acta Tropica*2004b*,* 89:171–186.

31. Muturi, J., Shililu, J., Jacob, B., Githure, J., Gu, W., Novak, R. Mosquito species diversity and abundance in relation to land use in a riceland agroecosystem in Mwea, Kenya. *Journal of VectorEcology*2006*,* 31:129–137.

32. Karaca, M., Ekici, A., Sen, O., Yesilnacar, I., Kindap, T. Irrigation induced climate change in Southeastern Turkey. Geophysical Research Abstracts 2008;10 EGU2008-A-08390.

33. Afrane, Y. A., Zhou, G., Lawson, B. W., Githeko, A. K., Yan, G. Effects of microclimatic changes caused by deforestation on the survivorship and reproductive fitness of *Anopheles gambiae* in western Kenya highlands. *American Journal of Tropical Medicine and Hygiene* 2006, 74:772–778.

34. Ijumba, J. N. and Lindsay, S. W. Impact of irrigation on malaria in Africa: paddies paradox. *Medical* and Veterinary *Entomology*2001, 15:1–11.

35. Toure, Y. T., Petrarca, V., Traore, S. F., Coulibaly, A., Maiga, H. M., Sankare, O., Sow, M., etal. Ecological genetic studies in the chromosomal form Mopti of *Anopheles gambiae* s.str. in Mali, west Africa. *Genetica*1994, 94:213–223.

36. Toure, Y. T., Dolo, G., Petrarca, V., Traore, S. F., Bouare, M., Dao, A.,Carnahan, J., etal. Mark-releaserecapture experiments with *Anopheles gambiae* s.l. in Banambani Village, Mali, to determine population size and structure. *Medical* and Veterinary *Entomology*1998,12:74–83.

607

608 37. Dolo, G., Briet, O. J. T., Dao, A., Traore, S. F., Bouare, M., Sogoba, N., Niare, O., etal.
609 Malaria transmission in relation to rice cultivation in the irrigated Sahel of Mali. *Acta*
610 *Tropica*2004, 89:147–159.
611

612 38. Lehmann, T., Hawley, W. A., Grebert, H., Danga, M., Atieli, F., Collins, F. H. The Rift
613 Valley complex as a barrier to gene flow for *Anophelesgambiae* in Kenya**.** *Journal of*
614 *Heredity* 1999, 90**:**613-621.
615

616 39. Lehmann, T., Licht, M., Elissa, N., Maega, B. T., Chimumbwa, J. M., Watsenga, F. T.,
617 Wondji, C. S., etal. Population structure of *Anopheles gambiae* in Africa. *Journal of*
618 *Heredity* 2003, 94:133-147.
619

620 40. Azrag, R. S. Microsatellite and mitochondrial genetic differentiation of *Anopheles*
621 *arabiensis* Patton (Diptera: Culicidae) along the River Nile, Northern Sudan. PhD Thesis
622 2007, Faculty of Science, University of Khartoum, Sudan.
623

624 41. Scott, J. A., Brogdon, W. G., Collins, F. H. Identification of single specimens of the
625 *Anopheles gambiae* complex by the polymerase chain reaction. *American Journal for*
626 *Tropical Medicine and Hygiene*1993, 49**:**520–529.
627

628 42. Arif, I. A., Khan, H. A., Shobrak, M., Al Homaidan, A. A., Al Sadoon, M., Al Farhan, A.
629 H. and Bahkali, A. H. Interpretation of electrophoretograms of seven microsatellite loci
630 to determine the genetic diversity of the Arabian Oryx. *Genetics and Molecular*
631 *Research*2010, 9 (1): 259- 265.
632
633 43. Van Oosterhout, C., Hutchinson, W. F., Wills, D. M., Shipley, P. Micro-checker:
634 software for identifying and correcting genotyping errors in microsatellite data.
635 *Molecular Ecology Notes*2004, 4:535–538.
636
637 44. Glaubitz, J. C. Convert: A user- friendly program to reformat diploid data for commonly
638 used population genetic software packages. *Molecular Ecology Notes*2004, 4(2): 309 –
639 310.
640
641 45. Slatkin, M. and Excoffier, L. Testing for linkage disequilibrium in genotypic data using
642 the EM algorithm. *Heredity*1996, 76:377-383.
643
644 46. Excoffier, L., Laval, G. and Schneider, S. Arlequin ver. 3.0: An integrated software
645 package for population genetics data analysis. *Evolutionary Bioinformatics Online*, 2005
646 1:47-50.
647

47. Benjamini, Y. and Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society* 1995, 57(1): 289- 300.

48. Weir, B.S. and Cockerham, C.C. Estimating F-statistics for analysis of population structure. *Evolution* 1984,38: 1358-1370.

49. Goudet, J. FSTAT version 2.9.3.2. A computer software to calculate Fstatistics. *Journal ofHeredity* 1995, 86:485-486.

50. Goodman, S. J. R-ST Calc: a collection of computer programs for calculating estimates of genetic differentiation from microsatellite data and determining their significance. *Molecular Ecology* 1997, 6: 881-885.

51. Francois, B. and Nicolas, L. M. The estimation of population differentiation with microsatellite markers. *Molecular Ecology*2002, 11:155-65.

52. Piry, S., Luikart, G., Cornuet, J. M. BOTTLENECK A Computer software program for detecting recent reductions in the effective population size using allele frequency data. *Journal of Heredity*1999, 90:502–503.

669 53. Luikart, G., Allendorf, F. W., Cornuet, J. M., Sherwin, W. B. Distortion of allele
670 frequency distributions provides a test for recent population bottlenecks. *Journal of*
671 *Heredity* 1998a, 89:238–247.
672
673 54. Pritchard, J. K., Stephens, M., Donnelly, P. Inference of population structure using
674 multilocus genotype data. *Genetics* 2000, 155:945–959.
675
676 55. Falush, D., Stephens, M. and Pritchard, J. K. Inference of population structure using
677 multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 2003,
678 164:1567–1587.
679
680 56. Evanno, G., Regnauto, S. and Goudet, J. Detecting the number of cluster of individuals
681 using the software STRUCTURE A simulation study. *Molecular Ecology*2005, 14, 2611-
682 20.
683
684 57. Jakobsson, M., and Rosenberg, N. A. CLUMPP: a cluster matching and permutation
685 program for dealing with label switching and multimodality in analysis of population
686 structure. *Bioinformatics*2007, 23: 1801-1806.
687

688 58. Earl, D. A., and vonHoldt, B. M. STRUCTURE HARVESTER: a website and program
689 for visualizing STRUCTURE output and implementing the Evanno method.
690 *Conservation Genetics Resources*2012, 4(2): 359-361.
691
692 59. Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M. and Kumar, S. MEGA5:
693 Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary
694 Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* 2011,
695 28910): 2731-2739**.**
696
697 60. Temu, E. A., Hunt, R. H., Coetzee, M. Microsatellite DNA polymorphism and
698 heterozygosity in the malaria vector mosquito *Anopheles funestus* (Diptera: Culicidae) in
699 east and southern Africa. *Acta Tropica*2004, 90:39-49.
700
701 61. Vicente, J. L., Sousa, C. A., Alten, B., Caglar, S. S., Falcutá, E., Latorre, J. M., Toty,
702 C., etal. Genetic and phenotypic variation of the malaria vector Anopheles atroparvus in
703 southern Europe. *Malaria Journal*2011, 10:5.
704
705 62. Slatkin, M. A measure of population subdivision based on microsatellite allele
706 frequencies.*Genetics*1995, 139:457-462.
707

63. Ohta, T. Linkage disequilibrium due to random genetic drift in finite subdivided populations. *Proceeding of the National Academy of Sciences* USA 1982, 79:1940-1944.

64. Chen, H., Minakawa, N., Beier, J. and Yan, G. Population genetic structure of *Anopheles gambiae* mosquitoes on Lake Victoria islands, west Kenya. *Malaria Journal* 2004, 3:48.

65. Kayondo, J. k., Mukwaya, L. G., Stump, A., Michel, A. P., Coulibaly, M. B., Besansky, N. J. and Collins, F. H. Genetic structure of *Anopheles gambiae* populations on islands in northwestern Lake Victoria, Uganda. *Malaria Journal* 2005, 4:59.

66. Kamau, L., Hawley W. A., Lehmann T., Orago, A. S., Cornel, A., Ke, Z., Collins F. H. Use of short tandem repeats for the analysis of genetic variability in sympatric populations of *Anopheles gambiae* and *Anopheles arabiensis*. *Heredity* 1998, 80:675–682.

67. Walton, C., Handley, J. M., Collins, F. H., Baimai, V., Harbach, R. E., Deesin, V., Butlin, R. K. Genetic population structure and introgression in *Anopheles dirus* mosquitoes in South-east Asia. *Molecular Ecology* 2001, 10:569-580.

68. Coluzzi, M. Spatial distribution of chromosomal inversions and speciation in anopheline mosquitoes. In C. Bargozzi, ed. Mechanisms of Speciation 1982, pp**.** 143–153**.** Alan R. Liss, New York.

69. Forbes, S. H., Hogg, J. T., Buchanan, F. C., Crawford, A. M. and Allendorf, F. W. Microsatellite evolution in congeneric mammals: Domestic and bighorn sheep. *Molecular and Biological Evolution* 1995, 12: 1106-1113.

70. Perez-Lezaun, A., Calaell, F., Mateu, E., Comas, D., Ruiz- Pacheco, R. and Bertranpetit, J. Microsatellite variation and the differentiation of modern humans. *Human Genetics*1997, 99(1): 1-7.

71. Schug, M. D., Mackay, T. and Charles, F. Low mutation rates of microsatellite loci in *Drosophila melanogaster.Nature Genetics*1997, 15: 99- 102.

72. Donnelly, M. J., Cuamba, N., Charlwood, J. D., Collins, F. H. and Townson, H. Population structure in the malaria vector, *Anopheles arabiensis* Patton, in East Africa. *Heredity* 1999, 83:408–417.

73. Failloux, A. B., Raymond, M., Ung, A., Chevillon, C. andPasteur, N. Genetic differentiation associated with commercial traffic in the Polynesian mosquito, *Aedes polynesiensis* Marks 1951**.** *Biological Journal of the LinneanSociety* 1997, 60**:**107-118.

74. Adams, P. C. G. Some observations on the flight of stained anophelines in Nkana, North Rhodesia. *Annals of tropical Medicine and Parasitology*1940*,* 34: 35-43.

75. Pasteur, N., Marquine, M., Rousset, F., Rousset, F., Failloux, A. B., Chevillon C. and Raymond, M. The role of passive migration in dispersal of resistance genes in *Culexpipiens* quinquefasciatus within French Polynesia. *Genetical Research*1995, 66: 139- 146.

76. Kamau, L., Mukabana W. R., Hawley, W. A., Lehmann, T., Irungu, L. W., Orago, A. A.and Collins F. H. Analysis of genetic variability in *Anopheles arabiensis* and *Anopheles gambiae* using microsatellite loci. *Insect Molecular Biology*1999, 8:287–297.

Figures

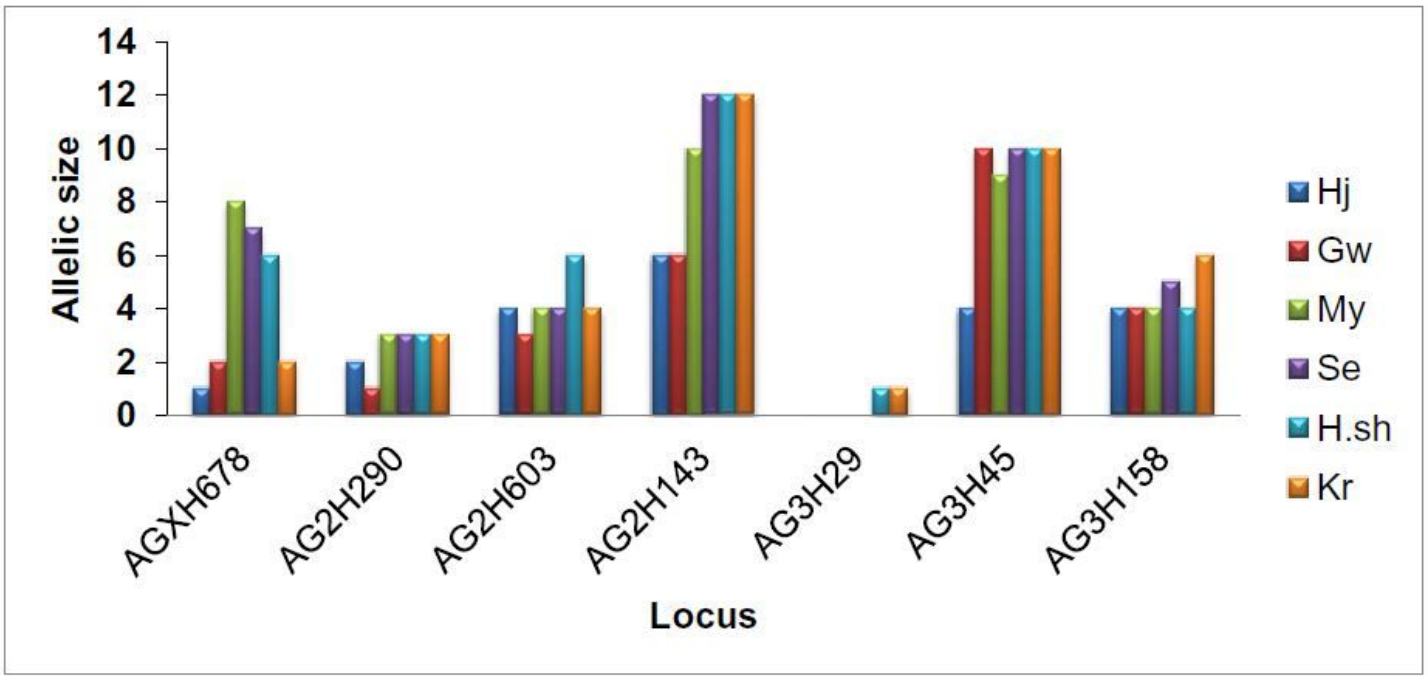

Figure 1

Number of alleles observed in each locus for each population.

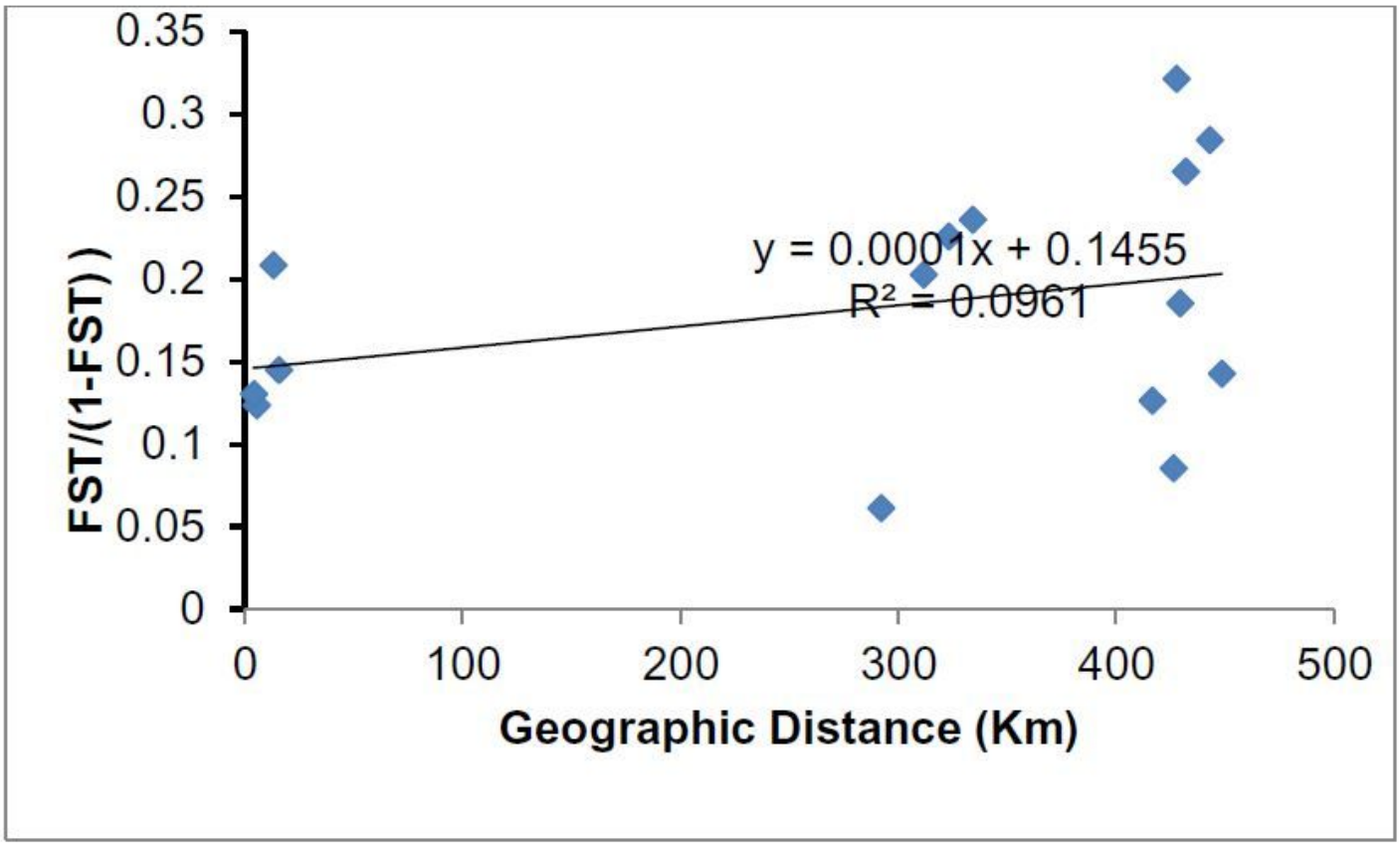

Figure 2

Scatter plot of the relationship between genetic and geographic distances. No significant correlation between genetic differentiation and geographic distance among An. arabiensis populations.

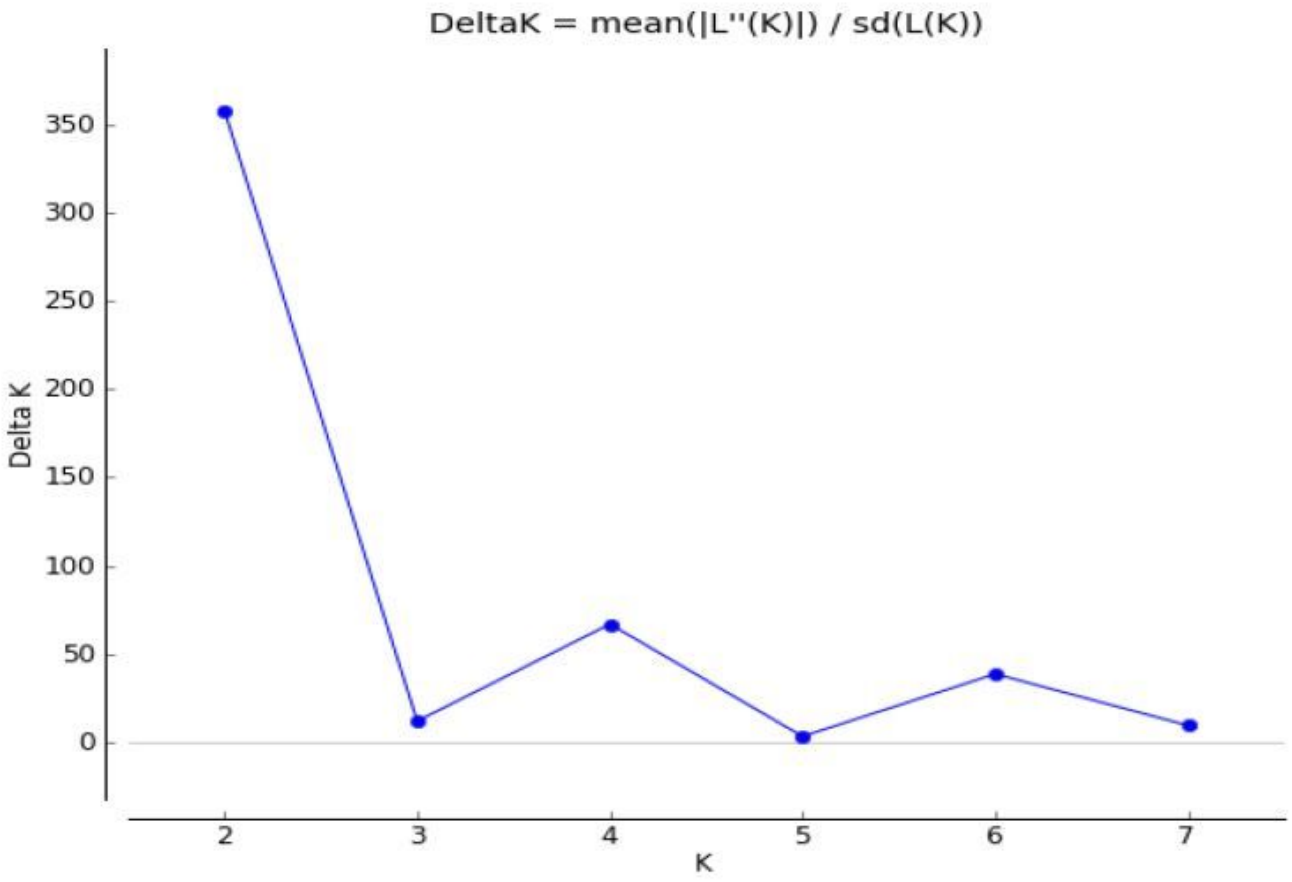

Figure 3

Graphical analysis using STUCTURE.HARVESTER Representation of the data set for the populations most likely K (K = 2).

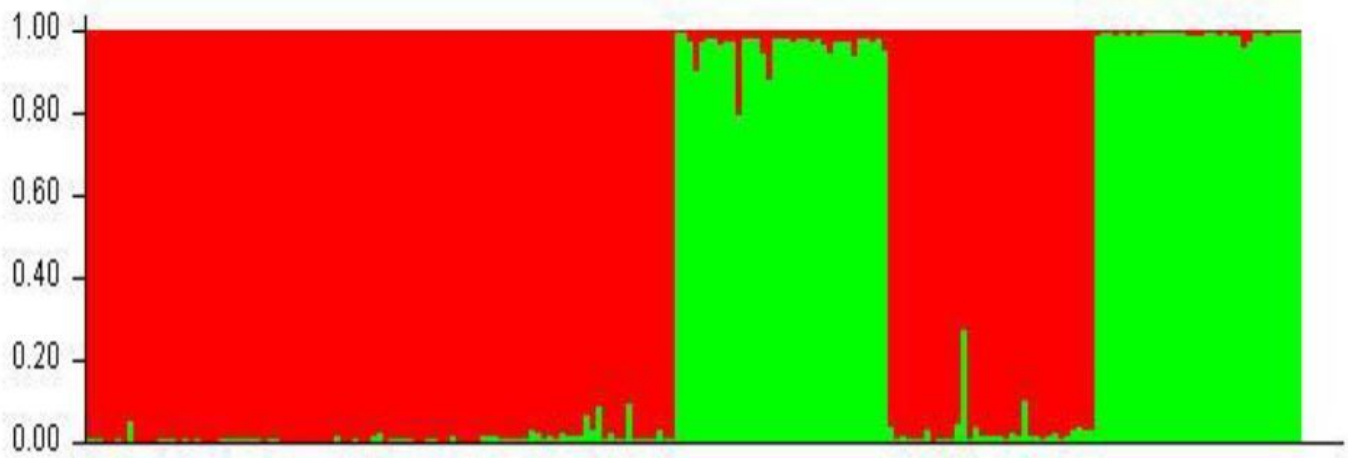

Figure 4

Graphical Bayesian cluster analysis using STRUCTURE. Representation of the data set for the most likely K (K = 2), where each color corresponds to a suggested cluster. Subpopulation A (red color) includes Hj, Gw, My and H.sh populations and subpopulation B (green color) includes Se and Kr populations.

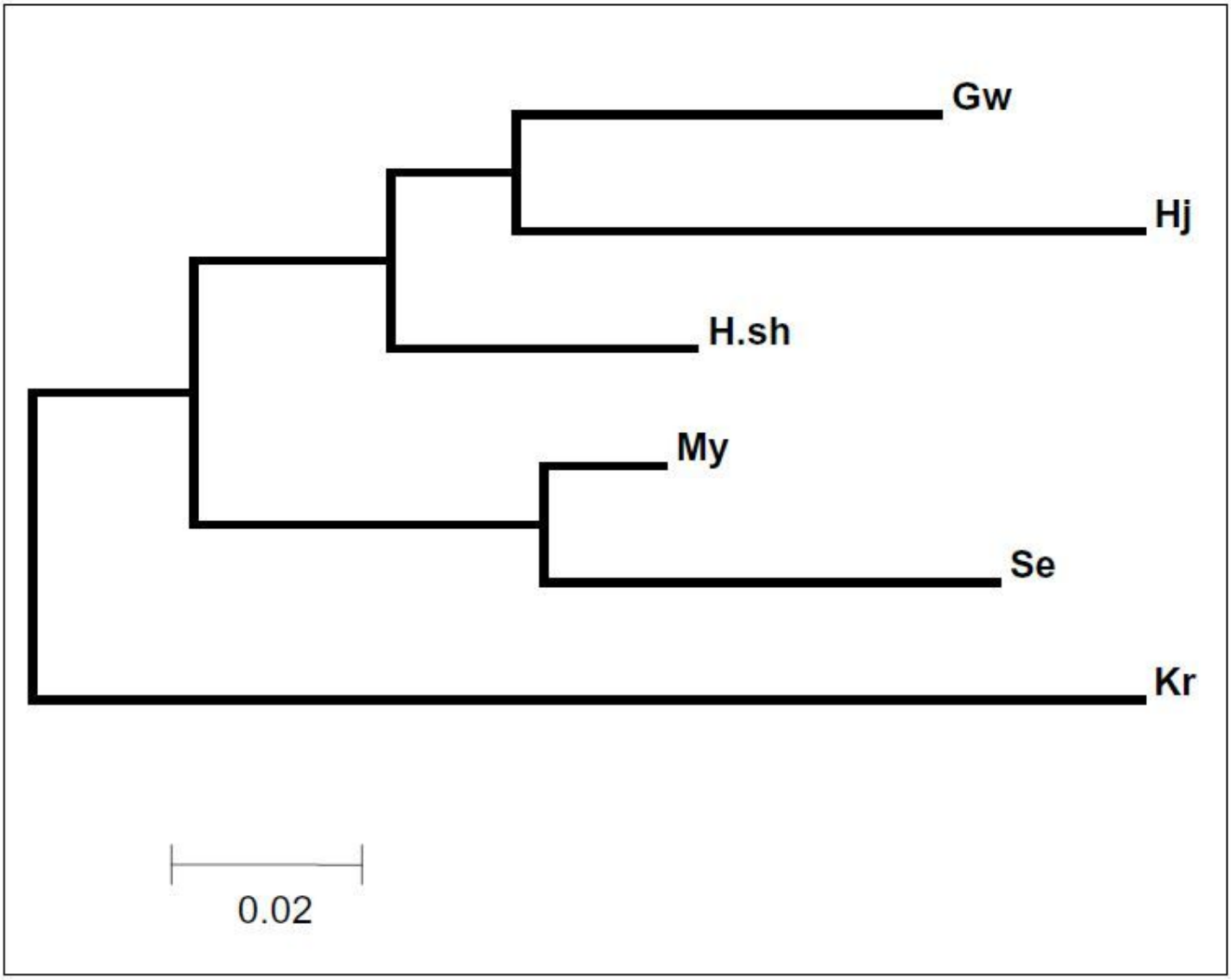

Figure 5

Phylogenetic tree based on Neighbour Joining method The phylogenetic tree of An. arabiensis populations from Sudan. Hj, GW and H.sh, My and Se are highly genetically related, while Kr appeared to be highly differentiated from other populations.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Listoftables.pdf