Population genetic structure of malaria vector *Anopheles stephensi* using mitochondrial Cytochrome oxidase II gene in Indian populations

Arvind Sharma, Arunaditya Deshmukh, Richa Sharma, Ashwani Kumar, Sayantan Mukherjee, G C Chandra, S K Gakhar

Centre for Biotechnology, Maharshi Dayanand University, Rohtak, 123 4001

*Received 1 November 2013; revised 17 June 2014*

The genetic differentiation in *A. stephensi* based on haplotype diversity using Restriction Fragment Length Polymorphism and by sequencing of CO II gene across different localities in India has been analyzed. The presence of only one DraI restriction site in CO II gene conferred to haplotype B indicating that the gene is very much conserved and the gene flow is not affected even by a major geographical distance barrier. The sequencing and analysis of various population parameters revealed seven haplotypes in all populations. The West Bengal population was found to be more genetically diverse than others. The geographic distance between populations was found to be contributing to the genetic differentiation. The sign of demographic expansion were found in three of the five populations. The local geographic barriers were found to be ineffective in prevention of gene flow.

**Keywords:** *Anopheles stephensi*, Cytochrome oxidase, Gene flow, Genetic differentiation, Haplotype diversity

The utility of mitochondrial DNA sequences particularly cytochrome oxidase subunit II (CO II) for inferring taxonomic, population and phylogenetic relationships and for exploring genetic variability is now well established. The reason for the adoption of mt DNA as a marker is because of its presence in almost all species in multiple copies per cell leading to easy amplification, having small genome size (no introns or mobile elements), very economical in function, having rapid evolution, maternal lineage which follows non mendelian inheritance without recombination events and the sequence is haploidic in nature as well. Mitochondrial genes are strongly conserved in animals with very short intergenic regions.

*An. stephensi* is an important vector of malaria in India, where its control has been hampered by insecticidal resistance and has a wide geographic distribution in the Indian sub continent. The population genetic structure of *An. stephensi* has recently been reviewed. However, no population genetic study appears to have been carried out in India on *An. stephensi* barring two studies, i.e. one in Rajasthan and the other on three forms of *An. stephensi* in Haryana. However, the variation between and within three ecological forms of *Anopheles stephensi* (type, intermediate, mysorensis) has been studied in Iran using CO I and CO II mitochondrial genes. The CO II gene has been extensively used for population genetic studies in other *Anophelines* species in different part of world. The populations of *Anopheles minimus* were examined using CO II gene for 23 sites in China, Thailand, and Vietnam. Similarly, the CO II gene was also used to explore the genetic variability and differentiation among nine populations of *An. nilii* in Sub Saharan Africa. Mitochondrial genetic differentiation across populations of the malaria vector *An. Lesteri* has been studied using CO II gene in China. In addition, the phylogenetic relationship of mosquitoes belonging to subfamily *Anophelinae* has been analyzed by using portions of the mitochondrial genes CO I and CO II. The phylogenetic inference in six primary malaria vectors in India was studied by using CO II in addition with other loci.

Realizing the importance of knowledge of population structure and gene flow in adequate implementation of vector control strategies, the genetic differentiation of mitochondrial DNA of the malaria vector *An. stephensi* of five distantly
located populations of India has been studied using mitochondrial CO II gene. In addition, the complete coding sequence of the mitochondrial CO II gene of *Anopheles stephensi* has been isolated and sequenced.

**Materials and Methods**

Mosquito collection and genomic DNA extraction—*Anopheles stephensi* samples were collected from five different localities in India, namely the districts (states); Rohtak (Haryana), Jabalpur (Madhya Pradesh), Kolkata (West Bengal), Hyderabad (Andhra Pradesh, none Telangana) and Trivandrum (Kerala). Specific spatial details of the collection sites are given in Table 1. Collection of mosquito samples was done by using hand-collection suction tube and torch light. To reduce the chance of sampling siblings, only a few individuals per site were selected. Specimens collected were confirmed as *An. stephensi* on morphological basis following the mosquito identification key. Individual mosquitoes were preserved in 70% ethanol and stored at -20 °C. Genomic DNA was extracted from the legs and wings of an individual mosquito as per Yan.

**PCR amplification, sequencing and RFLP analysis**—The CO II region was amplified by PCR using primers Leu- forward primer 5'-TCTAATATGGCAAAGATTAGTGCA-3' and Lys- reverse primer 5'-ACTTGCTTTCAGTTGNATGCA-3' with the following reaction conditions: denaturation at 95 °C (1 min), annealing at 56.6 °C (45 sec) and extension 72 °C (1 min). Each 25 µL PCR reaction mixture contained 250 µM concentration of each deoxyribonucleotide (dNTP), 3 pmol of each primer (forward and reverse), 2.5 mM MgCl2, 1U Taq DNA polymerase (BangloreGenei, India) with 1 X polymerase buffer and 1 µL template DNA (1/100th of the DNA extracted from a single mosquito). The PCR product was run on 1% agarose gel in 1X TAE buffer. PCR fragments were purified using a PCR purification kit (Genei) and cloned into the TOPO TA cloning vector (Invitrogen). Clones were sequenced using M13-forward and M13- reverse universal primers, in an automated fluorescence sequencing system 3730 DNA analyzer (Applied Biosystem). All the PCR products, were also subjected to Restriction Fragment Length Polymorphism (RFLP) analysis using Dra I restriction enzyme to observe any haplotypic diversity.

DNA sequence editing and data mining—DNA sequences obtained were edited by SEQSCAPE software 2.5 and aligned by Clustal W. The homology search was performed in the NCBI database with the BLASTN algorithm so as to confirm that the desired gene fragment has been amplified. The DNA fragment showing homology with the corresponding gene of *An. gambiae* (CO II) and with the *An. stephensi* (CO II) were considered for further analysis. Information on polymorphic sites, haplotype diversity (Hd), nucleotide diversity (π) and average number of nucleotide difference (K) based on the expected number of segregating sites in a population, was extracted using DnaSP 5.6. Statistical tests of Tajima, Fu and Li and Fu also implemented in DNAsp 5.0 were used to test for non-neutral evolution and deviation from MDE. A haplotype network was constructed using the median joining (MJ) algorithm in NETWORK 4.5. The distribution of pairwise nucleotide differences (Mismatch distribution) was calculated as an additional test for demographic expansion, using DNAsp. To test whether the observed distribution deviated significantly from those expected under the population expansion model, sum of square deviations (SSD) and Harpending’s Raggedness index were calculated using Arliquin program. The phylogenetic trees were constructed following four different approaches, i.e. neighbor joining (NJ), maximum parsimony (MP), maximum likelihood (ML), and Bayesian (data not shown) using mitochondrial CO II gene of *Anopheles* mosquitoes. Amongst all, the most appropriate evolutionary model was detected using computer program MODELTEST. Finally, NJ trees were constructed using computer program MEGA version 5.0 (utilizing the maximum composite likelihood model for nucleotide substitution and bootstrapping for testing the phylogeny). All positions containing gaps and missing data were eliminated.

---

Table 1—Showing location of different study sited of *An. stephensi*.

<table>
<thead>
<tr>
<th>S. No</th>
<th>Location</th>
<th>Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Haryana (Rohtak)</td>
<td>28°53’40.64’’N, 76°35’23.93’’E</td>
</tr>
<tr>
<td>2</td>
<td>West Bengal (Kolkata)</td>
<td>22°34’23.51’’N, 88°21’42.86’’E</td>
</tr>
<tr>
<td>3</td>
<td>Madhya Pradesh (Jabalpur)</td>
<td>23°10’04.67’’N, 79°56’45.65’’E</td>
</tr>
<tr>
<td>4</td>
<td>Andhra Pradesh (Hyderabad)</td>
<td>17°23’27.26’’N, 78°29’11.00’’E</td>
</tr>
<tr>
<td>5</td>
<td>Kerala (Trivandrum)</td>
<td>8°29’19.68’’N, 76°56’57.63’’E</td>
</tr>
</tbody>
</table>
Results

**RFLP Analysis**—The CO II gene from the individuals of each population digested with Dra I restriction enzyme yielded two fragments of approximately 341 and 344 bps. The agarose gel could not resolve 2 or 3 bps difference; therefore two juxtaposition bands were seen. The presence of only one DraI restriction site at 341 nucleotide position in CO II gene of *An. stephensi* conferred to haplotype B. All the samples appeared to be haplotype B showing the similar pattern of bands. However, to understand the population structure of the *An. stephensi* in detail, the sequencing was done and variation in COII gene was studied.

**Sequence variation**—The nucleotide sequence variations were determined along 674 bp of the CO II gene in a total of 26 individuals of *An. stephensi* belonging to five different parts of the country. In the total dataset, 58 mutations were found with 48 parsimony information sites. A total of 7 haplotypes were observed in all the populations with haplotype diversity 0.520 ± 0.115 (Table 2). The haplotype H3 was found to be common in all the populations except Haryana. Fifty eight variable sites (8.60%) were identified of which 12 were singleton variable sites (1.78%). The haplotype diversity ranged from 0.286 ± 0.196 (West Bengal) to 0.500 ± 0.265 (Haryana and Hyderabad), whereas, the nucleotide diversity ranged from 0.0006 (Kerala) to 0.0032 (West Bengal). The average number of nucleotide difference was highest for West Bengal samples (2.0000) and ranged from 0.4000 (Kerala) to 2.000 (West Bengal). Overall low polymorphism was found in the CO II gene. Across the whole data set, there were only 8 (2.5%) polymorphic sites across for CO II gene. This low number of variable sites resulted in low nucleotide diversity in almost all samples. The West Bengal population was found to be more genetic diverse than others whereas, Kerala population was found to be least diverse. Tajima’s D, Fu & Li’s F and D were all negative for all the populations reflecting either a selective sweep or a recent demographic expansion.

**Genetic differentiation and analysis of molecular variance (AMOVA)**—The genetic differentiation ($F_{ST}$) showed a large range ($F_{ST}$ = -0.036 to 0.919); however, half of the pairwise comparison were not significant (Table 3). The highest significant genetic differentiation was found between the Haryana and Kerala samples ($F_{ST}$ = 0.919), while the lowest significant differentiation was found between Hyderabad and Kerala ($F_{ST}$ = 0.107). This corresponds to the signal of Isolation by distance. A mental test found positive correlation between the genetic distance and the geographic distance ($r = 0.1758$, $p = 0.322$), indicating the role of geographic distance in genetic differentiation (Fig. 1). The results were further supported by Analysis of Molecular Variance (AMOVA) which revealed 68.32% variations among populations and only 31.68 % variance was found within populations (Table 4).

**Genealogy and phylogenetic relationship**—The neutrality tests were all found to be negative, but non-significant indicating demographic expansion. The haplotype network showed that almost all
An. stephensi haplotype derived from a single common ancestral haplotype (Fig. 2). The overall star like haplotype network suggests recent population expansion or a recent selective sweep\textsuperscript{18}. There are two very high frequencies, widespread haplotypes (H3 and H1) at the center of the network, surrounded by five low frequency haplotypes.

The Bayesian phylogenetic analysis of the CO II produced a maximum clade credibility tree that showed the presence of two distinct monophyletic groups (Fig. 3). The bootstrapping consensus of the tree was inferred from 1000 replication. The first group contains Haryana population and second group contains all other populations.

**Demographic expansion**—Out of the five populations studied, three (Haryana, Hyderabad and Kerala) showed unimodal pattern of mismatch distribution curves as expected in population expansion (Fig. 4).

<p>| Table 3—Genetic differentiation ($F_{ST}$) among An. stephensi populations. |</p>
<table>
<thead>
<tr>
<th>Populations</th>
<th>HR</th>
<th>HYD</th>
<th>MP</th>
<th>WB</th>
<th>KL</th>
</tr>
</thead>
<tbody>
<tr>
<td>HR</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HYD</td>
<td>0.902</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MP</td>
<td>0.910***</td>
<td>0.111***</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WB</td>
<td>0.825***</td>
<td>0.012</td>
<td>0.012</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>KL</td>
<td>0.919**</td>
<td>0.107**</td>
<td>0.041</td>
<td>-0.036</td>
<td>0</td>
</tr>
</tbody>
</table>

Level of significance is indicated by: * = $p<0.05$, ** = $p<0.01$, *** = $p<0.001$

<p>| Table 4—AMOVA of genetic variation in the An. stephensi samples. |</p>
<table>
<thead>
<tr>
<th>Source of variation</th>
<th>d.f.</th>
<th>Sum of squares</th>
<th>Variance components</th>
<th>Percentage of variations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Among populations</td>
<td>4</td>
<td>232.622</td>
<td>10.38809 Va</td>
<td>68.32</td>
</tr>
<tr>
<td>Within populations</td>
<td>21</td>
<td>101.148</td>
<td>4.81655 Vb</td>
<td>31.68</td>
</tr>
<tr>
<td>Total</td>
<td>25</td>
<td>333.769</td>
<td>15.20464</td>
<td>100</td>
</tr>
</tbody>
</table>

Fig. 1—Correlation between the genetic distance and geographic distance in An. stephensi populations.

Fig. 2—Haplotype networks of COII gene in An. stephensi population. Circles are roughly proportional to haplotype frequencies.

Fig. 3—Phylogenetic tree of An. stephensi populations based on Maximum likelihood method according to Tamura-Nei genetic distance method.
In the remaining two pupations (Jabalpur and West Bengal) the multimodal pattern was observed indicating population in equilibrium. None of the sums of squared deviation (SSD) was significant, indicating that the curves fit the sudden expansion model.

**Discussion**

Mitochondrial DNA evolves much faster than the nuclear genome and has been widely used for population genetics and phylogenetics. The mitochondrial cytochrome c oxidase subunit II (CO II) is the most frequently used marker for population studies. This is the first extensive study of *An. stephensi* genetic variation and differentiation from India using cytochrome oxidase II (COII) mitochondrial gene.

During the present study, only one haplotype B was observed across five different localities in India indicating that the COII gene region is very much conserved and the gene flow is not affected even by a major geographical distance barrier. However, earlier two haplotypes A and B were reported in two forms i.e., ‘mysorensis’ and ‘type’ forms respectively due to the presence of two Dra I sites at 51 residues apart in the CO II region of *An. stephensi* whereas the intermediate forms of *An. stephensi* had an only Haplotype B in the Iran population. The presence of first Dra I site is the deciding factor for the two Haplotypes i.e., either A or B, while the second

---

**Fig. 4**—Observed and expected mismatch distribution under population expansion model among the haplotypes of *An. stephensi* populations.
A restriction site was found to be conserved in both Iran and Indian population (present study). In the present study, it was observed that the second restriction site was found at 341th position, while the first Dra I site was absent, which otherwise was supposed to be present at 288th position. The sequencing and analysis of the population parameters revealed a total of seven haplotypes in five populations across India. The West Bengal population from Eastern India was found to be more diverse with seven segregating sites and nucleotide diversity of 0.0032. The least diversity was found in the Kerala population (Nucleotide diversity = 0.0006). This is in accordance of a previous study of An. baimaii from North-East India using COII gene, where North East population were found to be more molecular diverse than the other populations from India. The majority of variations were found among populations. The differentiation studies found significant and high genetic differentiation between half of the population pairs. No geographical barrier, except the Vindhyas mountainrange, which passes through central India separating northern and southern parts of India, exists between the populations. Moreover, large geographic distance exists between the populations and anopheline species in general are known to have limited flight range, thereby, indicating that geographic distance is the main reason of differentiation. This was further confirmed by mantel test which indicated positive correlation between the genetic distance and the geographic distance. The An. stephensi have a flight range of 3-5 km, therefore there dispersal over the long distances could be explained by passive migration of this vector. When a gene flow is inferred indirectly from F statistics, the assumption of equilibrium between drift and migration is often reasonable, since migration is a relatively fast homogenizing force. However, the genetic signature of population expansion can remain for a long time, obscuring true ecological population structure that may exist. In addition, the rate of approach to equilibriums is dependent on effective population size, so it would take an extremely long time for the very large population sizes of An. stephensi to reach mutation drift equilibrium. It would therefore be misleading to infer migration rates form the FST values.

Tajima’s D values are used to test against selective neutrality and population equilibrium. The values of Tajima’s D were negative but not significant, which indicated either population expansion or purifying selection. The background selection effect on genetic variability was confirmed by Fu and Li’s D* and F* statistics, where negative values supported the Tajima’s D results. Fu’s F, is very powerful to detect the increase in effective population size (population expansion) estimating departure from the neutrality. The negative values indicated departure from neutrality. If Fu and Li’s statistics are significant and Fu’s F, and Tajima’s D are not, then background selection is more likely to create the observed pattern. Tajima’s D, Fu and Li’s F and D, and Fu’s F, statistics were all found to be negative, but non-significant in all the population indicating recent demographic expansion. There were considerable deviations from a poisson fit in the mismatch distribution. A unimodal pattern of mismatch distribution curves was found in three populations i.e. Haryana, Hyderabad and Kerala the smooth unimodal distribution are consistent with some sort of population expansion as simulations have shown that stable populations never produce this type of profile. Nor can such a profile be the result of pooling data from structures subpopulation. This population expansion can be visualized by overall star like haplotype network where more frequent haplotypes were in the core surrounded by many low frequent haplotypes. The finding supports the hypothesis that a severe bottleneck effect occurred due to insecticide pressure or to periods of pronounced dry/cold weather. The phylogenetic analysis showed the presence of two clades in the overall samples with samples from Haryana in one clade and all other samples in the other, thereby indicating that Haryana population is more likely geographically isolated form all other populations.

Conclusion

This is the first and successful attempt made to study genetic differentiation based on haplotype diversity of An. stephensi across India using RFLP and mitochondrial gene. The more genetic diversity was found in the Eastern population. The variation was found to be due to vast geographic distance between the populations. The local geographical barriers were found to be ineffective in prevention of gene flow. The signs of recent population expansion were found in three populations which might be due to bottleneck in the past. To distinguish between the hypothesis of population expansion and selective sweep, it is essential to obtain data from same populations on nuclear loci. Understanding the
evolutionary history and genetic structure of these mosquitoes is important if reliable estimates of gene flow are to be made, which is crucial to the malaria control efforts.

Acknowledgement
Financial support received from the Department of Biotechnology, Govt. of India, acknowledged.

References