Sequence analysis of mitochondrial COII gene fragment from five Culicinae mosquito species

Devinder Singh, Navpreet Kaur Gill and Navneet Kaur Rai

Abstract

Sequence analysis of the mitochondrial COII gene has been used for molecular taxonomy in many insects and to reveal phylogenetic relationship among them. A ~500 to 550bp sequence of mitochondrial Cytochrome Oxidase II (COII) gene has been analyzed to construct molecular database and to establish the phylogenetic relation among the five Culicinae mosquito species (Aedes vittatus, Aedes aegypti, Culex bitaeniorhynchus, Culex vishnui and Culex quinquefasciatus) from the state of Punjab and adjoining areas (India). The sequences were found to be A+T rich and in substitution the rate of transitions was higher than the rate of transversions. Individuals of the same species grouped closely together in a Neighbor-Joining tree regardless of collection site. Conspecifics showed <2% divergence (range = 0% to 1.7%), whereas interspecific divergence was >2% with K2P (range = 4.1% to 15.0%).

Keywords: culicinae, identification, phylogenetic analysis, mitochondrial and divergence

1. Introduction

Mosquitoes are important vectors of pathogens that represent a major cause of human mortality and morbidity worldwide more than any other group of organisms. They carry various pathogens which mainly include the malaria parasite (Plasmodium), filarial parasite and arboviruses [1, 2]. In terms of mosquito diversity, India is ranked fifth after Brazil, Indonesia, Malaysia and Thailand [3]. Among the 3540 mosquito species recorded worldwide, 393 have been documented from India [4]. Although mosquitoes have been studied more extensively than most other insect groups because of their role as vectors of diseases, yet our taxonomic knowledge of these insects is far from complete. Therefore, species identification constitutes the first step in the surveillance and control of mosquito-borne diseases [5].

Detailed taxonomic studies have focused on various mosquito vectors [6] while other species have received little attention [7, 8]. Furthermore, closely related species of mosquitoes are nearly indistinguishable morphologically having different ecological niches and host preferences [9]. Such factors indicate that the identification of mosquitoes to a species or sometimes even at genus level is often difficult [10, 11]. As a result, DNA-based approaches for mosquito identification, molecular phylogeny, and genetic diversity have gained increasing acceptance [12, 13, 14, 15, 16, 17].

The use of molecular techniques in combination to morphological methods, has resolved some long-standing taxonomic questions [18]. The increased availability of molecular markers has facilitated the precise identification of mosquitoes, predominantly within sibling species [19, 20]. Mitochondrial DNA (mtDNA) is commonly used for molecular evolution studies in insects [21]. Due to its rapid evolution, maternal and non-recombinant mode of inheritance, mtDNA has been widely used to reconstruct the dispersal history of the insects [22], to investigate the population differentiation and evolutionary history of populations and subspecies [23, 24]. In particular, the mitochondrial cytochrome oxidase II gene (COII) has been successfully used to study the population genetic structure and population history for a wide range of insect species [25, 26, 27, 28].

Mosquitoes, because of their medical importance, are one of the most thoroughly studied groups of insects. Incongruously, despite the significant amount of morphological and molecular work that has been done, little progress has been made towards an understanding of the phylogenetic relationship between species of subfamily Culicinae. In the present study an attempt has been made to generate the genetic profile of five different mosquito species Aedes vittatus, Aedes aegypti, Culex quinquefasciatus, Culex bitaeniorhynchus and Culex vishnui using cytochrome oxidase subunit II to understand phylogenetic relationship among them.
2. Methodology

2.1 Mosquito collection

The adult mosquitoes were collected from different localities of various districts of Punjab and adjoining states. The mosquitoes were collected from outdoor shelters like gardens, nurseries and wild vegetation during morning and evening hours. Relevant collection details were recorded in the field for each collected specimen. Adults were killed with the help of ethyl acetate vapors and then mounted on a wedge of thick paper supported by entomological pin for identification purpose. The identification of mosquitoes was done by using standard taxonomic keys [29].

2.2 DNA isolation, amplification and sequencing

The mosquitoes were preserved in 100% Ethanol at -20°C for further studies. Insect legs were used for the DNA extraction process using liquid nitrogen for better PCR result. A set protocol was followed along with QIAgen and HiMedia extraction kits to ensure accuracy of the procedure. The mitochondrial COII gene fragment was amplified using forward primer C2-J-3138 (5'AGAGCTTCTCCTTTAATGGAAACA3') and reverse primer C2-N-3686 (3'CAATTGGTATAAACATGTGTTTG5') [30]. The thermal cycler conditions included an initial denaturation at 98°C for 2 minutes followed by 38 cycles at 98 °C for 30 seconds, annealing at 49°C for 40 seconds, elongation at 72°C for 1 minute and final elongation at 72°C for 7 minutes. 10μL PCR cocktail constituted of Phusion DNA polymerase enzyme 0.1U/10μL, 1.2μL 5X Buffer, 1.2μL dNTP, 50Mm MgCl₂ 0.2μL, 0.25μL primer and Nuclease free water (Thermo Fisher Scientific, India). The amplified samples were sent to SciGenom Labs, Cochin (Kerala) for DNA sequencing. The sequences were then submitted in the GenBank and corresponding accession numbers were obtained for all these sequences. The sequences retrieved for comparative study are provided with their accession number and location of the submission (Table 1).

Table 1: Detail of mosquito specimens used for molecular identification and sequence analysis.

<table>
<thead>
<tr>
<th>S. No</th>
<th>Species</th>
<th>Accession No.</th>
<th>Locality</th>
<th>Habitat</th>
<th>Sequences retrieved from Genbank</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Aedes vittatus</td>
<td>KX085420</td>
<td>Kathua (Jammu)</td>
<td>Human dwellings</td>
<td>Aedes vittatus Accession no. KC913572</td>
</tr>
<tr>
<td>2</td>
<td>Aedes aegypti</td>
<td>KX495183</td>
<td>Ludhiana (Punjab)</td>
<td>Tyre</td>
<td>Aedes aegypti Accession no. KC913582</td>
</tr>
<tr>
<td>3</td>
<td>Culex bitaeniorhynchus</td>
<td>KX524976</td>
<td>Tarn taran (Punjab)</td>
<td>Cattle sheds</td>
<td>Culex bitaeniorhynchus Accession no. KP687363</td>
</tr>
<tr>
<td>4</td>
<td>Culex vishnui</td>
<td>KX691870</td>
<td>Bahinda (Punjab)</td>
<td>Cattle sheds</td>
<td>Culex vishnui Accession no. KP04957</td>
</tr>
<tr>
<td>5</td>
<td>Culex quinquefasciatus</td>
<td>KX495184</td>
<td>Ludhiana (Punjab)</td>
<td>Human dwellings</td>
<td>Culex quinquefasciatus Accession no. KP687387</td>
</tr>
</tbody>
</table>

3. Results

3.1 Sequence analysis

Sampling stratagem and geographic region significantly influence the analysis and interpretation of data generated from the samples. PCR amplification of DNA of mosquito species with COII primers yielded a fragment of ~500 to 550 bp. Both nucleotide sequence and the percentage of particular nucleotide have been evaluated as these parameters are important for studying the variation among different species. The average percentage of each nucleotide for the studied fragment of COII gene has been observed to be T=39.6%, C=14.5%, A=34.6% and G=11.3%. Various other studies also reported higher A+T frequency to be a characteristic of insect mtDNA [31, 32, 33, 34, 21]. The A and T content at the first codon position are as high as 45% and 48% respectively. Because of codon preference the base composition of A+T across the whole dataset was biased at third codon position and totaled as 67.7%. The G and C content at the first codon position was found less than 1% in Aedes aegypti & Culex bitaeniorhynchus respectively. This base composition of mtDNA is highly correlated with codon usage, because mitochondrial protein genes exhibit a preference for using A+T rich codons [35]. The phenomenon seems characteristic of the insects while A+T content is found significantly low in crustaceans [36]. The G+C-rich codons include the codons encoding for Pro (P), Ala (A), Arg (R), and Gly (G), and the A+T-rich ones contain codons encoding for Phe (F), Ile (I), Met (M), Asn (N), and Tyr (Y) (Fig. 1). Across the whole dataset of 350bp, the overall number of conserved sites was 280 that clearly indicate that COII gene is highly conserved. The number of variable sites and parsimoniously informative site was found to be 70 and 64 respectively.

![Fig 1: showing relative content of various amino acids.](image-url)

"1387"
3.2 Species boundaries

Conspecific sequences showed <2% divergence range i.e. 0% to 1.7% whereas interspecific divergence was >2% with K2P divergence varying from 4.1% to 15.0% (Table 2). The interspecific distance between *Aedes vittatus* and *Aedes aegypti* was found to be 13.4%. Genetic distances among the species belonging to the genus *Culex* varied from 4.0% to 8.0%, while the intergeneric distance between species of *Culex* and *Aedes* varied from 9.0% to 15.0%, thereby indicating a hierarchical increase in K2P mean divergence across different taxonomic levels.

<table>
<thead>
<tr>
<th>S. No</th>
<th>Species</th>
<th>Intraspecific Distance</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td><em>Aedes vittatus</em></td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td><em>Aedes aegypti</em></td>
<td>1.7</td>
<td>13.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td><em>Culex quinquefaciatus</em></td>
<td>Nil</td>
<td>10.7</td>
<td>15.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td><em>Culex bitaeniorhynchus</em></td>
<td>0.9</td>
<td>9.0</td>
<td>13.6</td>
<td>6.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td><em>Culex vishnui</em></td>
<td>Nil</td>
<td>11.3</td>
<td>14.3</td>
<td>7.9</td>
<td>4.1</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2: Pairwise intraspecific and interspecific divergence (%).

Estimation of the Transition/Transversion rate bias is important not only to our understanding of the patterns of DNA sequence evolution, but also to reliable estimation of sequence distance and phylogeny reconstruction. The estimated transition to transversion ratio (R) under the K2P model found was 0.94. The rate of transitions was higher than the rate of transversions and T→C transitions were most frequent. The number of transitional substitution was found highest among *Culex quinquefaciatus* and *Aedes aegypti* while the lowest were found among *Culex bitaeniorhynchus* and *Aedes vittatus*. Nucleotide sequences of the COII fragment of these species were aligned by Clustal W in Mega 6.0. *Drosophila* sequence, retrieved from GenBank, was used as outgroup. Combining NJ tree and bootstrap analysis is the most appropriate method for evaluating phylogenetic trees using distance methods [37]. Nodes linking sequences of individuals of the same species had a high bootstrap value (90%–100%) whereas the species belonging to same or different genera had a bootstrap value of 55-91% (Fig. 2).

4. Discussion

Various studies have been carried out in mosquito species using sequences of nuclear rDNA genes, internal transcribed spacers of rDNA, mitochondrial DNA and various random amplified polymorphic DNA (RAPD), and RFLP markers [38, 39, 40, 22, 41, 42, 43]. Such studies have explicitly proved the importance of mitochondrial DNA based methods. The present study has been focused upon most conserved mitochondrial COII gene to analyze the relationships among the five species of the subfamily Culicinae. Smith et al. (1996) [44] used the entire COII sequence to infer the phylogeny of Calliphoridae within the Diptera. The mtDNA COII gene proved as a good phylogenetic marker for four species of the Annulipes Complex [45]. Therefore, COII is considered as a slow evolving protein coding region, thus its use can be limited in the higher level phylogenies. It was the first ever attempt on the sequence analysis of *Aedes vittatus* from India using the mitochondrial COII gene. The results obtained during the study reveal a pattern of low intraspecific and high interspecific variations. Maximum range of interspecific distance was observed between *Aedes vittatus* and *Aedes aegypti* whereas minimum range was found between *Culex bitaeniorhynchus* and *Culex vishnui*. The intraspecific variations were found highest among *Aedes aegypti* whereas it was observed zero in both *Culex vishnui* and *Culex quinquefaciatus* despite of different geographic locations. A similar range of intraspecific and interspecific distance range from 0.0% - 2.4% and 2.3% - 17.8% respectively was reported in 32 species of mosquitoes from Pakistan [46]. Intraspecific (0- 1.67%) and interspecific variations (2.3% - 21.8%) were also observed within 122 species from China in comparable range and transition distance was found to be significantly greater than the transversion distance when intraspecific sequence divergence occurs.
was <2% [4]. The phylogenetic analysis revealed paraphyly of Aedes with respect to Culex and as expected Drosophila medioiustrata branched out as outgroup. Three species belonging to genus Culex i.e. Culex bitancerohynchus, Culex vishnui and Culex quinquefasciatus formed a monophyletic clade and Culex bitancerohynchus was found closer to its sister species i.e. Culex vishnui whereas Aedes aegypti and Aedes vittatus formed sister clades.

In conclusion, the molecular phylogeny obtained in this work matches the classical morphological taxonomy reasonably well and confirms the robust nature of COII gene in distinguishing between these species. Thus, selection of a particular gene that evolve at the rate reflect the species differentiation level and helps to construct a proper phylogenetic relationship among the Culicines. The analysis will be valuable in studies involving molecular taxonomy of mosquito vectors. Sequence comparisons of the different geographic populations will give estimates of their genetic relatedness and provide some information on vector movement. Further studies propose an effort to construct database of Culicinae mosquitoes based on COII gene due to the absence of reference database for DNA based identification of this extensive subfamily in India that will provide priority in molecular identification of mosquito.

5. References


90:635-642.
35.Crozier RH, Crozier YC. The mitochondrial genome of the honeybee Apis mellifera: complete sequence and genome organization. Genetics. 1993; 133:97-117.
42.Porter CH, Collins FH. Species diagnostic difference in a