CHAPTER – III

BIOSYSTEMATICS OF CHOSEN REDUVIINE SPECIES BASED ON MITOCHONDRIAL CYTOCHROME C OXIDASE SUBUNIT I GENE SEQUENCES

Introduction

The genetic diversity of life underpins all biological studies, but it is also a harsh reality. In fact, since few taxonomists can critically identify more than 0.01% of the estimated 10-15 million species (Hammond, 1992; Hawksworth and Kalin-Arroyo, 1995), a community of 15000 taxonomists will be required, in perpetuity, to identify life if our reliance on morphological diagnosis is to be sustained.

Moreover, this approach to the task of routine species identification has significant limitations. Both phenotypic plasticity and genetic variability in the characters employed for species recognition can lead to incorrect identifications. This approach overlooks morphologically cryptic taxa, which are common in many groups (Knowlton, 1993; Jarman and Elliott, 2000). Since morphological keys are often effective only for a particular life stage or gender, many individuals cannot be identified. Finally, although modern interactive versions represent a major advance, the use of keys often demands such high levels of expertise that wrong diagnoses are common.

The limitations inherent in morphology-based identification systems and the dwindling pool of taxonomists signal the need for a new approach to taxon recognition. Microgenomic identification systems, which permit life’s discrimination through the analysis of a small segment of the genome, represent one extremely promising approach to the diagnosis of biological diversity. This concept has already gained broad acceptance among those working with the least morphologically tractable groups, such as viruses and bacteria (Nanney, 1982; Pace, 1997; Allander et al., 2001; Hamels et al., 2001). However, the
problems inherent in morphological taxonomy are general enough to merit the extension of this approach to all life. In fact, there are a growing number of cases in which DNA-based identification systems have been applied to higher organisms (Brown et al., 1999; Bucklin et al., 1999; Trewick, 2000; Vincent et al., 2000).

Mitochondria

The mitochondrion (plural mitochondria) is a membrane-bound organelle found in most eukaryotic cells (the cells that make up plants, animals, fungi, and many other forms of life (Henze, 2003). Mitochondria range from 0.5 to 1.0 micrometer (μm) in diameter. These structures are sometimes described as "cellular power plants" because they generate most of the cell's supply of adenosine triphosphate (ATP), used as a source of chemical energy (Campbell, 2006). In addition to supplying cellular energy, mitochondria are involved in other tasks such as signaling, cellular differentiation, cell death, as well as the control of the cell cycle and cell growth (McBride, 2006)

Mitochondria are known as energy suppliers for the cell, as it is in the mitochondria that the transfer of electrons down an electrochemical gradient to the final acceptance of oxygen occurs with the subsequent production of energy. The cytochrome, large oligometric protein, is embedded in the inner lipid bilayer membrane of the mitochondrion and acts as a carrier of the electrons (Keeton and Gould, 1985).

Mitochondrial DNA

Mitochondrial DNA (mtDNA or mDNA) is the DNA located in organelles called mitochondria, structures within eukaryotic cells that convert chemical energy from food into a form that cells can use, adenosine triphosphate (ATP). Mitochondrial DNA is only a small portion of the DNA in a eukaryotic cell; most of the DNA can be found in the cell nucleus, and in plants, the chloroplast as well (Sykes, 2003).
Nuclear and mitochondrial DNA are thought to be of separate evolutionary origin, with the mtDNA being derived from the circular genomes of the bacteria that were engulfed by the early ancestors of today's eukaryotic cells. This theory is called the endosymbiotic theory. Each mitochondrion is estimated to contain 2-10 mtDNA copies (Wiesner, 1992).

Due to its ease of enzymatic amplification with conserved primers, mtDNA is frequently used to infer genetic and biogeography subdivision within species (Kaplan et al., 1989). Mitochondrial genes are viewed as advantageous for phylogenetic analysis for several reasons. The mitochondrial genes are generally easier to amplify than nuclear genes (Simon et al., 1994), clonally inherited and non-recombining (Sunnucks and Hales, 1996; Zhang and Hewitt, 1996). It is also estimated to evolve 2–9 times faster than nuclear protein-coding genes (Monteiro and Pierce, 2001). Mitochondrial genes have been used as the source of data for studies of insect molecular phylogeny, genetic variation and phylogeography (Avise, 2000; Caterino et al., 2000; Simmons and Weller, 2001).

Among the mitochondrial genes the COI gene possesses additional interesting characteristics which make it particularly suitable as a molecular marker for evolutionary and genetic studies. A combination of highly conserved and variable region is so closely associated in mitochondrial genes (Gennis, 1992). The COI gene is the largest of the three mitochondria-encoded cytochrome oxidase subunits and one of the largest protein-coding genes in the mitochondrial genome (Clary and Wolstenholme, 1985; Morlais and Severson, 2002).

**COI, COII, COIII genes**

Cytochrome c oxidase I (COX1) is a large transmembrane protein complex found in bacteria and the mitochondrion of eukaryotes (Kosakyan, et al., 2012) and it is the main subunit of the cytochrome c oxidase complex (Rumbley, 1994).

Cytochrome c oxidase subunit II, also known as cytochrome c oxidase polypeptide II, is a protein that in humans is encoded by the MT-CO2 gene which is abbreviated as
COXII, COX2, COII, or MT-CO2, is the second subunit of cytochrome c oxidase (Capaldi, *et al.*, 1983). Cytochrome c oxidase subunit II is an oligomeric enzymatic complex which is a component of the respiratory chain and is involved in the transfer of electrons from cytochrome c to oxygen. In eukaryotes this enzyme complex is located in the mitochondrial inner membrane; in aerobic prokaryotes it is found in the plasma membrane. The enzyme complex consists of 3-4 subunits in prokaryotes to up to 13 polypeptides in case of mammals.

In Leigh's disease, there may be an abnormality or deficiency of cytochrome oxidase (Rumbley *et al.*, 1994).

Cytochrome c oxidase subunit III is one of the main transmembrane subunits of cytochrome c oxidase which is an oligomeric enzymatic complex that is located in the mitochondrial inner membrane of eukaryotes and in the plasma membrane of aerobic prokaryotes. The core structure of prokaryotic and eukaryotic cytochrome c oxidase contains three common subunits, I, II and III. In prokaryotes, subunits I and III can be fused and a fourth subunit is sometimes found, whereas in eukaryotes there are a variable number of additional small subunits (Mather, *et al.*, 1993).

**Cytochrome Oxidase Subunit I:**

The COI gene is one of the best-known mitochondrial genes. It contains both highly conserved and variable regions. The sequence of this gene is potentially useful for phylogenetic analysis over a wide taxonomic range and has been explored for that purpose. This COI gene sequence has been generally useful for reconstructing phylogenetic relationships among more closely related groups (Willis *et al.*, 1992; Beckenbach *et al.*, 1993; Brown *et al.*, 1994; Sperling and Hickey, 1994; Emerson and Wallis, 1995; Spicer, 1995; Smith and Bush, 1997). When compared to other cytochrome subunits, COI has a relatively large size and can therefore present the researcher with both highly conserved and
variable regions making it an especially valuable tool in molecular genetic studies (Morlais and Severson, 2002).

COI gene sequences have been used to address phylogenetic problems at a wide range of hierarchical levels, from species to orders (Caterino and Sperling, 1999). Furthermore, Gaunt and Miles (2002) found that the COI gene is better suited to conduct studies based on the molecular genetics assumption compared to other gene sequences, such as 16S rRNA and 18S rRNA. The mtDNA is the most commonly sequenced region in studies involving insect genetic variation and COI is one of the most frequently sequenced segments within the mtDNA.

Applications: Although COI gene may be matched by other mitochondrial genes in resolving such cases of recent genetic divergence, this gene is more likely to provide deeper phylogenetic insights than alternatives such as Cyt-b (Simmons and Weller, 2001), because changes in its amino acid sequence occur more slowly than those in this, or any other, mitochondrial gene (Lynch and Jarrell, 1993). Nonetheless, COI gene has proved to be a versatile tool with a variety of applications, for example, by facilitating the association between different developmental stages in insects (Ahrens et al., 2007). The approach has also proved to be an effective auxiliary tool in forensic science (Dawnay et al., 2007), on feeding ecology (Bourlat et al., 2008; Garros et al., 2008; Kuusk and Agusti, 2008) and habitat conservation initiative (Neigel et al., 2007; Ward et al., 2008) among other applications. Most importantly, the COI gene has proved to be especially useful in the study of taxonomically challenging taxa, where morphology-based identifications are frustrated due to genetic diversity (Hebert et al., 2004b; Smith et al., 2006; Quicke et al., 2006; Witt et al., 2006; Yassin et al., 2008).

Merits: The 13 protein-coding genes in the animal mitochondrial genome are better targets because indels are rare since most lead to a shift in the reading frame. There is no prior
reason to focus analysis on a specific gene, but the COI does have two important advantages. The universal primers for this gene are very robust, enabling recovery of its 5’ end from representative of most, if not all, animal phyla (Folmer et al., 1994; Zhang and Hewitt, 1997). The COI gene appears to possess a greater range of phylogenetic signal than any other mitochondrial gene.

In fact, the genetic variation of this gene is rapid enough to allow the discrimination of not only closely allied species, but also phylogenetic groups within a single species (Cox and Hebert, 2001; Wares and Cunningham, 2001).

The mitochondrial COI gene is the most popular marker in molecular genetics-partitions of this gene are frequently employed for higher arthropods phylogeny. In addition, COI is currently in the focus of special interest, its 5’ partition is used for the molecular genetics initiative (Hebert et al., 2003b; Stoeckle, 2003). The nucleotide sequence of this particular 640 nucleotide region (Folmer et al., 1994) shall be the unique identification code for all species to facilitate the correct determination of species and the discovery of new species (Moritz and Cicero, 2004). The DNA sequences of the mitochondrial COI gene can serve as a molecular marker for identifying all kinds of animals (Hebert et al., 2004a), especially cryptic species in tropical regions (Wilcox et al., 1997; Berkov, 2002; Monaghan et al., 2005; Hajibabaei et al., 2006).

**Demerits:** While some ‘‘COI-like’’ sequences may represent simple errors in manual editing and lack of quality control, others may actually be nuclear copies of mitochondrial derivedgenes that crossed into the nuclear genome and became non-functional and therefore, non-coding (Lopez et al., 1994). The rate of evolution of cox1 is very slow (Hebert et al., 2003a).

The aim of the present study is thus, to compare the nucleotide sequences from the mitochondrial COI gene of six species of Reduviinae belonging to four genera viz.,
Acanthaspis pedestris Stål, Acanthaspis quinquespinosa (Fabricius), Acanthaspis siva Distant, Empyrocoris annulata (Distant), Edocla slateri Distant and Velitra sinensis (Walker) encompassing an attempt to analyze the generic as well as inter- and intraspecific genetic variations to understand the phylogeny of the sub family Reduviinae.

**Materials and Methods**

**DNA isolation**

The adults of above said six species were morphologically identified. The body tissues (100mg) from each species were selected as the source of genomic DNA, stored in 90% ethyl alcohol and washed in ddH$_2$O and 750µl of 1X suspension buffer was added to the sample such that the final volume did not exceed 750µl. Then 5µl RNAse was added and mixed for 5-6 times. It was kept at 65°C for 10 min and the mixture was separated into two tubes. Thereafter, 0.5µl of lysis buffer was added to each tube and mixed for 5-6 times. It was kept at 65°C for 15min and cooled to room temperature. Centrifugation was carried out at 10,000 rpm for 1min at room temperature. The supernatant was collected and equal volume of isopropanol was added and mixed well. Centrifugation was carried out at 10,000 rpm for 15 min at room temperature and the supernatant was discarded. One ml of 70% ethanol was added to the pellets and again centrifuged at 10,000 rpm for 15minutes and the supernatant was discarded. Ethanol wash was repeated. Pellets were dried at 57°C for 10min and 20µl of glass-distilled water was added and kept at 65°C for 15 minutes. The DNA sample thus obtained was preserved and used for further analysis.

**PCR amplification**

PCR was carried out to amplify the partial mitochondrial COI gene. It was amplified using COI gene forward primer **LCOF:** 5’ GGT CAA CAA ATC ATA AAG ATA TTG G 3’ and reverse primer **HCOR:** 5’ TAA ACT TCA GGG TGA CCA AAA AAT CA 3’. The primer combination yielded a fragment size of ~700 bp, in all the species. The PCR was
performed on a BioRAD Thermal Cycler PTC-100 in a reaction mixture containing genomic DNA: ~20ng, dNTP mix (2.5mM each): 1µl, forward primer: 100ng, reverse primer: 100ng. 10XAssay buffer for Taq DNA polymerase: 1X, Taq DNA polymerase enzyme: 3units. The final volume of 50µl was made using glass distilled water. Thereafter, amplification at 94°C for 30 seconds, annealing at 55°C for 30 seconds, extension at 72°C for 40 seconds and final extension at 72°C for 10min were performed.

**Gel electrophoresis and PCR product purification**

The amplified PCR products were separated on 1.8% agarose gel along with 100 bp ladder and stained using ethidium bromide (0.5 mg/ml) and visualized under ultraviolet light. The PCR products were column purified by GeneiPure™ gel extraction kit.

**Gene sequence analysis**

The purified amplicons were sequenced using Ampli Taq Fs dye terminator cycle sequencing reaction kit (Applied Biosystems Inc., USA) in the ABI prism 3100 genetic analyzer. Each run is termed as a ‘Single Pass Analysis’ where the electropherogram represents a multicolour picture of sequence showing coloured peaks that indicate the corresponding bases. The ABI Genetic analyser uses the sequencing analysis software V.5.1 with kilobasecaller which displays the quality values for pure and mixed bases. The partial nucleotide sequence of partial COI gene obtained for each of the species was confirmed for its identity by Blast analysis at NCBI ([http://www.ncbi.nlm.nih.gov/blast/](http://www.ncbi.nlm.nih.gov/blast/)). The representative sequences were submitted to NCBI GenBank database and accession numbers are given in table 8.

**Gene sequence data analysis**

The DNA sequences of the six species were aligned by ClustalW program and imported into MEGA version 5.1 (Tamura *et al.*, 2011) (Figure 10) for phylogenetic reconstruction. The A+T and G+C composition of the nucleotide sequences of six species
and the codon frequencies were calculated by MEGA 5.1 (Tamura et al., 2011). These sequences were translated by ExPASy server http://www.expasy.org/translate (Gasteiger et al., 2003) and amino acid compositions were analyzed by statistical analysis of protein sequence server http://www.ebi.ac.uk/tools/saps (Brendel et al., 1992) and the structure of transmembrane protein segments was predicted using Hidden Markov Models by server http://www.cbs.dtu.dk/services/TMHMM (Krogh et al., 2001). The protein secondary structure was predicted by PSIPRED server http://www.bioinf.cs.ucl.ac.uk/psipred (Jones, 1999). The homology modeling was built using SWISS-MODEL workspace server http://www.swissmodel.expasy.org/workspace (Arnold et al., 2006). This model was validated with RAMPAGE server www.mordred.bioc.cam.ac.uk/~rapper/rampage.php (Lovell et al., 2003). The dihedral angles phi against psi of amino acid residues in protein structures were visualized with the help of Ramachandran Plot. It shows the possible conformation of phi and psi angles for a polypeptide chain (Lovell et al., 2003). The SWISS-MODEL Server produces theoretical models for proteins. The possible 3D structure was generated from theoretical model protein using Deep View-Swiss Pdb Viewer ver.4.0.3 software http://www.expasy.org.org/spdbv (Guex and Peitsch, 1997).

**Homology modeling**

To determine the quaternary structure of proteins present in six species homology modeling was done using SWISS-MODEL workspace server http://www.swissmodel.expasy.org/workspace. The model was developed based on experimentally determined structures of related family members as templates. The template structure was selected based on Blast E-value limit. A single chain model of quaternary structure was built based on the selected template, 1v54N (1.80 A) (Guex and Peitsch, 1997; Schwede et al., 2003; Arnold et al., 2006). The evaluation of the quality of structure was carried out by calculating QMEAN 4 score (Benkert, 2011). (The QMEAN4 score is a
composite score consisting of a linear combination of four statistical potential terms that estimates model reliability between 0-1). This score helps to determine the quality of the model obtained.

**Phylogenetic analysis**

The analysis involved nucleotide sequences of six species. Codon positions included were 1st+2nd+3rd+Noncoding. All positions containing gaps and missing data were eliminated. Evolutionary analyses were conducted in MEGA ver.5.1 (Tamura *et al.*, 2011) in three different methods namely, Maximum Likelihood, Maximum Parsimony and Neighbour-Joining Distance methods as described below.

**Maximum Likelihood method**

The evolutionary history was inferred from the phylogenetic tree constructed using the Maximum Likelihood method, based on the Tamura 3-parameter model (Tamura, 1992). The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches (Felsenstein, 1985). Initial tree(s) for the heuristic search were obtained automatically as follows. When the number of common sites was <100 or less than one fourth of the total number of sites, the maximum parsimony method is used. A discrete Gamma distribution was used to model evolutionary rate differences among sites (16 categories (+G, parameter = 1.3705). The rate variation model allowed for some sites is evolutionarily invariable ([+I], 40.1177% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site.

**Maximum Parsimony method**

The evolutionary history was inferred from the phylogenetic tree constructed using the Maximum Parsimony method. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches (Felsenstein, 1985). The MP tree was obtained using the Close-Neighbor-Interchange
algorithm (Nei and Kumar, 2000) with search level 3 in which the initial trees were obtained with the random addition of sequences (1000 replicates). Branch lengths were calculated using the average pathway method (Nei and Kumar, 2000) and are in the units of the number of changes over the whole sequence.

**Neighbor-Joining method**

The evolutionary history was inferred from the phylogenetic tree constructed using the Neighbor-Joining method (Saitou and Nei, 1987). The optimal tree with the sum of branch length=0.75503880 is shown. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1000 replicates) is shown next to the branches (Felsenstein, 1985). The tree is drawn to scale, with branch lengths (next to the branches) in the same units as those of the evolutionary distances used to infer the phylogenetic relationship. The evolutionary distances were computed using the Tamura 3-parameter method (Tamura, 1992) and are in the units of the number of base substitutions per site. The rate variation among sites was modeled with a gamma distribution (shape parameter = 1000).

**Results**

**Basic Sequence Statistics**

In the study of molecular evolution, it is necessary to know some basic statistical quantities, such as nucleotide composition (Table 9), codon frequencies (Table 10) and transition/transversion ratios (table 11) (Software MEGA ver 5.1) (Tamura et al., 2011).

**Nucleotide composition**

In *A. siva*, the nucleotide bases were found AT rich (70.0%) followed by *E. annulata* (69.4%), *E. slateri* (65.5%), *V. sinensis* (64%), *A. pedestris* (61.9%) and *A. quinquespinosa* (59.5). The average AT composition of nucleotide bases of the six reduviids was 65.1%. *A. quinquespinosa* was found GC rich (40.5%), followed by *A. pedestris* (38%), *V. sinensis*
(35.9%), *E. slatari* (34.6%), *E. annulata* (30.7%) and *A. siva* (29.9%). The average GC composition of six reduviids was 34.9%.

The nucleotide composition of first codon of *A. siva* had the highest percentage of AT (70.9%) followed by *E. annulata* (70.5%), *E. slatari* (67.8%), *A. pedestris* (64.3%), *V. sinensis* (64.1%) and *A. quinquespinosa* (63.6%). The average AT composition of first codon of six reduviids was 66.9%. The first codon nucleotide composition of *A. quinquespinosa* was found GC rich (36%) followed by *A. pedestris* (35.6%), *V. sinensis* (35.4%), *E. slatari* (32.2%), *E. annulata* (29.7%) and *A. siva* (29.4%). The average GC composition of six reduviids was 33.1%.

The nucleotide composition of second codon of *A. siva* was found AT rich (65.8%) followed by *E. annulata* (63.6%), *V. sinensis* (62.2%), *E. slatari* (61.1%), *A. pedestris* (57.1%) and *A. quinquespinosa* (55.3%). The average AT composition of six reduviids was 60.85%. The second codon of *A. quinquespinosa* was found GC rich (45.1%), followed by *A. pedestris* (42.8%), *E. slatari* (39%), *V. sinensis* (38.3%), *E. annulata* (36.3%) and *A. siva* (33.8%). The average GC composition of six reduviids was 39.2%.

The nucleotide composition of third codon of *E. annulata* (74.5%) followed by *A. siva* (73.2%), *E. slatari* (67.5%), *V. sinensis* (65.8%), *A. pedestris* (64.8%) and *A. quinquespinosa* (59.8%) was found AT rich. The average AT composition of third codon of six reduviids was 67.6%. The third codon of *A. quinquespinosa* was found GC rich (40.6%), followed by *A. pedestris* (35.7%), *V. sinensis* (34.2%), *E. slatari* (32.5%), *A. siva* (26.6%) and *E. annulata* (26%). The average GC composition of third codon of six reduviids was 32.6%. (Table 9)

**Codon frequencies**

Many amino acids are coded by more than one codon, thus multiple codons for a given amino acid are synonymous. However, many genes display a non-random usage of synonymous codons for specific amino acids. A measure of the extent of this non-
randomness is given by the Relative Synonymous Codon Usage (RSCU) (Sharp et al., 1986). The codon frequencies of six reduviids are given in table 4. The frequently used codons were AUU (14.5), UUU (11.7), AAU (10.5) and AUA (9.2) (table 10).

**Transition/Transversion ratios (R=si/sv)**

This is the ratio of the number of transitions to the number of transversions for a pair of sequences. Transition/Transversion ratio test of six reduviids is given in table 11.

The average transition/transversion ratio of six reduviids studied was 0.79. The average identical pair of six reduviids was 408, transitional pairs was 102 and transversional pairs was 130.

The first codon transition/transversion ratio of six reduviids analysed was 0.68. The average identical pairs of first codon nucleotides was 136, transitional pairs was 31 and transversional pairs was 46.

The second codon transition/transversion ratio of six reduviids was 0.93. The average identical pairs of second codon nucleotides was 141, transitional pairs was 35 and transversional pairs was 37.

The third codon transition/transversion ratio of six reduviids was 0.78. The average identical pairs of third codon nucleotides was 131, transitional pairs was 36 and transversional pairs was 46.

**Genetic distance estimation**

**Estimation of pairwise distance**

The evolutionary divergence between nucleotide sequences of six reduviids was estimated by MEGA 5.1 (Tamura et al., 2011) (Table 13). The longest evolutionary distance was observed between *A. quinquespinosa* and *V. sinensis* (1.182) followed by *E. slateri* and *V. sinensis* (1.180), *A. siva* and *V. sinensis* (1.165), *E. annulata* and *V. sinensis* (1.079) and *A. pedestris* and *V. sinensis* (1.025) and the shortest between *E. slateri* and *E. annulata* (0.088).
followed by *A. quinquespinosa* and *A. pedestris* (0.211), *A. siva* and *E. annulata* (0.269), *A. pedestris* and *E. annulata* (0.306), *A. siva* and *E. slateri* (0.334), *A. pedestris* and *E. slateri* (0.363), *A. quinquespinosa* and *E. slateri* (0.393), *A. siva* and *A. pedestris* (0.543) and *A. siva* and *A. quinquespinosa* (0.566).

**Models for estimating distances**

Evolutionary distances are fundamental for the study of molecular evolution and are useful for phylogenetic reconstruction and the estimation of divergence times (Nei and Kumar, 2000). The evolutionary distance between a pair of sequences is usually measured by the number of nucleotide substitutions occurring between them. Models with the lowest Bayesian Information Criterion (BIC) scores are considered to describe the substitution pattern, the best. The Tamura 3-parameter is the best-fit substitution model for estimating maximum likelihood distance and the estimated distances are given in table 12. The lowest BIC scores were observed in Tamura 3-parameter model (6313.7) followed by Tamura-Nei model (6316.4), General Time Reversible model (6318.3) and Kasegawa-Kishino-Yano model (6318.6).

**Substitution Pattern Disparity**

**(i) Homogeneity test for substitution pattern**

Disparity index measures the observed difference in substitution patterns for a pair of sequences. It works by comparing the nucleotide frequencies in a given pair of sequences and using the number of observed differences between sequences. The disparity index test showed homogeneity between *A. siva* and *E. annulata* (0±0.824) followed by *A. siva* and *A. pedestris* (0±2.474), *E. slateri* and *V. sinensis* (0.004±2.958), and *E. annulata* and *V. sinensis* (0±3.368) *A. siva* and *A. quinquespinosa* (0±3.391), *A. pedestris* and *E. slateri* (0±4.240), *A. quinquespinosa* and *E. slateri* (0±4.706), *A. pedestris* and *E. annulata* (0±5.297), *A. quinquespinosa* and *E. annulata* (0±6.172), *A. siva* and *E. slateri* (0.016±0.863), *A. siva* and
V. sinensis (0.016±1.332), A. quinquespinosa and V. sinensis (0.020±1.361), E. slateri and E. annulata (0.102±0.105), A. quinquespinosa and A. pedestris (0.126±0.162) and A. pedestris and V. sinensis (0.274±0.167) (Table 14).

(ii) Composition distance

Composition distance is a measure of the difference in nucleotide composition for a given pair of sequences. The MEGA software computes and presents the composition distance per site, which is given by the total composition distance between two sequences divided by the number of positions compared, excluding gaps and missing data. The composition distance analysis showed maximum similarity between E. slateri and E. annulata (0.105) followed by A. quinquespinosa and A. pedestris (0.162), A. pedestris and V. sinensis (0.167), A. siva and E. annulata (0.824), A. siva and E. slateri (0.863), A. siva and V. sinensis (1.332), A. quinquespinosa and V. sinensis (1.361), A. siva and A. pedestris (2.474), E. slateri and V. sinensis (2.958), E. annulata and V. sinensis (3.368), A. siva and A. quinquespinosa (3.391), A. pedestris and E. slateri (4.240), A. quinquespinosa and E. slateri (4.706), A. pedestris and E. annulata (5.297) and A. quinquespinosa and E. annulata (6.172) (Table 15).

Test of neutral evolution

This provides a test of selection based on the comparison of the number of synonymous (d_S) and nonsynonymous (d_N) substitutions between sequences. The variance of difference was computed using the bootstrap method (100 replicates). Analyses were conducted using the Nei-Gojobori method (Nei and Gojobori, 1986). The test of neutral evolution revealed minimum variation between A. quinquespinosa and E. annulata (0±3.625) followed by A. quinquespinosa and E. slateri (0±3.648), E. slateri and E. annulata (0.001±3.306), A. quinquespinosa and A. pedestris (0.003±2.994), A. siva and A. quinquespinosa (0.007±2.758), A. pedestris and E. slateri (0.060±1.900), A. pedestris and E.
annulata (0.091±1.704), A. siva and E. slateri (0.116±1.585), A. siva and V. sinensis (0.214±1.250), A. siva and A. pedestris (0.307±1.027), E. annulata and V. sinensis (0.372±0.895), A. siva and E. annulata (0.387±0.868), A. pedestris and V. sinensis (0.421±0.808), E. slateri and V. sinensis (0.661±0.440) and A. quinquespinosa and V. sinensis (0.825±0.222). They are significant at $P<0.05$ level (Table 16).

**Phylogenetic analysis**

The phylogenetic analysis of six reduviids namely, A. siva, A. quinquespinosa, A. pedestris, E. slateri, E. annulata and V. sinensis by three methods such as Maximum Likelihood, Maximum Parsimony and Neighbor-Joining was carried out. The phylogenetic tree branch lengths were also calculated.

In Maximum Likelihood method, the branch lengths of the phylogenetic tree observed were: A. siva (0.2490), A. quinquespinosa (0.1314), A. pedestris (0.0801), E. slateri (0.0789), E. annulata (0.0092), and V. sinensis (0.6855). Acanthaspis pedestris and A. quinquespinosa share a common node of branch length (0.436) and E. slateri and E. annulata share a common node of branch length (0.0206). The longest length observed in V. sinensis (0.6855) diverged independently as a separate node and the shortest length was observed in A. pedestris (0.0801) (Figure 11).

In Maximum Parsimony method, the branch lengths of the tree observed were: A. pedestris (46.917), A. quinquespinosa (68.417), A. siva (94.333), E. slateri (38.958), E. annulata (11.708), and V. sinensis (206.583). Acanthaspis pedestris and A. quinquespinosa share a common node of branch length (63. 792) and E. slateri and E. annulata share a common node of branch length (36. 042). The longest length observed in V. sinensis (206.583) diverged independently as a separate node and the shortest length was observed in A. pedestris (46.917) (Figure 12).
In Neighbour-Joining method, the branch lengths of the tree observed were: *A. pedestris* (0.0789), *A. quinquespinosa* (0.1292), *A. siva* (0.2290), *E. slateri* (0.0824), *E. annulata* (0.0046), and *V. sinensis* (0.6428). *Acanthaspis pedestris* and *A. quinquespinosa* share a common node of branch length (0.0613) and *E. slateri* and *E. annulata* share a common node of branch length (0.0341). The longest length observed in *V. sinensis* (0.6428) diverged independently as a separate node and the shortest length was observed in *A. pedestris* (0.0789) (Figure 13).

**COI genes**

*A. pedestris*

The partially sequenced COI gene amplicon of *A. pedestris* is given in figure 14. revealed an average size of 668bp. The A+T percentage was 61.9% and G+C percentage was 38%. The analysis also revealed the nucleotide frequencies: A-31.7%, T-30.2%, G-17.5%, C-20.5% (Table 9).

*In-silico* translation with the ExPASy translate tool (Gasteiger *et al.*, 2003) revealed a peptide of 218 amino acid sequences for *A. pedestris* with four TGA stop codons. The translation of partial nucleotide sequence and its deduced amino acid sequences are given in Figure 14. Because of the codon preference, the A+T composition in *A. pedestris* was particularly biased at the third codon position, which accounts for 64.8%. The A+T content at first and second codon positions were 64.3% and 57.1% respectively. The G+C composition in *A. pedestris* was the highest in second codon position, which accounts for 42.8%. The G+C content at first and third codon positions were 35.6% and 35.7% respectively (Table 18).

The amino acid compositions analysed by Statistical Analysis of Protein Sequence (Brendel *et al.*, 1992) revealed that *A. pedestris* had higher percentage of isoleucine (13.3%) followed by leucine (11.5%), glycine (9.2%), serine (8.3%), proline (7.3%), valine (7.3%), phenylalanine (6.9%), threonine (6.9%), alanine (6.4%), asparagine (6%), arginine (5%),
aspartic acid (4.1%), tyrosine (1.8%), histidine (1.4%), methionine (1.4%), glutamic acid (0.9%), glutamine (0.9%), cysteine (0.5%), tryptophan (0.5%) and lysine was completely absent (Table 19).

The transmembrane prediction using Hidden Markov Models (Krogh et al., 2001) of *A. pedestris* COI gene sequence of 218 amino acid residues revealed four transmembrane segments between amino acids 25-70, 80-110, 125-150 and 160-190 (Figure 20).

The secondary structure of protein predicted by PSIPRED server (Jones, 1999) from the translated amino acid sequences of *A. pedestris* had the following composition: strands (0%), alpha helix (61.5%) and 3,10 helix (2.3%) (Figure 26).

Distribution of amino acid residues analysed by Ramachandran Plot is given in tables 20 & 21 (Lovell et al., 2003). It revealed that *A. pedestris* had a number of residues: favoured region 174 (93%), allowed region 9(4.8%) and outlier region 4(2.1%) (Figure 32).

The amino acid residues ranging from 22 to 210 were subjected to homology modeling (Figure 44 & 45). The quaternary structure was determined based on template 1v54 (1.80A): HETERO 26-mer and the identity of the sequence with this template was 75.52% and E-value obtained was 1.16823e-60. The quality of the model was determined by QMEAN z-score and the score obtained was -3.888 and the model built was single chain (Figure 38).

**A. quinquespinosa**

The partially sequenced COI gene amplicon of *A. quinquespinosa* is given in figure 15 revealed an average size of 649bp. The A+T percentage was 59.5% and G+C percentage was 40.5% (Table 17). The analysis also revealed the nucleotide frequencies: A-29.6%, T-29.9%, G-17.1% and C-23.4% (Table 9).

*In-silico* translation with the ExPASy translate tool (Gasteiger et al., 2003) revealed a peptide of 192 amino acid sequences for *A. quinquespinosa* with twenty four TGA stop
codons. The translation of partial nucleotide sequence and its deduced amino acid sequences are given in figure 15. Because of the codon preference, the A+T composition in *A. quinquespinosa* was particularly biased at the first codon position, which accounts for 63.6%. The A+T content at second and third codon positions were 53.3% and 59.8% respectively. The G+C composition in *A. quinquespinosa* was the highest in second codon position, which accounts for 45.1%. The G+C content at first and third codon positions were 36% and 40.6% respectively (Table 18).

The amino acid compositions analysed by Statistical Analysis of Protein Sequence (Brendel *et al.*, 1992) revealed that *A. quinquespinosa* had higher percentage of leucine (16.1%) followed by glutamic acid (9.9%), serine (9.4%), glutamine (8.9%), proline (7.8%), tyrosine (7.3%), threonine (7.3%), phenylalanine (6.2%), histidine (6.2%), aspartic acid (5.2%), isoleucine (4.2%), valine (2.1%), glycine (2.6%), asparagine (2.1%), alanine (1.0%), methionine (0.5%), cysteine (1.6%), lysine (1.0%), tryptophan (0.5%) and arginine was completely absent (Table 19).

The transmembrane prediction using Hidden Markov Models (Krogh *et al.*, 2001) analysis of *A. quinquespinosa* COI gene sequence of 218 amino acid residues revealed the presence of transmembrane segments between the amino acids 25-60 (Figure 21).

The secondary structure of protein predicted by PSIPRED server (Jones, 1999) from the translated amino acid sequences of *A. quinquespinosa* had the following composition: strand (3.1%), of alpha helix (33.3%) and 3,10 helix (2.3%) (Figure 27).

Distribution of amino acid residues analysed by Ramachandran Plot is given in tables 20 & 21 (Lovell *et al.*, 2003). It revealed that *A. quinquespinosa* had a number of residues: favoured region 165 (86.8%), allowed region 18(9.5%) and outlier region 7(3.7%) (Figure 33).
The amino acid residues ranging from 22 to 210 were subjected to homology modeling (Figure 44 & 45). The quaternary structure was determined based on template 1v54 (1.80Å): HETERO 26-mer and the identity of the sequence with this template was 75.52% and E-value obtained was 1.16823e-60. The quality of the model was determined by QMEAN z-score and the score obtained was -4.274 and the model built was single chain (Figure 39).

**A. siva**

The partially sequenced COI gene amplicon of *A. siva* given in figure 16 revealed an average size of 644bp. The A+T percentage was 70% and G+C percentage was 29.9% (Table 17). The analysis also revealed the nucleotide frequencies: A-32.6%, T-37.4%, G-13%, C-16.9% (Table 9).

*In-silico* translation with the ExPASy translate tool (Gasteiger et al., 2003) revealed a peptide of 214 amino acid sequences for *A. siva* with eight TGA stop codons. The translation of partial nucleotide sequence and its deduced amino acid sequences are given in figure 16. Because of the codon preference, the A+T composition in *A. siva* was particularly biased at the third codon position, which accounts for 73.2%. The A+T content at first and second codon positions were 70.9% and 65.8% respectively. The G+C composition in *A. siva* was the highest in second codon position, which accounts for 33.8%. The G+C content at first and third codon positions were 29.4% and 26.6% respectively (Table 18).

The amino acid compositions analysed by Statistical Analysis of Protein Sequence (Brendel et al., 1992) revealed that *A. siva* had higher percentage of leucine (17.5%) followed by isoleucine (10.7%), phenylalanine (8.7%), tyrosine (8.3%), serine (7.8%), lysine (7.8%), asparagine (5.3%), glutamine (4.9%), proline (3.9%), histidine (3.4%), arginine (2.9%), glutamic acid (2.9%), aspartic acid (2.9%), cysteine (2.4%), threonine (2.4%), valine (2.4%), tryptophan (2.4%), glycine (1.9%), alanine (1.0%) and methionine (0.5%) (Table 19).
The transmembrane prediction using Hidden Markov Models (Krogh et al., 2001) analysis of *A. siva* COI gene sequence of 214 amino acid residues revealed three transmembrane segments between amino acid 50-100, 125-150 and 160-200 (Figure 22).

The secondary structure of protein predicted by PSIPRED server (Jones, 1999) from the translated amino acid sequences of *A. siva* had the following composition: strand (0%), alpha helix (55.3%), 3,10 helix (1.5%) (Figure 28).

Distribution of amino acid residues analysed by Ramachandran Plot is given in tables 20 & 21 (Lovell et al., 2003). It revealed that *A. siva* had a number of residues: favoured region 181(88.7%), allowed region 15(7.4%) and outlier region 8(3.9%) (Figure 34).

The amino acid residues were subjected to homology modeling (Figure 44 & 45). The quarternary structure was determined based on template 1v54 (1.80Å):HETERO 26-mer. The quality of the model was determined by QMEAN z-score and the score obtained was -4.982 and the model built was single chain (Figure 40).

**E. slateri**

The partially sequenced COI gene amplicon of *E. slateri* given in figure 17 revealed an average size of 729bp (Table 17). The A+T percentage was 65.5% and G+C percentage was 34.6%. The analysis also revealed the nucleotide frequencies: A-26.5%, T-39%, G-15.8%, C-18.8% (Table 9).

*In-silico* translation with the ExPASy translate tool (Gasteiger et al., 2003) revealed a peptide of 220 amino acid sequences for *E. slateri* with twenty three TGA stop codons. The translation of partial nucleotide sequence and its deduced amino acid sequences are given in figure 17. Because of the codon preference, the A+T composition in *E. slateri* was particularly biased at the first codon position, which accounts for 67.8%. The A+T content at second and third codon positions were 61.1% and 67.5% respectively. The G+C composition
in *E. slateri* was the highest in second codon position, which accounts for 39%. The G+C content at first and third codon positions were 32.2% and 32.5% respectively (Table 18).

The amino acid compositions analysed by Statistical Analysis of Protein Sequence (Brendel *et al.*, 1992) revealed that *E. slateri* had higher percentage of serine (14.1%) followed by phenylalanine (13.2%), tyrosine (11.4%), isoleucine (10.5%), asparagine (7.3%), arginine (6.8%), threonine (5.5%), cysteine (5.5%), leucine (5.0%), tryptophan (4.5%), proline (4.1%), lysine (3.2%), histidine (2.7%), glycine (1.8%), alanine (1.4%), valine (1.4%), aspartic acid (0.9%) and methionine (0.9%). Glutamic acid and glutamine were completely absent (Table 19).

The transmembrane prediction using Hidden Markov Models (Krogh *et al.*, 2001) analysis of *E. slateri* COI gene sequence of 220 amino acid residues revealed six transmembrane segments between amino acids 25-50, 60-160 and 165-200 (Figure 23).

The secondary structure of protein predicted by PSIPRED server (Jones, 1999) from the translated amino acid sequences of *E. slateri* had the following composition: strand (25%), alpha helix (1.8%) and 3,10 helix (0%) (Figure 29).

Distribution of amino acid residues analysed by Ramachandran Plot is given in tables 20 & 21 (Lovell *et al.*, 2003). It revealed that *E. slateri* had a number of residues: favoured region 170(78%), allowed region 27(12.4%) and outlier region 21(9.6%) (Figure 35).

The amino acid residues were subjected to homology modeling (Figure 44 & 45). The quaternary structure was determined based on template 1v54 (1.80A):HETERO 26-mer. The quality of the model was determined by QMEAN z-score and the score obtained was -4.462 and the model built was single chain (Figure 41).

**E. annulata**

The partially sequenced COI gene amplicon of *E. annulata* given in figure 18 revealed an average size of 646bp (Table 17) The A+T percentage was 61.4% and G+C
percentage was 30.7%. The analysis also revealed the nucleotide frequencies: A-28.2%, T-41.2%, G-14.4%, C-16.3% (Table 9).

*In-silico* translation with the ExPASy translate tool (Gasteiger *et al.*, 2003) revealed a peptide of 211 amino acid sequences for *E. annulata* with four TGA stop codons. The translation of partial nucleotide sequence and its deduced amino acid sequences are given in figure 18. Because of the codon preference, the A+T composition in *E. annulata* was particularly biased at the third codon position, which accounts for 74.5%. The A+T content at first and second codon positions were 70.5% and 63.6% respectively. The G+C composition in *E. annulata* was the highest in second codon position, which accounts for 36.3%. The G+C content at first and third codon positions were 29.7% and 26% respectively (Table 18).

The amino acid compositions analysed by Statistical Analysis of Protein Sequence (Brendel *et al.*, 1992) revealed that *E. annulata* had higher percentage of isoleucine (16.1%), leucine (14.7%), serine (9.5%), glycine (8.5%), proline (7.6%), phenylalanine (7.1%), alanine (6.6%), threonine (6.2%), asparagine (5.2%), valine (4.7%), arginine (4.3%), glutamine (1.9%), histidine (1.9%), tyrosine (1.4%), cysteine (0%), aspartic acid (3.3%), glutamic acid (0.9%) and lysine, methionine and tryptophan were completely absent (Table 19).

The transmembrane prediction using Hidden Markov Models (Krogh *et al.*, 2001) analysis of *E. annulata* COI gene sequence of 211 amino acid residues revealed four transmembrane segments between amino acids 25-75, 80-120 and 125-175 (Figure 24).

The secondary structure of protein predicted by PSIPRED server (Jones, 1999) from the translated amino acid sequences of *E. annulata* had the following composition: strand (0%), alpha helix (60.7%) and 3,10 helix (5.2%) (Figure 30).

Distribution of amino acid residues analysed by Ramachandran Plot is given in tables 20 & 21 (Lovell *et al.*, 2003). It revealed that *E. annulata* had a number of residues: favoured region 201(96.2%), allowed region 7(3.3%) and outlier region 1(0.5%) (Figure 36).
The amino acid residues were subjected to homology modeling (44 & 45). The quaternary structure was determined based on template 1v54 (1.80Å):HETERO 26-mer. The quality of the model was determined by QMEAN z-score and the score obtained was -3.933 and the model built was single chain (Figure 42).

**V. sinensis**

The partially sequenced COI gene amplicon of *V. sinensis* given in figure 19 revealed an average size of 687bp (Table 17). The A+T percentage was 64% and G+C percentage was 35.9%. The analysis also revealed the nucleotide frequencies: A-31.7%, T-32.3%, G-18.9%, C-17% (Table 9).

*In-silico* translation with the ExPASy translate tool (Gasteiger *et al.*, 2003) revealed a peptide of 223 amino acid sequences for *V. sinensis* with six TGA stop codons. The translation of partial nucleotide sequence and its deduced amino acid sequences are given in figure 19. Because of the codon preference, the A+T composition in *V. sinensis* was particularly biased at the third codon position, which accounts for 65.8%. The A+T content at first and second codon positions were 64.1% and 62.2% respectively. The G+C composition in *V. sinensis* was the highest in second codon position, which accounts for 38.3%. The G+C content at first and third codon positions were 35.4% and 34.2% respectively (Table 18).

The amino acid compositions analysed by Statistical Analysis of Protein Sequence (Brendel *et al.*, 1992) revealed that *V. sinensis* had higher percentage of isoleucine (13.0%), proline (9.0%), leucine (8.5%), serine (7.6%), valine (7.2%), asparagine (6.7%), alanine (5.8%), lysine (5.8%), arginine (5.4%), threonine (5.4%), glycine (5.4%), methionine (3.6%), aspartic acid (3.6%), phenylalanine (2.7%), glutamic acid (2.7%), glutamine (1.8%), histidine (1.8%), tryptophan (1.3%), tyrosine (1.3%) and cysteine (1.3%) (Table 19).
The transmembrane prediction using Hidden Markov Models (Krogh et al., 2001) analysis of *V. sinensis* COI gene sequence of 223 amino acid residues revealed two transmembrane segments between amino acids 10-30 and 160-175 (Figure 25).

The secondary structure of protein predicted by PSIPRED server (Jones, 1999) from the translated amino acid sequences of *V. sinensis* had the following composition: strand 6.3%, alpha helix (30%) and 3,10 alpha helix (0%) (Figure 31).

Distribution of amino acid residues analysed by Ramachandran Plot is given in tables 20 & 21 (Lovell et al., 2003). It revealed that *V. sinensis* had a number of residues in favoured region 190(86%), allowed region 20(9%) and outlier region 11(5%) (Figure 37).

The amino acid residues were subjected to homology modeling (Figure 44 & 45). The quaternary structure was determined based on template 1v54 (1.80Å): HETERO 26-mer. The quality of the model was determined by QMEAN z-score and the score obtained was -4.6574 and the model built was single chain (Figure 43).

**Discussion**

The sequence length of around 639-724 bp observed for the six species of Reduviinae is closer to that of the COI gene sequences of different species of Reduviinae (NCBI database, Genbank accession no: GU12616, GU012616.1 and GU012616.1). The length of this region is highly variable among different insects due to its high rate of nucleotide substitution, insertions/deletions and the presence of a variable number of tandem repeats (Fauron and Wolstenholme, 1980; Inohira *et al.*, 1997).

The mitochondrial genome has become not only a major resource for comparative genomics but also play an important role in metabolism, apoptosis, disease and aging (Boore, 1999). A similar overall A+T richness (65%) was reported for the family Reduviidae (Muraji *et al.*, 2001; Pfeiler *et al.*, 2006; Liu *et al.*, 2009; Li *et al.*, 2011) and for oriental Rhynocoris species (Baskar, 2010 and Singh, 2012). However, the richness of the A+T% ranging from
Biosystematics of chosen reduviine species based on mitochondrial COI gene sequences

59.5% (*A. quinquespinosa*) to 70% (*A. pedestris*) recorded for the reduviine species is lower than that of Apidae (80%) and Coleoptera (79.8%) (Jermiin and Crozier, 1994).

**Subfamily characters**

Though the six reduviine species have their own specific molecular characteristics, the data obtained from the molecular studies suggest that the following characters could be used as subfamily markers.

(i) **AT composition**

The highest percentage of AT composition observed for six reduviine species (61 to 70%) could be a diagnostic marker of the family Reduviinae. The richness of AT composition has been used as a marker in Reduviidae (Muraji *et al*., 2001; Pfeiler *et al*., 2006; Liu *et al*., 2009; Li *et al*., 2011; Singh, 2012).

(ii) **GC composition**

A narrow range of GC composition (29.9% to 40.5%) observed for six species of Reduviinae could be again considered as a subfamily character.

(iii) **Amino acid composition**

The least amount of methionine observed in all the six species of Reduviinae could be also considered as a subfamily character.

(iv) **Codon frequencies**

The most frequently occurred codon, AUU in six reduviine species could be again considered as a subfamily character.

**Generic characters**

Each genus has its own specific molecular characteristics. However, the data obtained for three species of *Acanthaspis* alone could be analysed for this purpose because all the other three genera are represented by a lone species.
(i) AT composition

The highest percentage of AT composition of three *Acanthaspis* species ranging from 63 to 70.9% could be used as a generic marker as it was used as a marker for *Rhynocoris* species (Singh, 2012).

(ii) GC composition

A narrow range of GC composition (29.4% to 35.6%) of three *Acanthaspis* species could be treated as a generic character.

(iii) Amino acid composition

The following closer values of amino acid, serine in *A. pedestris* (8.3), *A. quinquespinosa* (9.4) and *A. siva* (7.8) could be considered as a generic character for *Acanthaspis*. The highest amino acid residues were observed in favoured region for all the three species.

Species specific characters

(i) *Acanthaspis pedestris*

The AT composition (61.9%) biased at the third codon (64.8%), the GC composition (38%) biased at the second codon (42.8 %), higher percentage of isoleucine (13.3%), absence of lysine, four specific transmembrane segments between amino acids 25-70, 80-110, 125-150 and 160-190 with specific percentage of alpha helix (61.5 %), 3,10 helix (2.3%), absence of coil and the highest amount of amino acids residues (93%) found in favoured region by Ramachandran Plot could be considered as specific markers for *A. pedestris*.

(ii) *Acanthaspis quinquespinosa*

The AT composition (59.5%) biased at the first codon (63.6%), the GC composition (40.5%) biased at the second codon (45.1%), higher percentage of leucine (16.1%) and absence of arginine, a specific transmembrane segment between amino acids 25-60 with specific percentage of strand (3.1%), alpha helix (33.3%) and 3,10 helix (2.3%) and the
highest amount of amino acids residues (86.8%) found in favoured region by Ramachandran Plot could be considered as specific markers for *A. quinquespinosa*.

**(iii) Acanthaspis siva**

The AT composition (70%) biased at the third codon (73.2%), the GC composition (29.9%) biased at the second codon (33.8%), higher percentage of leucine (17.1%), three specific transmembrane segments between amino acids 50-100, 125-150 and 160-200 with specific percentage of alpha helix (55.3%) and 3,10 helix (1.5%), and the highest amount of amino acids residues (88.7%) found in favoured region by Ramachandran Plot could be considered as specific markers for *A. siva*.

**(vi) Edocla slateri**

The AT composition (65.5%) biased at the first codon (67.8%), the GC composition (34.6%) biased at the second codon (39%), higher percentage of serine (14.1%), absence of glutamic acid and glutamine, six specific transmembrane segments between amino acids 25-50, 60-160 and 165-200 with specific percentage of strand (25%), alpha helix (1.8%) and 3,10 helix (0%) and the highest amount of amino acid residues (78%) found in favoured region by Ramachandran Plot could be considered as specific markers for *E. slateri*.

**(v) Empyrocoris annulata**

The AT composition (61.4%) biased at the third codon (74.5%), the GC composition (30.7%) biased at the second codon (74.5%), higher percentage of the isolucine (16.1%), absence of lysine, methionine and tryptophan, four specific transmembrane segments between amino acids 25-75, 80-120 and 125-175 with specific percentage of strand (0%), alpha helix (60.7%) and 3,10 helix (5.2%) and the highest amount of amino acids residues (96.2%) found in favoured region by Ramachandran Plot could be considered as specific markers for *E. annulata*. 
(v) *V. sinensis*

The AT composition (64%) biased at the third codon (65.8%), the GC composition (35.9%) biased at the second codon (38.3%), higher percentage of the isolucine (13%), two specific transmembrane segments between aminoacids 10-30 and 160-175 with specific percentage of strand (6.3%), alpha helix (30%) and 3,10 helix (0%) and the highest amount of amino acids residues (86%) found in favoured region by Ramachandran Plot could be considered as specific markers for *V. sinensis*.

**Intergeneric and Intrageneric affinity**

(i) **Evolutionary distance**

The order of affinity between any two species based on evolutionary distance between sequences observed: *E. slateri* and *E. annulata* (0.088±0.245) > *A. quinquespinosa* and *A. pedestris* (0.211±0.099) > *A. siva* and *E. annulata* (0.269±0.132) > *A. siva* and *E. slateri* (0.334±0.112) > *A. quinquespinosa* and *E. annulata* (0.352±0.146) > *A. pedestris* and *E. annulata* (0.306±0.132) > *A. pedestris* and *E. slateri* (0.363±0.112) > *A. quinquespinosa* and *E. slateri* (0.393±0.157) > *A. siva* and *A. pedestris* (0.543±0.021) > *A. siva* and *A. quinquespinosa* (0.566±0.099) > *A. pedestris* and *V. sinensis* (1.025±0.068) > *E. annulata* and *V. sinensis* (1.079±0.079) > *A. siva* and *V. sinensis* (1.165±0.068) > *E. slateri* and *V. sinensis* (1.180±0.068) > *E. slateri* and *V. sinensis* (1.182±0.082) suggests affinity between species showing the inter- and intragenric affinity among these species.

(ii) **Homogeneity of substitution patterns between sequences**

The order of affinity between any two species based on homogeneity of substitution patterns between sequences obtained from disparity index test: *A. siva* and *A. pedestris* (0±2.474) > *E. annulata* and *V. sinensis* (0±3.368) > *A. siva* and *A. quinquespinosa* (0±3.391) > *A. pedestris* and *E. slateri* (0±4.240) > *A. quinquespinosa* and *E. slateri* (0±4.706) > *A. pedestris* and *E. annulata* (0±5.297) > *A. quinquespinosa* and *E. annulata* (0±6.172) > *E.
annulata and V. sinensis (0±3.368) > E. slateri and V. sinensis (0.004±2.958) > A. siva and E. slateri (0.016±0.863) > A. siva and V. sinensis (0.016±1.332) > A. quinquespinosa and V. sinensis (0.020±1.361) > A. quinquespinosa and A. pedestris (0.126±0.162) > A. pedestris and V. sinensis (0.274±0.167) also reveals the inter- and intragenric affinity among these species.

(iii) Nucleotide base composition

The order of affinity between any two species based on the distance in composition of nucleotide bases: E. slateri and E. annulata (0.105±0.099) > A. quinquespinosa and A. pedestris (0.162±0.042) > A. pedestris and V. sinensis (0.167±0.691) > A. siva and E. annulata (0.824±0.078) > A. siva and E. slateri (0.863±0.512) > A. siva and V. sinensis (1.332±0.023) > A. quinquespinosa and V. sinensis (1.361±0.791) > A. siva and A. pedestris (2.474±0.031) > E. slateri and V. sinensis (2.958±0.481) > E. annulata and V. sinensis (3.368±0.312) > A. siva and A. quinquespinosa (3.391±0.12) > A. pedestris and E. slateri (4.240±0.069) > A. quinquespinosa and E. slateri (4.706±0.245) > A. quinquespinosa and E. annulata (6.172±0.137) further shows the inter- and intrageneric affinity among these species.

(iv) Neutrality test

From the codon-based Z-test of neutrality between sequences of six reduviine species, the highest neutrality (affinity) observed between A. quinquespinosa and E. annulata (0±2.703) followed by E. slateri and E. annulata (0.001±1.714), A. quinquespinosa and A. pedestris (0.003±2.092), A. siva and A. quinquespinosa (0.007±2.557), A. pedestris and E. slateri (0.060±1.413), A. pedestris and E. annulata (0.091±1.604), A. siva and A. pedestris (0.307±1.002), A. siva and V. sinensis (0.214±1.100), E. annulata and V. sinensis (0.372±0.896), A. siva and E. annulata (0.387±0.740), A. pedestris and V. sinensis
(0.421±0.715), *E. slateri* and *V. sinensis* (0.661±0.410) and *A. quinquespinosa* and *V. sinensis* (0.825±0.153) again reiterates inter- and intrageneric affinity among these species.

(iv) Transmembrane segments

The affinity among *A. pedestris*, *A. quinquespinosa*, *E. slateri* and *E. annulata* as evidenced by the presence of closer value transmembrane segments between aminoacids 25-75; between aminoacids 125-190 among *A. pedestris*, *E. slateri*, *E. annulata* and *V. sinensis* and between aminoacids 125-150 between *A. pedestris* and *A. siva* reveal the inter- and intrageneric affinity among these species.

(v) Phylogenetic analysis

Divergence of *A. siva* diverges in a separate clade and *A. pedestris* and *A. quinquespinosa* grouping together in a single cluster in phylogenetic tree determined using Neighbour-Joining method; clustering of *A. pedestris* and *A. quinquespinosa* and *E. annulata* and *E. slateri* uniformly in all the three phylogenetic trees analysed, revealing intergeneric affinity. The intergeneric specificity is also revealed by the independent divergence of *V. sinensis* as a separate clade in all the three phylogenetic trees. Baskar (2010) also reported the intrageneric affinity of *R. marginatus*, *R. fuscipes*, *R. kumarii* and *R. longifrons* by Neighbour-Joining method. Furthermore, Singh (2012) also reported the inter- and intrageneric affinity of three morphs of *R. kumarii* and four ecotypes of *R. marginatus*. Such analysis was carried out for other members of the family Reduviidae with different genes namely COI and Cyt-b genes (Pfeiler *et al*., 2006; Liu *et al*., 2009). The greater intrageneric and interspecific affinity between *A. pedestris* and *A. quinquespinosa* and between *E. slateri* and *E. annulata* as evidenced by the disparity index analysis and supported by evolutionary distance analysis is worth mentioning. The study could be further extended analyzing other mitochondrial genes and by analyzing a reasonable number of reduviine species.
Conclusion

The parameters like AT composition, GC composition, codon frequencies, amino acid composition, transmembrane pattern, protein secondary structure, dihedral angles between amino acids and protein quaternary structure of mitochondrial cytochrome C oxidase subunit I gene analysis could be used either as a generic marker or species specific marker based on the common and unique characters observed.

Greater affinity between *A. pedestris* and *A. quinquespinosa* and *E. slateri* and *E. annulata* and lesser affinity between *V. sinensis* and rest of the species revealed by Maximum Likelihood, Maximum Parsimony and Neighbour-Joining distance methods and supported by neutrality test and evolutionary distance analysis is interesting.