Genome organization and comparative evolutionary mitochondriomics of rice earhead bug *Leptocorisa oratoria* (Fabricius)

Guru-Pirasanna-Pandi Govindharaj[1,*], M. Annamalai[1], Jaipal Singh Choudhary[2], G. Basana-Gowda[1], Totan Adak[1], Naiyar Naaz[2], Naveenkumar Patil[1], Enrico Ruzzier[3] and Prakash Chandra Rath[1]

[1]ICAR-National Rice Research Institute, Cuttack 753 006, India
[2]ICAR-Research Complex for the Eastern Region, Farming System Research Centre for Hill and Plateau Region, Ranchi 834 010, India
[3]World Biodiversity Association Onlus, C/o Museo Civico di Storia Naturale, Lungadige Porta Vittoria 9, 37129 Verona, Italy

The rice earhead bug, *Leptocorisa oratoria* (Fabricius, 1794) is a critical rice pest in India. No mitochondrial genome of *L. oratoria* has been sequenced earlier, and the mitochondrial data are crucial for phylogenetic and population genetic studies of this significant rice pest. In the present study, the genome of *L. oratoria* is 17,584 bp long with 73.57% AT content. We observed tandem repeat in the control region. Analyses from genetic distance, sliding window and K_a/K_s ratio revealed a purifying selection of 13 protein-coding genes, with *cox1* and *nad2* reporting the lowest and highest rate of evolution respectively. Phylogenetic analysis was reconstructed using 65 pentatomid mitogenomes with Bayesian inference and maximum likelihood methods. The results help differentiate the Coreoidea superfamily from Lygaeoidea, Aradoidea and Pentatomoidea. There were two topologies at the family level, i.e. one clade formed with Coreidae + Rhopalidae + Alydidae, and the rest of the families of Pentatomomorpha formed in separate clades. Further, *L. oratoria* produced an independent subclade from the earlier reported *Leptocorisa* sp. genome. This study provides a source mitogenome for *L. oratoria* species to study population demography, individual differences and phylogeography of hemipterans.

Keywords: Mitogenome, next generation sequencing, population genetics, phylogeny, rice earhead bug.

THE rice earhead bug, *Leptocorisa oratoria* (Fabricius, 1794) (Heteroptera : Coreoidea : Alydidae), is a crucial pest of rice reported from major rice-producing countries of Asia[1,2]. In India, *L. oratoria* is usually known as 'gundhi bug', as adults discharge an unpleasant odour as a defensive strategy. Both adults and nymphs of the bug suck the sap from rice grains during the milky stage[1,2]. *L. oratoria* infestation may cause 10–30% yield losses[3,4]; severe infestation leads to >50% yield loss[2]. Sri Lanka and the Philippines have experienced 60–70% yield loss due to *Leptocorisa* sp. in rice[5]. Apart from direct loss, *L. oratoria* affects seed quality and germination[6]. Among the nine species of *Leptocorisa* reported from Asia and Oceania, *L. acuta* and *L. chinensis* are found in temperate climates, whereas *L. oratoria* is prevalent in tropical climates[7]. *L. oratoria* is an indigenous pest species in India. The first report of an infestation in paddy fields was recorded from Madras Province at the beginning of the 20th century.

The mitogenome naturally consists of an exceptional non-coding region known as the control region with higher AT content[8–11], and areas for genome replication and initiation of transcription[12,13]. Mitogenome organization and content are commonly used in population genetics, phylogenetics, genomics and molecular evolution[11,14]. There is minimal information from the literature on *L. oratoria* geographical variability and genetic structure. Alydidae[15] has been estimated to contain over 56 genera, with 285 species worldwide[16]. Nevertheless, mitochondrial genome data exist for only a few species. Thus far, sequences for only *Riptortus pedestris* (Fabricius, 1775) and *Leptocorisa* sp. are available in NCBI GenBank. Phylogeny analyses at the family level and new genomics approaches have increased understanding of species evolution and phylogeny[17]. Next-generation sequencing (NGS) techniques have been widely used for sequencing bacterial genomes[18], arthropod genomes[19] and the mitochondrial genomes of living organisms or herbarium and museum specimens[20–22]. Data generated from these techniques have been instrumental in gaining accurate information about species population genetics and evolution[18–22].

In this study, we have arranged a mitogenome of *L. oratoria* from India using NGS to characterize its evolution and phylogeny. We report the nucleotide composition, assembly and overlap of genes, non-coding control region, codon usage and tRNA secondary structure. We have also compared the *L. oratoria* mitogenome with the existing mitogenomes of superfamily Coreoidea, viz. *Leptocorisa* sp.; *Riptortus pedestris* (Fabricius, 1775); *Aeschyntelus notatus* Hsiao, 1963; *Corius* sp. Leach 1815; *Myrmus lateralis*

*For correspondence. (e-mail: Guru.G@icar.gov.in)

Hsiao, 1964; *Acanthocoris* sp. Amyot & Serville, 1843; *Cletus punctiger* (Dallas, 1852); *Cletus rubidiventris* (Westwood, 1842); *Cloresmus pulchellus* Hsiao, 1963; *Enoplops potanini* (Jakovlev, 1890); *Hydaropsis longirostris* (Hsiao, 1963); *Leptoglossus membranaceus* (Fabricius, 1781); *Mictis tenebrosa* (Fabricius, 1787); *Molipteryx lunata* (Distant, 1900) and *Pseudomictis brevicornis* Hsiao, 1963. The phylogeny of Pentatomomorpha infraorder has been revised using the present study sequences and previously reported mitogenome sequences from the NCBI database. This mitogenome reference sequence can be exploited for biotype differences and other molecular studies in future *L. oratoria* research.

Materials and methods

Mitochondrial genome sequencing and assembly

Specimens of *L. oratoria* were collected from the experimental farms (20°26′54″; 85°56′25″) of ICAR-National Rice Research Institute (NRRI), Cuttack, India. Insect samples were identified following Barrion and Litsinger[23], and reconfirmed by the insect biosystematics laboratory taxonomist, Department of Entomology, University of Agricultural Science, Bengaluru, India. Voucher specimens were deposited in the Crop Protection Division, ICAR-NRRI (CPT/GB/VS-1-3).

A mitochondrial DNA isolation kit of Abcam (ab65321, Abcam, Cambridge, UK) was used for extracting the mitochondrial DNA from mature *L. oratoria*. Sequencing and Illumina library preparation were completed at Xcelris Labs Limited, Ahmedabad, India. Detailed information on the sequencing procedure is available in our earlier publication[24]. In brief, a library of paired-end sequencing was developed, and a high-sensitivity DNA chip-based Bioanalyzer 2100 (Agilent Technologies, California) was used to analyse the amplified library. Sequencing was performed onilluminaMiSeq Sequencing System (Illumia Inc., California) with a 2 × 150 bp kit. Assembly of sequences was done using NOVOPlasty (a seed extend-based assembler for whole-genome data) with default parameters. The reference data for *de novo* assembly were used from *Leptocorisa* sp. *cox1* gene of MT-2014 mitochondria (accession number KM244663). A single scaffold was obtained from the assembly with approximately 17.5 kb. The resultant contigs comprising the mitogenome sequence were verified by the NCBI blast algorithm.

Leptocorisa oratoria mitogenome annotation

The detailed procedure used for annotation is available from our earlier publication[24]. In short, all the genes while placing invertebrate mitochondria genetic code as the base in MITOS web server were predicted[25]. Protein-coding genes (PCGs) were cross-checked by the ORF Finder employed at NCBI. Again, all rRNAs and PCGs were manually interpreted and cross-checked by comparing them with the 16 mitogenomes of 15 other Coreoidea species downloaded from NCBI. MITOS-downloaded two-dimensional tRNA structures were rechecked using tRNAscanSEv 1.21 and a cut-off value of 15.0 of covariance employing Mito/Chloroplast mode[26]. CGview server (circular mitochondrial map), ClustalW (alignment), and DNA mfold server (secondary fold structure) were run with the following standard procedures[27–29]. AT skew = [A – T]/[A + T] and GC skew = [G – C]/[G + C] formulae were employed to calculate the composition of GC skew[30]. Manually, the overlapping regions and intergenic spacers between genes were estimated. For each PCG, and synonymous and non-synonymous substitution rates were calculated using DnaSP 5.0 (ref. 31). The sequenced *L. oratoria* mitogenomes were uploaded into NCBI GenBank as accession numbers OM177216 and OM177217.

Phylogenetic analysis

All available complete mitogenomes of Pentatomorpha infraorder (75 mitogenomes) were chosen for analysis. *L. oratoria* mitogenome (two numbers) was correlated with the existing mitogenomes of Pentatomorpha infraorder. All 13 PCG nucleotide sequences from 75 mitogenomes were used for phylogenetic analysis. Detailed information on the sequencing procedure is available from our earlier publication[24], using standard protocols[32–38].

Results and discussion

Sequencing and assembly of L. oratoria mitogenome

Libraries of genomic DNA were used to obtain the *L. oratoria* mitogenome sequence data. We acquired 24.5 (sample-1) and 24.4 (sample-2) million reads of 2 × 150 bp read size from the paired-run Illumina MiSeq sequencing platform. Mitogenomes of *L. oratoria* had minimal variation (<0.1) with high similarity. Hereafter, for clarity, only one sample is discussed in this study.

L. oratoria mitogenome organization

The whole mitochondrial genome of *L. oratoria* (accession numbers OM177216 and OM177217) is 17,584 and 17,583 bp closed circular in length, which matches with previously studied mitogenome of Coreidae (14,532 bp in *A. notatus* [EU427333] to 18,022 bp in *C. pulchellus* [MF497719]). Mitochondrial genomes include the 37 genes usually present in metazoan mitogenomes[13], consisting of 13 PCGs (*cox1-3*, *nad4l*, *nad1-6*, *atp6*, *atp8* and *cob*), 22 tRNA (two each for serine and leucine; one for each amino acid), and two rRNA (rrnS and rrnL) genes (Figure 1, Supplementary Tables 1 and 2). The places of the gene in *L. oratoria* mitochondrial genome according to characteristic and plesiomorphic form theorized for the Pancrustacea

arrangement[39], consistent with most hemipteran insects sequenced so far[40].

A total length of 3035 bp non-coding region was present between 14,550 and 17,584 bp of *rrnS* and *trnI* (Figure 1 and Supplementary Figure 1). *L. oratoria* genome organization was condensed by 900 bp, dispersed in 24 intergenic spacers from 1 to 579 bp, and contiguous genes overlying at six boundaries by 23 bases from 1 to 8 bp. An extended putative regulatory region and a lengthy repetitive region containing additional genes contributed to the unusually long mitogenome in *L. oratoria*. However, their functional role remains unclear. Similar observations were also reported in the mitogenome *N. lugens*[41,42].

Leptocorisa oratoria nucleotide composition and skewness

Genome-wide and strand-specific compositional biases were observed in the nucleotide composition. The composition of nucleotides in *L. oratoria* was biased towards A and T (73.95%; A = 44.97%, T = 28.98%, G = 10.61%, C = 15.44%), corresponding to previously studied mitogenomes of Coreoidea species (from 71.47% in *Leptoglossus membranaceus* to 76.59% in *Riptortus pedestris*). All PCGs had their AT content (74.05%) and tRNAs (78.49%), and rRNAs (77.17%) evaluated separately. The conserved region (75%) was also well within the range reported in the 16 Coreoidea superfamily species examined in the present study. PCGs of the majority (J-strand) and the minority (N-strand) were scrutinized separately. It was observed that the N-strand (four PCGs, 74.98%) had higher AT content than the J-strand (nine PCGs, 73.64% AT content). Additionally, there was a strong bias for T content (33.06% and 53.70% for J-strand and N-strand respectively) when compared with the corresponding values of 40.58% and 21.28% for A content in the PCGs. Similar observations have been reported in hemipteran mitogenomes and other insects[43–46].

The biasness in the data of nucleotides is reflected in codon usage. Due to code degeneracy, the study revealed that AT content was considerably higher in the third codon position, which was found in other insects, and T was over-represented in the second codon position[11,43,45]. Similarly, all three PCG codons on the N-strand showed higher G and C content in the J-strand. The base composition corresponds to the general mitogenome inclination towards a reduced G content[7]. The RSCU of NNA and NNU codons >1 directed at the third position of the U/A has a high frequency of codon usage in the *L. oratoria* mitochondrial genome (Supplementary Table 3). Bias towards AT-rich codons was also evident in the codons TTA–Leu (3.47), ATT–Met (1.80), AAA–Met (1.80), AAA–Lys (1.71), AAT–Asn (1.60), TTT–Phe (1.59), TAT–Tyr (1.54) and ATT–Ile (1.53). The overall number of non-stop codons was 3448 bp. Leu (14.52%), Phe (10.03%), Ile (9.17%), Ser (9.17%), Met (8.39%), Thr (5.57%) and Val (5.12%) were the most prevalent amino acids making up >53.61% of the total.

Several studies have been carried out on the insect mitogenome AT composition of various regions or genes to examine compositional heterogeneity of bases and site rate variation[24,40]. In the present study, AT composition of PCG in zerofold (P_{0FD}), twofold (P_{2FD}) and fourfold degenerate sites (P_{4FD}) was each ascertained for 16 Coreoidae species (*Leptocorisa* sp., *L. oratoria*, *R. pedestris*, *A. notatus*, *Corius* sp., *M. lateralis*, *Acanthocoris* sp., *C. punctiger*, *C. rubidiventris*, *C. pulchellus*, *E. potanini*, *H. longirostris*, *L. membranaceus*, *M. tenebrosa*, *M. lunata* and *P. brevicornis*) (Supplementary Figure 2). Furthermore, among the 16 coreoidae species considered in this study, the majority of PCGs showed much less variance in AT content at P_{0FD} than at P_{2FD} and P_{4FD} sites (excluding *atp8* gene).

GC skew and AT skew were estimated for all accessible entire mitochondrial genomes of Coreidae species. AT skewness of the mitogenome was somewhat positive (0.216), indicating the presence of more A than T. Our findings are consistent with other reports from Coreoidae species, which ranged from 0.091 in *Myrmus lateralis* to 0.204 in *Leptocorisa* sp.[40,47]. Although all of the chosen Coreoidea mitochondrial genome samples had negative values for GC skew (–0.164 to –0.253), it is interesting that the GC skew of *L. oratoria* was also negative (–0.164), showing the occurrence of more C than G. Similar skewness has been seen in hemipteran, lepidopteran, dipteran and other insect species[47–50].

Protein coding genes

The mitogenome of *L. oratoria* comprises all 13 PCGs (Supplementary Tables 1 and 2). The majority of PCGs begin with a representative ATN initiation codon (*nad2*, *atp8*, *nad6*, *cyb* and *nad1* use ATA; *atp6*, *cox3* and *nad4* use ATG and *nad4L* use ATT). Initiation codon, i.e. ATA or ATG containing methionine, is the most prevalent among insects[40,44]; however, the uncommon ATT initiation codon has also been reported in other Hemiptera[41,50]. Two PCGs (*cox2* and *atp8*) were found to have typical TAA termination codons. Other PCGs like *cox3*, *nad5*, *cyb* (TA) and *nad2*, *cox1*, *atp6*, *nad3*, *nad4*, *nad4L* and *nad1* (T) had incomplete termination. The *atp8* gene length of *L. oratoria* was 156 bp, which is similar to other insects (like 138 bp in *Petiole gall psyllid*, *Pachypsylla venusta* (Osten-Sacken, 1861)[51] and 183 bp in the gypsy moth *Lymantria dispar* (Linnaeus, 1758)[52] and mostly falling in and around 160 bp (ref. 41).

Ribosomal and transfer RNA genes, and other structure

Similar to other insect mitogenomes, *L. oratoria* has two ribosomal RNA genes, namely *rrnL* (situated between

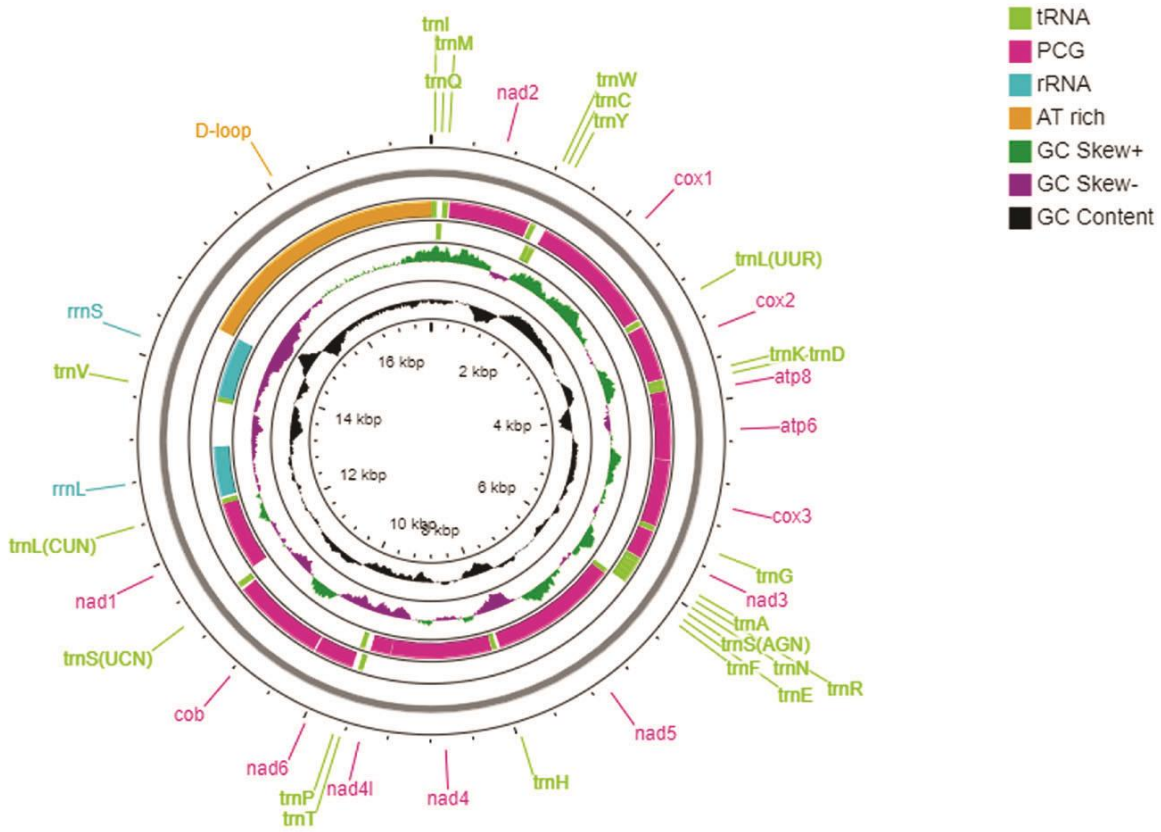

Figure 1. *Leptocorisa oratoria* mitogenome map. The first outer circle indicates protein-coding genes (PCGs), rRNA, tRNAs and control region. The second and third circles specify GC content and GC skew respectively.

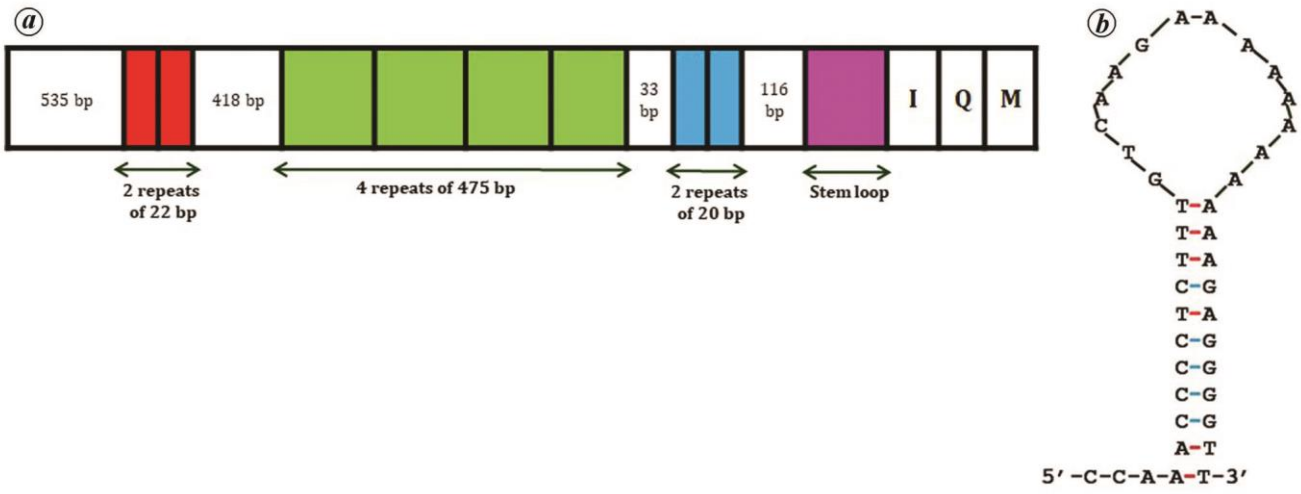

Figure 2. Depicted control region of *L. oratoria* mitogenome. ***a***, Diagram of control region. ***b***, Possible stem–loop structure from control region.

trnL and *trnV*) and *rrnS* (situated between *trnV* and the control region). The ribosomal genes *rrnS* and *rrnL* were 1233 and 785 bp respectively, in the range recorded in Coreoidea and hemipterans[40,53,54]. A typical total of 22 tRNA genes was decoded in the *L. oratoria* mitogenome, and the size ranged from 62 to 73 bp for *trnW*, *trnD*, *trnG*, *trnA* and *trnT*. All projected tRNAs had the characteristic cloverleaf secondary structure, except these dihydrouridine arms represented in *trnS* (AGN) having a straight forward loop as in most insects and other metazoans[41]. By adapting its structural conformation, *trnS* can fit into the ribosome and perform similarly to other tRNAs[55].

The anticodons observed in this study are comparable to those found in other Coreoidea species. Except for Ala, Arg, Met and Ser (AGN), *L. oratoria* possesses a 7 bp acceptor stem, a 5 bp anticodon stem and a 5 bp anticodon loop. The DHU arm (stem size 1–4 bp, loop size 3–11 bp) and T–C arm (stem size 3–5 bp, loop size 1–10 bp) frequently showed size inconsistency. In addition, 22 tRNA genes had 29 mismatched base pairs and G–U wobble pairings. Weak bonds were formed by 21 G–U wobble pairs and unmatched U–U base pairs in two *trnR* (acceptor arm) and *TrnY* (TC arm), one unmatched A–A base pair in *trnE* (anticodon arm), one unmatched G–G base pair in *trnA* (anticodon arm), one unmatched A–G base pair in *trnS* (DHU arm), and three unmatched A–C base pairs in *trnD* and *trnM* (acceptor arm) and *trnN* (TΨC arm).

L. oratoria control region and intergenic spacers

Similar to other insect mitogenomes, overlapping genes were observed in *L. oratoria*[50,56]. The present study revealed the presence of 11 overlapping regions. The most extended intergenic spacer (51 bp) was noted between *cytb* and *trnS*. Likewise, 44 bp intergenic spacer was observed between *trnF* and *nad5*, and 23 bp between *nad6* and *cytb* (Supplementary Tables 1 and 2). In addition, *L. oratoria* genome size, intergenic spacer length, and gene count differed significantly from those of other Hemiptera[50].

Control regions of *L. oratoria* are situated between clusters of *rrnS* and the *trnI–trnQ–trnM* genes at the conserved position (Figure 1 and Supplementary Figure 1). *L. oratoria* control region spans 3035 bp and has a high AT content (75%) within the range of other completely sequenced Hemipetran insects[50]. Both vertebrate and invertebrate mitochondrial genomes have a high AT content in the control region, associated with replication initiation[13]. The complete length of control region varies from 224 bp (Largidae) to 3035 bp (Alydidae, in the present study). Thus, non-coding regions are the main reason for the variation observed in the mitogenome size of insects[47]. The repeated region of the mitogenome contained three different types of tandem repeats. They are two repeats of 22 bp between 536 and 578 bp; four repeats of the longest type 475 bp between 998 and 2809 bp; the fourth repeat region is partial, and two repeats of 20 bp positioned at 2842–2280 bp (Figure 2 *a*). Approximately 39 bp putative stem–loop structure was found in the regulatory area, and no conserved functional motifs were found (Figure 2 *b*). Variable tandem repeats in control region have also been described previously, such as more long tandem repeats (LTRs) in *Dinorhynchus dybowsky* Jakovlev, 1876 (ref. 57) and fewer LTRs in various species of Hemiptera[58,59].

Gene evolutionary rate and nucleotide diversity

For all PCGs, the rate and ratio of non-synonymous (K_a) and synonymous substitutions (K_s) were determined (Supplementary Table 3). The fastest-evolving genes were *nad5* and *nad6*, followed by *nad2*, *nad4*, *nad4L* and *atp8*. The slowest rate was shown by *cox1*, which is appropriate given that it is a universal barcode marker for insect species[41,50,54]. The *Cox3* region gene can also be considered a candidate gene for the *L. oratoria* barcode indicator due to its slow evolutionary rate. However, due to their higher divergence rates, complete *nad* genes and *atp6* may be effective markers to ascertain intraspecific relationships. These genes can be used for population differentiation within the species. These observations are in line with earlier findings in other insects[60]. All PCGs, except *nad2*, *nad5* and *nad6* had K_a/K_s and JK_a/JK_s ratio <1, demonstrating that genes were developing under purifying selection and that all (except *nad2*, *nad5* and *nad6*) could be pooled to explore Coreoidea phylogeny.

The sliding window study revealed highly variable nucleotide diversity (ND; P_i) and substantial disparity between the 13 aligned PCGs of 16 Coreoidea mitogenomes. While *cox1*, *nad1*, *cox3* and *cytb* genes displayed substantially lower ND of 0.153, 0.165, 0.173 and 0.184 respectively, the genes *nad2*, *nad5*, *nad6* and *nad4* displayed relatively higher ND of 0.512, 0.466, 0.439 and 0.345 respectively. Additionally, pairwise genetic distance examinations provided the same results, with more considerable distances for the genes *nad2*, *nad5*, *nad6* and *nad4* being 1.101, 1.029, 0.729 and 0.521 respectively, and lower distances for the genes *cox1*, *nad1*, *cox3* and *cytb* being 0.173, 0.188, 0.199 and 0.214 respectively.

Base composition and skewness of mitochondrial genome of Coreoidea

The majority strand (J-strand) composition of the mitochondrial genome from *L. oratoria* was made up of the following bases: A = 7868 bp (44.7%), T = 5070 bp (28.8%), G = 1856 bp (10.6%) and *C* = 2701 bp (15.4%). The alignments of nucleotides of the *L. oratoria* mitochondrial genomes were strongly biased towards A and T (73.6%), like most hemipteran species[61]. Significant variation in the base composition of *L. oratoria* was observed compared to the studied Coreoidea samples. The base composition of the present study samples varied significantly, averaging 74.02%, ranging from 71.5% in *L. membranaceus* (MF497724) to 76.6% in *R. pedestris* (EU427344). The AT skew in the present study was 0.2163. Earlier studies have also claimed that the hemipteran AT skew ranged from –0.1812 (*Trialeurodes vaporariorum*) to 0.2765 (*Lycorma delicatula*)[24]. Likewise, the present sample GC skews matched with other hemipteran species ranging from –0.2827 (*Acyrthosiphon pisum*) to 0.2086 (*Neomaskellia andropogonis*). These results imply that the base composition and base skew values of Coreoidea mitogenomes are comparable to other hemipteran insects[41,61,62].

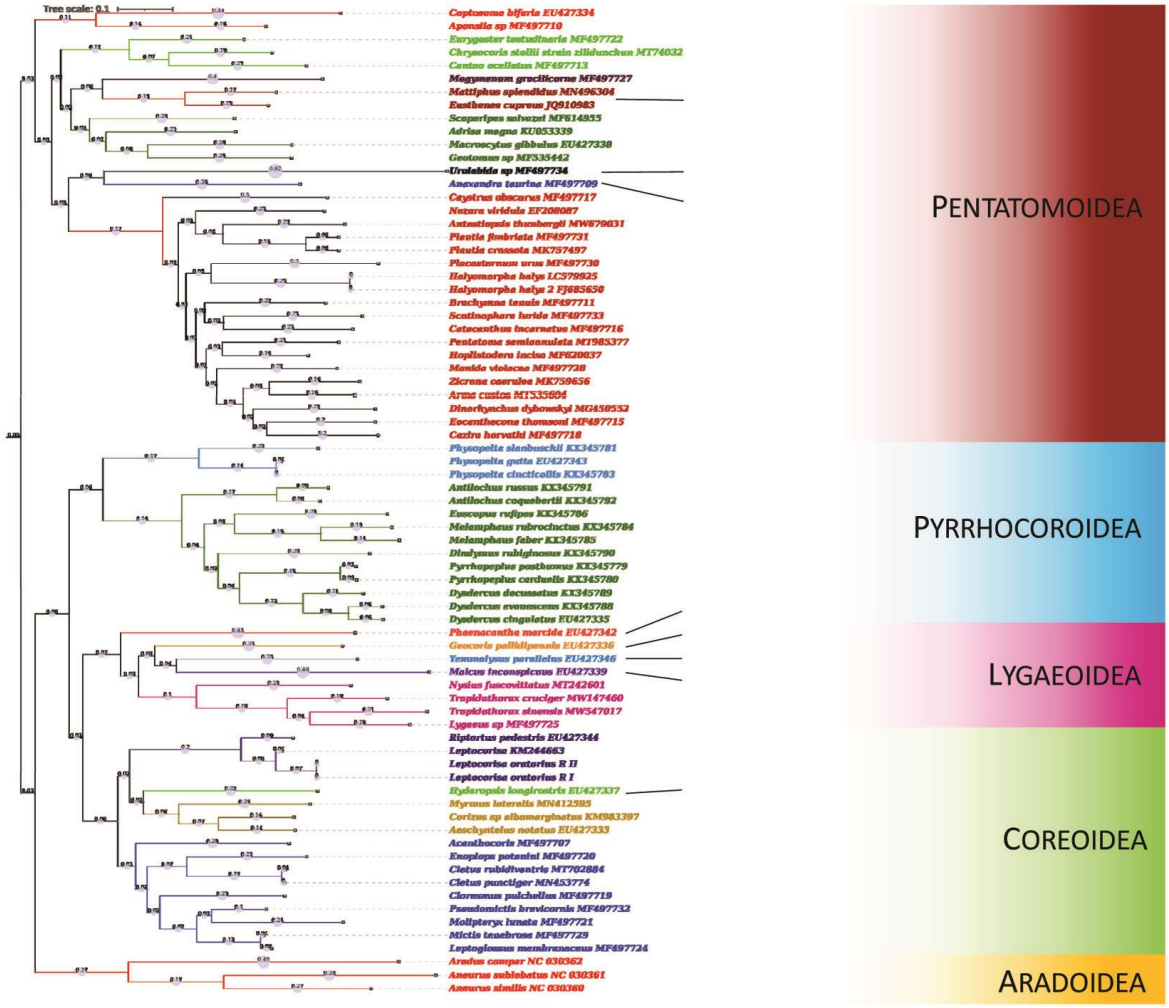

Figure 3. Phylogenetic tree of 75 species from Pentatomomorpha infraorder including two numbers of *L. oratoria* from the present study. The tree was obtained from ML analysis based on concatenated data of 13 protein-coding genes (PCGs). The probabilities of ML bootstrap values are indicated by the numbers at the nodes.

Phylogenetic relationship

Previously, phylogenetic relationship of insects has been mainly inferred based on partial mitochondrial gene sequences[41]. The present study used maximum likelihood (ML) and Bayesian inference (BI) methods, a model evolutionary technique based on a sequencing dataset containing 13 PCGs and 123 nucleotide sequences. Previous reports have suggested that PCGs show a higher significant positive correlation than ribosomal genes during a comparison of AT content of all PCGs and the *rrnS* and *rrnL* genes with the complete mitogenome[41]. As a result, PCGs appear to properly represent the overall evolution of mitochondrial genomes compared to rRNA genes. Likewise, a recent phylogenomic study of bugs suggests using Aradoidea as the basal group for most lineages of Pentatomomorpha; hence we used Aradoidea as the outgroup[62]. Phylogenetic analysis based on ML indicated that the Pentatomoidea superfamily (Pentatomidae, Cydnidae, Tessaratomidae, Scutelleridae, Acanthosomatidae, Plataspidae, Urostylididae) formed a separate clade with all other superfamilies (Pyrrhocoroidea, Coreoidea, Lygaeoidae). Furthermore, family-level phylogenetic analysis yielded two topologies in Pentatomoidea and Pyrrhocoroidea + Coreoidea + Lygaeoidae (Figure 3). Pentatomoidea was rooted with Plataspidae in one clade, while the remaining families (Pentatomidae, Cydnidae, Tessaratomidae, Scutelleridae, Acanthosomatidae, Urostylididae) formed another clade. Similarly, Largidae and Pyrrhocoridae formed a single clade that could be differentiated from the other clade consisting of Rhopalidae, Coreidae, Alydiade, Lygaeidae, Geocoridae, Malcidae, Berytidae and Colobathristidae. Coreidae is recovered as the sister group of Alydidae under phylogenetic analysis. Earlier reports also support our phylogenetic analysis[62,63]. Similarly, phylogenetic analysis based on BI method of all 75 mitogenomes of Pentatomomorpha infraorder was performed. Coreidae were differentiated from the other 15 families of Pentatomomorpha infraorder (Figure 4). Similar to

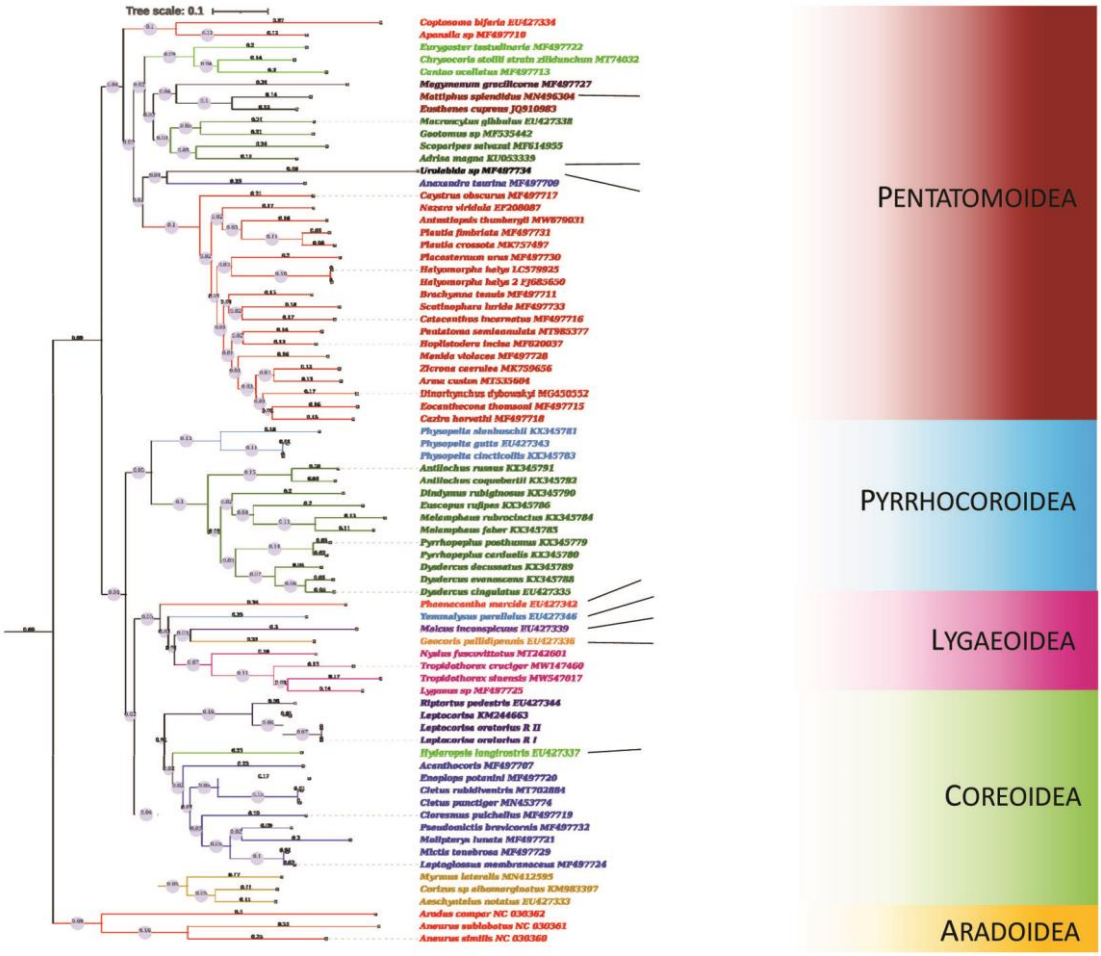

Figure 4. Phylogenetic tree of 75 species from Pentatomomorpha infraorder, including two numbers of *L. oratoria* from the present study. The tree was obtained from Bayesian analysis based on concatenated data of 13 PCGs. The probability of Bayesian bootstrap values is indicated by numbers at the nodes.

ML analysis, *L. oratoria* formed a separate subclade from the previously reported *Leptocorisa* sp. (KM244663) in Bayesian analysis.

Conclusion

This study reports a mitogenome of *L. oratoria* from India, with 37 distinctive metazoan mitochondrial genes and following an organizational pattern like many other hemipteran mitogenomes. Compared to other hemipterans and insects, *L. oratoria* also has different types of tandem repeats for which functional characterization should be established in future to determine their crucial role in this species. Moreover, phylogeny reconstruction based on 75 mitogenomes of the Pentomomorpha infraorder provides a good understanding of the evolution of this species, which will be useful for understanding another homo/heteroptera species. Future research on population demography and genetics, individual variations and phylogeographics of hemipteran insects will benefit from the findings of the present study. A phylogenetic reconstruction based on more *Leptocorisa* taxa is required to understand Alydidae evolution better.

1. Rao, J. and Prakash, A., Bio-deterioration of paddy seed quality due to insects and mites and its control using botanicals. Final report of ICAR Ad-hoch Scheme, Central Rice Research Institute, Cuttack, India, 1995.
2. Aktera, U. S., Islam, K. S., Jahan, M., Rahman, M. S., Talukder, F. U. and Hasan, M. A., Extent of damage of rice bug (*Leptocorisa acuta*) and its control with insecticides. *Acta Sci. Malays.*, 2020, **4**(2), 82–87; doi:10.26480/asm.02.2020.82.87.
3. Rai, A. B., Singh, J. and Rai, L., Evaluation of gundhi bug, *Leptocorisavaricornis* (F.) damage in rice. In International Symposium on Rice Research, Hyderabad, 1990.
4. Gupta, K. and Kumar, A., Field efficacy of certain insecticides against rice gundhi bug (*Leptocorisa acuta* (Thonberg)) under agro-climatic condition of Allahabad, India. *Int. J. Curr. Microbiol. Appl. Sci*., 2017, **6**(8), 343–345.
5. Nugaliyadde, L., Dissanayake, N., Mitrasena, J. and Wijesundera, D. S., Advances of pest and disease management of rice in Sri Lanka: a review. In Annual Symposium of the Department of Agriculture, Sri Lanka, 2000, vol. 2, pp. 409–422.
6. Jahn, G. C., Domingo, I., Liberty, M., Almazan, P. and Pacia, J., Effect of rice bug *Leptocoris aoratorius* (Hemiptera: Alydidae) on

rice yield, grain quality and seed viability. *J. Econ. Entomol.*, 2004, **97**(6), 1923–1927; https://doi.org/10.1093/jee/97.6.1923.
7. Lessinger, A. C. *et al.*, The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae). *Insect Mol. Biol.*, 2000, **9**(5), 521–529; https://doi.org/10.1046/j.1365-2583.2000.00215.x.
8. Lewis, J. A., Huq, A., Liu, W. and Jacob, A., Induction of gene expression by intracellular interferon-γ: abrogation of the species specificity barrier. *Virology*, 1995, **212**(2), 438–450; https://doi.org/10.1006/viro.1995.1501.
9. Zhang, D. X., Szymura, J. M. and Hewitt, G. M., Evolution and structural conservation of the control region of insect mitochondrial DNA. *J. Mol. Evol*., 1995, **40**(4), 382–391; https://doi.org/10.1007/BF00164024.
10. Shao, R., Campbell, N. J. and Barker, S. C., Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthiraptera). *Mol. Biol. Evol.*, 2001, **18**(5), 858–865; https://doi.org/10.1093/oxfordjournals.molbev.a003867.
11. Choudhary, J. S., Naaz, N., Prabhakar, C. S., Rao, M. S. and Das, B., The mitochondrial genome of the peach fruit fly, *Bactrocera zonata* (Saunders) (Diptera: Tephritidae): Complete DNA sequence, genome organization, and phylogenetic analysis with other tephritids using next generation DNA sequencing. *Gene*, 2015, **569**(2), 191–202; https://doi.org/10.1016/j.gene.2015.05.066.
12. Taanman, J. W., The mitochondrial genome: structure, transcription, translation and replication. *Biochim. Biophys. Acta*, 1999, **1410**(2), 103–123; https://doi.org/10.1016/S0005-2728(98)00161-3.
13. Boore, J. L., Animal mitochondrial genomes. *Nucleic Acids Res*., 1999, **27**, 1767–1780; doi:10.1093/nar/27.8.1767.
14. Cameron, S. L., Beckenbach, A. T., Dowton, M. P. and Whiting, M. F., Evidence from mitochondrial genomics on interordinal relationships in insects. *Arthropod Syst. Phylogen*., 2006, **64**(1), 27–34.
15. Ashlock, P. D. and Slater, A., Family Lygaeidae Schilling, 1829 (= Infericornes Amyot and Serville, 1843; Myodochidae Kirkaldy, 1899; Geocoridae Kirkaldy, 1902): the seed bugs and chinch bugs. In *Catalog of the Heteroptera, or True Bugs of Canada and the Continental United States*, CRC Press, Boca Raton, Florida, 2019, pp. 167–245.
16. Schuh, R. T. and Slater, J. A., *True Bugs of the Eorld (Hemiptera: Heteroptera): Classification and Natural History*, Cornell University Press, Ithaca, New York, 1995, pp. 609–610.
17. Chilana, P., Sharma, A. and Rai, A., Insect genomic resources: status, availability and future. *Curr. Sci*., 2012, **102**(4), 571–580.
18. Ribeiro, F. J. *et al.*, Finished bacterial genomes from shotgun sequence data. *Genome Res*., 2012, **22**(11), 2270–2277; http://www.genome.org/cgi/doi/10.1101/gr.141515.112.
19. Kirkness, E. F. *et al.*, Genome sequences of the human body louse and its primary endosymbiont provide insights into the permanent parasitic lifestyle. *Proc. Natl. Acad. Sci. USA*, 2010, **107**(27), 12168–12173; https://doi.org/10.1073/pnas.1003379107.
20. Knaus, B. J., Cronn, R., Liston, A., Pilgrim, K. and Schwartz, M. K., Mitochondrial genome sequences illuminate maternal lineages of conservation concern in a rare carnivore. *BMC Ecol*., 2011, **11**(1), 1–4; https://doi.org/10.1186/1472-6785-11-10.
21. Ma, P. F., Guo, Z. H. and Li, D. Z., Rapid sequencing of the bamboo mitochondrial genome using Illumina technology and parallel episodic evolution of organelle genomes in grasses. *PLoS ONE*, 2012, **7**(1), e30297; https://doi.org/10.1371/journal.pone.0030297.
22. Coates, B. S., Assembly and annotation of full mitochondrial genomes for the corn rootworm species, *Diabrotica virgifera virgifera* and *Diabrotica barberi* (Insecta : Coleoptera : Chrysomelidae), using next generation sequence data. *Gene*, 2014, **542**(2), 190–197; https://doi.org/10.1016/j.gene.2014.03.035.
23. Barrion, A. T. and Litsinger, J. A., Dichogaster nr. Curgensis Michaelsen (Annelida : Octochaetidae): an earthworm pest of terraced rice in the Philippine Cordilleras. *Crop Prot*., 1997, **16**(1), 89–93; https://doi.org/10.1016/S0261-2194(96)00058-0.
24. Govindharaj, G. P. *et al.*, Genome organization and comparative evolutionary mitochondriomics of brown planthopper, *Nilaparvata lugens* biotype 4 using next generation sequencing (NGS). *Life*, 2022, **12**(9), 1289; https://doi.org/10.3390/life12091289.
25. Bernt, M. *et al.*, MITOS: improved *de novo* metazoan mitochondrial genome annotation. *Mol. Phylogenet. Evol*., 2013, **69**(2), 313–319; https://doi.org/10.1016/j.ympev.2012.08.023.
26. Lowe, T. M. and Eddy, S. R., tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.*, 1997, **25**(5), 955–964; https://doi.org/10.1093/nar/25.5.955.
27. Grant, J. R. and Stothard, P., The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Res*., 2008, **36**(2), W181–W184; https://doi.org/10.1093/nar/gkn179.
28. Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S., MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.*, 2013, **30**(12), 2725–2729; https://doi.org/10.1093/molbev/mst197.
29. Choudhary, J. S., Naaz, N., Lemtur, M., Das, B., Singh, A. K., Bhatt, B. P. and Prabhakar, C. S., Genetic analysis of Bactrocer-azonata (Diptera : Tephritidae) populations from India based on *cox1* and *nad1* gene sequences. *Mitochondrial DNA Part A*, 2018, **29**(5), 727–736; https://doi.org/10.1080/24701394.2017.1350952.
30. Perna, N. T. and Kocher, T. D., Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J. Mol. Evol*., 1995, **41**(3), 353–358; https://doi.org/10.1007/BF0018-6547.
31. Librado, P. and Rozas, J., DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 2009, **25**(11), 1451–1452; https://doi.org/10.1093/bioinformatics/btp187.
32. Katoh, K. and Standley, D. M., MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol*., 2013, **30**(4), 772–780; https://doi.org/10.1093/molbev/mst010.
33. Talavera, G. and Castresana, J., Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.*, 2007, **56**(4), 564–577; https://doi.org/10.1080/10635150701472164.
34. Bu, R. *et al.*, Tillage and straw-returning practices effect on soil dissolved organic matter, aggregate fraction and bacteria community under rice-rice-rapeseed rotation system. *Agric. Ecosyst. Environ.*, 2020, **287**, 106681; https://doi.org/10.1016/j.agee.2019.106681.
35. Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T. and Calcott, B., PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Mol. Biol. Evol*., 2017, **34**(3), 772–773; https://doi.org/10.1093/molbev/msw260.
36. Guindon, S. and Gascuel, O., A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol*., 2003, **52**(5), 696–704; https://doi.org/10.1080/10635150390235520.
37. Huelsenbeck, J. P. and Ronquist, F., MRBAYES: bayesian inference of phylogenetic trees. *Bioinformatics*, 2001, **17**(8), 754–755.
38. Letunic, I. and Bork, P., Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res*., 2019, **47**(W1), W256–W259; https://doi.org/10.1093/nar/gkz239.
39. Crease, T. J., The complete sequence of the mitochondrial genome of *Daphnia pulex* (Cladocera: Crustacea). *Gene*, 1999, **233**(1–2), 89–99; https://doi.org/10.1016/S0378-1119(99)00151-1.
40. Yuan, M. L., Zhang, Q. L., Guo, Z. L., Wang, J. and Shen, Y. Y., The complete mitochondrial genome of *Corizus tetraspilus* (Hemiptera: Rhopalidae) and phylogenetic analysis of Pentatomomorpha. *PLoS ONE*, 2015, **10**(6), e0129003; https://doi.org/10.1371/journal.pone.0129003.
41. Zhang, K. J. *et al.*, The complete mitochondrial genomes of two rice planthoppers, *Nilaparvata lugens* and *Laodelphax striatellus*: conserved genome rearrangement in Delphacidae and discovery of new characteristics of atp8 and tRNA genes. *BMC Genomics*, 2013, **14**(1), 1–2; https://doi.org/10.1186/1471-2164-14-417.

42. Anant, A. K. *et al.*, Genetic dissection and identification of candidate genes for brown planthopper, *Nilaparvata lugens* (Delphacidae: Hemiptera) resistance in farmers' varieties of rice in Odisha. *Crop Prot.*, 2021, **144**, 105600; https://doi.org/10.1016/j.cropro.2021.105600.
43. Cha, S. Y. *et al.*, The complete nucleotide sequence and gene organization of the mitochondrial genome of the bumblebee, *Bombus ignitus* (Hymenoptera: Apidae). *Gene*, 2007, **392**(1–2), 206–220; https://doi.org/10.1016/j.gene.2006.12.031.
44. Jiang, S. T., Hong, G. Y., Yu, M., Li, N., Yang, Y., Liu, Y. Q. and Wei, Z. J., Characterization of the complete mitochondrial genome of the giant silkworm moth, *Eriogyna pyretorum* (Lepidoptera: Saturniidae). *Int. J. Biol. Sci*., 2009, **5**(4), 351; doi:10.7150/ijbs.5.351.
45. Song, N. and Liang, A. P., Complete mitochondrial genome of the small brown planthopper, *Laodelphax striatellus* (Delphacidae: Hemiptera), with a novel gene order. *Zool. Sci.*, 2009, **26**(12), 851–860; https://doi.org/10.2108/zsj.26.851.
46. Chen, M. M. *et al.*, Complete mitochondrial genome of the Atlas moth, *Attacus atlas* (Lepidoptera: Saturniidae) and the phylogenetic relationship of Saturniidae species. *Gene*, 2014, **545**(1), 95–101; https://doi.org/10.1016/j.gene.2014.05.002.
47. Hua, J., Li, M., Dong, P., Cui, Y., Xie, Q. and Bu, W., Comparative and phylogenomic studies on the mitochondrial genomes of Pentatomomorpha (Insecta: Hemiptera: Heteroptera). *BMC Genomics*, 2008, **9**(1), 1–5; https://doi.org/10.1186/1471-2164-9-610.
48. Hou, W. R. *et al.*, A complete mitochondrial genome sequence of Asian black bear *Sichuan subspecies* (*Ursus thibetanus mupinensis*). *Int. J. Biol. Sci*., 2007, **3**(2), 85; doi:10.7150/ijbs.3.85.
49. Hong, G., Jiang, S., Yu, M., Yang, Y., Li, F., Xue, F. and Wei, Z., The complete nucleotide sequence of the mitochondrial genome of the cabbage butterfly, *Artogeia melete* (Lepidoptera: Pieridae). *Acta Biochim. Biophys. Sin.*, 2009, **41**(6), 446–455; https://doi.org/10.1093/abbs/gmp030.
50. Lv, L., Peng, X., Jing, S., Liu, B., Zhu, L. and He, G., Intraspecific and interspecific variations in the mitochondrial genomes of *Nilaparvata* (Hemiptera: Delphacidae). *J. Econ. Entomol.*, 2015, **108**(4), 2021–2029; https://doi.org/10.1093/jee/tov122.
51. Thao, M. L., Baumann, L. and Baumann, P., Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha). *BMC Evol. Biol.*, 2004, **4**(1), 1–3; https://doi.org/10.1186/1471-2148-4-25.
52. Zhu, Y. J., Zhou, G. L., Fang, R., Ye, J. and Yi, J. P., The complete sequence determination and analysis of *Lymantria dispar* (Lepidoptera: Lymantriidae) mitochondrial genome. *Plant Quarantine*, 2010, **24**(4), 6–11.
53. Valero, M. C., Ojo, J. A., Sun, W., Tamò, M., Coates, B. S. and Pittendrigh, B. R., The complete mitochondrial genome of *Anoplocnemis curvipes* F. (Coreinea, Coreidae, Heteroptera), a pest of fresh cowpea pods. *Mitochondrial DNA, Part B*, 2017, **2**(2), 421–423; https://doi.org/10.1080/23802359.2017.1347829.
54. Huang, Y. X. and Qin, D. Z., First mitogenome for the tribe Saccharosydnini (Hemiptera: Delphacidae: Delphacinae) and the phylogeny of three predominant rice planthoppers. *Eur. J. Entomol*., 2018, **30**, 115.
55. Ohtsuki, T., Kawai, G. and Watanabe, K., The minimal tRNA: unique structure of *Ascaris suum* mitochondrial tRNASerUCU having a short T arm and lacking the entire D arm. *FEBS Lett.*, 2002, **514**(1), 37–43; https://doi.org/10.1016/S0014-5793(02)02328-1.
56. Sheffield, N. C., Song, H., Cameron, S. L. and Whiting, M. F., Nonstationary evolution and compositional heterogeneity in beetle mitochondrial phylogenomics. *Syst. Biol.*, 2009, **58**(4), 381–394; https://doi.org/10.1093/sysbio/syp037.
57. Zhao, Q., Wang, J., Wang, M. Q., Cai, B., Zhang, H. F. and Wei, J. F., Complete mitochondrial genome of *Dinorhynchus dybowskyi* (Hemiptera: Pentatomidae: Asopinae) and phylogenetic analysis of Pentatomomorpha species. *J. Insect Sci*., 2018, **18**(2), 44; https://doi.org/10.1093/jisesa/iey031.
58. Lee, W., Kang, J., Jung, C., Hoelmer, K., Lee, S. H. and Lee, S., Complete mitochondrial genome of brown marmorated stink bug *Halyomorpha halys* (Hemiptera: Pentatomidae), and phylogenetic relationships of hemipteran suborders. *Mol. Cells*, 2009, **28**(3), 155–165; https://doi.org/10.1007/s10059-009-0125-9.
59. Li, H., Liu, H., Shi, A., Štys, P., Zhou, X. and Cai, W., The complete mitochondrial genome and novel gene arrangement of the unique-headed bug *Stenopirates* sp. (Hemiptera: Enicocephalidae). *PLoS ONE*, 2012, **7**(1), e29419; https://doi.org/10.1371/journal.pone.0029419.
60. Castellana, S., Vicario, S. and Saccone, C., Evolutionary patterns of the mitochondrial genome in Metazoa: exploring the role of mutation and selection in mitochondrial protein-coding genes. *Genome Biol. Evol.*, 2011, **3**, 1067–1079; https://doi.org/10.1093/gbe/evr040.
61. Wang, Y., Chen, J., Jiang, L. Y. and Qiao, G. X., Hemipteran mitochondrial genomes: features, structures and implications for phylogeny. *Int. J. Mol. Sci.*, 2015, **16**(6), 12382–12404; https://doi.org/10.3390/ijms160612382.
62. Zhao, L., Wei, J., Zhao, W., Chen, C., Gao, X. and Zhao, Q., The complete mitochondrial genome of *Pentatoma rufipes* (Hemiptera, Pentatomidae) and its phylogenetic implications. *ZooKeys*, 2021, **1042**, 51; doi:10.3897/zookeys.1042.62302.
63. Johnson, K. P. *et al.*, Phylogenomics and the evolution of hemipteroid insects. *Proc. Natl. Acad. Sci. USA*, 2018, **115**(50), 12775–12780; https://doi.org/10.1073/pnas.1815820115.

ACKNOWLEDGEMENTS. We thank the Director, ICAR-National Rice Research Institute, Cuttack for support. This work was funded through ICAR-National Rice Research Institute in-house project no. 3.6.

Received 2 December 2022; revised accepted 22 April 2023

doi: 10.18520/cs/v125/i4/407-415