Gene Rearrangements in the Mitochondrial Genomes of Whiteflies (Hemiptera: Aleyrodinae): Plesiomorphies, Synapomorphies and Autapomorphies

Avas Pakrashi
Zoological Survey of India
VIKAS KUMAR
Zoological Survey of India
Dhirti Banerjee
Zoological Survey of India
Kaomud Tyagi (kumud.tyagi5@gmail.com)
Zoological Survey of India https://orcid.org/0000-0003-1064-9826
C. M. Kalleshwaraswamy
College of Agriculture Shimoga: University of Agricultural and Horticultural Sciences

Research Article

Keywords: Aleurodicus rugioperculatus, Aleurodicinae, Mitogenome, whitefly, gene rearrangement

Posted Date: September 28th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-906638/v1

Abstract

Mitochondrial genome rearrangements have been used for defining historical relationships, but there have been incidences of convergences at different taxonomic levels. Here, we sequenced complete mitogenome of *Aleurodicus rugioperculatus* (Aleyrodidae: Aleurodicinae) to examine gene rearrangements and phylogenetic relationships within the family Aleyrodidae. We identified five gene blocks (I-V) in the whitefly ancestor that are shared plesiomorphies retained in different whitefly lineages. Gene block I is conserved in all whiteflies except three species (*Tetraleurodes acaciae* and two *Bemisia* species). Conversely, we detected 83 derived gene boundaries within the family. Mapping these gene boundaries onto a phylogenetic tree revealed that 16 were symplesiomorphies for two subfamilies; 9 were synapomorphies at different taxonomic levels, and 28 autapomorphies for individual species. Bayesian Inference (BI) and Maximum Likelihood (ML) phylogenetic analyses yielded similar topologies supporting the monophyly of Aleyrodinae and Aleurodicinae. Exclusion of PCG third codon positions from phylogenetic analyses improved both node support and consistency with prior analyses. To understand the significance of gene order convergence on phylogeny of the whiteflies, more species-level data is required.

Introduction

The small, circular insect mitochondrial genomes (mitogenomes) are usually ranging from 15-18 kb in size [1, 2]. It usually contains 37 genes [13 protein-coding genes (PCGs), 22 transfer RNA genes (tRNAs) and two ribosomal RNA genes (rRNAs)]. Additionally, there is usually a non-coding region known as control region/s (CR) which contain essential regulatory elements responsible for transcription and replication [3]. Mitogenomes have widely used in phylogenetic studies, population genetics, comparative evolution, divergence time analysis [4,5], and to explore gene rearrangement patterns/ estimating of ancestral genome reconstruction [6-8].

Whiteflies belong to the family Aleyrodidae, (Suborder Sternorrhyncha, Order Hemiptera) [9]. The family Aleyrodidae is further classified into two subfamilies, Aleurodicinae and Aleyrodinae. Whiteflies are inconspicuous phytophagous insects, usually found on the lower surfaces of leaves. Whiteflies are an economically important group that infest a wide variety of plant families [10]. To date, 1556 species and 161 whitefly genera have been described from throughout the world [11]. India's whitefly diversity is nearly one quarter of global diversity with 406 species from 60 genera [12].

Aleurodicus rugioperculatus is commonly known as the rugose spiraling whitefly, and was originally described from Belize in 2004 as a pest on coconut [13]. It was subsequently reported as a pest on gumbo limbo (*Burera simaruba*) from Miami Dade County, in Florida (United States of America) [14]. This species was recently (2017) introduced to the states of Tamil Nadu and Kerala in India and has been collected from coconuts, mango and guava [15]. Subsequently it was also reported from other Indian states, Andhra Pradesh, Assam, Goa, West Bengal, Maharashtra, Telangana, Meghalaya and Gujarat, as pests of coconut, banana, sapota, maize, oil palm, mango, cashew and many ornamental plants [16-18].

Given its range of potential hosts and impact on Indian agriculture this pest needs accurate identification tools, including in-depth molecular data, is required to apply successful control strategies.

Currently, there are twelve complete or partial mitogenomes of whiteflies (family Aleyrodidae) available [19-28]. Out of the twelve available mitogenomes, two are from the subfamily Aleurodicinae (*Aleurodicus dispersus*, *Aleurodicus dugesii*) and the remaining 10 from subfamily Aleyrodinae (*Aleurochiton aceris*, *Aleurocanthus camelliae*, *Aleurocanthus spiniferus*, *Bemisia afer*, *Bemisia tabaci*, *Crenidorsum turpiniae*, *Neomaskellia andropogonis*, *Pealius machili*, *Trialeurodes vaporariorum*, and *Tetraleurodes acaciae*). Six of the available mitogenomes are partial due to the absence of tRNAs: *Ac. camelliae*, *Ac. spiniferus, Ach. aceris*, *Ad. dispersus*, *N. andropogonis*, and *Te. acaciae*.

Here, we sequenced and annotated the complete mitogenome of the rugose spiraling whitefly, *Ad. rugioperculatus* (Aleurodicinae). The objectives of the present study were to: i) characterize the complete mitogenome of *Ad. rugioperculatus*; ii) compare mitogenomes evolution across whiteflies to understand patterns of gene rearrangements and elucidate ancestral, shared and derived gene boundaries.

Material And Methods

Ethics statement

There is no specific permission required to collect the whitefly specimens.

Sample collection, and DNA Isolation

Whiteflies specimens were collected in September 2020 from the College of Agriculture Campus, Navile, Shivamogga district, Karnataka state of India (13°58'15.654"N, 75°34'47.952"E). Specimens were identified morphologically using available taxonomic keys by KS [13]. Specimens were preserved in absolute alcohol and stored at −80˚C in the Centre for DNA Taxonomy, Molecular Systematics Division, Zoological Survey of India, Kolkata. All the specimens were washed individually prior to extraction five to six times with PBS solution to eliminate the contamination. Genomic DNA was extracted from a single specimen using the DNeasy DNA Extraction kit (Qiagen, Valencia, CA). The dsDNA high-sensitivity kit was used for checking the DNA quantity by Qubit Fluorometer (Thermo Fisher Scientific, MA, U00SA).

Mitochondrial genome sequencing, assembly, and annotation

Whole genome sequencing was carried out on the Illumina platform (Illumina Hiseq 2500) using 2x150 paired end chemistry. Libraries were constructed using the TruSeq DNA Library Preparation kit. Around 50 GB data were generated and low-quality reads were removed by NGS-Toolkit. The mitogenome was assembled from whole genome reads data using GetOrganelle (ver. 1.7.4). The location of protein coding genes (PCGs), transfer RNAs (tRNAs), and ribosomal RNAs (rRNAs) were estimated using MITOS web-server (http://mitos.bioinf.uni-leipzig.de/index.py) [29] (Table 1). Subsequently BLASTn, BLASTp, and ORF Finder (https://www.ncbi.nlm.nih.gov/orffinder/) were used to confirm gene boundaries for the PCGs

and rRNAs. tRNA secondary structures were predicted in MITOS [29], and ARWEN 1.2 [30]. Control region secondary structures were predicted in mFold [31], implemented in the UNAFold web server using. The data for the newly sequenced species were deposited in GenBank (accession numbers in Table 2).

Genome visualization, and comparative analysis

CGview was used to visualize the complete mitogenome of *Ad. rugioperculatus*. Nucleotide composition, Relative Synonymous Codon Usage (RSCU), and AT- GC skew statistics were calculated in MEGAX [32]. Nucleotide skew was calculated as AT skew = (A−T) / (A+T) and GC skew = (G−C) / (G+C) [33]. ENC values were calculated in DnaSP6.0 [34]. Pairwise comparisons of gene order, and the evolutionary sequence of rearrangements within whiteflies were evaluated by CREx (Common Interval Rearrangement Explorer) [35] while the reconstruction of the ancestral gene orders at each tree node was performed using TreeREx [36].

Phylogenetic analysis

A phylogeny of whiteflies was constructed using the PCGs from 13 Aleyrodidae mitogenomes, (one newly sequenced and 12 previously published) with *Sitobion avenae* (Aphididae) used as an outgroup [37]. Each PCGs was individually aligned using MAFFT as implemented in the TranslatorX [38]. Poorly aligned regions were masked using GBlocks (as implemented in TranslatorX) with default settings. Datasets were concatenated using SequenceMatrix v1.7.845 [39]. Four datasets were used to assess the utility of datasets and the impact of substitution rate on the phylogeny: (1) 13 PCGs-GBlock+all codon positions (10955bp), (2) 13 PCGs+GBlock+all codon positions (8532bp), (3) 13 PCGs-GBlock+3rd codon position excluded (7304bp), (4) 13 PCGs+GBlock+3rd codon position excluded (5688bp). PartitionFinder version 2.1.1 [40] was used to identify best partition and substitution models with the 'greedy' algorithm based on the following predefined partitions: codon positions for each PCG (39 partitions for datasets 1 and 3, 26 for datasets 2 and 4) (Table S1). Phylogenetic trees were inferred with maximum likelihood (ML) implemented in IQTree [41], and Bayesian inference (BI) implemented in MrBayes 3.2 [42]. For Maximum likelihood analysis, bootstrap analysis of 1,000 replicates was performed in IQ tree ver.1.6.11 [41]. For BI analysis, two simultaneous runs of 1000000 generations using the best fit models. The BI analysis was stopped after reaching stationary phase with average standard deviation of split frequencies below 0.01. Trees were sampled in every 100 generations, with the first 25% discarded as burn in. Phylogenetic trees were visualized and edited using FigTree v.1.4.4 [43] (https://tree.bio.ed.ac.uk/software/figtree/).

Results And Discussion

1) Characterization of complete mitogenome of *Aleurodicus rugioperculatus*

The complete mitogenome of *Aleurodicus rugioperculatus* (GenBank accession no. MW649000) is a 15,060 bp circular DNA molecule (Fig 1, Table 1). It contains 37 genes, including 13 PCGs, two rRNAs, 22 tRNAs and a non-coding control region (CR). Out of the 37 genes, 23 were located on the majority strand

and 14 on minority strand. A+T content was 86.5% (38.6% A + 47.9% T) and G+C content 13.5% (7.9%G + 5.6% C). A+T content was lowest in the control region (75.23%) followed by PCGs (85.21 %), rRNAs (88.1%) and tRNAs (88.24%) (Table S1).

The mitogenome of *Ad. rugioperculatus* includes 13 PCGs (total length of 10,840 bp) with negative AT skew (0.141) and positive GC skew (0.226). All PCGs used ATN start codons (ATA for *nad1*, *nad2*, *nad4* and *nad6*; ATG for *cox1*, *cox2*, *cox3*, *atp6, nad4L* and *cytb*; and ATT for *nad3*, *nad5* and *atp8*) (Table 1). Most PCGs have TAA stop codons except *cox1, cox2* and *nad1* which have incomplete termination codons (T). Incomplete termination codons are observed in many other insect mitogenomes and are completed by post-transcriptional polyadenylation [44].

The average A+T content across PCGs was 85.21%, highest in *nad6* (92.72%) and lowest in *cox1* (77.80%) (Table S1). Base composition at each codon position (Ist, IInd, IIrd) of the concatenated PCGs was 77.73%, 72.6%, and 95.2% respectively. RSCU analysis of the 3,613 codons in *Ad. rugioperculatus* revealed that TTT(Phenylalanine), AAT(Asparagine), TAT(Tyrosine), ATT(Isoleucine), TTA(Leucine), and AAA (Lysine) were the most frequently used codons. RSCU also indicated that almost all frequently used codons ended with either A or T.

The large ribosomal RNA (*rrnL*) gene is located at the conserved position between *trnL1* and *trnV* and is 1,286 bp in length (88.26 % A+T content). The small ribosomal RNA (*rrnS*) gene was located at the conserved position between *trnV* and *trnQ* and is 779 bp in length (87.93% A+T content). rRNA secondary structures were compared to that of *Drosophila melanogaster* [45,46]. *rrnL* contains six domains with 47 helices (domain III is absent in *Ad. rugioperculatus* as is found in other arthropod mitogenomes) (Fig 2), whereas *rrnS* contains three domains with 31 helices (Fig 3).

Ad. rugioperculatus contains 22 tRNAs with a total length of 1,420 bp (range 57 to 70). Fourteen tRNA genes were coded on the majority strand and eight on the minority strand. The A+T content of tRNAs was 88.24% with negative AT skew (0.031) and positive GC skew (0.006). Most tRNAs showed the typical cloverleaf secondary structure, except *trnS1* and *trnS2* where the DHU stem and loop was absent (Fig. S1).

A total of 11 intergenic spacer regions, totaling 99 bp, and varying from 1 to 37 bp in length were detected. The longest intergenic spacer (37 bp) was between *nad6* and *trnP.* A total 15 overlapping regions, covering 179 bp (ranging from 1 to 87 bp in length) were detected (Table 1).

The control region (CR) is located between *trnQ* and *trnI,* and is 805 bp in length. It consists of i) a 161 bp A+T rich region along with poly T-stretch; ii) a stem/loop structure flanking TATA motif and GAAT motifs; iii) a 221 bp A+T rich region; iv) two tandem repeats (R1, 120 bp); and v) a terminal 28 bp A-rich region. The A+T content was 93.54 % with positive AT skew (0.19) and GC skew (0.15) (Fig. 4 A).

Comparative studies

A comparative study was carried out using the 13 mitogenomes of whiteflies (Table 2) including the newly generated species to explore the codon usage bias, gene rearrangement and phylogeny.

Codon usage bias

Codon usage bias in the 13 whitefly mitogenomes was compared to the effective number of codons (ENC) (Fig. 4 B). The ENC value ranged from 29.35 to 56.02, with an average of 42.96. ENC values in the subfamily Aleyrodinae were significantly higher (t-test, $P < 0.05$) than those of Aleurodicinae, ranging from 34.31 to 56.02 and from 29.35 to 33.93, respectively. That more codons were used by members of the Aleyrodinae in comparison to Aleurodicinae, suggests weak selection constraints on mitogenomes of Aleyrodinae.

The neutrality analysis revealed that GC3 content is higher in subfamily Aleyrodinae than in Aleurodicinae ranging from 8.6-40.3 and 4.2-9.9, respectively (Table S3). A significant positive correlation between GC12 and GC3 was observed in both subfamilies ($Y= 0.355x+23.96$, $R^2 = 0.791$) (Fig. 4C) suggesting directional mutation pressure on all codon positions. The slope of linear regression was less than 1 in whiteflies, implying that mutational pressure played a minor role in their mitogenome evolution. Other factors like natural selection, and translational bias accounted for 65% of observed codon usage bias.

Gene rearrangement

Mitogenome rearrangements can be explained by transposition, inversion, inverse transposition of individual genes or gene blocks. Gene rearrangement in *Ad. rugioperculatus* was analysed by comparing common gene intervals to those of the ancestral insect's inferred gene order in CREx. CREx analysis identified the transpositions of two tRNAs (*trnQ* and *trnY*) away from the ancestral gene blocks *I-Q-M* and *W-C-Y* (Table S4).

Comparison of gene order across the 13 available whiteflies was conducted. Gene order in 6 whitefly species is partial due one or more missing tRNAs. The genes *trnA*, *trnN*, *trnR* were not detected in *N. andropogonis*; *trnI* in *Ac. camelliae* and *Ach. aceris*; *trnS1* in *Ac. spiniferus* and *Te. acaciae*; *trnS2*, *trnQ* in *Ad. dispersus*. We mapped gene boundaries with respect to the inferred ancestral insect gene order to identify ancestral vs derived, and unique vs shared gene boundaries (Fig. 5). A total of 83 gene boundaries were detected, of which 37 were ancestral and 46 derived boundaries (Table S5). The ancestral 37 gene boundaries were grouped into 5 gene blocks (I-V) that were conserved in different whitefly lineages. Gene block I (*cox1-trnL2-cox2-trnK-trnD-atp6-atp8*) was conserved in all taxa except *Te. acaciae*, *B. afer* and *B. tabaci*. *trnD* was translocated from gene block I in these species resulting in the derived gene boundaries *trnR-trnD* (55) and *trnD-trnI* (56) in *Te. acaciae*; and *trnV-trnD* (61) and *trnD-trnQ* (62) in *Bemisia* species. Gene block II (*atp6-cox3-trnG-nad3-trnA-trnR-trnN-trnS1-trnE*) is conserved in only two species of the subfamily Aleurodicinae (*Ad. rugioperculatus* and *Ad. dugesii*). Gene block II is the most variable block across the remaining taxa. A partial set (*cox3-trnG-nad3-trnA-trnR)* of gene block II is found in the subfamily Aleyrodinae except *Te. acaciae* and *N. andropogonis*. Gene block III (*trnE-trnF-*

nad5-trnH-nad4-nad4L-trnT-trnP-nad6-cytb-trnS2-nad1-trnL1) is conserved in 8 species (*Ad. dugesii*, *Ad. rugioperculatus*, *Ac. camelliae*, *Ac. spiniferus, C. turpiniae*, *B. afer*, *B. tabaci*, *Te. acaciae*). The genes *nad1-trnL1* are translocated out of gene block III in *P. machili* and *Ach. aceris*; as are *trnS2* in *Tr. vaporariorum*; *trnP* in *N. andropogonis*; and *nad6-cytb-nad1-trnL1* in *Ad. dispersus*.

Gene block IV (*trnL1-rrnL-trnV-rrnS*) is conserved in 6 species (*Ad. dispersus*, *Ad. dugesii*, *Ad. rugioperculatus*, *Tr. vaporariorum*, *Ach. aceris*, *P. machili*). *trnV* is translocated out of gene block IV in *Ac. spiniferus*, *Ac. camelliae*, *Te. acaciae*, *C. turpiniae*; *rrnS* in *B. afer*, *B. tabaci.* Gene block V (*rrnS-trnI-trnQ-trnM-nad2-trnW-trnC-trnY-cox1*) is not conserved in any whitefly species due to multiple transposition and inversions.

The thirty-seven ancestral boundaries were mapped to the phylogenetic tree which revealed that 16 boundaries are found in all members of the family Aleyrodidae. Eleven additional ancestral gene boundaries are retained in the subfamily Aleurodicinae. The ancestral boundaries 4 (*trnk-trnD*) and 5 (*trnD-atp8*) are plesiomorphies found in all Aleyrodidae except *Bemisia* species and *Te. acaciae.* Gene boundary 7 (*atp6-cox3*) is a plesiomorphy for the subfamily Aleurodicinae but is lost in subfamily Aleyrodinae due to the transposition of *cox3* gene. Gene boundary 27 (*trnL-rrnL*) is a plesiomorphy for the subfamily Aleyrodinae, whereas 21 (*trnT-trnP*) and 22 (*trnP-nad6*) are plesiomorphies for the Aleyrodidae except for *Ad. dispersus* and *N. andropogonis*. Gene boundary 24 (*cyt-trnS2*) is conserved Aleyrodidae except for *Ad. dispersus* and *Tr. vaporariorum*.

Out of 47 derived gene boundaries, 19 were synapomorphies and 28 autapomorphies. Synapomorphic gene boundaries were mapped to the phylogenetic tree demonstrating that only nine were shared at the nodes. Gene boundaries 46 (*trnW-trnY*) and 47 (*trnC-cox1*) are shared across the family Aleyrodidae with a few exceptions, *N. andropogonis* (46,47 absent) and *Tr. vaporariorum* (46 absent). Moreover, gene boundaries 63-65 (*trnS2-trnR*, *cox3-trnN*, *trnQ-nad1*) are synapomorphies for the clade *Ach. aceris*+*P. machili*. Gene boundaries 61 (*trnV-trnD*), 62 (*trnD-trnQ*) are synapomorphies for the Genus *Bemisia*. Gene boundaries 38 (*atp6-trnE*), 39 (*rrnL-rrnS*) were synapomorphies for the clade (*Ac. spiniferus* + *Ac. camelliae*) + (*Te. acaciae* + *C. turpiniae*).

Gene boundary 49 (*trnS1-trnM*) is an autapomorphy for *Ac. camelliae*; 52-58 (*trnV-trnN*, *trnN-nad3*, *cox3-trnR*, *trnR-trnD*, *trnD- trnI*, *trnI-trnA*, *trnA-trnM*) for *Te. acaciae*; 59 (*trnS1-trnI*) for *C. turpiniae*; 66 (*rrnS-trnM*) for *Ach. aceris*; 67-72 (*cox3-nad3*, *cytb-nad1*, *rrnS-trnS2*, *trnS2-trnQ*, *trnW-trnG*, *trnG-trnY*) for *Tr. vaporariorum*; 73-78 (*trnT-nad6*, *rrnL-trnP*, *trnP-nad3*, *cox3-trnC*, *trnY-rrnS*, *trnW-cox1*) for *N. andropogonis*; 79-83 (*trnR-rrnS*, *nad1-nad6*, *cytb-trnS1*, *trnN-trnE*, *trnP-trnI*) for *Ad. dispersus*.

TreeREx

We carried out TreeREx analysis using the 13 PCGs+2RNAs ML-4 phylogenetic tree (Table S5). To eliminate biases due to partial mitogenomes, the 7 missing tRNAs were excluded from all species in this analysis. TreeRex analysis revealed the four inversions at the node of A0 towards *Ad. dispersus*; one inversion (*rrnS*) and one inverse transposition (*nad3 cox3*) at the node of A10 towards *N. andropogonis*.

One inverse transposition (*nad3 cox3*) occurred at node A9 towards A8 which contains (*Ach. aceris* + *P. machili*) + (*Bemisia* species) + (*C. turpinae* + *Te. acaciae*) + (*Ac. camelliae* + *A. spiniferus*) in which *nad3* and *cox3* were inverse transpositioned. One transposition (*nad1 rrnL rrnS*) occurred at node A8 towards to A7 which contains *Bemisia* + (*C. turpinae* + *Te. acaciae*) + (*Ac. camelliae* + *Ac. spiniferus*) (Fig. 6A).

Investigating ancestral gene blocks and mapping new gene boundaries through character coding and TreeREx revealed similar results. Gene order in subfamily Aleurodicinae was more conserved relative to the ancestral insect than was subfamily Aleyrodinae. Gene blocks II and V are the most variable gene regions whereas gene block I is the most conserved one in the Aleyrodidae. These mapped gene boundaries require further investigation across more species within each genus to determine whether they represent species-level autapomorphies or genus-level synapomorphies.

Phylogeny

In the present study, phylogenetic analyses were carried out using 4 datasets of derived from the 13 PCGs of whiteflies. Eight phylogenetic trees were constructed by Bayesian Inference and Maximum likelihood which have broadly similar topologies with few discrepancies (Figs. 6B, S2-S9). Both inference methods support the monophyly of both subfamilies Aleurodicinae and Aleyrodinae, consistent with previous studies [27]. *Ad. rugioperculatus* was sister to *Ad. dugesii* and they share an identical gene order. In contrast, *Ad. dispersus* and *Ad. dugesii* are in one clade but differ in their gene order. *Tr. vaporariorum* was sister to *Aleurochiton* + *Pealius* in all the analyses except ML-1, 2 and BI-2, where it was sister to *Bemisia*. This result is well supported in earlier phylogenies of whiteflies [27]. The genera *Bemisia*, and *Aleurocanthus* were recovered as monophyletic in all the analyses. *Ach. aceris* was sister to *P. machili* in all the analyses which was also support by earlier phylogenies. *Te. acaciae* was sister to *C. turpinae* in all analyses. Finally, to resolve the phylogenetic relationships of *Trialeurodes* more species from this genus will need to be sequenced.

Conclusion

Whiteflies belong to the order Hemiptera forming the family Aleyrodidae with two subfamilies Aleyrodinae and Aleurodicinae. Until now, 12 either complete or partial mitogenome data of family Aleyrodidae have been available. Here, we sequenced and annotated the second complete mitogenome from the subfamily Aleurodicinae and conducted a comparative analysis. The mitogenome is 15060 bp with 86.5% A + T content and 13.5% G + C content. Gene order in this species is similar to that of another Aleurodicinae species, *Ad. dugesii*. Gene rearrangements have been detected in the mitogenomes of whiteflies, located in several gene blocks between (*cox2-trnK-trnD-atp6-atp8*), (*cox3-trnG-nad3-trnA-trnR*), (*trnL1-rrnL-trnV-rrnS*), (*trnI-trnQ-trnM-nad2-trnW-trnC-trnY-cox1*). Five ancestral gene blocks (I-V) and 83 derived gene boundaries (37 ancestral, 19 derived, 27 unique) were identified in whiteflies. Ancestral gene block I was conserved of the whole family Aleyrodidae except for *Tetraleurodes acaciae* and *Bemisia* species. The ancestral gene block IV was conserved across the subfamily Aleurodicinae. Out of the 37 ancestral boundaries (plesiomorphies), 15 were shared across the two subfamilies. Nineteen derived gene

boundaries were mapped on to the whitefly phylogeny but only 9 were synapomorphic. We have identified 27 species-level derived boundaries (autapomorphies) which require further investigation. Eight phylogenies were constructed using Bayesian Inference (BI) and Maximum Likelihood (ML) that resulted in similar topologies. It supported the monophyly of both subfamilies Aleyrodinae and Aleurodicinae with good support values. Exclusion of third codon positions, and GBlock masking improved nodal support. The PCGs based phylogenetic study demonstrated that mitogenomes were informative in addressing the phylogenetic questions in whiteflies. More mitogenome sequencing data is needed to provide comprehensive taxonomic sampling and better understand the evolution of whiteflies.

Declarations

Data availability

Annotated mitogenome assemblies of *Aleurodicus rugioperculatus* are deposited in NCBI GenBank under the following accession number MW649000.

Acknowledgements

KT, VK and AP are thankful to the Director of Zoological Survey of India (ZSI), Ministry of Environment, Forests and Climate Change (MoEFCC), Govt. of India for providing necessary permissions and facilities. This work was financially supported by Zoological Survey of India (ZSI) in-house project, 'National Faunal Genome Resources (NFGR)'. We are thankful to Prof. Stephan L. Cameron, College of Agriculture, Entomology, Purdue University, West Lafayette, United States for his quick help in language editing of the manuscript.

Author contributions

Conceptualization: Kaomud Tyagi, Vikas Kumar;

Specimen Collection: C. M. Kalleshwaraswamy

Data curation: Avas Pakrashi, Kaomud Tyagi,

Formal analysis: Avas Pakrashi, Kaomud Tyagi;

Funding acquisition: Vikas Kumar;

Investigation: Kaomud Tyagi, Vikas Kumar;

Methodology: Kaomud Tyagi, Vikas Kumar,

Project administration: Vikas Kumar;

Resources: Vikas Kumar;

Software: Kaomud Tyagi, Vikas Kumar,

Supervision: Vikas Kumar, Kaomud Tyagi, C. M. Kalleshwaraswamy;

Validation: Vikas Kumar, Kaomud Tyagi;

Visualization: Vikas Kumar, Kaomud Tyagi;

Writing: Avas Pakrashi, Kaomud Tyagi, Vikas Kumar, C. M. Kalleshwaraswamy.

Competing interests

The authors declare no competing interests.

References

1. Song N, Liang A-P, Bu C-P (2012) A Molecular Phylogeny of Hemiptera Inferred from Mitochondrial Genome Sequences. PLoS ONE 7(11):e48778
2. Cameron SL (2014) Insect mitochondrial genomics: Implications for evolution and phylogeny. Annual Review of Entomology 59:95–117
3. Zhang DX, Hewitt GM (1997) Insect mitochondrial control region: A review of its structure, evolution and usefulness in evolutionary studies. Biochemical Systematics Ecology 25(2):99–120
4. Hao YJ, Zou YL, Ding YR, Xu WY, Yan ZT, Li XD, Fu WB, Li TJ, Chen B (2017) Complete mitochondrial genomes of *Anopheles stephensi* and *An. dirus* and comparative evolutionary mitochondriomics of 50 mosquitoes. Scientific Reports. 7(1): 1–13
5. Yang L, Dai J, Gao Q, Yuan G, Liu J, Sun Y, Sun Y, Wang L, Qian C, Zhu B, Liu C, Wei G (2020) Characterization of the complete mitochondrial genome of *Orthaga olivacea* Warre (Lepidoptera Pyralidae) and comparison with other Lepidopteran insects. PLoS One 15(3):1–20
6. Tyagi K, Chakraborty R, Cameron SL, Sweet AD, Chandra K, Kumar V (2020) Rearrangement and evolution of mitochondrial genomes in Thysanoptera (Insecta). Sci Rep 10(1):1–16
7. Moreno-Carmona M, Cameron SL, Prada Quiroga CF (2021) How are the mitochondrial genomes reorganized in Hexapoda? Differential evolution and the first report of convergences within Hexapoda. Gene 791:145719
8. Tyagi K, Kumar V, Poddar N, Prasad P, Tyagi I, Kundu S, Chandra K (2020) The gene arrangement and phylogeny using mitochondrial genomes in spiders (Arachnida: Araneae). International Journal of Biological Macromolecules 146:488–496
9. David BV, Subramaniam TR (1976) Studies on some Indian Aleyrodidae. Records of Zoological Survey of India. 70: 133–233
10. Mound LA, Halsey SH (1978) Whitefly of the world: a systematic catalogue of the Aleyrodidae (Homoptera) with host plant and natural enemy data. British Museum (Natural History), London and John Wiley and Sons, Chichester

11. Martin JH, Mound LA (2007) An annotated checklist of the world's whiteflies (Insecta: Hemiptera: Aleyrodidae). Zootaxa 1492(1):1–84
12. Anjum H, Ahmed SI (2019) An updated and consolidated review on Indian Aleyrodids fauna (Hemiptera: Aleyrodidae: Insecta) along with their host plant families and distributional records. Records of Zoological Survey of India 119(4):381–417
13. Martin JH (2004) The whiteflies of Belize (Hemiptera: Aleyrodidae) Part 1-introduction and account of the subfamily Aleurodicinae Quaintance & Baker. Zootaxa 681:1–119
14. Evans GA (2008) The whiteflies (Hemiptera: Aleyrodidae) of the world and their host plants and natural enemies. USDA/Animal Plant Health Inspection Service (APHIS), pp 9–23
15. Sundararaj R, Selvaraj K (2017) Invasion of rugose spiraling whitefly, *Aleurodicus rugioperculatus* Martin (Hemiptera: Aleyrodidae): a potential threat to coconut in India. Phytoparasitica 45(1):71–74
16. Selvaraj K, Venkatesan T, Sumalatha BV, Kiran CM (2019) Invasive rugose spiralling whitefly *Aleurodicus rugioperculatus* Martin a serious pest of oil palm *Elaeis guineensis* in India. Journal of Oil Palm Research 31(4):651–656
17. Pathak S (2019) First Record of Rugose Spiraling Whitefly, *Aleurodicus rugioperculatus* Martin (Hemiptera: Aleyrodidae) an Invasive Pest in Assam. Pantnagar Journal of Research 7(2):120–122
18. Mondal P, Ganguly M, Bandyopadhyay P, Karmakar K, Kar A, Ghosh DK (2020) Status of Rugose Spiraling Whitefly *Aleurodicus rugioperculatus* Martin (Hemiptera: Aleyrodidae) in West Bengal with notes on host plants, natural enemies and management. Journal of Pharmacognosy Phytochemistry 9(1):2023–2027
19. Ming-Xing L, Zhi-Teng C, Wei-Wei Y, Yu-Zhou D (2017) The complete mitochondrial genome of a spiraling whitefly, *Aleurodicus dispersus* Russell (Hemiptera: Aleyrodidae), Mitochondrial DNA Part A. 28(2): 165–166
20. Thao LL, Baumann L, Baumann P (2004) Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha). BMC Evol Biol 4(1):1–13
21. Wang HL, Lei T, Liu YQ (2019) Complete mitochondrial DNA genome of whitefly species (Hemiptera: Aleyrodidae) from *Litchi chinensis*. Mitochondrial DNA Part B 4(2):2765–2766
22. Chen SC, Wang X-Q, Li P-W, Hu X, Wang J-J, Peng P (2016) The Complete Mitochondrial Genome of *Aleurocanthus camelliae*: Insights into Gene Arrangement and Genome Organization within the Family Aleyrodidae. International Journal of Molecular Sciences 17(11):1843
23. Chen ZT, Mu LX, Wang JR, Du YZ (2016) Complete mitochondrial genome of the citrus spiny whitefly *Aleurocanthus spiniferus* (Quaintance) (Hemiptera: Aleyrodidae): Implications for the phylogeny of whiteflies. PLoS One 11:e0161385
24. Wang HL, Xiao N, Yang J, Wang XW, Colvin J, Liu SS (2016) The complete mitochondrial genome of *Bemisia afer* (Hemiptera: Aleyrodidae). Mitochondrial DNA Part A 27(1):98–99
25. Tay WT, Elfekih S, Court L, Gordon KH, De Barro PJ (2016) Complete mitochondrial DNA genome of *Bemisia tabaci* cryptic pest species complex Asia I (Hemiptera: Aleyrodidae). Mitochondrial DNA Part A 27(2):972–973

26. Li S, Wang WR, Zhou YF, Zhong LK, Jiang Y, Meng ZH (2020) The complete mitochondrial genome sequence of *Crenidorsum turpiniae* (Hemiptera: Aleyrodidae). Mitochondrial DNA Part B 5(4):3859–3860
27. Zhang ZT, Yan X, Yang WJ, Jin DC (2020) Characterization of the complete mitochondrial genome of *Pealius machili* (Hemiptera: Aleyrodidae) with phylogenetic analysis. Mitochondrial DNA Part B 5(2):1463–1464
28. Khamis FM, Ombura FLO, Ajene IJ, Akutse KS, Subramanian S, Mohamed SA, Dubois T, Tanga CM, Ekesi S (2021) Mitogenomic analysis of diversity of key whitefly pests in Kenya and its implication to their sustainable management. Sci Rep 11(1):1–11
29. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, Pütz J, Middendorf M, Stadler PF (2013) MITOS: Improved de novo metazoan mitochondrial genome annotation. Molecular Phylogenetics Evolution 69(2):313–319
30. Laslett D, Canbäck B (2008) ARWEN: A program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. Bioinformatics 24(2):172–175
31. Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31(13):3406–3415
32. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular evolutionary genetics analysis across computing platforms. Molecular Biology Evolution 35(6):1547–1549
33. Perna NT, Kocher TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J Mol Evol 41(3):353–359
34. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19(18):2496–2497
35. Bernt M, Merkle D, Ramsch K, Fritzsch G, Perseke M, Bernhard D, Schlegel M, Stadler PF, Middendorf M (2007) CREx: Inferring genomic rearrangements based on common intervals. Bioinformatics 23(21):2957–2958
36. Bernt M, Merkle D, Middendorf M (2008) An algorithm for inferring mitogenome rearrangements in a phylogenetic tree. RECOMB International Workshop on Comparative Genomics. 143–157
37. Zhang B, Zheng J, Liang L, Fuller S, Ma CS (2016) The complete mitochondrial genome of *Sitobion avenae* (Hemiptera: Aphididae). Mitochondrial DNA Part A 27(2):945–946
38. Abascal F, Zardoya R, Telford MJ (2010) TranslatorX: Multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res 38(suppl_2):W7–W13
39. Vaidya G, Lohman DJ, Meier R (2011) SequenceMatrix: Concatenation software for the fast assembly of multi-gene datasets with character set and codon information. Cladistics 27(2):171–180
40. Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B (2017) Partitionfinder 2: New methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Molecular Biology Evolution 34(3):772–773

41. Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. Nucleic Acids Res 44(W1):W232–W235
42. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP (2012) Mrbayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61(3):539–542
43. Rambaut A (2020) FigTree. Version 1.4.4 Institute of Evolutionary Biology. University of Edinburgh, Edinburgh
44. Fenn JD, Cameron SL, Whiting MF (2007) The complete mitochondrial genome sequence of the Mormon cricket (*Anabrus simplex*: Tettigoniidae: Orthoptera) and an analysis of control region variability. Insect Mol Biol 16(2):239–252
45. Woese CR, Gutell R, Gupta R, Noller HF (1984) Detailed analysis of the higher-order structure of 16S-like ribosomal ribonucleic acids. Microbiological Reviews 47(4):621–669
46. Gillespie JJ, Johnston JS, Cannone JJ, Gutell RR (2006) Characteristics of the nuclear (18S, 5.8S, 28S and 5S) and mitochondrial (12S and 16S) rRNA genes of *Apis mellifera* (Insecta: Hymenoptera): Structure, organization, and retrotransposable elements. Insect Molecular Biology 15(5):657–686

Tables

Table 1: **The mitochondrial genes of *Aleurodicus rugioperculatus***. The standard nomenclature was used for all the genes. IGN represents (+) values as intergenic nucleotides and (-) values as overlapping regions. Control region represented by CR.

Gene	Strand	Location	Size bp	Anticodon	Start	Stop	IGN
trnI	+	320-387	68	GAT	-	-	-2
trnM	+	386-451	66	CAT	-	-	0
nad2	+	452-1426	975	-	ATA	TAA	-1
trnW	+	1426-1489	64	TCA	-	-	2
trnY	-	1492-1557	66	GTA	-	-	0
trnC	-	1558-1620	63	TGC	-	-	1
cox1	+	1622-3166	1545	-	ATG	T(AA)	0
trnL2	+	3167-3231	65	TAA	-	-	0
cox2	+	3232-3892	661	-	ATG	T(AA)	0
trnK	+	3893-3962	70	CTT	-	-	3
trnD	+	3966-4030	65	GTC	-	-	0
atp8	+	4031-4180	150	-	ATT	TAA	-7
atp6	+	4174-4827	654	-	ATA	TAA	0
cox3	+	4828-5610	783	-	ATG	TAA	8
trnG	+	5619-5680	62	TCC	-	-	3
nad3	+	5684-6037	354	-	ATT	TAA	9
trnA	+	6047-6108	62	TGC	-	-	-2
trnR	+	6107-6172	66	TCG	-	-	0
trnN	+	6173-6240	68	GTT	-	-	-3
trnS1	+	6238-6295	58	AGA	-	-	0
trnE	+	6296-6358	63	GAA	-	-	-13
trnF	-	6346-6411	66	TTC	-	-	0
nad5	-	6412-8082	1671	-	ATA	TAA	-3
trnH	-	8080-8141	62	GTG	-	-	2
nad4	-	8144-9421	1278	-	ATA	TAA	-7
nad4l	-	9415-9711	297	-	ATG	TAA	1
trnT	+	9713-9782	70	TGT	-	-	-2
trnP	-	9781-9846	66	TGG	-	-	37

nad6	+	9884-10309	426	-	ATA	TAA	-1
cytb	+	10309-11445	1137	-	ATG	TAA	-2
trnS2	+	11444-11500	57	TGA	-	-	20
nad1	-	11521-12430	910	-	ATA	T(AA)	-3
trnL1	-	12428-12493	66	TAG	-	-	13
rrnL	-	12507-13792	1286	-	-	-	-87
trnV	-	13706-13769	64	TAC	-	-	-6
rrnS	-	13764-14542	779	-	-	-	-40
trnQ	-	14503-14574	72	CAA	-	-	0
CR		14575-15060	486	-	-	-	0
		1-319	319	-	-	-	-

Table 2: The mitogenomes of whitefly and outgroup were used in this study. Asterisk indicates that the data of this species is generated in this study.

Sl. No.	Subfamily	Species	Accession No.
1.	Aleurodicinae	*Aleurodicus dispersus*	KR063274
1.		*Aleurodicus dugesii*	AY521251
1.		*Aleurodicus rugioperculatus**	MW649000
1.	Aleyrodinae	*Aleurocanthus camelliae*	KU761949
1.		*Aleurocanthus spiniferus*	NC_029155
1.		*Aleurochiton aceris*	AY572538
1.		*Bemisia afer*	KR8191744
1.		*Bemisia tabaci*	NC_006279
1.		*Crenidorsum turpiniae*	NC_0509306
1.		*Neomaskellia andropogonis*	AY572539
1.		*Pealius machili*	MT01558
1.		*Tetraleurodes acaciae*	AY521262
1.		*Trialeurodes vaporariorum*	NC_006280
1.	Out Group	*Sitobion avenae*	NC_024683

Figures

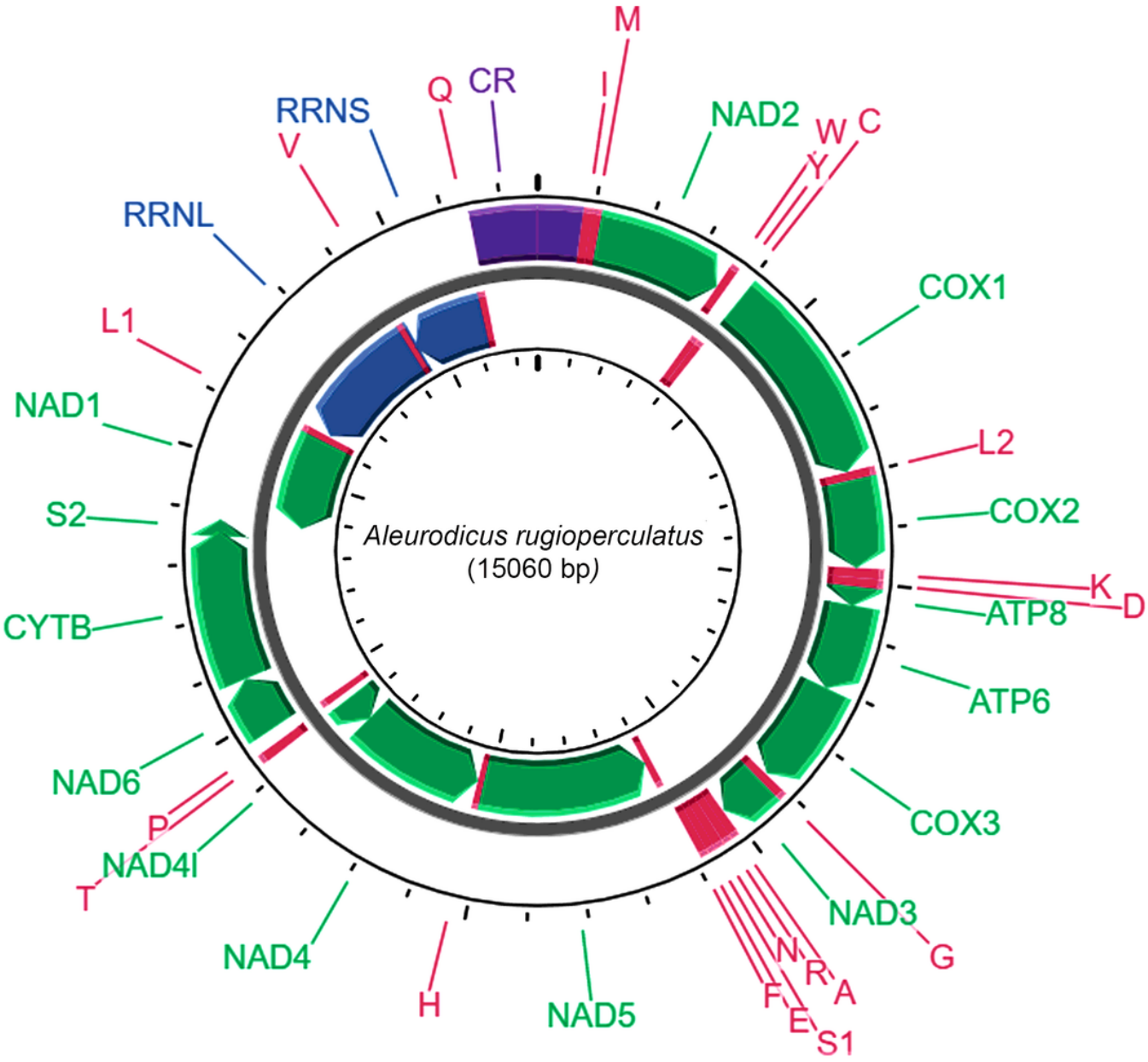

Figure 1

CG view of the mitogenome of Aleurodicus rugioperculatus.

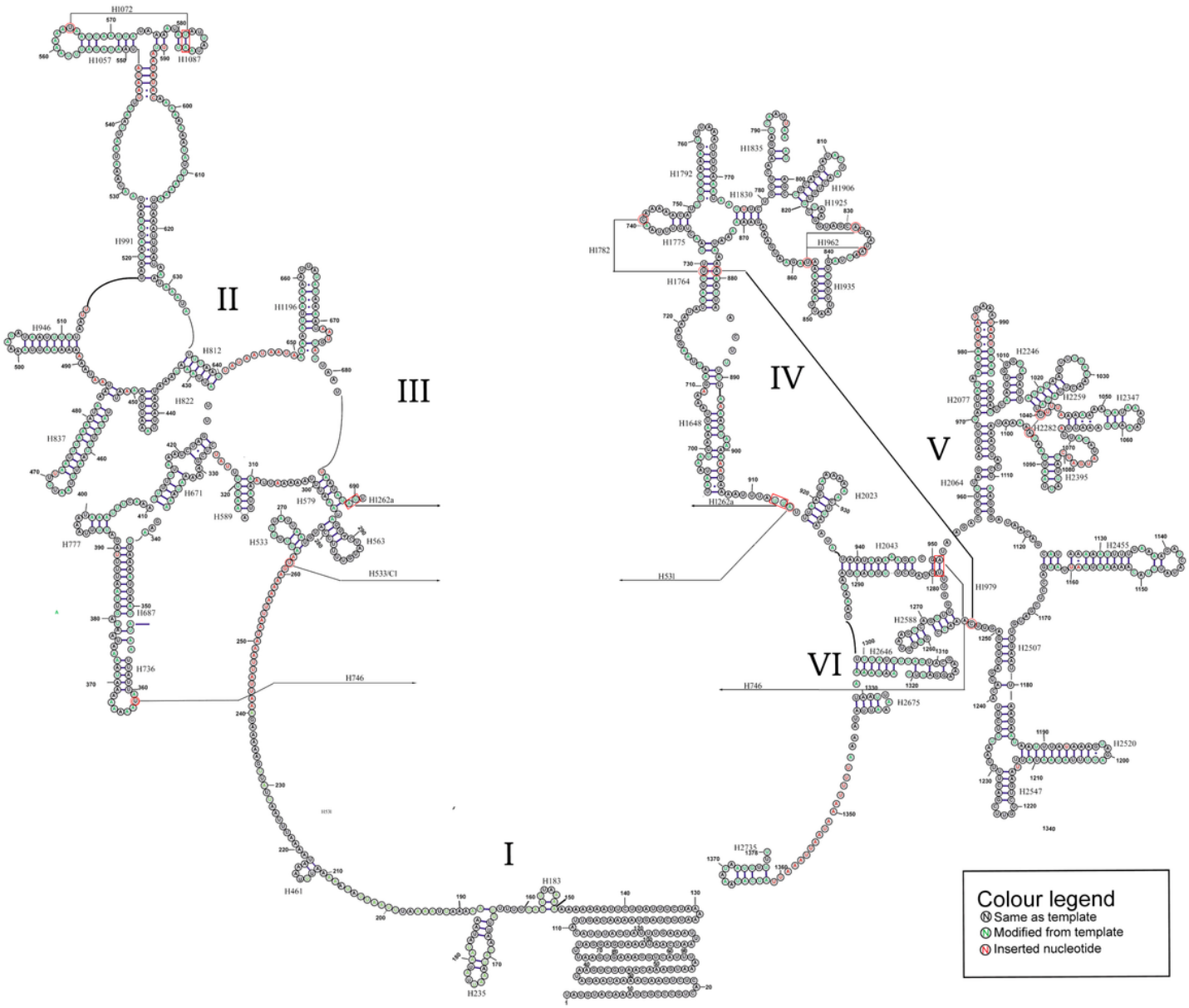

Figure 2

Secondary structure of the large subunit ribosomal RNA.

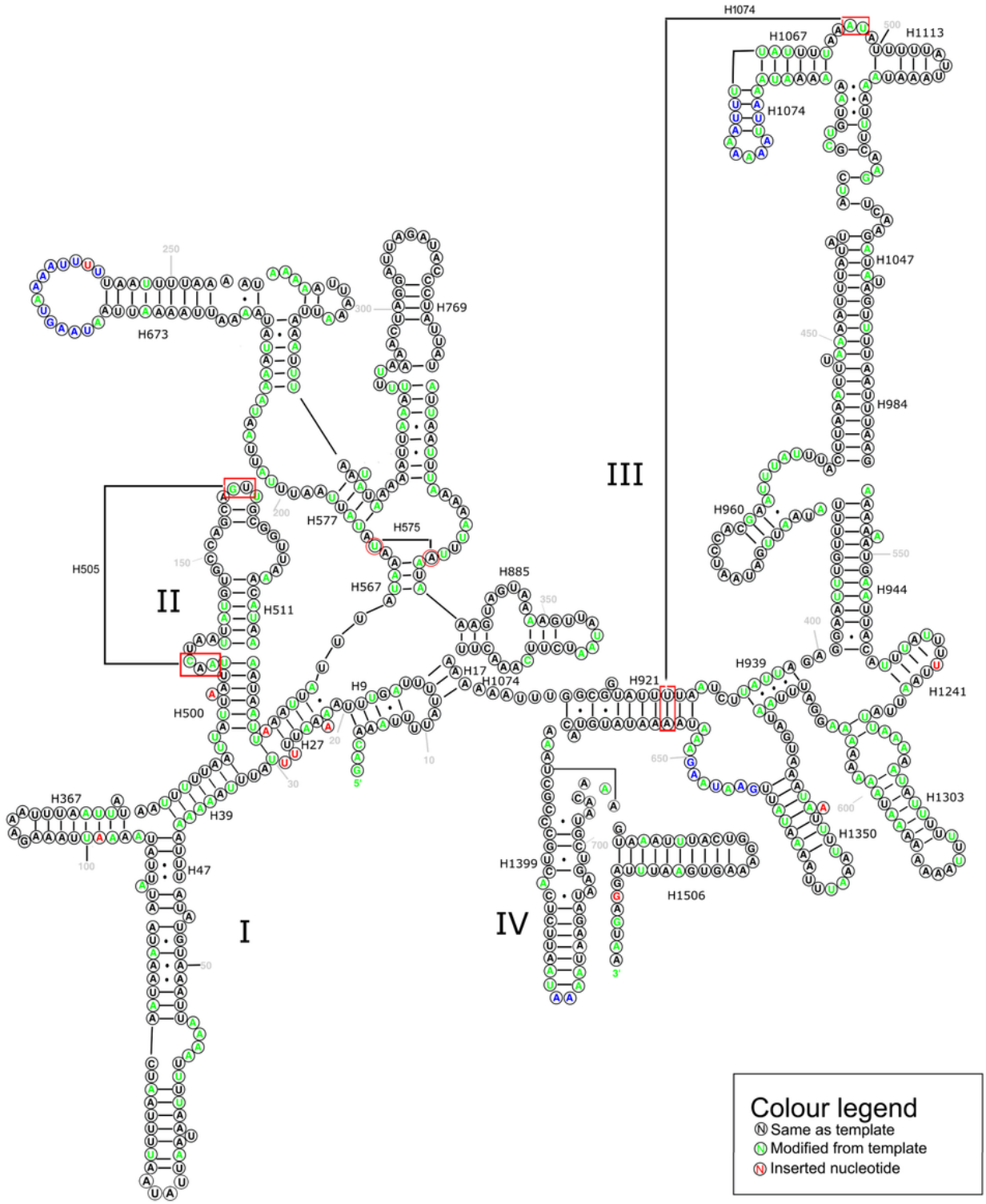

Figure 3

Secondary structure of the small subunit ribosomal RNA.

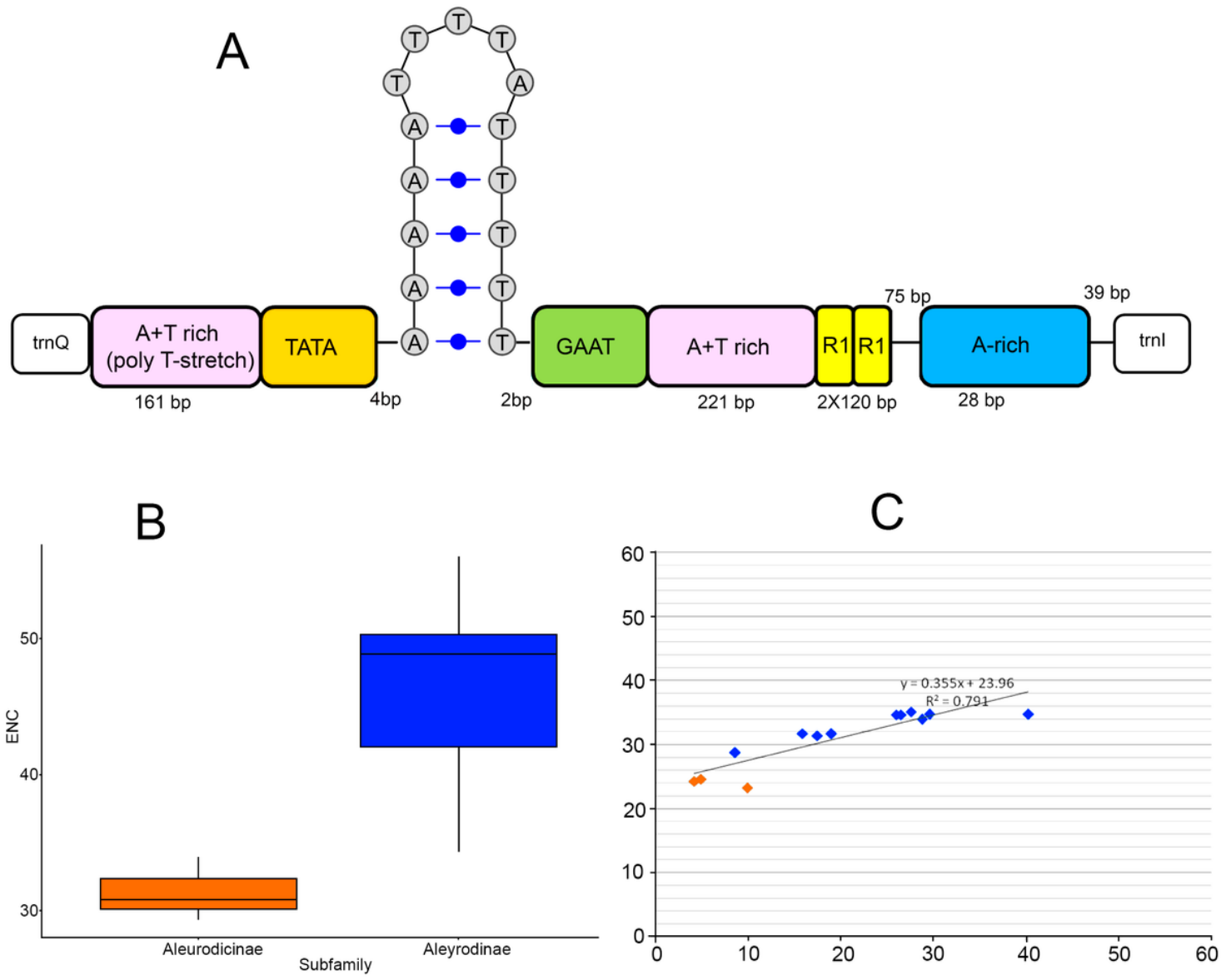

Figure 4

A) Secondary structure of control region; B) Boxplot of ENC and GC3 of two subfamilies of Aleyrodidae; C) Neutrality plot of 13 species of whiteflies.

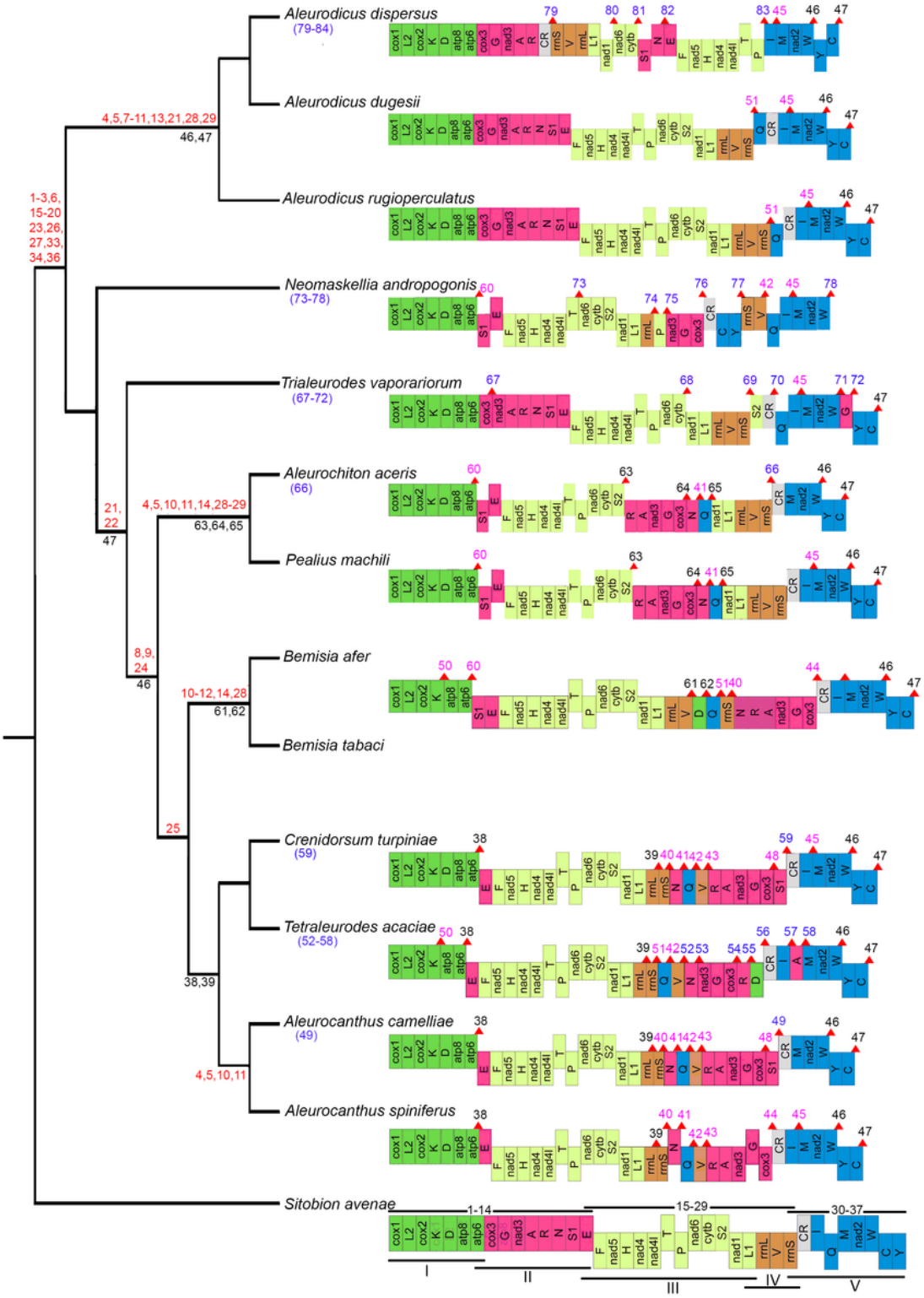

Figure 5

Linear representation of the gene order of 13 whiteflies and ancestor. The ancestral gene blocks were represented on the ancestor with I-V and highlighted by different colour; ancestral boundaries represented by red colours on the node of phylogenetic tree; shared derived boundaries by black colour on the node of phylogenetic tree and with pink colour on to the gene order; unique boundaries by blue colour.

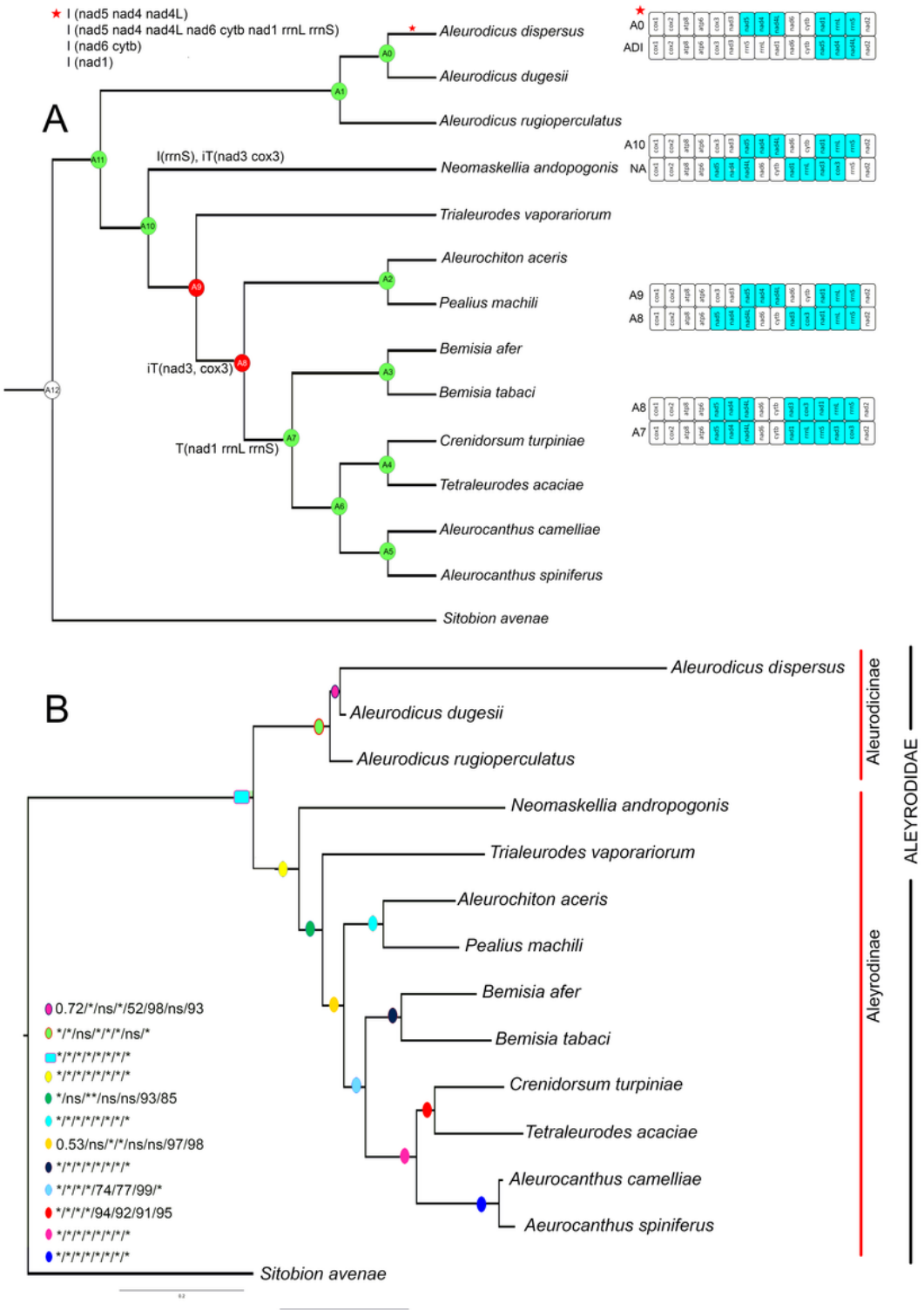

Figure 6

A) Cladogram showed the output of TreeREx analysis. B) BI-3 phylogeny along with the support values of 8 trees superimposed on the node.

Supplementary Files

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

- Supplementary.docx