Research Article

www.dergipark.org.tr/en/pub/mediterranean

MEDITERRANEAN AGRICULTURAL SCIENCES
(2024) 37(2): 67-71
DOI: 10.29136/mediterranean.1494881

Assembly and annotation of the first complete mitochondrial genome of *Phenacoccus solenopsis* Tinsley (Hemiptera: Pseudococcidae)

Payman EHSAS[1], Merve AYGUN[2], Cengiz IKTEN[2], Hilal Sule TOSUN[2]

[1]Akdeniz University, Faculty of Agriculture, Department of Agricultural Biotechnology, 07070, Antalya, Türkiye
[2]Akdeniz University, Faculty of Agriculture, Department of Plant Protection, 07059, The Campus, Konyaalti, Antalya, Türkiye

Corresponding author: H. S. Tosun, e-mail: hilaltosun@akdeniz.edu.tr
Author(s) e-mail: payman5790@gmail.com, merweaygun.666@hotmail.com, cikten@akdeniz.edu.tr

ARTICLE INFO

Received: June 3, 2024
Received in revised form: June 5, 2024
Accepted: June 26, 2024

***Keywords*:**

Phenacoccus solenopsis
Mitochondrial genome assembly
High-throughput DNA sequencing

ABSTRACT

The cotton mealybug, *Phenacoccus solenopsis* Tinsley (Hemiptera: Pseudococcidae), is a significant economic pest with a global distribution. Despite the increasing volume of literature on mitochondrial genome sequencing, complete mitochondrial genome sequences have not been reported for *P. solenopsis* yet. Here, we assembled the complete mitochondrial genome of *P. solenopsis* using high throughput DNA sequencing technology. The genome is 14,831 bp in length and is comprised of 37 genes: 13 protein-coding genes (PCGs), 22 transfer RNA (tRNA) genes, and two ribosomal RNA (rRNA) genes. The total length of all PCGs is 9,678 bp, accounting for 65.2% of the overall mitogenome. Among the PCGs, *atp8* is the smallest gene (99 bp) and *nad5* is the largest (1,593 bp). Of the 13 PCGs, seven (*nad3, cox3, atp6, atp8, cox2, cox1, nad2*) are encoded on the majority strand (J-strand), while six (*nad1, nad6, cob, nad4L, nad4, nad5*) are on the minority strand (N-strand). The analysis revealed a predominant A+T base composition, making up 90.3% of the total genome. However, the non-coding regulatory control region (CR) is missing due to the non-overlapping endpoints of the linear assembly. This assembly provides comprehensive information for investigating the evolutionary relationships between scale insects and for the precise identification of insects.

1. Introduction

The superfamily Coccoidea, known as scale insects, resides in the suborder Sternorrhyncha (Hemiptera) and contains around 8500 species. Pseudococcidae (mealybug), the second largest family in Coccoidea, includes 2143 species broadly dispersed around the globe (García et al. 2016). Most mealybug species are considered economically important pests since they are phloem feeders on several plant parts. Furthermore, they are known to play an important vector role for several plant viruses. One such pest, *Phenacoccus solenopsis* Tinsley 1898, is a highly invasive example of the family Pseudococcidae reported inmore than 70 countries (EPPO 2024) causing detrimental damage to cotton plants (Figure 1).

Despite the existence of vast amount of literature on pest status, little is known about the genetics and phylogeny of the Pseudococcidae. The current taxonomy of the family is mainly focused on the morphological characters measured on female adults (Gullan and Cook 2007). However, precise identification of mealybugs may become difficult and cumbersome using only morphological characters. On the other hand, mitochondrial genome information can present an opportunity to rapidly resolve genetic structure, evolutionary relationships and precise identification of insects (Cameron 2014). However, complete mitochondrial genome information in the family Pseudococcidae is still scarce, and absent for *P. solenopsis*. Therefore, we utilized high throughput DNA sequencing technology to assem mitogenome of the *P. solenopsis* collected from Denizli province in Türkiye and annotated it in detail.

2. Materials and Methods

2.1. Sample collection and DNA extraction

Cotton mealybug samples were collected from a *Gossypium hirsitum* L. cultivated field in Adaköy (37,95887°N, 29,00747°E) in the Sarayköy district in Denizli province, Türkiye in August 2023. The samples were immediately stored in absolute ethanol and brought to the laboratory for long term storage at -20°C until DNA extraction. The samples were initially identified morphologically, and more precise identification was carried out based on BLAST comparison of mtCOI sequences in NCBI database. For total DNA, a single female specimen was subjected to the CTAB extraction protocol (Doyle and Doyle 1987). The resulting total DNA quality was checked by running it on 1% agarose gel and then, stored at -20°C until library preparation.

A whole genome shotgun sequencing approach was utilized with 1 µg of DNA as input material. Enzymatic digestion was carried out using fragmentase enzyme (New England Biolabs, Inc) producing 300-1200 bp DNA fragments. DNA fragments were then cleaned using AMPure size selection kits (Beckman Coulter, Brea, CA). The remaining protocol followed NEBNext® V Ultra II DNA Library Prep Kit (New England

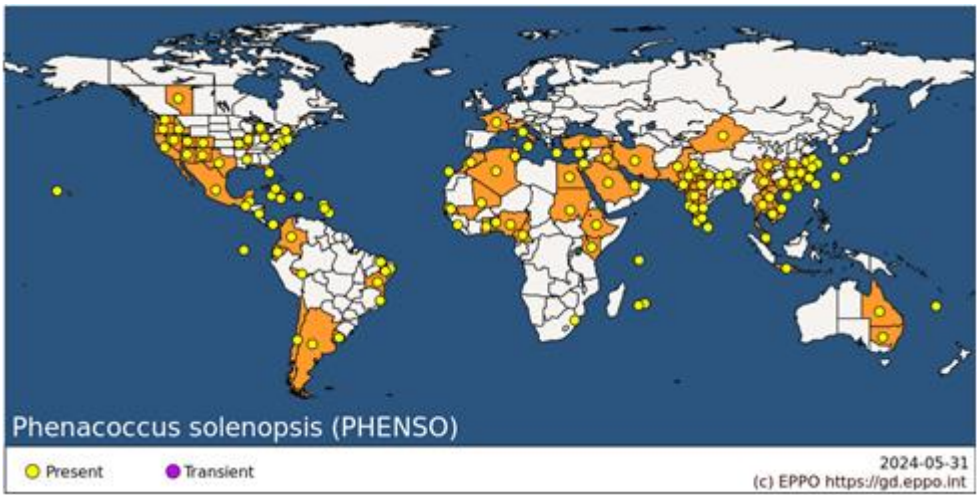

Figure 1. Global distribution of *Phenacoccus solenopsis*.

Biolabs, Inc.) according to the manufacturer's recommendations and index sequences were added to the genomic libraries. The resulting library, with an average insert size of 400 bp, was sequenced using the paired-end 150 sequencing method on the Illumina NovaSeq 6000 platform by Macrogen Ltd.

2.2. Mitochondrial assembly

The raw reads were at first subjected to demultiplexing and debarcoding protocol using "je demultiplex" software (Girardot et al. 2016) with default values and then, cleaned from low quality reads using "trim galore" (Martin 2011) and finally analyzed for quality metrics using "FastQC" (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/).

The clean reads were assembled with "MEGAHIT" (Li et al. 2015) and "metaSPAdes" (Nurk et al. 2017) with default values resulting in several nuclear and mitochondrial contigs. Among these contigs, mitochondria associated ones were baited using complete mitogenomes from closely related taxa on "bbduk" (Bushnell 2020) of BBTools Packages (ttps://jgi.doe.gov/data-and-tools/bbtools/). The baited contigs were then used as a second round bait file for the selection of reads from the original cleaned fastq file followed by final assembly construction by "MEGAHIT" and "metasSPAdes". The resulting two route assemblies were checked against each other for consistency and further BLAST checked for similarities to mitogenome of closely related taxa.

2.3. Mitochondrial genome annotation and analyses

The assembled mitogenome was initially annotated for PCGs, rRNA, and tRNA genes using MITOS v2 web server (Bernt et al. 2013). Since initial annotation by MITOS2 was lacking for some PCGs (*atp8*, *nad4L*, *nad6*), the ORF Finder server (https://www.ncbi.nlm.nih.gov/orffinder/) was used for capturing missing PCGs. Furthermore, the ORF finder was used for verification of boundaries of PCGs, *rrnS* and *rrnL* reported by MITOS2. In parallel to PCGs, several tRNA's were not captured by MITOS2, hence a BLAST search of closely related taxa was utilized to find missing tRNA's, leading to improved annotation quality. The graphic representation of *P. solenopsis* mitogenome were visualized by Proksee (https://proksee.ca/), an updated version of the CGView web server (Grant and Stothard 2008).

3. Results and Discussion

3.1. General Features

The mitogenome sequence of *Phenacoccus solenopsis* was assembled into a single contig of 14831 bp in length (Figure 2). However, the non-coding regulatory control region (CR) was apparently missing from the final mitogenome assembly as the end point of linear assembly was not overlapping on each other. Furthermore, CR is known to have a repetitive AT-rich composition in many insect species and close taxa, and yet was not apparent in the finished mitogenome. The ambiguity about CR most probably stemmed from an underrepresentation of AT-rich sequences in Illumina read data (Shen et al. 2015) leading to a downsized final assembly of mitogenome. Furthermore, in order to avoid mis-assemblies in the final data, assembly programs are prone to skip repetitive regions from the final assembly. There are only three full mitogenomes available for mealybugs in the GenBank database (OY390719; NC_066716; OX465514) for comparison and none of them clearly indicated the presence of CR region in their annotation. However, close examination of two accessions revealed the presence of a probable CR region of approximately 500 bp and 2100 bp for *Planococcus citri* Risso and *Balanococcus diminitus* Leonardi, respectively. On the other hand, there was no apparent CR region for *Phenacoccus manihoti* Matile-Ferrero mitogenome sequence. For other *Coccoidea* species, the annotation for the CR region was either completely missing or variable between 310 bp and 403 bp (NC_ 085772; NC_ 070232; NC_ 067791; NC_ 063660; NC_ 057479). Despite not having certain length information in the current study, the putative CR *P. solenopsis* was likely to lie between rrnL and trnM genes.

3.2. Protein-coding genes (PCGs)

The annotation of the newly assembled mitogenome revealed a total of 37 genes, including 13 protein-coding genes (PCG), two ribosomal RNA (rRNA) genes and 22 transfer RNA (tRNA) genes (Figure 2). Overall nucleotide composition of *P. solenopsis* mitochondria had a significant A+T bias (90.3 %) even in the absence of AT-rich CR in the final assembly (Table 1). The length of all PCG's totaled 9678 bp and accounted for 65.2% of the overall mitogenome (Table 1). Among the PCG's, *atp8* was the smallest gene (99 bp) whereas *nad5* was the largest gene (1593 bp) which is similar in other known Coccoidea

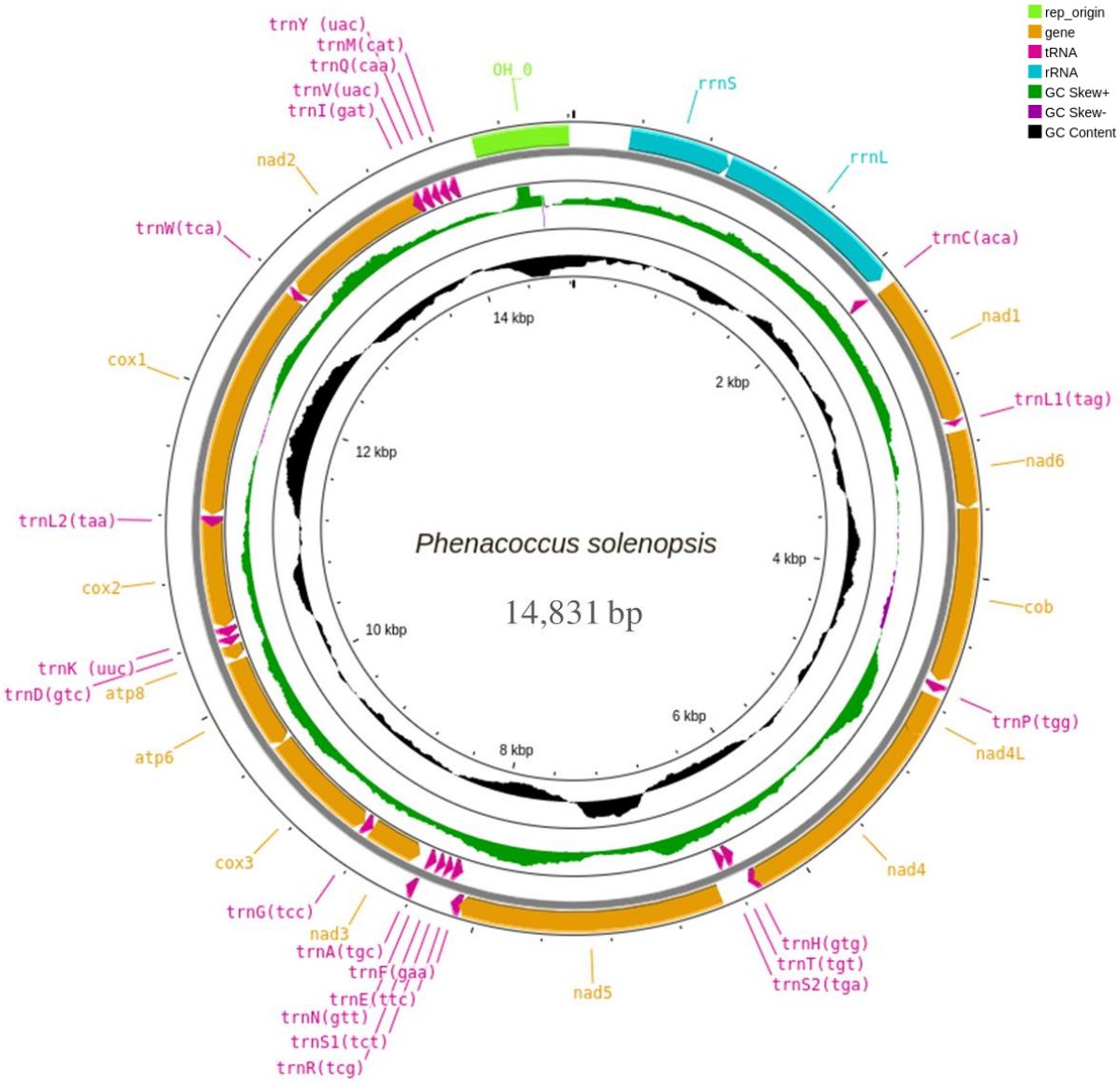

Figure 2. Circular map of *Phenacoccus solenopsis* mitochondrial genome.

mitogenomes. All PCGs were initiated with the standard ATN start codon for translation (Table 2). ATT was the most common start codon utilized by 5 PCGs followed by ATA, ATG and ATC, respectively (Table 2). For termination of translation, all PCGs used TAA sequence as the stop codon (Table 2). Out of 13 PCG's, 7 (*nad3*, *cox3*, *atp6*, *atp8*, *cox2*, *cox1*, *nad2*) were found on the majority strand (J-strand), whereas 6 PCG were located *(nad1*, *nad6*, *cob*, *nad4L*, *nad4*, *nad5*) on the minority strand (N-strand) (Figure 2; Table 2). In both strands, base composition was skewed towards A+T for all PCG's, the lowest being *cox1* gene (82.6%) and the highest *nad4L* gene (93.3%) (Table 1). High A+T bias in mitochondria is a general feature in many insect species. This composition is generally attributed to the lower nitrogen requirement of A-T base pairs than C-G base pairs. As scale insects feed on plant sap, where nitrogen content is the lowest in plants, this extreme A+T content may be a reflection of evolutionary adaptation to low nitrogen concentration on feeding site (Lu et al. 2020).

3.3. Ribosomal RNA (rRNA) genes and Transfer RNA genes (tRNA)

The current mitogenome assembly reflected standard two adjacent ribosomal RNA genes. Both genes were located between putative CR region and *trnC* gene on N-strand (Figure 2). The lengths of two ribosomal genes were 618 bp and 1142 bp for small subunit ribosomal RNA gene (*rrnS*) and large subunit ribosomal RNA gene (*rrnL*) respectively (Table 1). A + T content of the *rrnS* and *rrnL* were in the range between 90.1 and 90.8% and similar to that of the PCGs (Table 1).

Initially, only 14 tRNAs (*trnM*, *trnI*, *trnW*, *trnL2*, *trnD*, *trnG*, *trnN*, *trnR*, *trnS1*, *trnE*, *trnF*, *trnT*, *trnH*, *trnP*) were systematically inferred through the MITOS2 annotation procedure, while the remaining 8 tRNA's (*trnY*, *trnQ*, *trnV*, *trnK*, *trnA*, *trnS2*, *trnL1*, *trnC*) were identified through multiple alignments with sequences from previously characterized scale insect mitogenomes. In the *P. solenopsis* mitogenome, 17 of the predicted tRNA genes were located on the J-strand and the rest

on the N-strand with varying lengths between 45 bp and 71 bp (Figure 2).

In a typical animal mitochondrial genome, all tRNA genes, except *trnS* (GCT), are folded into clover leaf-like secondary structures where four arms (AA-arm, D-arm, AC-arm and T-arm) are concurrently forming (Wolstenholme 1992). In the current research, only 10 of the predicted tRNA genes (*trnM*, *trnI*, *trnW*, *trnL2*, *trnD*, *trnN*, *trnS1*, *trnE*, *trnF*, *trnH*, *trnP*) had the typical clover leaf secondary structure and the remaining tRNAs (*trnG*, *trnR*, *trnT*, *trnY*, *trnQ*, *trnV*, *trnK*, *trnA*, *trnS2*, *trnL1*, *trnC*) had irregular secondary structures where the D-arm/T-arm, or both were absent. Heavy tRNA truncation was prevalent in nematodes where 20 of the 22 tRNAs lacking T-arm or D-arm were reported in the literature (Wolstenholme et al. 1987). Furthermore, arm reduction in tRNA was exemplified in other arthropod species, such as arachnids and mites (Domes et al. 2008). In fact, loss or reduction of D-arm or T-arm in tRNA genes were a common feature in closely related scale insect species such as *Ceroplastes japonicus* Green and *Saissetia coffeae* Walker (Deng et al. 2019; Lu et al. 2020). Therefore, the arm reduction in tRNA genes reported here is not a specific event for *P. solenopsis*. Furthermore, rather than MITOS2 auto-annotation, the positions of 8 tRNA were only located by multiple sequence alignment in the current mitogenome assembly. Whether the results of unusual secondary structures in tRNA genes or limited identification ability of tRNA search programs, the overall tRNA results indicate a heavy truncation of tRNA genes for the *P. solenopsis* mitogenome.

Rather than nuclear machinery, mitochondria orchestrate its own transcription and translation systems. Hence, dysfunctional tRNAs encoded by the mitochondrial genome could be detrimental to organisms unless a replacement from the nuclear genome could be transported to the mitochondria to assist a deficient tRNA function. Furthermore, the post-transcriptional tRNA editing scheme is proposed as a likely mechanism for some species (Lavrov et al. 2000; Segovia et al. 2011). The current results add up another case for truncated/deficient tRNA genes in scale insect's mitochondria even these species function normally. Since the limited mitochondrial genome is available, it cannot be assured tRNA truncation is a common feature for all scale insects. Hence, more research is needed to explore mitogenomes in this taxon.

Table 1. Nucleotide composition and size of *P. solenopsis* genes

Genes	A	T	G	C	AT	Size (bp)
rrnS	0.44	0.46	0.06	0.04	0.90	618
rrnL	0.44	0.47	0.06	0.03	0.91	1142
nad1	0.34	0.54	0.08	0.03	0.88	903
nad6	0.37	0.55	0.04	0.03	0.92	465
cob	0.34	0.52	0.06	0.07	0.86	1059
nad4L	0.43	0.51	0.05	0.01	0.93	267
nad4	0.35	0.56	0.06	0.03	0.91	1332
nad5	0.35	0.55	0.07	0.03	0.91	1593
nad3	0.40	0.51	0.03	0.06	0.92	354
cox3	0.40	0.52	0.03	0.05	0.92	753
atp6	0.41	0.49	0.03	0.06	0.91	618
atp8	0.40	0.48	0.02	0.09	0.89	99
cox2	0.43	0.45	0.04	0.07	0.89	696
cox1	0.37	0.46	0.07	0.10	0.83	1539
nad2	0.42	0.53	0.01	0.04	0.94	981
All genes	0.38	0.51	0.06	0.05	0.89	
PCG	0.37	0.52	0.06	0.05	0.89	
All Genome	0.43	0.48	0.06	0.03	0.90	

Table 2. Mitogenome structure of *P. solenopsis*, direction, position and start/stop codons

Gene	Direction	Position	Start codon	Stop codon
nad1	N	2154-3056	ATG	TAA
nad6	N	3122-3586	ATT	TAA
cob	N	3588-4646	ATG	TAA
nad4L	N	4742-5008	ATT	TAA
nad4	N	4969-6300	ATA	TAA
nad5	N	6522-8114	ATT	TAA
nad3	J	8454-8807	ATT	TAA
cox3	J	8871-9623	ATG	TAA
atp6	J	9624-10241	ATA	TAA
atp8	J	10249-10347	ATC	TAA
cox2	J	10472-11167	ATT	TAA
cox1	J	11212-12750	ATA	TAA
nad2	J	12800-13780	ATA	TAA

4. Conclusion

In this study, we sequenced and assembled the first mitogenome of *P. solenopsis* and described its mitochondrial features. The results indicated the presence of gene arrangement and reduction of tRNA genes. Nucleotide bias was aberrant toward A+T bases, a common feature in other scale insect mitogenomes (Cameron 2014; Deng et al. 2019; Han et al. 2021). The ability to use high throughput sequencing technology provides the opportunity to reach fast and accurate insect mitogenomes. However, they are not without drawbacks. Despite high coverage rates presented by Illumina data, current assembly was missing in the CR region in the final *P. solenopsis* mitogenome due to, most probably, the existence of high repetitive and AT content in the region. Hence, non-coding CR regions rich in repetitive sequences should be PCR amplified with the primers developed from *trnY* and *rrnS* site and, then Sanger sequencing could be utilized to reveal the CR region details in *P. solenopsis*. Overall, in its current form, *P. solenopsis* mitogenome can provide useful data for the investigation of phylogenetic relationships between different populations and across different scale insect lineages.

Acknowledgements

We thank the anonymous reviewers for their efforts in reading the manuscript of this article and shared invaluable comments.

References

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 and Evolution 69(2): 313-319. doi: org/10.1016/j.ympev.2012.08.023.

Bushnell B (2020) BBDuk. Jt Genome Inst. Available online: https://jgi.doe.gov/data-and-tools/bbtools/bb-toolsuserguide/bbduk-guide. Accessed 10 May, 2024.

Cameron SL (2014) Insect mitochondrial genomics: Implications for evolution and phylogeny. Annual Review of Entomology 59(1): 95-117. doi: org/10.1146/an- nurev-ento-011613-162007.

Deng J, Lu C, Huang X (2019) The first mitochondrial genome of scale insects (Hemiptera: Coccoidea). Mitochondrial DNA Part B Resources 4: 2094-2095. doi: org/10.1080/23802359.2019.1622464.

Domes K, Maraun M, Scheu S, Cameron SL (2008) The complete mitochondrial genome of the sexual oribatid mite Steganacarus magnus: Genome rearrangements and loss of tRNAs. BMC Genomics 9: 532. doi: org/10.1186/1471-2164-9-532.

Doyle JJ, Doyle JL (1987) A Rapid DNA Isolation Procedure for Small Quantities of Fresh Leaf Tissue. Phytochemical Bulletin 19: 11-15.

EPPO (2024) Statistical database. https://gd.eppo.int/taxon/PHENSO/distribution. Accessed 31 May, 2024.

García Morales M, Denno BD, Miller DR, Miller GL, Ben Dov Y, Hardy NB (2016) ScaleNet: a literature-based model of scale insect biology and systematics. Database doi: org/10.1093/database/bav118, https://scalenet.info/. Accessed 25 May, 2024.

Girardot C, Scholtalbers J, Sauer S, Su SY, Furlong, EE (2016) Je, a versatile suite to handle multiplexed NGS libraries with unique molecular identifiers. BMC Bioinformatics 17: 419. doi: org/10.1186/s12859-016-1284-2.

Grant JR, Stothard P (2008) The CGView Server: a comparative genomics tool for circular genomes. Nucleic Acids Research 36(2): 181-184. doi: org/10.1093/nar/gkn179.

Gullan PJ, Cook LG (2007) Phylogeny and higher classification of the scale insects (Hemiptera: Sternorrhyncha: Coccoidea). Zootaxa 1668: 413-425.

Han Xu, Xiaochen Liu, Hu L, Sanan Wu (2021) The mitogenome of the scale insect *Didesmococcus koreanus* Borchsenius, 1955 (Coccoidea: Coccidae). Mitochondrial DNA Part B 6(4): 1298-1299.

Lavrov DV, Brown WM, Boore JL (2000) A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede *Lithobius forficatus*. Proceedings of the National Academy of Sciences 97(25): 13738-13742. doi: org/10.1073/pnas.250402997.

Li D, Liu C, Luo R, Sadakane K, Lam T (2015) Application note MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31(10): 1674-1676.

Lu C, Huang X, Deng J (2020) The challenge of Coccidae (Hemiptera: Coccoidea) mitochondrial genomes: The case of *Saissetia coffeae* with novel truncated tRNAs and gene rearrangements. International Journal of Biological Macromolecules 158: 854-864. doi: org/10.1016/j.ijbiomac.2020.04.257.

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet Journal 17: 10-12. doi: org/10.14806/ej.17.1.200.

Nurk S, Meleshko D, Korobeynikov A, Pevzner PA (2017) metaSPAdes: a new versatile metagenomic assembler. Genome Research 27(5): 824-834. doi: org/10.1101/gr.213959.116.

Segovia R, Pett W, Trewick S, Lavrov DV (2011) Extensive and evolutionarily persistent mitochondrial tRNA editing in velvet worms (phylum Onychophora). Molecular Biology and Evolution 28(10): 2873-2881. doi: org/10.1093/molbev/msr113.

Shen J, Cong Q, Grishin NV (2015) The complete mitochondrial genome of *Papilio glaucus* and its phylogenetic implications. Meta Gene 5: 68-83. doi: org/10.1016/j.mgene.2015.05.002.

Wolstenholme DR, Macfarlane JL, Okimoto R, Clary DO, Wahleithner JA (1987) Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. Proceedings of the National Academy of Sciences 84(5): 1324-1328. doi: org/10.1073/pnas.84.5.1324.

Wolstenholme DR (1992) Animal mitochondrial DNA: structure and evolution. International review of cytology 141: 173-216. doi: org/10.1016/S0074-7696(08)62066-5.