The mitochondrial genomes of two walnut pests, *Gastrolina depressa depressa* and *G. depressa thoracica* (Coleoptera: Chrysomelidae), and phylogenetic analyses

Qi Qi Wang 1,2,3, Guanghui Tang 1,*

1  Key Laboratory of State Forestry Administration on Management of Forest Bio-Disaster, College of Forestry, Northwest A&F University, Yangling, 712100, China
2  Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing, 100101, China
3  University of Chinese Academy of Sciences, Beijing, 100049, China
* Correspondence: tanggh@nwafu.edu.cn

In this study, the mitochondrial genomes (mitogenomes) of two walnut leaf insect pests, *Gastrolina depressa depressa* and *G. depressa thoracica*, were sequenced by Sanger sequencing technology. The mitogenome of *G. depressa thoracica* was complete at 16,109 bp in length, while the mitogenome of *G. depressa depressa* (14,277 bp) was partial. The genomic analyses indicated that both mitogenomes have the typical gene content and arrangement. The formerly identified elements, ‘TAGTA’ between trnSer(UCN) and nad2, and ‘ATGATAA’ between atp8 and atp6, were found to be more conserved than that between nad4L and nad4, which was ‘ATGTTAA’ in Coleoptera excluding Polyphaga. Phylogenetic analyses of the 13 protein-coding genes from 36 species of Coleoptera well supported a close affinity between the subfamily Chrysomelinae including *G. depressa thoracica* and *G. depressa depressa* and Galerucinae, as well as a sister relationship of ((Eumolpinae + Cryptocephalinae) + Cassidinae) within Chrysomelidae.

**Introduction**

Both *Gastrolina depressa thoracica* and *G. depressa depressa* belong to Chrysomelinae (Coleoptera:
Chrysomeloidea) and are the major insect pests of walnut trees in China. The larvae and adults feed on walnut leaves, and seriously influence walnut growth and yield. They are different from other species of Chrysomelinae with typical flat back in the boundary between Palaearctic and oriental regions. *G. depressa thoracica* has a black prothorax, while *G. depressa depressa* has a yellow prothorax (Chen, 1974). Ge et al. (2003) comprehensively compared the morphological differences and the distributions of *G. depressa thoracica* and *G. depressa depressa*. Overall, in China *G. depressa depressa* are widely distributed in most areas south of the Yellow River, while *G. depressa thoracica* live north of the Yangtze River. They both have filamentous antennae, and similar structure of the hind wing, claw and upper lip, but there are large differences in the structure of lower lip (cilia: long and dense in *G. depressa thoracica*, short and sparse in *G. depressa depressa*) and lower jaw (between the anterior chin and lower lip: triangle in *G. depressa thoracica*, circular soliton in *G. depressa depressa*). However, there is no information about their mitochondrial genomes (mitogenomes) or other molecular data to support their classification.

Mitogenomes have been widely used as molecular markers for phylogenetics and phylogeographics because of maternal inheritance, conserved gene order and orientation, low recombination rate and high mutation rate (Avise, 1989; Crozier & Crozier, 1993; Jin et al., 2004). In general, the mitogenomes range from 14 to 36 kb in length, encoding 13 protein-coding genes (PCGs), 22 transfer RNA genes (tRNAs), 2 ribosomal RNA genes (rRNAs) and a large non-coding region (control region or A+T-rich region) (Wolstenholme, 1992; Boore et al., 1999). Recent advances in sequencing technology have enriched insect genome and mitogenome datasets and accelerated insect molecular studies. Insect genome and mitogenome data are now widely used in species identification, population genetics and molecular evolution (Ma et al., 2012; Timmermans et al., 2014).

Coleoptera has as many as 380,000 described species and an estimated total number of about three million members. Curculionoidea is one of the most highly diversified infraorders of Polyphaga (Crowson, 1960; Ødegaard, 2000; Footit & Adler, 2009) and includes six superfamilies, among which the superfamilies of Curculionoidea and Chrysomeloidea are the major plant-eating beetles (Grimaldi & Engel, 2005). Several studies have reported the phylogenetic relationships within Chrysomeloidea by using morphological data (Farrell & Sequeira, 2004; Gómez-Zurita et al., 2007), 18S rRNA (Hunt et al., 2007), 28S rDNA (Marvaldi et al., 2009), and partial mitochondrial genes (Bocak et al., 2014; Li et al., 2016). In this study, we sequenced the mitochondrial genomes of *G. depressa thoracica* and *G. depressa depressa* and clarified the mitogenome differences between them. The nucleotide and amino acid sequences of the 13 PCGs of these two walnut pest insects were then aligned with data for 34 other beetles species for a phylogenetic analysis.

Materials and Methods

Sample collection

Adults of *G. depressa thoracica* and *G. depressa depressa* were originally collected from the experimental station of Northwest A&F University, Shanyang County, Shaanxi Province, China (109°88'E, 33°53'N), and identified according to key morphological characteristics. All samples were stored in 100% ethanol at −20℃.
DNA extraction and PCR amplification

Total genomic DNA was extracted from an individual insect using a standard phenol-chloroform extraction (Tamura & Aotsuka, 1988). The specific primers of each mitochondrial gene (Table 1) were designed from the conserved regions after multiple alignments of coleopteran insect mitochondrial sequences. PCR was performed with a 3-step program: an initial denaturation at 94 °C for 2 min; followed by 35 cycles at 94 °C for 30 s, annealing temperature (Table 1) for 30 s, 72 °C for 1 min; and a final extension at 72 °C for 10 min. PCR products (1.1–4.1 Kb in length) were analyzed by 1% agarose gel electrophoresis, and sequenced.

Genome assembly and annotation

All sequences were blasted in NCBI (https://www.ncbi.nlm.nih.gov/) and assembled using the program SeqMan in DNAStar package v7.1 (DNAStar Inc., Madison, WI, USA). The 13 PCGs were identified using ORF Finder (available on NCBI) and the rRNAs were determined by comparison with other coleopteran mitogenomes. The tRNAs and their cloverleaf secondary structures were predicted using tRNAscan-SE Search Online Server (Lowe & Eddy, 1997), while undefined tRNAs were further compared with tRNAs of other species, including Atrijuglans hetuoei (Wang et al., 2016), Dastarcus helophoroides (Zhang et al., 2015) and Anopheles minimus (Hua et al., 2016). The secondary structure of all tRNAs were drawn using RNAviz v2.0 (De Rijk et al., 2003). The composition of nucleotide sequences was described by skewness according to the following formulas (Perman et al., 1995): AT-skew = (A–T)/(A+T) and GC-skew = (G–C)/(G+C). The A+T content and relative synonymous codon usage (RSCU) were calculated using MEGA v5.1 (Tamura et al., 2011).

Phylogenetic analysis

The newly sequenced mitogenomes of *G. depressa thoracica* and *G. depressa depressa* were aligned with 32 mitochondrial genomes of Chrysomeloidea available in GenBank, with *Tetraphalerus bruchi* (Coleoptera: Archostemata: Ommatidae) and *Abax parallelepipedus* (Coleoptera: Adephaga: Carabidae) as outgroups (Table S1). The nucleotide and amino acid sequences of 13 PCGs from all 36 species were aligned separately using ClustalW implemented in MEGA v5.1 (Tamura et al., 2011). Gblocks v0.91b (Talavera & Castresana, 2007) was used to refine the final alignments and identify the conserved sequences (or conserved motifs). After Gblocks analysis, the final matrix consisted of 7289 nucleotides and 2340 amino acids. The best-fit models (TVM + I + G for the nucleotide dataset and MtREV + I + G + F for amino acid dataset) were selected using MrModeltest v2.3 (Nylander, 2004) and ProtTest v2.4 (Abascal et al., 2005) with specific parameters (Table S2).

Phylogenetic trees of PCGs were constructed using maximum likelihood (ML) and Bayesian inference (BI); these two methods (ML and BI) have different algorithms in phylogenetic analyses. In ML analyses, phylogenetic trees of nucleotide and amino acid sequences were constructed by PhyML v3.0 (Guindon et al., 2010) based on the best-fit model (as above) with 1000 replicates. For BI analysis MrBayes v3.2.6 (Huelsenbeck & Ronquist, 2001) was used to compute the probability distribution and get a sharper and stronger prediction; four simultaneous Markov chains ran for 10 million generations and sampled every 1000 generations after discarding the first 25% “burn-in” trees. Node support was assessed by the value
Results and Discussion

Gene content and nucleotide composition

The mitogenomes of *G. depressa thoracica* and *G. depressa depressa* were sequenced. The complete *G. depressa thoracica* mitogenome had 16, 109 bp, including 13 PCGs, 2 rRNAs and 22 tRNAs, and an A+T-rich region (Fig 1, Table 2). In the mitogenome of *G. depressa thoracica*, there were four intergenic spacers (a total length of 23 bp), ranging from 1 bp to 17 bp, and the longest intergenic spacer was located between trnSer(AGN) and nad1. Furthermore, 15 pairs of genes overlapped each other, with a length ranging from 1 to 17 bp.

The sequenced mitogenome of *G. depressa depressa* was partial and 14, 277 bp long, and was comprised of 13 PCGs, 21 tRNAs, complete rrnL and partial rrnS (Fig 1, Table 2). The whole A+T-rich region, trnI and partial rrnS failed to be sequenced although we have tried several pairs of species-specific primers. There were 3 intergenic spacers (23 bp), ranging from 1 bp to 17 bp, and the longest intergenic spacer was detected between trnSer(AGN) and nad1. Seventeen pairs of genes overlapped each other, with a length ranging from 1 to 20 bp.

Two 7-bp long overlaps (ATGATAA) were detected in *G. depressa depressa* and *G. depressa thoracica*, which were also found in many other Polyphaga insects (Fig 2). The overlaps were located between atp8 and atp6 on the H-strand and between nad4L and nad4 on the L-strand respectively. The overlapped sequences were thought to be translated as a bictron (Stewart & Beckenbach, 2005). Another 5 bp long motif sequence (TATA) was detected between trnSer(UCN) and nad1 in mitogenomes of *G. depressa depressa* and *G. depressa thoracica*, which was also present in other coleopterans (Fig 3). This consensus sequence has been proposed as the possible binding site of mtTERM because it is located at the end of the H-strand coding region in the circular mitogenome (Taunman, 1999). The sequenced motifs between atp8 and atp6, and between trnSer(UCN) and nad1 were relatively conserved in four suborders (41 Polyphaga, 24 Adephaga, 2 Archostemata and 2 Myxophaga) after mitogenomic comparisons. However, the motif ‘ATGATAA’ between nad4 and nad4L was only found in the mitogenomes of Polyphaga insects, while ‘ATGTTAA’ was identified in insect mitogenomes from the other three suborders, Adephaga, Archostemata and Myxophaga (Fig S1).

Protein-coding genes

In the mitogenomes of *G. depressa thoracica* and *G. depressa depressa*, a total of 3731 and 3678 amino acids were encoded by 11, 182 bp and 11, 026 bp nucleotides, respectively. All genes had the same orientations and organization. The 13 PCGs ranged from 156 bp (atp8) to 1722 bp (nad5) for *G. depressa thoracica* and from 156 bp (atp8) to 1728 bp (nad5) for *G. depressa depressa* (Table 2). Except for nad1 which used TGG as a start codon, all other PCGs had the typical ATN as the start codon. For example, nad2, cox1, cox2 and nad5 have ‘ATT’ as start codon, while the start codon ‘ATA’ was used for nad3 and nad6, ‘ATG’ for atp6, cox3, nad4, nad4L and cob, and ‘ATC’ for atp8 in *G. depressa thoracica* mitogenome. In *G. depressa depressa*, start codon ‘ATT’ was used in six genes (nad2, cox1, cox2, atp8, nad5 and nad6); ‘ATG’ for atp6, cox3, nad4, nad4L and cob, and ‘ATC’ for nad3. The stop codons for the
two mitogenomes were TAA, TAG and incomplete termination codons, like TA or T (Table 3). The incomplete stop codons could be completed as TAA through post-transcriptional polyadenylation (Ojala et al., 1981; Boore, 2004).

The A+T content, AT-skew and GC-skew of G. depressa thoracica, G. depressa depressa and three other Chrysomelidae were calculated, respectively. All PCGs have a high A+T percentage (70%) (Table 3). This indicated that PCGs have the high background mutational pressure toward AT nucleotides at the third codon position (Kim et al., 2014).

The relative synonymous codon usage (RSCU) showed that the most frequently used amino acids in these two mitogenomes were Leu, Ile, Phe and Met. TTA for Leu, ATT for Ile, TTT for Phe and ATA for Met were the most popular codons in G. depressa thoracica and G. depressa depressa mitogenomes (Fig 4, Fig 5). The sum of the most popular amino acids, like Leu, Ile, Phe and Met, varied from 38% (Agasicles hygrophila) to 42.23% (G. depressa depressa). All PCGs are rich with A or T nucleotides, which was also found in other coleopteran insects (Kim et al., 2009; Du et al., 2016).

The codon families in PCGs were encoded by H and L strands (Fig 6), respectively. The amino acid abundance of PCGs on H and L strands have different skewness, which led to an amino acid usage unbalance in PCGs. PCGs on the H-strand were more TA-skewed than CG-skewed, whereas the PCGs on the L-strand had a higher frequency of T and G, in which the first and second codons were skewed toward A nucleotide (Pons et al., 2010). Similarly, the third codon was also A/T biased (Fig 6). Overall, this nucleotide preference lead to 13 PCGs having the higher percentage of Leu and a lower percentage of Cys. However, L-strand encoded nad genes (Table S3) have a higher percentage of Cys than any genes encoded by H-strand. Four-fold degenerate codon usage was obviously A/T biased in the third position, and two-fold degenerate codon usage showed a similarly biased pattern, with A/T favored over C/G in the third position (Fig 6). Both patterns were in agreement with the AT-biased content exhibited by PCGs. The most obvious bias of amino acid usage was due to structural/functional requirements of PCGs, which is well represented by the distribution of the Cys codon family. The richness of Cys in mitochondrial NADHs helps to establish intra- and inter-chain disulfide bridges (S-S), which are essential structures (Mishmar et al., 2006).

**Transfer RNA genes**

Twenty-two tRNAs were found in G. depressa thoracica and ranged from 62 bp to 71 bp in length with 14 tRNAs on the H-strand and 8 on the L-strand. Except for trnSer(AGN), all tRNAs were folded into the typical clover-leaf structure (Fig S2), containing an amino acid acceptor arm (7 bp), TPC arm (3-5 bp), dihydourouridine (DHU) arm (3-4 bp), anticodon arm (3-5 bp) and a variable extra arm. In the mitogenome of G. depressa depressa, 21 tRNAs were identified (Fig S2) and the length of these genes were between 62 bp to 70 bp, with 13 and 8 tRNAs encoded by the H-strand and L-strand, respectively. Twenty tRNAs had the typical clover-leaf structures, containing an amino acid acceptor arm (7 bp), TPC arm (3-5 bp), DHU arm (3-5 bp), anticodon arm (3-5 bp) and a variable extra arm. However, trnSer(AGN) did not have the DHU arm, and formed a simple-loop structure, which is a common phenomenon in many insect mitogenomes (Wolstenholme, 1992).

In tRNAs, the typical DHU arm is encoded by 3-4 nucleotides; the AC arm and the TPC arm vary from...
3-5 bp, and the variable loops range from 4-6 bp. In general, the size of the variable loop and the D-loop usually affects the length of the tRNA (Navajas et al., 2002). In the comparison of both mitochondrial genomes (Fig S2), the major base pairs, A-T and G-C in stem regions are Watson-Crick pairs, but G-U wobble and mismatched pairs can also be observed. A total of 18 and 12 G-U wobbles were observed in the mitogenomes of *G. depressa thoracica* and *G. depressa depressa*, with 10, 4, 3, 1 and 3, 4, 4, 1 base pairs in AA stem, D stem, AC stem and T stem of the tRNAs, respectively. In these wobbles, three were in the D stems of *trnGln*, *trnPro* and *trnHis*, and one was in the AA stem of *trnCys*. Except for *trnlle*, three [one AG in *trnTrp*, and two UU in *trnLeu(UUR)* and *trnLeu(CUN)*] and two [one AG in *trnTrp* and one UU in *trnLeu(UUR)*] mismatched base pairs were found in the tRNAs of *G. depressa thoracica* and *G. depressa depressa* mitogenomes, respectively. Among the six mismatched base pairs, four were located in the amino acid acceptor arms, while one UU pair was located in the anticodon arm of *trnLeu(CUN)*. These mismatches, which are apparently deleterious, can be repaired and corrected by a putative RNA editing process (Masta & Boore, 2004). From protozoa to plants and metazoan animals, tRNA editing events have been observed in a variety of mitochondrial tRNAs (Laforest et al., 1997; Alfonzo et al., 1999; Leigh & Lang, 2004; Grewe et al., 2009; Knoop, 2011).

The percentage of identical nucleotides (%INUC) for both insects was also calculated (Table S4). *TrnD, trnN, trnR, trnE, trnW, trnS(UCN)* and *trnP* showed %INUC >90% in both insects, the first six of which were located on the H strand (Table 2). Only *trnF* and *trnL(CUN)* on the L strand exhibited %INUC <80%, indicating the level of conservation was positively H strand-biased.

**Phylogenetic analyses**

Phylogenetic trees were constructed based on 13 mitochondrial PCGs nucleotide sequences from 36 species (Fig 7A) and their corresponding amino acid sequences (Fig 7B) by maximum likelihood (ML) and Bayesian Inference (BI) analyses. Overall, BI analyses provided more resolution with strong supports. *G. depressa thoracica* and *G. depressa depressa* formed one clade (BPP/BS=1/100), and had a closer relationship with species from Galerucinae than from *G. intermedia* from Chrysomelinae.

The family Chrysomelidae was divided into three major clades in both nucleotide and amino acid sequence trees using ML and BI analyses (Fig 7). Chrysomelinae and Galerucinae consistently formed a clade, and the other four subfamilies, Bruchinae, Criocerinae, Donaciinae, and Spilopyrinae, were clustered in a clade. The phylogenetic analyses in this study were also well supported by previous analyses (Farrell, 1998; Reid, 2000; Farrell & Sequeira, 2004; Gómez-Zurita et al., 2007; Bocak et al., 2014; Song et al., 2017), in which the clade of Eumolpinae + Cryptocephalinae + Cassidinae were close to Galerucinae and Chrysomelinae, with the latter observed in all the phylogenetic trees.

In the phylogenetic analyses of amino acid sequences, Orsodacninae had a closer relationship with Lepturinae + Neoclydinae in the superfamily Cerambycidae. This implied that Orsodacninae might be evaluated outside Chrysomelidae (Haddad & Mckenna, 2016). Reid (1995) was also against placing Orsodacninae within the Cerambycidae based on mitochondrial nucleotide sequences analyses. The representative species from Lamiinae formed a clade in the four phylogenetic trees, while the interrelationships within Cerambycidae remained unclarified due to insufficient mitogenomes and genomes from these insects. Certainly, as more mitogenomes and genomes of insect species are available...
In databases, phylogenetic analyses will be more reliable and convincing.

**Conclusions**

In this study, the complete mitogenome of *G. depressa thoracica* (16, 109 bp) and the partial mitogenome of *G. depressa depressa* (14, 277 bp) were sequenced and analyzed. In the mitogenomes of *G. depressa thoracica* and *G. depressa depressa*, the overall mitochondrial gene order and orientation were identical with some exceptions. For example, the motifs, ‘TAGTA’ between *trnSer(UCN)* and *nad2*, and ‘ATGATAAA’ between *atp8* and *atp6*, were found to be more conserved than ‘ATGATAAA’ between *nad4* and *nad4L* in the mitogenomes of Polyphaga, which was ‘ATGTAAA’ in the Adephaga, Myxophaga and Archostemata. Phylogenetic analyses showed that *G. depressa thoracica* and *G. depressa depressa* consistently formed a clade. Within Chrysomeloidea, the sister relationships of ((Eumolpinae + Cryptocephalinae) + Cassidinae), and the close affinity of Galerucinae + Chrysomelinae were confirmed with high nodal supports. We believe that the mitogenomes of *G. depressa thoracica* and *G. depressa depressa* will be useful for further studies of molecular classification, and coleopteran mitogenome architecture and phylogenetics.

**References**


Hua YQ, Ding YR, Yan ZT, Si FL, Luo QC, Chen B. 2016. The complete mitochondrial genome of *Anopheles minimus* (Diptera: Culicidae) and the phylogenetics of known *Anopheles* mitogenomes. *Insect Science* 23:353–365


reveal the global phylogeography and dispersal routes of the migratory locust. *Molecular Ecology* 21:4344–4358


Rambaut A. 2009. FigTree v1.3.1. Available at http://tree.bio.ed.ac.uk/software/ (accessed on 31 September 2017)


