Complete Mitochondrial Genome of *Dinorhynchus dybowskyi* (Hemiptera: Pentatomidae: Asopinae) and Phylogenetic Analysis of Pentatomomorpha Species

Qing Zhao,1 Juan Wang,2 Meng-Qing Wang,2 Bo Cai,3 Hu-Fang Zhang,1,4 and Jiu-Feng Wei1

1Department of Entomology, Shanxi Agricultural University, Taigu, Shanxi 030801, China, 2Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing 100193, China, 3Hainan Entry-Exit Inspection and Quarantine Bureau, Haikou, Hainan 570311, China, and 4Corresponding author, e-mail: zh_hufang@sohu.com

Subject Editor: Igor Sharakhov

Received 18 December 2017; Editorial decision 6 March 2018

Abstract

*Dinorhynchus dybowskyi* (Hemiptera: Pentatomidae: Asopinae) is used as a biological control agent against various insect pests for its predatory. In the present study, the complete mitochondrial genome (mitogenome) of the species was sequenced using the next-generation sequencing technology. The results showed that the mitogenome is 15,952 bp long, including 13 protein-coding genes (PCGs), 22 transfer RNAs (tRNAs), two ribosomal RNAs (rRNAs), and a control region. Furthermore, the gene order and orientation of this mitogenome are identical to those of most heteropterans. There are 21 intergenic spacers (of length 1–28 bp) and 13 overlapping regions (of length 1–23 bp) throughout the genome. The control region is 1,291 bp long. The start codon of the PCGs is ATN, except *cox1* (TTG), and stop codon is TAA, except *nad1* (TAG). The 22 tRNAs exhibit a typical cloverleaf secondary structure, except *trnS1*, which lacks a dihydrouridine (DHU) arm and *trnV*, where the DHU arm forms a simple loop. The analyses based on nucleotide sequences of the 13 PCGs by Bayesian Inference and maximum likelihood methods. The results support the monophyly of five superfamilies Aradoidea, Pentatomoidae, Pyrrhocoroidea, Lygaeoidea, and Coreoidea. Within Pentatomoidae, the relationship observed is as follows: (Plataspidae + Urostylididae) + (Pentatomidae + (Acanthosomatidae + (Cydnidae + (Scutelleridae + (Dinidoridae + Tessaratomidae))))), and *D. dybowskyi* was placed in Pentatomidae and close to *Eurydema gebleri*.

Key Words: *Dinorhynchus dybowskyi*, mitogenome, next-generation sequencing, phylogenetic analysis, biological control

The mitochondrial genome (mitogenome) of an insect is a circular double-stranded DNA molecule, 14–20 kb in size. Generally, it contains 13 protein-coding genes (PCGs), 22 transfer RNA (tRNA) genes, two ribosomal RNA genes (rRNAs) (12S rRNA and 16S rRNA), and a control region (Boore 1999, Cameron 2008). Because of the high content of A + T, the control region is also known as the (A + T)-rich region. However, recent studies have reported that the control region does not always contain the highest content of A + T in the mitochondrial genome. Therefore, it is not recommended to associate the term (A + T)-rich area with the control region (Hua et al. 2008). During recent decades, the mitochondrial genome of insects has been extensively used in population genetics, molecular phylogeography, phylogenetic analysis, and evolutionary biology (Simon et al. 2006, Cameron 2014, Yuan et al. 2015a, Wang et al. 2017a, Zhu et al. 2017).

Pentatomomorpha is one of the most common and important groups of Heteroptera, and distributed worldwide. It comprises more than 14,000 known species belonging to 40 families (Weirauch and Schuh 2011, Rider 2015). Most Pentatomomorpha species are economically important as agricultural pests, such as *Tessaratomina papillosa* Drury (Hemiptera: Heteroptera: Pentatomomorpha: Tessaratomidae) and *Poecilocoris latus* Dallas (Hemiptera: Heteroptera: Pentatomomorpha: Scutelleridae). However, some are predatory and usually used as biological control agents, such as the Asopinae species (Hemiptera: Pentatomomorpha) (Panizzí et al. 2000, De Clercq et al. 2003, Castro et al. 2015). Studies on the phylogenetic relationships within Pentatomomorpha and Pentatomomorpha have used morphological characters (Gaspard 1991), mitochondrial genomic data (Hua et al. 2008, Yuan et al. 2015b, Wu et al. 2017), nuclear ribosomal RNA (Lis et al. 2017), or a combination of all the data (Grazia et al. 2008).

Until now, only 47 complete or nearly complete mitogenomes of Pentatomomorpha and 17 complete mitogenomes of Pentatomomorpha have been sequenced (GenBank, 1 November 2017) (Accession no. in Table 1). In particular, the mitogenome of just one Asopinae species *Picromerus griseus* has been sequenced completely (Zhao et al. 2015a, Yuan et al. 2015b, Wang et al. 2017a).
Therefore, sequencing the mitogenome of Asopinae species is essential to understand the evolution of Pentatomomorpha and Pentatomomorpha at the genomic level.

As the insects rapidly develop resistance to pesticides, biological control has been increasingly employed in agriculture and forestry (Chang et al. 2003). *Dimorhinychus dybowskyi* Jakovlev, belonging to the subfamily Asopinae (Hemiptera: Pentatomidae), feeds on the larvae of Lepidoptera and Coleoptera. Furthermore, the species has been reported to prey on the fifth instar larva of *Anthera pernyi* (Guerin-Meneville 1855) (Lepidoptera: Bombycoidea: Saturniidae) (Yang 1962). These observations indicate that *D. dybowskyi* can be used as a biological control agent. In the present study, the mitochondrial genome of *D. dybowskyi* was sequenced and annotated. Furthermore, the results were used to elucidate the phylogenetic position of this species (Gapud 1991, Zhao et al. 2017a).

### Materials and Methods

**Sample Collection, DNA Extraction, and PCR Amplification**

The specimens of *D. dybowskyi* were collected in Huoshankou Forest Park (44.08° N, 128.73° E), Ning’an County, Heilongjiang Province, China, on 10 August 2015. The genomic DNA was sequenced and annotated.

**Table 1. List of species used to construct the phylogenetic tree**

| Infraorder   | Superfamily       | Family          | Species                  | Genbank no. | Length  |
|--------------|-------------------|-----------------|--------------------------|-------------|---------|
| Pentatomomorpha | Aradoidea         | Aradida         | Aneurus similis          | JQ780816    | 16,477  |
| Pentatomomorpha | Aradoidea         | Aradida         | Aneurus sublobatus       | NC_030361   | 16,091  |
| Pentatomomorpha | Aradoidea         | Aradida         | Aradacantha beissi       | HQ441233    | 15,528  |
| Pentatomomorpha | Aradoidea         | Aradida         | Aratus sp.               | JQ780818    | 16,814  |
| Pentatomomorpha | Aradoidea         | Aradida         | Baeotrychynchus hsaao     | HQ441232    | 15,250  |
| Pentatomomorpha | Aradoidea         | Aradida         | Libiocoris beissi        | NC_030363   | 15,168  |
| Pentatomomorpha | Aradoidea         | Aradida         | Neuroctenus parus        | NC_012459   | 15,354  |
| Pentatomomorpha | Coreoidea         | Coreida         | Leptocoris sp.           | KM244663    | 15,322  |
| Pentatomomorpha | Coreoidea         | Coreida         | Riportortus pedestris     | NC_012462   | 17,191  |
| Pentatomomorpha | Coreoidea         | Coreida         | Anoplocnemis curvipes    | KY906099    | 16,345  |
| Pentatomomorpha | Coreoidea         | Coreida         | Clavigralla tomentosicollis | KY274846   | 16,089  |
| Pentatomomorpha | Coreoidea         | Coreida         | Hydaropsis longirostris | NC_012456   | 16,321  |
| Pentatomomorpha | Coreoidea         | Coreida         | Aescryntelus notatus      | NC_012446   | 14,532  |
| Pentatomomorpha | Coreoidea         | Coreida         | Cortica sp. 'albomarginatus' | KM983397   | 14,989  |
| Pentatomomorpha | Coreoidea         | Coreida         | Stictopleurois subervidus | NC_012488   | 15,319  |
| Pentatomomorpha | Lygaeoidea        | Berytidae       | Yennymalus paralellus     | NC_012464   | 15,747  |
| Pentatomomorpha | Lygaeoidea        | Lygaeidae       | Phaenacantha maricina     | NC_012460   | 14,540  |
| Pentatomomorpha | Lygaeoidea        | Lygaeidae       | Gecoris pallidennis       | EU427336    | 14,592  |
| Pentatomomorpha | Lygaeoidea        | Lygaeidae       | Kleidocryx vexatae       | KJ84365     | 14,688  |
| Pentatomomorpha | Lygaeoidea        | Malidae         | Chauliota fallax         | NC_020772   | 15,739  |
| Pentatomomorpha | Lygaeoidea        | Malidae         | Malacis inconspicuus      | NC_012458   | 15,575  |
| Pentatomomorpha | Lygaeoidea        | Rhyparochromidae | Panaoros albomaculatus    | NC_031364   | 16,345  |
| Pentatomomorpha | Pentatomoidea     | Acanthosomatidae| Acanthosoma labiduraeides | JQ746370    | 16,678  |
| Pentatomomorpha | Pentatomoidea     | Acanthosomatidae| Sastrapaga edessoides     | JQ746376    | 16,358  |
| Pentatomomorpha | Pentatomoidea     | Cydnidae        | Macrocyctus gibbulus      | NC_012457   | 14,620  |
| Pentatomomorpha | Pentatomoidea     | Dinorididae     | Cordius chinensis         | JQ739179    | 14,648  |
| Pentatomomorpha | Pentatomoidea     | Dinorididae     | Megynenrum brevicorne     | JQ739181    | 14,584  |
| Pentatomomorpha | Pentatomoidea     | Pentatomidae    | Dolyocoris baccarum       | KJ07135     | 15,976  |
| Pentatomomorpha | Pentatomoidea     | Pentatomidae    | Dimorhinychus dybowskyi   | MG450552    | 15,952  |
| Pentatomomorpha | Pentatomoidea     | Pentatomidae    | Erthesina fullo           | JQ746373    | 14,611  |
| Pentatomomorpha | Pentatomoidea     | Pentatomidae    | Eurydemia gebleri         | NC_027489   | 16,005  |
| Pentatomomorpha | Pentatomoidea     | Pentatomidae    | Graphosoma rubrolineatum  | NC_033875   | 15,633  |
| Pentatomomorpha | Pentatomoidea     | Pentatomidae    | Halymorpha halys          | FJ683650    | 16,518  |
| Pentatomomorpha | Pentatomoidea     | Pentatomidae    | Nezana viridula           | NC_011755   | 16,889  |
| Pentatomomorpha | Pentatomoidea     | Pentatomidae    | Pentatomidae sp.          | KM244699    | 15,498  |
| Pentatomomorpha | Pentatomoidea     | Pentatomidae    | Rabicenia intermedia      | KP207596    | 14,967  |
| Pentatomomorpha | Pentatomoidea     | Plataspidae     | Copotosoma bifaria        | NC_012449   | 16,179  |
| Pentatomomorpha | Pentatomoidea     | Plataspidae     | Magacopta crinitaria      | NC_013542   | 15,647  |
| Pentatomomorpha | Pentatomoidea     | Scutelleridae   | Encorysidae grandid       | JQ746371    | 14,611  |
| Pentatomomorpha | Pentatomoidea     | Scutelleridae   | Lamprocoris sp.           | JQ746374    | 16,143  |
| Pentatomomorpha | Pentatomoidea     | Scutelleridae   | Poreciocoris nepalensis   | JQ746375    | 14,677  |
| Pentatomomorpha | Pentatomoidea     | Tessaratomidae  | Dalcantha dilatata        | JQ910981    | 15,350  |
| Pentatomomorpha | Pentatomoidea     | Tessaratomidae  | Eusthenes currex          | NC_022449   | 16,229  |
| Pentatomomorpha | Pentatomoidea     | Urostylidae     | Urostylis sp.             | JQ746379    | 15,582  |
| Pentatomomorpha | Pentatomoidea     | Urostylidae     | Urochela quadrinodotata   | NC_020144   | 15,687  |
| Pentatomomorpha | Pyrrhocoridea     | Largidae        | Physopelta gutta          | NC_012432   | 14,935  |
| Pyrrhocoridea   | Pyrrhocoridea     | Largidae        | Dysdercus cingulatus      | NC_012421   | 16,249  |

Cimicomorpha

Phymatidae

Phymata americana

NC036011

Cimicomorpha

Phymatidae

Carcinocoris binghami

NC036012
extracted from the thoracic muscle of a single specimen using the Genomic DNA Extraction Kit (Sangon Biotech, Shanghai, China), following instruction of the manufacturer.

The mitochondrial genome of *D. dybowskyi* was sequenced using the next-generation sequencing technology. By the whole-genome shotgun method, paired-end libraries were constructed and sequenced on an Illumina MiSeq platform at the Personal Biotechnology Company (Shanghai, China). The target insert size was 500 bp. The adapter sequences were removed and low-quality bases were trimmed using the Trimmomatic version 0.36 (Bolger et al. 2014). These targeted sequences were assembled using the A5-miseq v2015022 (Coil et al. 2015) and Spades v3.9.0 (Bankevich et al. 2012) software.

**Genome Annotation and Sequence Analysis**

The sequences were assembled using the Geneious version 9.1.4 software (Kearse et al. 2012). PCGs boundaries were identified with the ORF finder (http://www.ncbi.nlm.nih.gov/orf/gorf.html). After sequencing, the mitogenome was annotated manually and by automated methods. The automated annotation was accomplished using MITOS (Bertol et al. 2013). The confirmation of the tRNA genes was determined using the tRNAscan-SE program (http://lowelab.ucsc.edu/tRNAscan-SE/) (Lowe and Eddy 1997). Furthermore, the unidentified tRNAs were compared with sequences from other species. The secondary structures of *rrnL* and *rrnS* was determined using the Mfold Web Server (http://mfold.rna.albany.edu/?q=mfold) and based on the genome of *Eurydema gebleri* (Yuan et al. 2015a). The control region was identified through the boundary of the neighboring genes.

Comparative analysis and spread correction were performed with Mega 6.0 software (Tamura et al. 2013), to obtain the complete mitogenome of *D. dybowskyi*. The codon usage of the 13 PCGs was calculated using Mega 6.0. The AT- and GC-skews were computed using the following formulas: AT-skew = (A% - T%) / (A% + T%) and GC-skew = (G% - C%) / (G% + C%) (Hassanin et al. 2005). The related species were analyzed using the Basic Local Alignment Search Tool (BLAST) searches on the National Center for Biotechnology Information (NCBI) database.

**Phylogenetic Analysis**

The phylogenetic analysis was carried out using the mitochondrial genome of *D. dybowskyi*, 47 Pentatomomorpha taxa, and two Cimicomorpha species (Table 1). The nucleic acids of 13 PCGs were extracted and aligned using Geneious and Mega 6.0, respectively. Subsequently, the 13 mitochondrial PCGs were aligned using MUSCLE (Edgar 2004) with default settings, and the resulting alignments were concatenated into a nucleotide matrix.

The concatenated set of nucleotide sequences were used in the phylogenetic analysis by the Bayesian inference (BI) and maximum likelihood (ML) methods. The optimal partitioning schemes and

---

### Table 2. The best schemes of partition and substitution models used for each partition

| Optimal partition | Model | Initial partition |
|-------------------|-------|-------------------|
| Partition 1       | GTR+I+G | atp6-1, nad3-1    |
| Partition 2       | GTR+I+G | atp6-2, cox2-2, cox3-2, cytb-2, nad3-2 |
| Partition 3       | GTR+G   | atp6-3, atp8-3    |
| Partition 4       | GTR+I+G | atp8-1, nad2-1, nad6-1 |
| Partition 5       | GTR+I+G | atp8-2, nad2-2, nad6-2 |
| Partition 6       | GTR+I+G | cox1-1            |
| Partition 7       | GTR+I+G | cox1-2            |
| Partition 8       | GTR+I+G | cox1-3, cox2-3, cox3-3, cytb-3, nad3-3 |
| Partition 9       | GTR+I+G | cox2-1, cox3-1, cytb-1 |
| Partition 10      | GTR+I+G | nad1-1, nad4-1, nad4l-1, nad5-1 |
| Partition 11      | GTR+I+G | nad1-2, nad4-2, nad4l-2, nad5-2 |
| Partition 12      | HKY+I+G | nad1-3, nad4l-3, nad5-3 |
| Partition 13      | HKY+G   | nad2-3, nad6-3    |
| Partition 14      | GTR+I+G | nad4-3            |

---

*Fig. 1. Mitochondrial genome map of *D. dybowskyi*. Protein coding and ribosomal genes are shown with standard abbreviations. The gene sequence is located in the outside of the circle, and the protein sequence is located in the inner circle.*
corresponding nucleotide substitution models for each dataset were determined using PartitionFinder v1.1.1 (Lanfear et al. 2012) and were used in the subsequent phylogenetic analyses (Table 2). The BI analysis was conducted using MrBayes 3.2.5 (Ronquist et al. 2012), with four (three heated and one cold) independent Markov chains run for 10,000,000 metropolis-coupled Markov chain Monte Carlo iterations. Table 3

### Table 3. Summary of the *D. dybowskyi* mitogenome

| Feature | Strand | Position | Length (bp) | Initiation codon | Stop codon | Anticodon | Intergenic nucleotide |
|---------|--------|----------|-------------|------------------|------------|-----------|-----------------------|
| trnI    | N      | 1–66     | 66          | GAT              | −3         |           |                       |
| trnQ    | J      | 64–132   | 69          | TTG              | 14         |           |                       |
| trnM    | N      | 147–213  | 67          | CAT              | 22         |           |                       |
| nad2    | N      | 236–1,192| 957         | ATT              | TAA        | 16        |                       |
| trnW    | N      | 1,209–1,273| 65         | TCA              | −8         |           |                       |
| trnC    | J      | 1,266–1,331| 66         | GCA              | 15         |           |                       |
| trnY    | J      | 1,347–1,411| 65         | GTA              | 5          |           |                       |
| coxl    | N      | 1,417–2,958| 1,542      | TTG              | TAA        | −5        |                       |
| trnl2   | N      | 2,954–3,018| 65         | TAA              | 18         |           |                       |
| cox2    | N      | 3,037–3,720| 684        | ATT              | TAA        | −23       |                       |
| trnk    | N      | 3,698–3,770| 73         | CTG              | 5          |           |                       |
| trnD    | N      | 3,776–3,843| 68         | GTC              | 9          |           |                       |
| atp8    | N      | 3,853–4,002| 150        | ATA              | TAA        | −7        |                       |
| atp6    | N      | 3,996–4,670| 675        | ATG              | TAA        | 8         |                       |
| cox3    | N      | 4,679–5,467| 789        | ATG              | TAA        | 5         |                       |
| trnG    | N      | 5,473–5,534| 62         | TCC              | 21         |           |                       |
| nad3    | N      | 5,556–5,900| 345        | ATA              | TAA        | −14       |                       |
| trnA    | N      | 5,887–5,951| 65         | TGC              | 3          |           |                       |
| trnR    | N      | 5,955–6,020| 66         | TCG              | 7          |           |                       |
| trnN    | N      | 6,028–6,096| 69         | GTT              | −1         |           |                       |
| trnS1   | N      | 6,096–6,163| 68         | GCT              | −1         |           |                       |
| trnE    | N      | 6,163–6,230| 68         | TTC              | −2         |           |                       |
| trnF    | J      | 6,229–6,293| 65         | GAA              |           |           |                       |
| nad5    | J      | 6,294–8,003| 1,710      | ATT              | TAA        | 1         |                       |
| trnh    | J      | 8,005–8,070| 66         | GTG              | 3          |           |                       |
| nad4    | J      | 8,074–9,402| 1,329      | ATG              | TAA        | −7        |                       |
| nad4l   | J      | 9,396–9,683| 288        | ATG              | TAA        | 2         |                       |
| trnt    | N      | 9,686–9,751| 66         | TGT              |           |           |                       |
| trnp    | J      | 9,752–9,817| 66         | TGG              | 3          |           |                       |
| nad6    | N      | 9,821–10,306| 486       | ATG              | TAA        | −8        |                       |
| cytB    | N      | 10,299–11,435| 1,137     | ATG              | TAA        | 3         |                       |
| trnS2   | N      | 11,439–11,508| 70        | TGA              | 28         |           |                       |
| nad1    | J      | 11,537–12,463| 927      | ATA              | TAG        | −6        |                       |
| trnl1   | J      | 12,458–12,523| 66        | TAG              | −3         |           |                       |
| rrnL    | J      | 12,521–13,773| 1,353    | TAC              | 1          |           |                       |
| trnV    | J      | 13,797–13,863| 67        | TAC              | 120        |           |                       |
| rrnS    | J      | 13,865–14,661| 797      | TAC              | 1290       |           |                       |

### Table 4. Base composition of the *D. dybowskyi* mitogenome

| Region | A% | C% | G% | T% | A+T% | G+C% | AT skew | GC skew |
|--------|----|----|----|----|------|------|---------|---------|
| Whole genome | 41.34 | 14.07 | 10.83 | 33.76 | 75.10 | 24.90 | 0.10 | −0.13 |
| nad2   | 40.86 | 11.29 | 10.03 | 37.83 | 78.68 | 21.32 | 0.04 | −0.06 |
| cox1   | 33.53 | 16.47 | 14.92 | 35.08 | 68.61 | 31.39 | −0.02 | −0.05 |
| cox2   | 40.35 | 15.06 | 12.28 | 32.31 | 72.66 | 27.34 | 0.11 | −0.10 |
| atp8   | 43.33 | 10.00 | 8.67 | 38.00 | 81.33 | 18.67 | 0.07 | −0.07 |
| atp6   | 36.74 | 14.52 | 10.37 | 38.37 | 75.11 | 24.89 | −0.02 | −0.17 |
| cox3   | 36.25 | 15.34 | 14.07 | 34.35 | 70.60 | 29.40 | 0.03 | −0.04 |
| nad3   | 36.23 | 13.91 | 13.62 | 36.23 | 72.46 | 27.54 | 0.00 | −0.01 |
| nad5   | 28.65 | 10.76 | 12.98 | 47.60 | 76.26 | 23.74 | −0.25 | 0.09 |
| nad4   | 27.39 | 11.96 | 11.81 | 48.83 | 76.22 | 23.78 | −0.28 | −0.01 |
| nad4l  | 25.69 | 8.68 | 13.54 | 52.08 | 77.78 | 22.22 | −0.34 | 0.22 |
| cytB   | 39.51 | 11.93 | 8.23 | 40.33 | 79.84 | 20.16 | −0.01 | −0.18 |
| nad1   | 34.92 | 16.09 | 12.31 | 36.68 | 71.59 | 28.41 | −0.03 | −0.13 |
| rrnL   | 27.40 | 8.95 | 14.46 | 49.19 | 76.39 | 23.41 | −0.29 | 0.24 |
| rrnS   | 33.00 | 8.28 | 14.93 | 43.79 | 76.79 | 23.21 | −0.14 | 0.29 |
Carlo generations; sampling trees every 100 generations. The first 25% of samples were discarded as burn-in and the remaining trees were used to calculate posterior probabilities in a 50% majority rule consensus tree. The ML analysis was conducted with RAxML v 8.0.2 (Stamatakis 2015), using the best-fit model presented by PartitionFinder, and node confidence was assessed with 1,000 bootstrap replications.

Results

Mitochondrial Genomic Structure
The complete mitogenome of *D. dybowskyi* is a circular double-stranded molecule of length 15,952 bp (GenBank accession number MG450552) with high A + T nucleotide content (41.34% A, 33.76% T, 14.07% C, and 10.83% G) similar to that of other hemipteran mitogenomes (Wang et al. 2017b, Zhao et al. 2017a, Zhao et al. 2017b). It contains 13 PCGs, 22 tRNAs, two rRNAs (rrnL and rrnS), and a non-coding control region. (Fig. 1, Table 3). The order and orientation of the mitochondrial genes are identical to those of most true bugs, and this is considered to be an ancestral arrangement (Hua et al. 2008, Yuan et al. 2015a, Zhao et al. 2017a). The AT- and GC-skews of the *D. dybowskyi* mitogenome are 0.101 and -0.130, respectively, indicating that the content of A + T nucleotides is higher than that of G + C nucleotides. The A + T contents of the 13 PCGs was compared; the lowest and highest A + T contents were 68.61% (*cox1*) and 81.33% (*atp8*), respectively (Table 4). The nucleotide composition and high skewness of the *D. dybowskyi* mitogenome was also reflected in the codon usage of the PCGs.

Comparative Analysis of Protein-Coding Genes
In the *D. dybowskyi* mitogenome, nine PCGs are coded on the J-strand (majority strand) and four PCGs are coded on the N-strand (minority strand). The start codon of most PCGs is ATN, except *cox1* (TTG). This unconventional start codon has also been reported in other heteropterans (Hua et al. 2008; Zhao et al. 2017a,b). Furthermore, the stop codon of most PCGs is TAA, except *nad1* (TAG).

Excluding the start and termination codons, the 13 PCGs consist of 3,673 codons. The most abundant amino acid codons are UUU (F), UUA (L), AUU (I), and AUA (M), which constitute 33.24% of the total amino acid codons. The content of A + T was higher than that of G + C (Fig. 2). Furthermore, the codon usage values reflected a significant bias toward A and T nucleotides (Table 5).

Transfer and Ribosomal RNAs
The mitogenome of *D. dybowskyi* consists of 22 tRNAs of size 62–73 bp. Eight tRNA genes (*trnQ*, *trnC*, *trnY*, *trnF*, *trnH*, *trnP*, *trnL*, *trnM*, Table 5). The content of A + T was higher than that of G + C (Fig. 2). Furthermore, the codon usage values reflected a significant bias toward A and T nucleotides (Table 5).

![Codon usage in the *D. dybowskyi* mitogenome.](image)

**Table 5.** Codon usage in the mitochondrial genome of *D. dybowskyi*

| Codon  | Count | RSCU | Codon  | Count | RSCU | Codon  | Count | RSCU | Codon  | Count | RSCU | Codon  | Count | RSCU |
|--------|-------|------|--------|-------|------|--------|-------|------|--------|-------|------|--------|-------|------|
| UUU(F) | 242   | 1.62 | UCU(S) | 103   | 2.23 | UAU(Y) | 151   | 1.62 | UGU(C) | 44    | 1.8  |
| UUC(F) | 56    | 0.38 | UCC(S) | 28    | 0.61 | UAC(Y) | 35    | 0.38 | UGC(C) | 5     | 0.2  |
| UUA(L) | 357   | 4.22 | UCA(S) | 92    | 1.99 | UAA(*) | 12    | 1.85 | UGA(W) | 81    | 1.67 |
| UUG(L) | 34    | 0.4  | UCG(S) | 4     | 0.09 | UAG(*) | 1     | 0.15 | UGG(W) | 16    | 0.33 |
| CUU(L) | 47    | 0.56 | CCA(P) | 63    | 1.92 | CAU(H) | 63    | 1.62 | CGU(R) | 34    | 2.57 |
| CUC(L) | 8     | 0.09 | CCC(P) | 27    | 0.82 | CAC(H) | 15    | 0.38 | GCG(G) | 8     | 0.16 |
| CUA(L) | 54    | 0.64 | CCA(P) | 37    | 1.13 | CAA(Q) | 43    | 1.72 | GAG(E) | 7     | 0.28 |
| CUG(L) | 8     | 0.09 | CCG(P) | 4     | 0.12 | CAG(Q) | 7     | 0.28 | GGG(G) | 1     | 0.08 |
| AUU(L) | 329   | 1.69 | ACC(T) | 76    | 1.69 | AUA(N) | 154   | 1.71 | AGG(S) | 38    | 0.82 |
| AUC(L) | 61    | 0.31 | ACC(T) | 18    | 0.4  | AAC(N) | 26    | 0.29 | AGC(G) | 13    | 0.28 |
| AUA(M) | 293   | 1.81 | ACA(T) | 83    | 1.84 | AAA(K) | 87    | 1.6  | AGA(S) | 89    | 1.93 |
| AUG(M) | 30    | 0.19 | ACG(T) | 3     | 0.07 | AAG(K) | 22    | 0.4  | AGG(S) | 2     | 0.04 |
| GUU(V) | 79    | 1.88 | GCU(A) | 51    | 1.51 | GAU(D) | 57    | 1.65 | GGU(G) | 57    | 1.15 |
| GUC(V) | 2     | 0.05 | GCC(A) | 21    | 0.62 | GAC(D) | 12    | 0.35 | GGC(G) | 8     | 0.16 |
| GUA(V) | 81    | 1.93 | GCA(A) | 57    | 1.69 | GAA(E) | 76    | 1.71 | GGA(G) | 105   | 2.12 |
| GUG(V) | 6     | 0.14 | GCG(A) | 6     | 0.18 | GAG(E) | 13    | 0.29 | GGG(G) | 28    | 0.57 |
Fig. 3. Predicted secondary structure of tRNAs gene in the D. dybowskyi mitogenome.

Fig. 4. Predicted secondary structure of the rrnL gene in the D. dybowskyi mitogenome.
The tRNAs (trnL1, and trnV) are located on the J-strand and the remaining 14 tRNA genes are located on the N-strand (Table 3). Nineteen tRNAs have a typical cloverleaf secondary structure, except trnS1, which lacks a dihydrouridine (DHU) arm; and trnV, where the DHU arm forms a simple loop (Fig. 3). There are 21 intergenic spacer regions, with a total length of 212 bp, and the largest spacer (28 bp) region is located between nad1 and trnS1. There are 13 intergenic overlapping regions of size 1–23 bp, with a total length 36 bp, and the largest overlapping region is located between cox2 and trnA. The total length of the 22 tRNAs is 1,478 bp, and these anticodons have a high A + T content (81.3%). In tRNAs, the AT- and GC-skews are positive and negative, respectively.

Two rRNAs are located on the J-strand. The rrnL gene is 1,290 bp long with an A + T content of 78.29%, while the rrnS gene is 802 bp long with an A + T content of 78.35%. The rrnL and rrnS genes are separated from each other by trnV. For rRNAs, the AT-skews are negative and the GC-skews are positive. The A + T content of rrnL and rrnS is 79.09% and 76.79%, respectively (Table 4). The complete secondary structure of rrnL and rrnS was determined (Figs. 4 and 5).

Control Region

The control region regulates the replication and transcription of mitognome (Zhang and Hewitt 1997). In some arthropod mitogenomes, the control region has been found in several or all of the following motifs: the tandem repeat sequences, a long sequence of Ts, an (A + T)-rich region, and a stem-loop structure (Cook 2005). In the mitogenome of D. dybowskyi, the control region is located between rrnS and trnI, with a length of 1,291 bp and a high A + T content (73%)—A 36.3%,
Fig. 6. The control region of *D. dybowskyi* mitogenome. (A) Structure of the control region. (B) A potential stem-loop structure found in the control region.

Fig. 7. Inferred phylogenetic relationships among Pentatomomorpha based on the concatenated nucleotide sequences of 13 mitochondrial protein-coding genes using Bayesian Inference (BI). Numbers on branches are Bayesian posterior probabilities.
T 36.6%, C 17.0%, and G 10%. The AT- and GC-skews are negative in the control region, indicating that the content of T and C is marginally higher than that of A and G. The tandem repeat sequences consist of four types of repeat units. The longest is type I, which is approximately 460 bp long, and this is interrupted by a 45 bp non-coding region. The remaining three tandem repeats are adjacent to each other, and their length ranged from 28 to 68 bp (Fig. 6A). A potential stem-loop structure of approximately 30 bp is present in the control region, and no conserved functional motifs were identified (Fig. 6B).

**Phylogenetic Relationships**

The phylogenetic analyses were performed by two inference methods—BL and ML. The phylogenetic relationships among Pentatomomorpha were reconstructed, and the result based on the 13 PCGs strongly supports the relationship between five superfamilies: Aradoidea + (Pentatomoidea + (Lygaeoidea + (Pyrrhocoroidea + Coreoidea))) (Figs. 7 and 8). In Pentatomomorpha, Aradoidea, as a sister group of Trichophora, was located at the base of the phylogenetic tree. Furthermore, Trichophora was divided into two clades, one clade consisted of Pentatomoidea, which is monophyletic, while the other consisted of Coreoidea, Pyrrhocoroidea, and Lygaeoidea. According to the traditional taxonomic placement, Pyrrhocoroidea and Coreoidea were indicated as sister groups. In Pentatomomorpha, Urostylididae and Plataspidae were placed as sister groups, and were separated from the other families of Pentatomoidea. Pentatomidae was grouped as sister to the remainder of Pentatomoidea.

**Fig. 8.** Inferred phylogenetic relationships among Pentatomomorpha based on the concatenated nucleotide sequences of 13 mitochondrial protein-coding genes using maximum likelihood (ML). Numbers on branches are bootstrap percentages.
Furthermore, Acanthosomatidae and Scutelleridae were monophyletic, and Dinidioridae and Tessaratomidae were placed as sister groups. *D. dybowskyi* was placed in Pentatomidae and close to *E. gelberii*.

**Discussion**

To the best of our knowledge, the present study is the first to sequence and annotate the complete mitogenome of *D. dybowskyi*. A comparative analysis of 47 Pentatomomorpha mitogenomes showed that the gene content, gene arrangement, base composition, codon usage, and RNA structures are highly conserved in Pentatomomorpha, especially within family (Hua et al. 2008, Lee et al. 2009, Dai et al. 2012, Song et al. 2013, Yuan et al. 2015a).

The mitochondrial genome of *D. dybowskyi* is 15,952 bp long, which is within the range of the mitogenome of Pentatomomorpha genomes (14,532 bp in *Rhopalus latus* (Jakovev) to 17,191 bp in *Riptortus pedestris* (Fabricius)). The mitochondrial genome of *D. dybowskyi* resembles that of known ancestral species in terms of structural organization and composition (Hua et al. 2008, Wang et al. 2017b, Zhao et al. 2017b).

The most frequently occurring start codon among the 12 PCGs is ATN, the exception is TTG in **cox1** of most pentatomid mitogenomes (Dai et al. 2012, Shi et al. 2012, Li et al. 2013, Yuan et al. 2015a, Zhao et al. 2017b). However, in *D. dybowskyi*, the start codons ATT, ATG, and ATA occur with the same frequency, the start codon of **cox1** is TTG. Most PCGs end with TAA; however, in some species, **nad1**, **cox2**, or some other genes end with TAG or a single T (Liu et al. 2012, Song et al. 2013, Yuan et al. 2015b). The incomplete termination codons have been reported to be completed by posttranscriptional polyadenylation (Anderson et al. 1981). Furthermore, it is important that incomplete stop codons undergo completion to TAA during the mRNA assembly (Boore 1999).

In many insects, the majority of tRNAs have a canonical cloverleaf secondary structure; however, abnormal tRNAs have been reported in the mitogenome of pentatomoids. For example, **trnS1** and **trnV** sometimes lacks the stem of the DHU arm (Dai et al. 2012, Shi et al. 2012, Cameron 2014, Yuan et al. 2015a) or the DHU arm forms a loop in these genes (Lee et al. 2009; Li et al. 2013; Zhao et al. 2017a,b). In the mitogenome of *D. dybowskyi*, **trnS1** lacks the stem of the DHU arm, and in **trnV**, the DHU arm forms a simple loop.

The length of the common overlap regions vary among families or species (1–44 bp) (Li et al. 2014), and this variation is also present in the mitogenome of *D. dybowskyi*. Furthermore, the position of the maximum overlap is not conserved. For instance, *Aradacanthia heissi* has the maximum overlap between **trnT** and **trnP** (Shi et al. 2012), *Eusthenes cupreus* has the maximum overlap between **cox3** and **trnG** (Song et al. 2013), *Coridius chimensis* has the maximum overlap between **cox3** and **trnG** (Liu et al. 2012), and *Urochela quadrimotata* has the maximum overlap between **trnW** and **trnC** (Dai et al. 2012). The size of the common non-coding regions in the mitogenome ranges from 1 to 28 bp, and this region is not conserved, as well. A special non-coding region, which is also the longest non-coding region, is located between **rrnS** and **trnl**. The length of the control region ranges from 224 bp (Largidae) to 2,400 bp (Alydidae). Thus, the variation in the size of mitogenomes is mainly due to the non-coding regions (Hua et al. 2008).

In the present study, the phylogenetic analyses based on 13 PCGs strongly support the relationship between five superfamilies: Aradoidea + (Pentatomomoeida + (Lygaeoidea + (Pyrrhocoroidea + Coreoidea))). This is in concordance with the findings of previous phylogenetic studies based on morphological and molecular data (Grazia et al. 2008, Henry 1997, Hua et al. 2008, Yuan et al. 2015a). Within Pentatomoidea, the results of the present study support the relationship of (Plataspidae + Urostylididae) + (Pentatomidae + (Acanthosomatidae + (Cydnidae + (Scutelleridae + (Dinidioridae + Tessaratomidae))))) and this is consistent with the findings of previous studies based on morphological and molecular data (Gapud 1991; Grazia 2008; Lis et al. 2012, 2017).

The species *Graphosoma rubrolineta*, which belongs to the subfamily Podopinae, and *D. dybowskyi*, which represents the subfamily Asopinae, were mix with the other Pentatomidae species. This suggests that each subfamily in Pentatomidae might not be monophyletic. Therefore, further studies are necessary to sequence the genome of other species of Asopinae and other subfamilies, which will enhance our understanding of the molecular phylogeny of Pentatomoidea.

The phylogenetic relationships within Pentatomomorpha and Pentatomoidea based on mitochondrial genomic data are consistent with those based on the traditional morphological classification. This indicates that the mitochondrial genomics is a useful tool to elucidate the phylogenetic relationships at taxonomic levels. As more mitogenomes are sequenced at various taxonomic levels, especially among closely related species, our understanding of mitogenomic evolution and phylogenetic relationships in pentatomooids will be enhanced.

**Acknowledgments**

We thank Professor Yuan Huang (Shanxi Normal University) for providing much guidance for using the software, analyzing data, and presentation. This project was supported by the National Science Foundation of China [31501876], Natural Science Foundation of Shanxi Province [201610D021122], Scientific and Technological Project of Shanxi Province [20153011010-7], State Quality Inspection Administration of Science and Technology project [2016IK094], and the Key Research and Development Program of Hainan Province [ZDYF2017036].

**References Cited**

Anderson, S., A. T. Bankier, B. G. Barrell, M. H. de Bruijn, A. R. Coulson, J. Drouin, I. C. Eperon, D. P. Nierlich, B. A. Roe, F. Sanger, et al. 1981. Sequence and organization of the human mitochondrial genome. Nature. 290: 457–465.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prijibelski, et al. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19: 435–477.

Bernt, M., A. Donath, F. Jühlning, F. Externbrink, C. Florentz, G. Fritzsch, J. Pütz, M. Middendorf, and P. F. Stadler. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol. Phylogen. Evol. 69: 313–319.

Bolger, A. M., M. Logue, and B. Usadel. 2014. Trimmmomatic: a flexible trimer for Illumina sequence data. Bioinformatics. 30: 2142–2120.

Boore, J. L. 1999. Animal mitochondrial genomes. Nucleic Acids Res. 27: 1767–1780.

Cameron, S. L. 2014. Insect mitochondrial genomics: implications for evolution and phylogeny. Annu. Rev. Entomol. 59: 95–117.

Cameron, S. L., and M. F. Whiting. 2008. The complete mitochondrial genome of the tobacco hornworm, *Manduca sexta*, (Insecta: Lepidoptera: Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths. Gene. 408: 112–123.

de Castro, A. A., W. de S. Tavares, J. Collatz, A. I. de A. Pereira, J. E. Serrão, M. de Castro, A. A., W. de S Tavares, J. Collatz, A. I. de A Pereira, J. E. Serrão. 2015. Antibiosis of tomato, *Solanum lycopersicum* (Solanaceae) plants to the Asopinae predator *Supputus cincticeps* (Heteroptera: Pentatomoidea). Invert. Surviv. J. 12: 179–187.

Chang, J. H., J. Y. Choi, B. R. Jin, J. Y. Roh, J. A. Olszewski, S. J. Seo, J. C. Zanuncio. 2015. An improved baculovirus insecticide producing occlusion bodies that contain *Bacillus thuringiensis* insect toxin. J. Invertebr. Pathol. 84: 30–37.
Xie, Q., W. Bu, and L. Zheng. 2005. The Bayesian phylogenetic analysis of the 18S rRNA sequences from the main lineages of Trichophora (Insecta: Heteroptera: Pentatomomorpha). Mol. Phylogenet. Evol. 34: 448–451.

Yang, W. I. 1962. Hemiptera, Pentatomidae, vol. 2, 138 pp. In Economic insect fauna of China. Academia Sinica, Science Press, Beijing.

Yuan, M. L., Q. L. Zhang, Z. L. Guo, J. Wang, and Y. Y. Shen. 2015a. Comparative mitogenomic analysis of the superfamily Pentatomomorpha (Insecta: Hemiptera: Heteroptera) and phylogenetic implications. BMC Genomics. 16: 460.

Yuan, M. L., Q. L. Zhang, Z. L. Guo, J. Wang, and Y. Y. Shen. 2015b. The complete mitochondrial genome of Corizus tetraspilus (Hemiptera: Rhopalidae) and phylogenetic analysis of Pentatomomorpha. Plos One. 10: e0129003.

Zhang, D. X., and G. M. Hewitt. 1997. Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. Biochem. Syst. Ecol. 25: 99–120.

Zhang, Q. L., M. L. Yuan, and Y. Y. Shen. 2013. The complete mitochondrial genome of Dobycess baccarum (Insecta: Hemiptera: Pentatomidae). Mitochondrial DNA. 24: 469–471.

Zhao, Q., J. F. Wei, W. Q. Zhao, B. Cai, X. Du, and H. F. Zhang. 2017a. The first mitochondrial genome for the subfamily Asopinae (Heteroptera: Pentatomidae) and its phylogenetic implications. Mitochondrial DNA B Resour. 2: 804–805.

Zhao, W. Q., Q. Zhao, M. Li, J. F. Wei, X. H. Zhang, and H. F. Zhang. 2017b. Characterization of the complete mitochondrial genome and phylogenetic implications for Eurydema maracandica (Hemiptera: Pentatomidae). Mitochondrial DNA B Resour. 2: 550–551.

Zhu, X. Y., Z. Z. Xin, Y. Wang, H. B. Zhang, D. Z. Zhang, Z. F. Wang, C. L. Zhou, B. P. Tang, and Q. N. Liu. 2017. The complete mitochondrial genome of Clostera anachoreta (Lepidoptera: Notodontidae) and phylogenetic implications for Noctuoidea species. Genomics 109: 221–226.