The Mitochondrial Genomes of Phytophagous Scarab Beetles and Systematic Implications

Nan Song 1,3,* and Hao Zhang 2,*

1College of Plant Protection, Henan Agricultural University, Zhengzhou 450002, China, 2Department of Ideological and Political Theory Course, Henan Vocational and Technological College of Communication, Zhengzhou 450015, China, and 3Corresponding author, e-mail: songnan@henau.edu.cn

*These authors contributed equally to this work.

Subject Editor: Ted MacRae

Received 17 April 2018; Editorial decision 18 July 2018

Abstract

In this study, we newly sequenced five mitogenomes of representatives of phytophagous scarab beetles (Coleoptera: Scarabaeidae) by using next-generation sequencing technology. Two species have complete (or nearly complete) mitogenome sequences, namely Popillia mutans Newman (Coleoptera: Scarabaeidae) and Holotrichia obliterata Faldermann (Coleoptera: Scarabaeidae). The remaining three species have the partial mitogenomes, and the missing genes are mainly located adjacent to the control region. The complete (or nearly complete) mitogenomes have the same genome structure as most of the existing Scarabaeidae mitogenomes. We conducted phylogenetic analyses together with 24 published mitogenomes of Scarabaeoidea. The results supported a basal split of coprophagous and phytophagous Scarabaeidae. The subfamily Sericinae was recovered as sister to all other phytophagous scarab beetles. All analyses supported a non-monophyletic Melolonthinae, which included two different non-sister clades. The Cetoniinae was recovered as sister to a clade including Rutelinae and Dynastinae. Although the Rutelinae was rendered paraphyletic by Dynastinae in the Bayesian trees inferred under the site-heterogeneous CAT-GTR or CAT-MTART model, discordant patterns were given in some of ML trees estimated using the homogeneous GTR model.

Key words: Scarabaeidae, mitochondrial genome, next-generation sequencing, phylogeny

The scarab beetles comprise a large group of insects classified as the family Scarabaeidae (Insecta: Coleoptera), with over 30,000 species of beetles worldwide. From the point of view of feeding habits, these insects can be divided into coprophagous and phytophagous (including saprophagous) lineages. The species of coprophagous Scarabaeidae are often found in dung and in excrement, which are also called as dung beetles with various nesting behaviors (e.g., the rolling and tunneling behaviors). Due to their significant economic and ecological importance, dung beetles have been one of the most intensively studied coleopteran groups (Nichols et al. 2008, Brown et al. 2010). In comparison, researches on the phytophagous Scarabaeidae are limited. In particular, there remain many controversies on the taxonomy and phylogeny of the phytophagous clade of Scarabaeidae.

The phytophagous scarab beetles as an independent lineage have been recognized since Erichson (1847). The majority of phytophagous scarabs are also called Pleurosticti (Erichson 1847). The monophyly of Pleurosticti is often supported by morphological studies (Richter 1958, Balthasar 1963, Browne and Scholtz 1998). Based on the partial nuclear 18S and 28S ribosomal DNA sequences, Smith et al. (2006) recovered the phytophagous scarabs as a monophyletic group, which contained the following groups: Melolonthinae, Cetoniinae, Rutelinae, Dynastinae, Anomalini, and Adoretini.

Despite the great progress made in recent years, many issues on the higher-level phylogeny of Scarabaeidae remain to be addressed (Sanmartin and Mmartin-Piera 2003, Smith et al. 2006). Ahrens (2005) recovered the Melolonthinae as a paraphyletic group based on the analysis of morphological characters. Ahrens et al. (2014) retrieved the Melolonthinae as paraphyletic again, based on a multi-locus sequence analysis. On the contrary, other authors considered the Melolonthinae as a separate clade and to be elevated to the rank of family. Cherman and Morón (2014) included the Rutelinae, Dynastinae, and Melolonthinae to form the family Melolonthidae. Moreover, the family status of Melolonthinae was favored by the subsequent study of Cherman et al. (2016). Apart from the monophyly of Melolonthidae, the phylogenetic position of Sericinae was also problematic. Machatschke (1959) recovered the monophyletic
Sericinae, including the tribes Diphucephalini, Ablaberini, and Sericinae. However, the sericines were more often regarded as a tribe of Melolonthinae in recent literature (Ahrens 2005, Smith 2006, Ahrens et al. 2014). Elucidating the status of subgroups included in the family Scarabaeidae may contribute to the understanding of the evolution of this group, which further improves the conservation strategies of insects and the control measures of associated pests. Thus, it is necessary to investigate the constitution and the phylogeny of Scarabaeidae using additional types of data and increased taxon sampling.

The mitochondrial genome (mitogenome) is a useful maker in resolving the phylogeny at different taxonomic levels of Coleoptera (Timmermans et al. 2010, 2016; Coates 2014; Gillett et al. 2014; Crampton-Platt et al. 2015; Breeschnitzen et al. 2016; Nie et al. 2018). Recent mitochondrial phylogenetic studies have applied the next-generation sequencing (NGS) technologies to reconstruct the insect mitogenomes (e.g., Gillett et al. 2014, Crampton-Platt et al. 2015, Song et al. 2016). Compared with traditional PCR and Sanger sequencing, the approaches implemented by NGS are fast and cost-effective due to their capability to assemble a large number of mitogenomes from mixtures of species samples simultaneously.

Herein, we reconstructed five mitogenomes of the phytophagous scarab beetles from pooled genomic DNA. Combined with existing Scarabaeoidea mitogenomes, we inferred the phylogeny of Scarabaeidae to 1) test the monophyly of Melolonthinae, 2) to examine the position of Sericinae, and 3) to investigate the relationship between Rutelinae and other scarab beetles.

**Materials and Methods**

**Taxon Sampling and DNA Extraction**

Taxa were collected from Zhengzhou, Henan province, China. No specific permits were required for the insect specimens collected for this study. The field studies did not involve endangered or protected species. All sequenced insects are common beetle species in China and not included in the ‘List of Protected Animals in China’.

Two species of Melolonthinae (i.e., Holotrichia obita and Holotrichia sp.), two of Sericinae (i.e., Serica orientalis and Serica sp.) and one of Rutelinae (i.e., Popilla mutans) were utilized for DNA extraction and further sequencing. The adult specimens were preserved in 100% ethanol. Total genomic DNA was extracted from the thoracic leg muscle tissue, with the TIANamp Micro DNA Kit (Tiangen Biotech Co., Ltd) following the manufacturer’s protocol. Vouchers are deposited at the Entomological Museum of Henan Agricultural University. DNA concentration was measured by Nucleic acid protein analyzer (Quawell Technology Inc.).

**Library Construction and Illumina Sequencing**

Five scarab beetle species were respectively added into five different sequencing pools. Besides scarab beetle, the separate pool included other 29 unrelated samples. Following Gillett et al. (2014), each sample in the mixture was devised to have a largely identical DNA concentration (1.5 μg) to maximize the efficiency of genome assembling. The Illumina libraries were prepared with Illumina TruSeq™ DNA Sample Prep Kit (Illumina, San Diego, CA) and insert size of 350 bp. Each library was sequenced on the Illumina HiSeq X Ten platform (Beijing Novogene Bioinformatics Technology Co., Ltd, China) generating 150 bp paired-end reads.

**Sequence Quality Control and Assembly**

FastQC was used for the quality control checks on raw sequence data (Andrews 2010). All reads were filtered with NGS QC Toolkit using default settings (Patel and Jain 2012). In this step, reads containing adapters and ploy-N, and low-quality reads were removed from raw data. At the same time, Q20, Q30, GC-content, and sequence duplication level of the cleaned data were calculated. All the downstream analyses were based on clean data with high quality (avg. Q20 > 90%, and avg. Q30 > 85%). De novo assembly were performed with IDBA-UD v. 1.1.1 (Peng et al. 2012). The assemblies were constructed using 200 for the setting of minimum size of contig, and an initial k-mer size of 40, an iteration size of 10, and a maximum k-mer size of 90.

**Mitogenome Identification and Annotation**

The mitochondrial scaffolds were identified by mitochondrial baiting, with local-blasting implemented in BioEdit (Hall 1999) against the genome data assembled by IDBA-UD. The bait gene fragments (i.e., mitochondrial cox1, cytb, and rrnS) were amplified using the universal primers identical to those of Song et al. (2016). The preliminary mitogenome annotation were conducted using the MITOS webservice, under default settings and the invertebrate genetic code for mitochondria (Bernt et al. 2013). The gene boundaries were further checked and refined by alignment with the published mitogenome sequences of Scarabaeidae (see details in Supp. Table S1). To check the quality of the mitogenome sequences assembled, the mappings to the mitochondrial contigs were performed using Bovine R11.

The secondary structures of mitochondrial rrnl and rrnS were predicted, mainly with reference to the models of Apis mellifera (Gillespie et al. 2006) and of Drosophila virilis (Cannone et al. 2002), using the method of ‘Comparative sequence analysis’ in the study of Ouvrard et al. (2000).

**Multiple Alignment**

For the mitochondrial protein-coding genes, we first removed the stop codon of each sequence. Then, the nucleotide matrix of each protein-coding gene was translated into amino acids under the invertebrate mitochondrial genetic code in MEGA 6 (Tamura et al. 2013) and aligned based on their amino acid sequences with Muscle as implemented in MEGA. The amino acid matrix was back-translated into the corresponding nucleotide matrix. For the mitochondrial tRNA and rRNA genes, each of them was aligned using MAFFT (version 7) (Katoh and Standley 2013) under the iterative refinement method incorporating the most accurate local pairwise alignment information (E-INS-i). The resulting alignments were checked in MEGA. Gaps were striped by Gap Strip/Squeeze v2.1.0 with 40% Gap tolerance (http://www.hiv.lanl.gov/content/sequence/GAPSTREEZE/gap.html). Finally, all alignments were concatenated using FASconCAT_v1.0 (Kuck and Meusemann 2010) to construct the full dataset of PCGRNA.

To reduce the bias effect introduced by synonymous change on the phylogenetic analysis, protein-coding genes were degenerated by using the Perl script Degen v1.4 to construct the dataset of PCGDeGen (Regier et al. 2010, Zwick et al. 2012). The degenerated protein-coding genes were combined with RNA genes to compile the dataset of PCGDeGenRNA.

Possible sequence saturation at each codon position and at each gene partition of the concatenated alignment was individually examined by DAMBE 5 (Xia 2013). Estimates of nonsynonymous (dN) and synonymous (dS) substitution rates of protein-coding genes were obtained by the method of Yang and Nielsen (Yang and Nielsen 2000) using the program yn00 as implemented in PAML 4.9 (Yang 2007). The Excel 2016 was used to perform one-way analysis of variance (ANOVA) analysis in order to test for significant differences.
of substitution rates between different lineages, and the significance level was set to be 0.05.

Phylogenetic Analysis

We estimated the phylogeny of Scarabaeidae with two different approaches: maximum likelihood (ML) analysis under homogeneous models and Bayesian analysis under heterogeneous models. The following datasets were utilized in the phylogenetic analysis:

1) PCG: the nucleotide sequences of 13 protein-coding genes;
2) PCG AA: the deduced amino acid sequences of the protein-coding genes;
3) PCGDegen: the nucleotide sequences of 13 protein-coding genes degenerated by Degen script;
4) PCGRNA: the nucleotide sequences of 13 protein-coding genes combined with RNA genes; and
5) PCGDegenRNA: the sequences of PCGDegen combined with RNA genes.

Prior to ML analyses based on the datasets of PCG, PCGRNA and PCG AA, the PartitionFinder (Lanfear et al. 2012) was run to select the optimal partitioning strategies and best-fitting models for each dataset under the Bayesian Information Criterion (BIC), using a greedy search with RAxML (Stamatakis 2015). ML tree searches were performed in IQ-TREE (Nguyen et al. 2015), with various data partition schemes and best-fitting models determined by PartitionFinder. The detailed information on the partitions and the best models selected are listed in Supp. Table S2. Branch support analysis was conducted using 1,000 ultrafast bootstrap replicates. For the datasets of PCGDegen and PCGDegenRNA, un-partitioned analyses were performed, due to the altered variation of alignments by Degeneracy recoding scheme.

The parallel version of PhyloBayes (pb_mpi1.5a implemented on a HP server) was used for Bayesian tree calculation (Lartillot and Philippe 2004, Lartillot et al. 2009). The CAT-GTR model was used for nucleotide analyses, while the CAT-MTART model for amino acids. Two independent chains were run in parallel Table 1. The organization of the newly sequenced mitogenomes

| Gene region | Popillia mutans | Holotrichia oblitae | Serica sp. | Serica orientalis | Holotrichia sp. |
|-------------|----------------|---------------------|-----------|-----------------|----------------|
| Start       | Stop           | Start               | Stop      | Start           | Stop           |
| Partial control region | - | - | - | - | - |
| trnI        | 1              | 65                  | 1         | 65              | 1              |
| trnQ        | 69             | 137                 | 63        | 131             | 719            |
| trnM        | 137            | 205                 | 131       | 200             | 787            |
| nad2        | 206            | 1213                | 201       | 1208            | 856            |
| trnW        | 1229           | 1299                | 1226      | 1292            | 1864           |
| trnC        | 1292           | 1354                | 1285      | 1347            | 1924           |
| trnY        | 1355           | 1419                | 1353      | 1417            | 1985           |
| cox1        | 1412           | 2956                | 1383      | 2951            | 2047           |
| trnL-UUR    | 2952           | 3016                | 2952      | 3018            | 3581           |
| cox2        | 3017           | 3718                | 3019      | 3738            | 3647           |
| trnK        | 3705           | 3775                | 3704      | 3774            | 4332           |
| trnD        | 3778           | 3844                | 3774      | 3839            | 4403           |
| atp8        | 3845           | 4000                | 3840      | 3995            | 4469           |
| atp6        | 3994           | 4668                | 3989      | 4660            | 4618           |
| cox3        | 4668           | 5454                | 4660      | 5446            | 5289           |
| trnG        | 5455           | 5519                | 5447      | 5510            | 6076           |
| nad3        | 5520           | 5873                | 5511      | 5864            | 6139           |
| trnA        | 5872           | 5937                | 5863      | 5928            | 6491           |
| trnR        | 5937           | 6001                | 5929      | 5993            | 6556           |
| trnN        | 6004           | 6067                | 5993      | 6056            | 6618           |
| trnS-AGN    | 6068           | 6134                | 6057      | 6124            | 6684           |
| trnE        | 6135           | 6202                | 6125      | 6189            | 6751           |
| trnF        | 6201           | 6265                | 6188      | 6253            | 6813           |
| nad5        | 6265           | 7980                | 6254      | 7972            | 6878           |
| trnH        | 7981           | 8043                | 7973      | 8037            | 8594           |
| nad4        | 8043           | 9380                | 8037      | 9374            | 8657           |
| nad4H       | 9374           | 9664                | 9368      | 9638            | 9988           |
| trnT        | 9667           | 9731                | 9661      | 9725            | 10281          |
| trnP        | 9732           | 9796                | 9726      | 9790            | 10346          |
| nad6        | 9798           | 10301               | 9783      | 10295           | 10411          |
| cyt b       | 10301          | 11443               | 10295     | 11437           | 10908          |
| trnS-UCN    | 11442          | 11506               | 11436     | 11501           | 12049          |
| nad1        | 11523          | 12473               | 11519     | 12469           | 12132          |
| trnL-CUN    | 12475          | 12538               | 12471     | 12532           | 13084          |
| rmL         | 12539          | 13827               | 12510     | 13804           | 13147          |
| trnV        | 13828          | 13897               | 13803     | 13872           | -              |
| rms5        | 13897          | 14695               | 13873     | 14672           | -              |
| Complete or partial control region | 14696 | 16192 | 14673 | 15968 | - |

Dash (-) indicates no application or the missing genome region.
Fig. 1. (A) The secondary structure of rrl gene predicted for Popillia mutans. (B) The secondary structure of rns gene predicted for Popillia mutans.
and started from a random topology. The ‘bpcomp’ program contained in the package of PhyloBayes was used to calculate the largest (maxdiff) and mean (meandiff) discrepancy observed across all bipartitions. The program ‘tracecomp’ was also used to summarize the discrepancies and the effective sizes estimated for each column of the trace file. When the maximum ‘maxdiff’ value was lower than 0.1 and minimum effective size was higher than 100, the Bayesian runs were recognized to be reached good convergence. A consensus tree was calculated from the saved trees by bpcomp program after checking for stationarity, with the default options.

Hypothesis Testing
For all five datasets the alternative hypotheses were tested, using the Shimodaira-Hasegawa (SH) test (Shimodaira and Hasegawa 1999) and the approximately unbiased (AU) test (Shimodaira 2002). The specific hypotheses concerning four phylogenetic issues were assessed: 1) the basal divergence between coprophagous and phytophagous Scarabaeidae, 2) the Sericinae as sister to all other phytophagous taxa, 3) the monophyly of Melolonthinae, 4) the monophyly of Rutelinae. Additionally, we compared the tree topologies resulting from phylogenetic inferences using different datasets under ML and Bayes criteria. The site-log-likelihood values were calculated under the GTR+G model using TREE-PUZZLE (Version 5.3) (Schmidt et al. 2002). And then, the obtained values were used as input for the software CONSEL (Shimodaira and Hasegawa 2001). All constraint likelihood trees were generated by IQ-TREE (Nguyen et al. 2015) using the settings described above.

Results
Mitogenome Assembly
Five Illumina HiSeq runs resulted in 72,395,049, 72,720,510, 74,104,974, 81,772,168, and 89,829,307 raw paired reads, respectively. After filtering, 66,451,916, 67,980,632, 72,278,440, 81,189,879, and 89,122,956 clean paired reads were generated for corresponding libraries. For each species, at least one Sanger sequence was obtained and utilized to identify the potential mitochondrial contig (Supp. Table S3). Local blast searching results showed that each baiting fragment had its best blast match with the targeted mitochondrial contig (i.e., Identities ≥ 99% and E-value = 0.00). In contrast, other contigs assembled had the lower identities and the higher e-values (Supp. Table S3). Moreover, the mitogenome identified was represented by a single large-contig (contig length > 13 kb in most cases). The sequencing coverages corresponding to five mitogenomes ranged from 171-fold to 370-fold.

Mitogenome Organization
The mitogenome organization of five scarab beetle species sequenced are summarized in Table 1. P. mutans and H. oblitera had the complete or nearly complete mitogenomes, which contained the full 37 mitochondrial genes and the complete or partial control region. The other three mitogenomes (i.e., Holotrichia sp., S. orientalis and Serica sp.) were partial, and the missing fragments were located in the putative control region and the adjacent regions.

The complete or nearly complete mitogenomes reconstructed showed the same gene arrangement proposed as ancestral for insects (Cameron 2014). The nucleotide composition of new mitogenomes are summarized in Table 2. Five Illumina HiSeq runs resulted in 72,395,049, 72,720,510, 74,104,974, 81,772,168, and 89,829,307 raw paired reads, respectively. After filtering, 66,451,916, 67,980,632, 72,278,440, 81,189,879, and 89,122,956 clean paired reads were generated for corresponding libraries. For each species, at least one Sanger sequence was obtained and utilized to identify the potential mitochondrial contig (Supp. Table S3). Local blast searching results showed that each baiting fragment had its best blast match with the targeted mitochondrial contig (i.e., Identities ≥ 99% and E-value = 0.00). In contrast, other contigs assembled had the lower identities and the higher e-values (Supp. Table S3). Moreover, the mitogenome identified was represented by a single large-contig (contig length > 13 kb in most cases). The sequencing coverages corresponding to five mitogenomes ranged from 171-fold to 370-fold.

Table 2. Estimation of synonymous and nonsynonymous substitution rates by yn00 implemented in PAML

| Item         | Subfamily | Species                      | Group | dN  | dS  | dN/dS |
|--------------|-----------|------------------------------|-------|-----|-----|-------|
| Ingroup      | Aphodiinae| Aphodius sp.                 | 0     | 0.1553 | 4.3560 | 0.0356 |
|              | Scarabaeinae | Sarophorus sp.              | 1     | 0.1304 | 4.2417 | 0.0307 |
|              | Scarabaeinae | Scarabaeidae sp. BMNH 1274750 | 1     | 0.1386 | 4.2943 | 0.0323 |
|              | Scarabaeinae | Scarabaeidae sp. BMNH 1274752 | 1     | 0.1294 | 4.2988 | 0.0301 |
|              | Scarabaeinae | Scarabaeidae sp. BMNH 1274753 | 1     | 0.1423 | 4.2896 | 0.0332 |
|              | Scarabaeinae | Xindium sp.                 | 1     | 0.1390 | 4.2969 | 0.0324 |
|              | Cetoniinae | Leucocelis sp.              | 2     | 0.1146 | 4.3489 | 0.0264 |
|              | Cetoniinae | Myderumsp.                   | 2     | 0.1280 | 4.4122 | 0.0290 |
|              | Cetoniinae | Osmoderma opicum            | 2     | 0.1372 | 4.3791 | 0.0313 |
|              | Cetoniinae | Protaetia brevitarsis       | 2     | 0.1179 | 4.3100 | 0.0274 |
|              | Dynastinae | Cyphonistes vallatus        | 3     | 0.1463 | 4.5169 | 0.0324 |
|              | Melolonthinae | Ashenophilis sp.        | 4     | 0.1228 | 4.3011 | 0.0273 |
|              | Melolonthinae | Cheirotomonis jansoni     | 4     | 0.1574 | 4.4242 | 0.0356 |
|              | Melolonthinae | Holotrichia oblitera       | 4     | 0.1389 | 4.4084 | 0.0315 |
|              | Melolonthinae | Holotrichia sp.           | 4     | 0.1402 | 4.3853 | 0.0320 |
|              | Melolonthinae | Polyphylla laticollis mandshurica | 4 | 0.1665 | 4.4430 | 0.0375 |
|              | Melolonthinae | Rhopaea magnicornis        | 4     | 0.1289 | 4.3641 | 0.0295 |
|              | Melolonthinae | Schizonycha sp.            | 4     | 0.1266 | 4.3786 | 0.0289 |
|              | Rutelinae | Adoretus sp.                | 5     | 0.1143 | 4.2931 | 0.0266 |
|              | Rutelinae | Popilia mutans             | 5     | 0.1226 | 4.3500 | 0.0282 |
|              | Rutelinae | Popilia sp.                 | 5     | 0.1258 | 4.3227 | 0.0291 |
|              | Sericinae | Plephysylla sp.            | 6     | 0.1253 | 4.3275 | 0.0290 |
|              | Sericinae | Serica orientalis          | 6     | 0.1236 | 4.3431 | 0.0285 |
|              | Sericinae | Serica sp.                  | 6     | 0.1257 | 4.3992 | 0.0286 |
| Outgroup     | Geotrupidae | Anoplotrupes stercorosus   | 7     | 0.1420 | 4.2712 | 0.0332 |
|              | Geotrupidae | Bolboceratex sp.           | 7     | 0.1520 | 4.5252 | 0.0336 |
|              | Glaresidae | Glaphyridae comosus        | 8     | 0.1411 | 4.3209 | 0.0326 |
|              | Trogidae | Trox sp.                    | 10    | 0.1309 | 4.2923 | 0.0305 |
sequences were similar to those published scarab beetle species, with the significant A + T bias in the major strand. For the protein-coding genes, five new mitogenomes used the typical ATN as start codons, except for the \textit{nad3} (using ACA) and \textit{nad4} (using TGT) in \textit{H. oblita}. Ten of 13 protein-coding genes terminated with the conventional stop codons TAG or TAA, while the \textit{cox2} (in \textit{P. mutans} and \textit{Holotrichia} sp.), \textit{cox3} (in \textit{P. mutans}, \textit{S. orientalis}, \textit{Serica} sp., and \textit{Holotrichia} sp.), and \textit{nad3} (in \textit{S. orientalis}) had an incomplete stop codon T.

The secondary structures of mitochondrial tRNA genes were predicted. Supp. Fig. S1 illustrates the tRNA secondary structures of the complete mitogenome of \textit{P. mutans} (Supp. Fig. S1A) and the nearly complete mitogenome of \textit{H. oblita} (Supp. Fig. S1B). All tRNA genes had the typical clover-leaf structure except for the trnS-AGN. The DHU arm was reduced or missing in the trnS-AGN.

For the \textit{P. mutans} and \textit{H. oblita}, the complete mitochondrial rRNA genes were identified between \textit{trnL-CUN} and \textit{trnV} and between \textit{trnV} and the control region, with the length of 1,270 or 1,289 bp for \textit{rrnL} and 799 or 800 bp for \textit{rrnS}. The secondary structures of mitochondrial rRNA genes were inferred (Fig. 1A and B for \textit{rrnL} and \textit{rrnS} of \textit{P. mutans}, and Supp. Fig. S2A and B for \textit{rrnL} and \textit{rrnS} of \textit{H. oblita}). Because the \textit{P. mutans} and \textit{H. oblita} have the basically identical secondary structure predictions for two mitochondrial rRNA genes, the following description will focus on those of \textit{P. mutans}. The secondary structure of \textit{rrnL} consisted of five canonical structural domains (II, IV-VI) and 44 helices (Fig. 1A and Supp. Fig. S2A). Domain III was absent, which was the same as the published insect mitochondrial \textit{rrnL} secondary models (Cannone et al. 2002, Gillespie et al. 2006). When compared with other scarab beetles, domains IV-VI located in the 3′ end of the \textit{rrnL} molecule were highly conserved. The three domains contained the helices with more than 75% identical positions across species (e.g., H1755, H1835 and H1906). The mitochondrial \textit{rrnS} secondary model predicted for \textit{P. mutans} consisted of 27 helices (Fig. 1B and Supp. Fig. S2B), which formed four typical domains (labeled I, II, III and IV). Across taxa analyzed, domains III and IV were more conserved than domains I and II.

The full control region located between \textit{rrnS} and \textit{trnI} was identified only in the mitogenome of \textit{P. mutans} (Supp. Fig. S1A) and the nearly complete mitogenome of \textit{H. oblita} (Supp. Fig. S1B). All rRNA genes had the typical clover-leaf structure except for the \textit{trnS-AGN}. The DHU arm was reduced or missing in the \textit{trnS-AGN}.
Phylogenetic Analysis

ML Analysis

Although results of the ML analyses based on various datasets were similar, the trees differed to a greater extent in the relationships of Melolonthinae and Rutelinae. The monophyly of Scarabaeidae was strongly supported (e.g., BP = 82 in PCGRNA-ML-tree, Fig. 2). The coprophagous scarab beetles comprising the clade Aphodiinae + Scarabaeinae were recovered as the sister group to the phytophagous lineages. The phytophagous clade contained five subfamilies, namely the Sericinae, Melolonthinae, Cetoniinae, Rutelinae, and Dynastinae. In all ML analyses but PCG_AA, the Sericinae was recovered outside of the assemblage of Melolonthinae, and as a sister lineage to all other phytophagous scarabs. In the ML tree from the PCG_AA dataset, the Sericinae was recovered as a sister group to a species from Melolonthinae (i.e., Holotrichia sp.), with weak node support value (BP = 74). The Melolonthinae as a non-monophyletic group was divided into two different non-sister clades.

The following phylogenetic results changed depending on the dataset. 1) For the relationship of melolonthine lineages, Holotrichia sp. formed a sister-group to Cheirotomus jansoni in the analyses from the PCG, PCGRNA, PCGDegen and PCGDegenRNA datasets (Fig. 2 and Fig. 3A, C, and D), whereas Holotrichia sp. was sister to the clade Sericinae in the analysis from the PCG_AA dataset (Fig. 3B); 2) The second major melolonthine clade containing the other five melolonthine species was sister to a large clade including Dynastinae, Rutelinae, and Cetoniinae, the relationships among melolonthine species were different across analyses: in the ML analyses using the PCG, PCGRNA and PCGDegenRNA datasets was Rhopaea magnicornis sister to Polyphylla laticollis mandshurica, and both together sister to Holotrichia obliterata (Fig. 2 and Fig. 3A and D), in contrast, in other two analyses (i.e., PCG_AA and PCG_Degen) was R. magnicornis sister to a clade P. laticollis mandshurica + H. obliterata (Fig. 3B and C); 3) the monophyly of Rutelinae was supported by the PCG, PCG_AA, and PCGRNA datasets, whereas the degenerated datasets (i.e., PCGDegen and PCGDegenRNA) retrieving the C. vallatus (Dynastinae) as sister to Adoretus sp. led to a non-monophyletic Rutelinae (Fig. 3C and D).

Bayesian Analysis

Bayesian tree reconstructions were performed under the site-heterogeneous CAT-GTR or CAT-MTART model (Figs. 4 and 5). The monophyly of Scarabaeidae were recovered in all Bayesian analyses, except for that from the PCGDegenRNA dataset. The nested position of the outgroup Glaphyrus comosus resulted in a non-monophyletic Scarabaeidae in the PCGDegenRNA Bayes tree, but branch support measures indicated that this reconstruction was weakly supported (PP = 0.57). Monophyletic groupings of taxa corresponding to the

Fig. 3. Maximum likelihood trees inferred from the datasets of (A) PCG, (B) PCG_AA, (C) PCGDegen, and (D) PCGDegenRNA using IQ-TREE. Branch support values are presented near each node. Scale bar represents substitutions/site.
subfamily level were congruent with the relationships found in the ML topologies. The Aphodiinae was well supported as a sister-group to Scarabaeinae in the coprophagous Scarabaeidae clade, with high posterior probabilities (PP ≥ 0.98). Among the phytophagous lineages, the Sericinae was consistently recovered as sister to the remaining phytophagous lineages. The non-monophyletic Melolonthinae branched next to the Sericinae. The terminal clades still consisted of the Cetoniinae, Rutelinae, and Dynastinae. The subfamily Rutelinae was rendered paraphyletic by an internal clade comprising *C. val- latus* + *Adoretus* sp. in all Bayesian analyses but for that from the PCG_AA dataset. The PCG_AA Bayes tree retrieved Dynastinae and Rutelinae to both be monophyletic.

Alternative Hypothesis Testing
The hypothesis that the basal division of Scarabaeidae into two clades (coprophagous and phytophagous) was strongly supported by the statistical testing with different datasets (Table 3). Moreover, hypothesis testing consistently supported Sericinae as an independent lineage and sister to all other phytophagous lineages. This hypothesis received the highest likelihood values in all analyses but for that from the PCG_AA dataset. Although the monophyly of Melolonthinae cannot be addressed unambiguously, the likelihood value for a tree constraining Melolonthinae as non-monophyletic was often higher than that constraining a monophyletic Melolonthinae. The paraphyly of Rutelinae was rejected by the PCG dataset (*P* < 0.05 in both SH and AU tests, Table 3), whereas the monophyly of Rutelinae was consistently supported by all testing. Based on five datasets and two inference methods, a total of 10 trees were produced (Figs. 2–5).

The results of statistical testing of tree topology indicated that the topology resulting from the partitioned ML analyses of the PCGRNA and PCG datasets (Fig. 2 and 3A) was more likely to present the true tree, with respect to their highest likelihood values (Supp. Table S5). Whereas, some of the Bayesian trees were rejected (e.g., PCGRNA-Bayes-tree, *P* < 0.05 in all SH or AU tests).

Discussion
The present study adds five new mitogenome sequences to the phytophagous scarab beetles, with the goal of investigating the higher-level relationships within Scarabaeidae. Our results demonstrate that NGS techniques adopted here can generate mitogenomes in a more efficient and cost-effective way, and that the mitogenomic data can provide insights into the subfamily relationships of Scarabaeidae. Although there have been numerous studies attempting to reconstruct the evolutionary history of scarab beetles (Howden 1982; Browne and Scholtz 1995, 1996, 1998; Ahrens 2005; Smith et al. 2006; Hunt et al. 2007; Ahrens et al. 2014; Bocak et al. 2014; Eberle et al. 2014), no authors utilized the mitogenomes alone to estimate the phylogeny of Scarabaeidae. This inspired us to perform this analysis to further assess the phylogenetic usefulness of mitogenomes. The backbone relationships recovered by the current mitogenomic data are concordant with previous morphological and molecular studies, namely that the Scarabaeidae comprises the traditionally recognized coprophagous and phytophagous lineages (Browne and Scholtz 1999, Smith et al. 2006, Ahrens et al. 2014, Bocak et al. 2014). The former clade
includes the subfamilies Aphodiinae and Scarabaeinae, while the latter consists of Sericinae, Melolonthinae, Rutelinae, Dynastinae, and Cetoniinae. We acknowledge the effect of the sparse taxon sampling of several lineages on the recovery of relationships within this megadiverse insect group. There is an immediate need of an extensive sequencing program of scarab beetles. This article contributes to stimulate interest in phylogenomic study of Scarabaeidae by using whole mitogenome sequences.

The monophyly of Rutelinae and the placement of \textit{C. jansoni} (Melolonthinae) were variable. Two ML analyses based on the degenerated datasets under the homogeneous model and four out of five Bayesian analyses under the site-heterogeneous model supported a paraphyletic Rutelinae. Recovery of the monophyly of Rutelinae depended on whether the single representative of Dynastinae (i.e., \textit{C. vallatus}) clustered with one exemplar of Rutelinae (i.e., \textit{Adoretus} sp.) or not. The majority of analyses retrieved a sister-group relationship of \textit{C. jansoni} and \textit{Holotrichia} sp., both of which were outside the main clade of Melolonthinae. Three species (i.e., \textit{C. jansoni}, \textit{Holotrichia} sp. and \textit{C. vallatus}) leading to the unstable reconstructions exhibited the obviously long branch lengths, particularly in the Bayesian trees. In the analysis of substitution rates of protein-coding genes, we found the rate heterogeneity in the current data (Table 2). Moreover, three species motioned above had the greatly accelerated rates, compared with other scarab beetles. Long branches associated with rate heterogeneity may result in the incongruence between analyses.

With regard to the monophyly of the subfamily Melolonthinae, there has been no consensus opinion among coleopterists. Based on the nuclear gene datasets, Smith et al. (2006) did not provide definitive results on the monophyly of the Melolonthinae and suggested additional phylogenetically informative characters needed to tackle this problem. Machatschke (1959) proposed the monophyly of the subfamily Sericinae, which rendered the paraphyly of the Melolonthinae. This hypothesis was confirmed by another morphological study by Ahrens (2005). In this paper, the Sericinae was consistently recovered as an independent clade and placed as the most basal branching lineage within the phytophagous scarab beetles. Thus, our data supported the point of the elevated status of Sericini to be the subfamily rank (Sericinae). Meanwhile, a non-monophyletic Melolonthinae was supported in all analyses.

The sister-group relationship between Rutelinae and Dynastinae was always recovered by previous studies (Howden 1982, Browne and Scholtz 1999, Hunt et al. 2007, Bocak et al. 2014). Smith et al. (2006) indicated that some taxa of the Dynastinae might be moved to the subfamily Rutelinae. In the PCGDegen and PCGDegenRNA ML analyses and in all Bayesian analyses but PCG_AA, the Rutelinae was recovered as a paraphyletic group with respect to
the single representative of Dynastinae. Despite this, the preferred trees (i.e., PCG-ML-tree and PCGRNA-ML-tree, see results in Supp. Table S5) supported the monophyly of Rutelinae (BP = 95 in both trees). Future studies should focus on adding whole mitogenome sequences of other underrepresented taxa, especially Dynastinae, in order to advance understanding of the subfamily level relationships of the phytophagous clades. The sister group relationship between the clade comprising Rutelinae and Dynastinae and the subfamily Cetoniinae was strongly supported in all analyses (BP ≥ 88 and PP ≥ 0.98). This arrangement is in line with the result of Bocak et al. (2014), but conflict with Browne and Scholtz (1999) and Hunt et al. (2007). The latter two researches recovered a close affinity of the clade Rutelinae + Dynastinae and the Melolonthinae (Browne and Scholtz 1999, Hunt et al. 2007).

Supplementary Data

Supplementary data are available at Journal of Insect Science online.

Acknowledgments

This research is supported by grants from the National Natural Science Foundation of China (31402002), Key Scientific Research Projects of Henan Province (14B210036, 16A210029, 18B210006), and Agricultural Science and Technology Innovation Project of Henan Province (30600309).

References Cited

Ahrens, D. 2005. The phylogeny of Sericini and their position within the Scarabaeidae based on morphological characters (Coleoptera: Scarabaeidae). Syst. Entomol. 31: 113–144.

Ahrens, D., J. Schwarzer, and A. P. Vogler. 2014. The evolution of scarab beetles tracks the sequential rise of angiosperms and mammals. Proc. Biol. Sci. 281: 20141470.

Andrews, S. 2010. FastQC: A quality control tool for high throughput sequence data. Bioinformatics Group, Babraham Institute, Cambridge, United Kingdom. (www.bioinformatics.babraham.ac.uk/projects/fastqc).

Balthasar, V. 1963. Monographie der Scarabaeidae und Aphodiidae der palaearktischen und orientalischen Region. Lamellicornia. Band 1. Der Tschechoslowakischen Akademie der Wissenschaften. Verlag, Prague.

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

Bocak, L., C. Barton, A. Crampton-Platt, D. Chesters, D. Ahrens, and A. P. Vogler. 2014. Building the Coleoptera tree-of-life for >8000 species: composition of public DNA data and fit with Linnaean classification. Syst. Entomol. 39: 97–110.

Breeschoten, T., C. Doorenweerd, S. Tarasov, and A. P. Vogler. 2016. Phylogenetics and biogeography of the dung beetle genus Onthophagus inferred from mitochondrial genes. Mol. Phylogenet. Evol. 105: 86–95.

Brown, J., C. H. Scholtz, J. L. Janeau, S. Grellier, and P. Podwojewski. 2010. Dung beetles (Coleoptera: Scarabaeidae) can improve soil hydrological properties. Appl. Soil Ecol. 46: 9–16.

Browne, J., and C. H. Scholtz. 1999. Phylogeny of the families of Scarabaeoidea (Coleoptera) based on characters of the hindwing articulation, hindwing base and wing venation. Syst. Entomol. 24: 145–173.

Browne, J., and C. H. Scholtz. 1996. Morphology of the hind wing articulation and wing base of the Scarabaeoidea (Coleoptera) with some phylogenetic implications. Bonn. Zool. Monogr. 40: 1–200.

Browne, J., and C. H. Scholtz. 1998. Evolution of the scarab hind wing articulation and wing base: a contribution toward the phylogeny of the Scarabaeoidea (Scarabaeidae: Coleoptera). Syst. Entomol. 23: 307–326.

Table 3. Hypothesis testing: results of the SH and AU tests, with rejection at $P < 0.05$.

| Hypothesis tested | Log-likelihood | P values for Log-likelihood | Log-likelihood | P values for Log-likelihood | Log-likelihood | P values for Log-likelihood | Log-likelihood | P values for Log-likelihood |
|-------------------|----------------|----------------------------|----------------|----------------------------|----------------|----------------------------|----------------|----------------------------|
| Hypothesis        |                |                            |                |                            |                |                            |                |                            |
| Coprophagous +    | −163937.18     | 0.433                      | −163907.95     | 0.915                      | −163926.75     | 0.639                      | −163995.82     | 0.031                      |
| Phytophagous       | −66075.52      | 0.312                      | −66085.88      | 0.728                      | −66098.73      | 0.298                      | −66098.73      | 0.298                      |
| Sericinae + all    | −163907.95     | 0.915                      | −163926.75     | 0.639                      | −163995.82     | 0.031                      | −163995.82     | 0.031                      |
| other phytophagous| −66075.52      | 0.312                      | −66085.88      | 0.728                      | −66098.73      | 0.298                      | −66098.73      | 0.298                      |
| Rutelinae as a    | −163937.18     | 0.433                      | −163907.95     | 0.915                      | −163926.75     | 0.639                      | −163995.82     | 0.031                      |
| monophyletic      | −66075.52      | 0.312                      | −66085.88      | 0.728                      | −66098.73      | 0.298                      | −66098.73      | 0.298                      |
| group              | −163937.18     | 0.433                      | −163907.95     | 0.915                      | −163926.75     | 0.639                      | −163995.82     | 0.031                      |
| Melolonthinae as  | −163995.82     | 0.031                      | −163995.82     | 0.031                      | −163995.82     | 0.031                      | −163995.82     | 0.031                      |
| a monophyletic    | −66075.52      | 0.312                      | −66085.88      | 0.728                      | −66098.73      | 0.298                      | −66098.73      | 0.298                      |
| group              | −163995.82     | 0.031                      | −163995.82     | 0.031                      | −163995.82     | 0.031                      | −163995.82     | 0.031                      |
| Melolonthinae as  | −163995.82     | 0.031                      | −163995.82     | 0.031                      | −163995.82     | 0.031                      | −163995.82     | 0.031                      |
| a monophyletic    | −66075.52      | 0.312                      | −66085.88      | 0.728                      | −66098.73      | 0.298                      | −66098.73      | 0.298                      |
| group              | −163995.82     | 0.031                      | −163995.82     | 0.031                      | −163995.82     | 0.031                      | −163995.82     | 0.031                      |
