Complete Mitochondrial Genome of *Papilio protenor* (Lepidoptera, Papilionidae) and Implications for Papilionidae Taxonomy

Nai-Yi Liu,1,* Yun-He Wu,1,2,* Xiao-Jun Yang,3 Jing Wu,4 Si-Zhu Zheng,5 Shu-Yan Wang,1 and Jie Fang1,6,*

1School of Life Sciences, Anhui University, Hefei 230601, China, 2Kunming College of Life Science, University of China Academy of Science, Kunming 650204, China, 3Jiangsu Entry-Exit Inspection and Quarantine Bureau, Lab for plant quarantine, Nanjing 210001, China, 4Nanjing entry exit inspection and Quarantine Bureau, Nanjing 211106, China, 5Suzhou entry exit inspection and Quarantine Bureau, Suzhou 215021, China, 6Corresponding author, e-mail: ahufangjie@126.com

*Nai-Yi Liu, Yun-He Wu and Jie Fang contributed equally to this work.

Subject Editor: Joanna Chiu

Received 10 May 2017; Editorial decision 4 September 2017

**Abstract**

The complete mitochondrial genome (mitogenome) of *Papilio protenor* (Lepidoptera: Papilionidae) was sequenced and annotated in this study. The mitogenome comprised a typical circular, double-stranded DNA molecule of 15,268 bp in length including 13 protein-coding genes (PCGs), two ribosomal RNA genes, 22 transfer RNA (tRNA) genes, and an A+T-rich region. The gene order and mitogenome orientation were similar to those of all known Papilionidae species. The nucleotide composition of the *P. protenor* mitogenome exhibits considerable A+T bias (80.5%) and the AT (~0.019) and GC (~0.231) skewness is slightly negative. All of the PCGs except for cytochrome c oxidase (*COI*) start with a canonical ATN start codon, whereas the *COI* gene is tentatively designated by the CGA codon. Of the 13 PCGs, 11 contain the complete stop codon TAA or TAG, whereas *COI* and *COII* were terminated with a single T nucleotide. All tRNAs exhibit the typical cloverleaf structure, except for *tRNA*\(^{\text{Ser(AGN)}}\), which does not contain the dihydrouridine arm. The 458 bp A+T-rich region is comprised of nonrepetitive sequences including the motif ATAGA followed by a poly-T stretch and a microsatellite-like (AT)\(_n\) element preceded by the ATTTA motif. Phylogenetic analysis of the 13 PCGs data using Bayesian inference, maximum likelihood methods, and maximum parsimony support the view that the subfamily Parnassiinae is regarded as an independent subfamily within Papilionidae and that Zerynthiini should be treated as one of the two clades of the subfamily Parnassiinae along with Parnasiini. In addition, the analysis strongly supports the monophyly of the subfamily Parnassiinae.

**Key words:** *Papilio protenor*, mitochondrial genome, secondary structure, Papilionidae taxonomy

Lepidoptera (moths and butterflies) is the second largest order in Insecta, accounting for more than 16,000 species (Dai et al. 2015). Among these, the Papilionidae family comprises over 570 species worldwide (Shen et al. 2015). However, despite the huge taxonomic diversity, current information regarding the lepidopteran mitogenome is very limited. To date, complete or nearly complete mitogenomes have been sequenced for approximately 100 species of lepidoptera whereas only 24 complete mitogenomes are available in GenBank (http://www.ncbi.nlm.nih.gov/genbank/) for Papilionidae species.

The insect mitochondrial genome (mitogenome) consists of a closed double-strand circular DNA molecule, ranging from 14 to 19 kb in length, which encodes an extremely highly conserved set of 37 genes, including 13 protein-coding genes (PCGs) [ATPase subunits (ATP) 6 and 8, cytochrome c oxidase subunits 1–3 (*COI–III*), cytochrome (*Cyt*) B, and NADH dehydrogenase subunits (ND) 1–6 and 4L], two ribosomal RNA genes (12S rRNA and 16S rRNA, encoding the small and large rRNA subunits, respectively), and 22 tRNA genes (Wolstenholme 1992, Boore 1999). In addition, it contains an A+T-rich region, which takes part in the regulation and initiation of mitogenome replication and transcription (Shadel and Clayton 1993). Compared with nuclear genes, the mitogenome is a compact structure, exhibits maternal inheritance, lacks genetic recombination, and undergoes a relatively fast evolutionary rate of variation. Because of these unique features, mitogenome genes and genomes have been widely used as a molecular marker to study molecular evolution, population genetics, phylogeography, and phylogenetics (Harrison 1989, Boore 1999, Ingman et al. 2000, Hurst and Jiggins 2005, Cameron et al. 2006, Timmermans et al. 2014).
Swallowtail butterflies (Papilionidae) are among the most popular insect taxa and have greatly contributed to studies of ecology, behavior, and evolution in insects. Although numerous studies have been published regarding Swallowtail butterflies, the taxonomic status of Parnassiinae and Zerynthiinae has been particularly controversial at the subfamily family. Many authors suggest that the clades Parnassiini and Zerynthiini should be treated as two independent subfamilies (Talbot and Sewell 1939, Wang 2001). On the other hand, some authors support that Parnassiini should be regarded as a tribe of the subfamily Parnassiinae, which would thus consist of two clades: Parnassiini (including Archon, Hypermenestra, and Parnassius), and Zerynthiini (including Allancastria, Sericinus, Zerynthia, Luehdorfia, and Bhutanitis), based on morphological and molecular data (Ehrlich 1938, Igarashi 1984, Huang et al. 2015). However, these views have been challenged by several authors, who believe that the subfamily Parnassiinae should not belong to Papilionidae at all but rather should be upgraded as a separate family, including the subfamilies Parnassiinae and Zerynthiinae (Li and Zhu 1992). Furthermore, the monophyly of the subfamily of Parnassiinae is increasingly controversial (Häuser 1993, Yagi et al. 1999, Caterino et al. 2001). Taxonomic disputes mainly focus on: 1) taxonomic status of the subfamily Parnassiinae; 2) taxonomic status of the subfamily Zerynthiinae; and 3) phylogenetic relationships with the subfamily Parnassiinae.

To address these issues, in the present study, we determined the complete mitogenome sequence of Papilio proter (Lepidoptera: Papilionidae) and compared its gene structure with that of other known Papilionidae species. Furthermore, we performed a phylogenetic analysis based on the available complete mitogenome sequences to provide insight into the phylogenetic relationship of Papilionidae.

Materials and Methods
Sample Collection and DNA Extraction
Adult individuals of P. proter were collected in Yaoluoping nature reserve, Yuexi County, Anhui province, China, in July 2015. After collection, the samples were immediately preserved in 100% ethyl alcohol and stored at −20°C prior to DNA extraction. Whole genomic DNA was extracted from the thorax muscle tissues of a single sample using a standard phenol-chloroform protocol (Sambrook and Russell 2001). DNA yield and quality were examined by 1.5% agarose gel/ethidium bromide.

Polymerase Chain Reaction Amplification and Sequencing
To amplify the whole mitogenome of P. proter, we designed three long polymerase chain reaction (PCR) primers using Primer Premier 5.0 software (Singh et al. 1998) and were based on three fragments of COI, ND5, and 16S rRNA of P. proter that were released by the National Center for Biotechnology Information. All PCRs were performed in a 50-μl reaction volume containing 33.5-μl sterilized distilled water, 5-μl 10 × Taq buffer (Mg2+ plus, TaKaRa, Shiga, Japan), 8-μl dNTPs, 1-μl each primer, 0.5-μl LA Taq polymerase (5 U/μl, TaKaRa), and 1-μl DNA template. The conditions for PCR amplification were as follows: initial denaturation at 95°C for 5 min, followed by 30 cycles at 98°C for 10 s, 68°C for 5 min (depending on primer combinations and putative length of the fragments), and a final extension step of 72°C for 10 min. The quality of the PCR products was detected by 1.5% agarose-gel electrophoresis, and products were purified using a DNA gel extraction kit (TaKaRa). The purified PCR products were sequenced after cloning into the pMD-19T Vector (TaKaRa, Dalian, China). Two vector-specific primers and internal primers were used to complete the sequences by primer walking.

Sequence Analysis and Gene Annotation
The mtDNA sequences of P. proter were assembled using the Sequencher program from the Lasergene package DNAStar (Madison, WI). The 13 PCGs, 2 rRNAs, and the A+T-rich region were annotated manually and determined by comparison with the close relative insects for which published complete mitogenome sequences were available. The nucleotide sequences of the 13 PCGs were translated into amino acid sequences based on the invertebrate mitochondrial genetic code. Nucleotide composition and relative synonymous codon usage (RSCU) were calculated using MEGA version 5.0 (Tamura et al. 2011). tRNA genes were identified using the tRNAscan-SE Search Server (http://lowelab.ucsc.edu/tRNAscan-SE/) (Lowe and Eddy 1997). The AT and CG skew were calculated according to the formula: AT skew = [A% − T%]/[A% + T%], CG skew = [C% − G%]/[C% + G%] (Perna and Kocher 1995).

Phylogenetic Analysis
To study the phylogenetic relationships among Papilionidae, phylogenetic trees were reconstructed based on the 13 PCGs of the complete mitogenomes of the 36 known butterfly sequences. The mitogenomes of Argyrosis hyperbius (NC_015988.1) and Timelaea maculata (NC_021090.1) were selected as outgroups. The concatenated nucleotide alignments of the 13 PCGs yielded a nucleotide matrix of 11,017 bp in length that was used for phylogenetic analysis with Bayesian inference (BI), maximum likelihood (ML), and maximum parsimony (MP) methods.

The ML analysis was performed by RAxML with 1,000 bootstrap replicates and using the rapid bootstrap feature (random seed value 12,345) (Stamatakis et al. 2008). BI analysis was implemented with MrBayes version 3.1.2 (Ronquist and Huelsenbeck 2003). MrModeltest 2.3 (Posada and Crandall 1998) was used in

![Fig. 1. Map of the circular P. proter mitochondrial genome. Gene names on the thick line side indicated that these genes were located on H strand, whereas the others were located on L strand. Color codes for different genes were detail listed on the map.](image-url)
combination with PAUP* 4b10 to select an appropriate nucleotide substitution model for BI. The GTR + I + G model was chosen as the best-fit model following the Akaike information criterion. In Bayesian analysis, The Monte Carlo Markov chain length was run for 10,000,000 generations. When the average standard deviation of split frequencies was below 0.01, we considered that stationarity was reached and stopped the run. Each set was sampled every 1,000 generations with a burn-in of 25%. Posterior probabilities over 0.9 were considered as strongly supported (Mutanen et al. 2010). MP analysis was performed with PAUP* 4b10 first under the heuristic search strategy with all sites weighted equally, gaps treated as missing data, and 1,000 random addition sequences and tree bisection-reconnection branch swapping. To assess the support for branching events, nonparametric bootstrapping was performed with 1,000 pseudoreplicates under a heuristic search strategy and 100 random-addition sequences in each pseudoreplicate.

Results and Discussion

Genome Structure and Organization

The P. proteanor mitogenome is a typical circular, double-stranded DNA molecule of 15,268 bp in length (GenBank accession No.KY272622.1), consistent with other known Papilionidae mitogenomes, which range from 15,185 bp in P. machaon to 16,094 bp in Agehana maraho. It contains 37 genes (13 PCGs, 22 tRNAs, and 2 rRNAs) and an A+T-rich region (Fig. 1). Among these, 14 genes are encoded on the L strand: four PCGs (ND1, ND4, ND4L, and ND5), two rRNA genes (16S rRNA and 12S rRNA), and eight tRNA genes (tRNAGlu, tRNACys, tRNA Tyr, tRNAPhe, tRNAHis, tRNAPro, tRNALeu(CUN), and tRNAVal). The remaining 23 genes and the A+T-rich region are encoded on the H strand (Table 1). The gene order and orientation of P. proteanor is identical to known Papilionidae mitogenomes, belonging to the gene order tRNAMet-tRNAIle-tRNAGln. However, the arrangement is different from the ancestral arrangement of tRNA Ile-tRNAGln-tRNAMet such as found in the species Thitarodes renzhiensis and Ahamus yunnanensis in the Hepialoidea (Cao et al. 2012).

The nucleotide composition (A: 39.5%, T: 41.0%, C: 12%, G: 7.5%) of the P. proteanor mitogenome shows a marked A+T bias (80.5%) (Table 2). This result is consistent with the A+T bias of the known Papilionidae mitogenomes, wherein the composition of A+T ranges from 79.5% in Papilio dardanus to 81.5% in Luehdorfia tai-bai (Table 2). The AT skew and CG skew were −0.019 and −0.231, respectively, in the P. proteanor mitogenome (Table 2), indicating

### Table 1. Characteristics of P. proteanor mitochondrial DNA genome

| Gene name     | Coding strand | Start position | End position | Intergenic nucleotides | Overlapping nucleotide | Size (bp) | Start codon | Stop codon |
|---------------|---------------|----------------|--------------|------------------------|------------------------|-----------|-------------|------------|
| tRNAMet       | H             | 1              | 69           | -                      | 1                      | 69        | -           | -          |
| tRNALeu       | H             | 69             | 132          | -                      | 3                      | 64        | -           | -          |
| tRNAGlu       | L             | 130            | 197          | 53                     | -                      | 68        | -           | -          |
| ND2           | H             | 251            | 1263         | -                      | 1                      | 1013      | ATC         | TAA        |
| tRNAtop       | H             | 1263           | 1327         | -                      | 8                      | 65        | -           | -          |
| tRNACys       | L             | 1320           | 1384         | -                      | -                      | 65        | -           | -          |
| tRNAtyr       | L             | 1385           | 1449         | 2                      | -                      | 65        | -           | -          |
| COXI          | H             | 1452           | 2982         | -                      | -                      | 1531      | -           | -          |
| tRNALeu(UUR)  | H             | 2983           | 3050         | -                      | 68                     | -         | -           | -          |
| COX2          | H             | 3051           | 3732         | -                      | 682                    | ATG       | T           |            |
| tRNAsyn       | H             | 3733           | 3803         | -                      | 1                      | 71        | -           | -          |
| tRNAasp       | H             | 3803           | 3869         | -                      | 67                     | -         | -           | -          |
| ATP8          | H             | 3870           | 4037         | 7                      | 168                    | ATT       | TAA         |            |
| ATP6          | H             | 4031           | 4711         | 6                      | 681                    | ATG       | TAA         |            |
| COX3          | H             | 4718           | 5506         | 3                      | 789                    | ATG       | TAA         |            |
| tRNAGly       | H             | 5510           | 5573         | -                      | 66                     | -         | -           | -          |
| ND3           | H             | 5576           | 5929         | -                      | 354                    | ATC       | TAG         |            |
| tRNAlys       | H             | 5928           | 5991         | -                      | 64                     | -         | -           | -          |
| tRNAArg       | H             | 5991           | 6054         | -                      | 64                     | -         | -           | -          |
| tRNAAsn       | H             | 6054           | 6119         | -                      | 66                     | -         | -           | -          |
| tRNAser(AGN)  | H             | 6120           | 6180         | 1                      | 61                     | -         | -           | -          |
| tRNAglu       | H             | 6128           | 6249         | -                      | 68                     | -         | -           | -          |
| tRNAPro       | L             | 6248           | 6311         | -                      | 64                     | -         | -           | -          |
| ND5           | L             | 6306           | 8030         | 15                     | 1725                   | ATT       | TAA         |            |
| tRNAHis       | L             | 8046           | 8110         | -                      | 65                     | -         | -           | -          |
| ND4           | L             | 8110           | 9450         | -                      | 1341                   | ATG       | TAA         |            |
| ND4L          | L             | 9450           | 9734         | 2                      | 285                    | ATG       | TAA         |            |
| tRNAThr       | H             | 9737           | 9800         | -                      | 64                     | -         | -           | -          |
| tRNAPro       | L             | 9801           | 9864         | 2                      | 64                     | -         | -           | -          |
| ND6           | H             | 9867           | 10400        | 7                      | 534                    | ATA       | TAA         |            |
| Cytb          | H             | 10408          | 11556        | 2                      | 1149                   | ATG       | TAA         |            |
| tRNAser(UUN)  | H             | 11539          | 11623        | 16                     | 65                     | -         | -           | -          |
| ND1           | L             | 11640          | 12581        | 1                      | 939                    | ATG       | TAG         |            |
| tRNALeu(CUN)  | L             | 12583          | 12651        | 4                      | 70                     | -         | -           | -          |
| lrRNA(16S)    | L             | 12656          | 13976        | -                      | 1321                   | -         | -           | -          |
| lrRNAVal      | L             | 13977          | 14039        | 4                      | 63                     | -         | -           | -          |
| lrRNA(12S)    | L             | 14044          | 14810        | -                      | 767                    | Me         | -          | -          |
| D-loop        | H             | 14811          | 15268        | -                      | 458                    | -         | -           | -          |
the presence of a greater number of T and C nucleotides than of A and G nucleotides. These results are similar to those for other Papilionidae butterfly species mitogenomes, with the values ranging from 0.006 in *Agehana maraho* to −0.048 in *Atrophaneura alcinous* and from −0.238 in *Teinopalpus aureus* to −0.187 in *Parnassius apollo* (Table 2).

### Protein-Coding Genes

The PCG regions of the *P. protenor* mitogenome, containing 13 PCGs, were consistent with those of other known Papilionidae mitogenomes. All the PCGs except for COI start with a canonical start codon, ATN (ATP6, COIII, ND4, ND4L, CytB, and ND1 with ATG; ATP8 and ND5 with ATT; and ND2 and ND3 with ATC), whereas the COI gene is tentatively designated by the CGA codon. With respect to the stop codon, 11 of the PCGs terminate with the complete stop codon TAA or TAG whereas COI and COII are terminated with a single T nucleotide. These incomplete stop codons are common in metazoan mitogenomes (Wolstenholme 1992) and are corrected via post-transcriptional polyadenylation during the mRNA maturation process (Ojala et al. 1981).

The RSCU value of the *P. protenor* mitogenome is summarized in Table 3 and Fig. 2. Excluding the initiation and stop codons, the 13 PCGs were 11,121 bp in length, encoding 3,707 amino acid residues. The codons CUG, CCG, ACG, and AGC were not present in these PCGs. The codons UUA (12.2), AUU (11.3), UUU (9.3), AAU (6.8), and AUA (7.3) were the five most frequently used codons in the *P. protenor* mitogenome, accounting for 40.1% of the total usage. RSCU analysis also demonstrated that the RSCU of the NNU and NNA

### Table 2. Composition and skewness in different Papilionidae mitogenomes

| Species                  | A%  | T%  | C%  | G%  | A+T% | AT skew | GC skew |
|--------------------------|-----|-----|-----|-----|------|---------|---------|
| Papilio bianor           | 39.7| 40.9| 11.8| 7.7 | 80.6 | −0.015  | −0.210  |
| Papilio maackii          | 39.8| 40.9| 11.7| 7.6 | 80.7 | −0.014  | −0.212  |
| Papilio syjanus          | 39.5| 41.1| 11.7| 7.7 | 80.6 | −0.019  | −0.205  |
| Papilio machson          | 38.9| 41.1| 11.8| 7.9 | 80.3 | −0.031  | −0.198  |
| Papilio xanthus          | 39.7| 40.7| 12.0| 7.6 | 80.4 | −0.012  | 0.224   |
| Papilio protenor         | 39.5| 41.0| 12.0| 7.5 | 80.5 | −0.019  | −0.231  |
| *Teinopalpus aureus*     | 39.7| 40.1| 12.5| 7.7 | 79.9 | −0.005  | −0.238  |
| *Agehana maraho*         | 40.5| 40.0| 12.3| 7.2 | 80.5 | 0.006   | −0.262  |
| *Troiedes aequus*        | 38.5| 41.7| 12.2| 7.6 | 80.2 | −0.040  | −0.232  |
| *Luehdorfa taibai*       | 40.4| 41.1| 11.1| 7.4 | 81.5 | −0.009  | −0.202  |
| *Papilio helenus*        | 39.9| 40.2| 12.4| 7.5 | 80.1 | −0.005  | −0.250  |
| *Papilio polytes*        | 39.6| 41.5| 11.4| 7.5 | 81.1 | −0.023  | −0.207  |
| *Papilio dardanus*       | 39.2| 40.3| 12.8| 7.7 | 79.5 | −0.014  | −0.249  |
| *Graphium chironides*    | 39.6| 40.8| 11.9| 7.7 | 80.4 | −0.015  | −0.214  |
| *Graphium timur*         | 39.8| 40.5| 11.7| 7.9 | 80.4 | −0.008  | −0.196  |
| *Atrophaneura alcinous*  | 38.6| 42.5| 11.5| 7.4 | 81.1 | −0.048  | −0.217  |
| *Sericinus montesus*     | 40.1| 40.8| 11.6| 7.4 | 80.9 | −0.009  | −0.221  |
| *Lampropetera curus*     | 39.9| 40.6| 11.7| 7.9 | 80.5 | −0.008  | −0.193  |
| *Parnassius epaphus*     | 40.0| 41.4| 11.1| 7.5 | 81.4 | −0.017  | −0.193  |
| *Parnassius bremeri*     | 40.2| 41.1| 11.2| 7.6 | 81.3 | −0.011  | −0.191  |
| *Parnassius imperator*   | 40.0| 41.1| 11.4| 7.5 | 81.1 | −0.012  | −0.207  |
| *Parnassius apollo*      | 40.0| 41.3| 11.1| 7.6 | 81.3 | −0.016  | −0.187  |
| *Parnassius cephalus*    | 40.2| 41.2| 11.1| 7.5 | 81.4 | −0.012  | −0.194  |

The examined species in this study is indicated by bold font.

### Table 3. Codon usage in the PCGs of *P. protenor* mitogenome

| Codon | Count | RSCU | Codon | Count | RSCU | Codon | Count | RSCU | Codon | Count |
|-------|-------|------|-------|-------|------|-------|-------|------|-------|-------|
| UU(U) | 343   | 1.82 | UCU(S)| 128   | 3.15 | UAU(Y) | 178   | 1.91 | UGU(C)| 31    | 1.82 |
| UU(C)| 33    | 0.18 | UCC(S)| 16    | 0.39 | UAC(Y) | 8     | 0.09 | UGC(C)| 3     | 0.18 |
| UU(A)| 451   | 4.99 | UCA(S)| 61    | 1.5  | UAA(*)| 0     | 0    | UGA(W)| 92    | 1.94 |
| UU(G)| 19    | 0.21 | UCG(S)| 4     | 0.1  | UAG(*)| 0     | 0    | UGG(W)| 3     | 0.06 |
| CU(U)| 50    | 0.55 | CCC(P)| 16   | 0.52 | CAC(H)| 61    | 1.82 | CGU(R)| 19    | 1.46 |
| CU(C)| 1     | 0.01 | CCC(P)| 16   | 0.52 | CAC(H)| 61    | 1.82 | CGU(R)| 19    | 1.46 |
| CU(A)| 21    | 0.23 | CCA(P)| 24   | 0.79 | CAA(Q)| 63    | 1.94 | CGA(R)| 29    | 2.23 |
| CU(G)| 0     | 0    | CCG(P)| 0    | 0    | CAG(Q)| 2     | 0.06 | CGG(R)| 3     | 0.23 |
| AU(U)| 420   | 1.87 | ACU(T)| 81   | 2.13 | AAU(N)| 251   | 1.87 | AGU(S)| 26    | 0.64 |
| AU(C)| 30    | 0.13 | ACC(T)| 7    | 0.18 | AAC(N)| 17    | 0.13 | AGC(S)| 0     | 0     |
| AU(A)| 271   | 1.89 | ACA(T)| 64   | 1.68 | AAA(K)| 91    | 1.82 | AGA(S)| 82    | 2.02 |
| AU(G)| 16    | 0.11 | ACG(T)| 0    | 0    | AAG(K)| 9     | 0.18 | AGG(S)| 8     | 0.2   |
| GU(U)| 69    | 2.04 | GCU(A)| 79   | 2.61 | GAU(D)| 55    | 1.72 | GGU(G)| 52    | 1.08 |
| GU(C)| 4     | 0.12 | GCC(A)| 8    | 0.26 | GAC(D)| 9     | 0.28 | GGC(G)| 5     | 0.1   |
| GU(A)| 60    | 1.78 | GCA(A)| 28   | 0.93 | GAA(E)| 65    | 1.78 | GGA(G)| 116   | 2.4   |
| GU(G)| 2     | 0.06 | GCC(A)| 6    | 0.2  | GAG(E)| 8     | 0.22 | GGG(G)| 20    | 0.41 |

*A total of 3,707 codons were analyzed excluding all initiation termination codons. *Stop codon.*
codons are >1, indicating that the third positions of U/A exhibit a high frequency of codon usage in the *P. protenor* mitogenome. In addition, our analysis also showed that tRNA<sup>Leu(UUR)</sup> (12.7%), tRNA<sup>Ile</sup> (12.1%), tRNA<sup>Phe</sup> (10.2%), tRNA<sup>Met</sup> (7.7%), and tRNA<sup>Asn</sup> (7.3%) were the most frequent amino acids in the *P. protenor* mitochondrial proteins, accounting for 50.0% of the total. The observed frequencies are similar to those detected in some other Papilionidae species.

**Intergenic Spacers and Overlapping Sequences**

The *P. protenor* mitogenome contains 118 bp of intergenic spacer regions in total, spread over 14 noncoding regions and ranging from 1 to 53 bp in length. Most of the spacers are shorter than 10 bp; only three spacers are more than 10 bp. The three intergenic spacers are located between tRNA<sup>Gln</sup> and ND2 (53 bp), tRNA<sup>Ser(UCN)</sup> and ND1 (16 bp), and ND5 and tRNA<sup>His</sup> (15 bp), respectively. The tRNA-Gln-ND2 has a high AT content (86.8%) and has been reported in other lepidopteran mitogenomes but has not been found in non-lepidopteran insects to date. Therefore, this spacer is regarded as a constitutive synapomorphic feature of lepidopteran mitogenomes (Cameron and Whiting 2008). The tRNA<sup>Ser(UCN)</sup>-ND1 intergenic spacer is inserted along with an ATACTAA motif that is found in all sequenced Papilionidae butterflies. The motif has been reported as the possible binding site for the transcription termination peptide (mtTERM protein) (Taanman 1999).

The mitogenome of *P. protenor* has a total overlap of 35 bp between genes in 13 locations, ranging from 1 to 8 bp. The longest overlapping sequence (8 bp, AAGCCTTA) is located between tRNA<sup>Trp</sup> and tRNA<sup>Trp</sup>. The second longest overlapping sequence (7 bp, ATGATAAA) is located between AT8 and ATP6. The overlapping sequence has been found in most of the lepidopteran insect mitogenomes and is reported to possibly contribute to form the structure of a hairpin loop required for posttranslational modifications (Fenn et al. 2007). The third longest overlapping sequence (6 bp) is located between tRNA<sup>Trp</sup> and ND5; all other overlapping sequences are shorter than 5 bp.

**Ribosomal RNA Genes**

The two rRNA genes, 12S rRNA and 16S rRNA, in the *P. protenor* mitogenome are located between tRNA<sup>Glu(UUR)</sup> and tRNA<sup>Vau</sup> and between tRNA<sup>Vau</sup> and the A+T-rich region, respectively. The length of the 16S rRNA gene is 1321 bp whereas that of 12S rRNA is 767 bp. The A+T contents of the 16S rRNA and 12S rRNA genes are 83.5% and 85.0%, respectively. Both of these values are well within the range reported for other lepidopteran insects (Salvato et al. 2008).

**Transfer RNA Genes**

The mitogenome of *P. protenor* contains 22 tRNA genes, ranging in length from 61 bp (tRNA<sup>Ser(AGN)</sup>) to 71 bp (tRNA<sup>Leu(CUN)</sup>). These 22 tRNA genes are scattered throughout the genome. Among these, 14 genes are encoded on the H-strand and 8 on the L-strand. All tRNA genes had the typical cloverleaf secondary structures with respective anticodons, except for the tRNA<sup>Ser(AGN)</sup> gene, in which a simple loop was substituted for a dihydrouridine arm in the *P. protenor* mitogenomes. All tRNA genes had a common length of 7 bp for the aminoacyl stem and an invariable size of 7 bp for the anticodon loop. There are 14 base pair mismatches present in the tRNA secondary structures of *P. protenor* mitogenome, including eight wobble G-U pairs, one U-G pairs, two U-U pairs, and one C-A pairs. The predicted secondary structures of the tRNAs in the *P. protenor* mitogenome were shown in Fig. 3. The nucleotide composition of these 22 tRNA genes (1,446 bp in total length) is AT biased (81.3%) and presents negative AT-skew (−0.016). The AT content of the tRNAs is higher than that in the 13 concatenated PCGs but lower than that in the rRNA genes and A+T-rich region.

**A+T-rich Region**

The A+T-rich region of *P. protenor* is 458 bp in length and is located between the 12S rRNA and tRNA<sup>Met</sup> loci. This region contains the highest AT content (93.2%) of any region of the *P. protenor*
mitogenome, with a positive AT skew (0.030) and negative GC skew (−0.353). These values are within the range of other known lepidopteran mitogenomes.

Some conserved features of other lepidopteran mitogenomes have been found in the A+T-rich region of the *P. protenor* mitogenome. Among these, the origin of minority or light strand replication (*Q*_n)

![Secondary structures for 22 typical tRNAs of the *P. protenor* mitogenome.](image1)

![A+T rich region in mitochondrial genomes of *P. protenor*.](image2)
contains the motif ATAGA followed by a poly-T stretch downstream of the 12S rRNA. Although the length of the poly-T differs among species, the ATAGA motif is prevalent in lepidopteran mitogenomes (Cameron and Whiting 2008, Tang et al. 2014). The poly-T has been considered to act as a transcription control site and the initiation site of replication (Zhang and Hewitt 1997). A poly-A commonly reported in lepidopteran mitogenomes was also observed upstream of the tRNAMet, representing a microsatellite-like (AT)₆ element preceded by the structural motif ATTTA (Fig. 4). In addition, two tandem 17 bp repeat elements are also found within the A+T-rich region of the P. protenor mitogenome.

**Phylogenetic Analyses**

In this study, the sequences of the 13 PCGs encoded by the mitochondrial genome were concatenated to reconstruct phylogenetic relationships among the Papilionidae and Pieridae of Lepidoptera using the BI, ML, and MP methods, and these multiple analyses yielded the same topology, the combination of which may result in a more complete analysis (Kristensen and Skalski 1999). With the addition of the P. protenor mitogenome sequence, this study utilized 37 species of lepidopterans, with complete sequences of 13 mitogenome PCGs which included seven Parnassinae mitochondrial sequences that are currently available for analysis (Fig. 5). The analysis produces two major clades. The first clade comprised 22 species of the family Papilionidae, which included two subfamilies (Parnassinae + Papilioninae), with strong bootstrap value (Bayesian posterior probabilities [BPP] = 100%; bootstrap branch support for maximum likelihood [MLBS] = 100%; bootstrap branch support for maximum parsimony [MPBS] = 100%). The Papilioninae subfamily (BPP = 100%; MLBS = 60%; MPBS = 80%) is composed of four clades (Papilionini + Teinopalpini + Troidini + Lampropterini), whereas the Parnassinae subfamily (BPP = 100%; MLBS = 100%; MPBS = 91%) is composed of two clades (Parnassini + Zerynthini). The second clade comprised 14 species of the family Pieridae, which included three subfamilies (Coliaclinae + Pierinae + Dismorphiinae), with high support (BPP = 100%; MLBS = 100%; MPBS = 100%). These two clades shared a sister relationship.

In addition, the phylogenetic position of Parnassini and Zerynthini is highly debated. In our study, the tribe Parnassiini clustered with Zerynthiini as a sister group, and phylogenetic analysis revealed that Parnassiini is a subfamily of Papilionidae, whereas Zerynthini and Parnassiini can be regarded as two clades of the subfamily Parnassinae. This result is consistent with those reported earlier based on morphological and molecular data (Ehrlich 1958, Igarashi 1984, Nazari et al. 2007, Huang et al. 2015). The phylogenetic results demonstrate strong support for the monophyly of the Parnassinae subfamily, which is concordant with some previous studies (Miller 1987, Nazari et al. 2007), although other studies have shown the group to be nonmonophyletic (Caterino and Sperling 1999, Omoto et al. 2004).

Although our phylogenetic analyses, which is based on the mitochondrial DNA sequence of 13 PCGs, provide evidence for the phylogenetic position of Parnassini and Zerynthini, additional analyses involving sequence information from both mitogenomes and nuclear genomes of Parnassinae members are required to determine the accurate phylogeny of Papilionidae.
Acknowledgments

This work was support by the “12th Five-Year Plan” of Science and Technology Project of China (2012BAK11B03), General Administration of Quality Supervision, Inspection and Quarantine project (2014IK023, 2015SK156) and Suzhou science and technology plan projects (SN201430). The author declare no competing interest. We thank Wenbo Li (School of life science, Anhui university, Anhui, China) for his help on our research.

References Cited

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

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.

Cameron, S. L., S. C. Barker, and M. F. Whiting. 2006. Mitochondrial genomes and the new insect order Mantophasmatodea. Mol. Phylogenet. Evol. 38: 274–279.

Cao, Y. Q., C. Ma, J. Y. Chen, and D. R. Yang. 2012. The complete mitochondrial genomes of two ghost moths, Thitarodes renzhiensis and Thitarodes yunnanensis: the ancestral gene arrangement in Lepidoptera. BMC Genomics. 13: 276.

Caterino, M. S., and F. A. Sperling. 1999. Papilio phylogeny based on mitochondrial cytochrome oxidase I and II genes. Mol. Phylogenet. Evol. 11: 122–137.

Caterino, M. S., R. D. Reed, M. M. Kuo, and F. A. Sperling. 2001. A partitioned likelihood analysis of swallowtail butterfly phylogeny (Lepidoptera: Papilionidae). Syst. Biol. 50: 106–127.

Dai, L., C. Qian, C. Zhang, L. Wang, G. Wei, J. Li, B. Zhu, and C. Liu. 2015. Characterization of the complete mitochondrial genome of Ceruca menicana and comparison with other lepidopteran insects. Plos One. 10: e0132951.

Ehrlich, P. R. 1958. The comparative morphology, phylogeny and higher classification of the butterflies (Lepidoptera: Papilionoidea). University of Kansas, Laurens, USA.

Fenn, J. D., S. L. Cameron, and M. F. Whiting. 2007. The complete mitochondrial genome sequence of the Mormon cricket (Anabrus simplex: Tettigonidae: Orthoptera) and an analysis of control region variability. Insect Mol. Biol. 16: 239–252.

Häuser, C. L. 1993. Critical comments on the phylogenetic relationships within the family Papilionidae (Lepidoptera); Nota Lepid. 16: 34–43.

Harrison, R. G. 1989. Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. Trends Ecol. Evol. 4: 6–11.

Hurst, G. D., and F. M. Jiggins. 2005. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. Proc. Biol. Sci. 272: 1525–1534.

Igarashi, S. 1984. The classification of the Papilionidae mainly based on the morphology of their immature stage. T. Lepid. Soc. JPN. 34: 41–96.

Ingman, M., H. Kaessmann, S. Piähto, and U. Gyllensten. 2000. Mitochondrial genome variation and the origin of modern humans. Nature. 408: 708–713.

Kristensen, N. P., and A. W. Skalski. 1999. Lepidoptera, moths and butterflies. Walter de Gruyter, Berlin and New York, DE.

Li, C. L., and B. Y. Zhu. 1992. Atlas of Chinese butterflies. Shanghai Far East Publisher, Shanghai, China.

Lowe, T. M., and S. R. Eddy. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 25: 955–964.

Miller, J. S. 1987. Phylogenetic studies in the Papilioninae (Lepidoptera, Papilionidae). Bull. Am. Mus. Nat. Hist. 186: 385–512.

Mutanen, M., N. Wahlberg, and L. Kaila. 2010. Comprehensive gene and taxon coverage elucidates radiation patterns in moths and butterflies. Proc. Biol. Sci. 277: 2839–2848.

Narazaki, V., E. V. Zakharov, and F. A. Sperling. 2007. Phylogeny, historical biogeography, and taxonomic ranking of Paparnassia (Lepidoptera, Papilionidae) based on morphology and seven genes. Mol. Phylogenet. Evol. 42: 131–156.

Ojala, D., J. Montoya, and G. Attardi. 1981. rRNA punctuation model of RNA processing in human mitochondria. Nature. 290: 470–474.

Omoto, K., T. Katoh, A. Chichivarkhn, and T. Yagi. 2004. Molecular systematics and evolution of the “Apollo” butterflies of the genus Paparnassia (Lepidoptera: Papilionidae) based on mitochondrial DNA sequence data. Gene. 326: 141–147.

Perna, N. T., and T. D. Kocher. 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. 41: 353–358.

Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics. 14: 817–818.

Ronquist, F., and M. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics. 19: 1572–1574.

Salvato, P., M. Simonato, A. Battisti, and E. Negrisolo. 2008. The complete mitochondrial genome of the bag-shelter moth Ochrogaster lunifer (Lepidoptera, Notodontidae). BMC Genomics. 9: 1908–1914.

Sambrook, J., and D. Russell. 2001. Molecular cloning: a laboratory manual (3rd ed). Cold Spring Harbor Laboratory Press, New York, USA.

Shadel, G. S., and D. A. Clayton. 1993. Mitochondrial transcription initiation. Variation and conservation. J. Biol. Chem. 268: 16083–16086.

Shen, J., Q. Cong, and N. V. Grishin. 2015. The complete mitochondrial genome of Papilio glauces and its phylogenetic implications. Meta Gene. 5: 68–83.

Singh, V. K., A. K. Mangalam, S. Dwivedi, and S. Naik. 1998. Primer pre-

"mier: program for design of degenerate primers from a protein sequence. Biotechniques. 24: 318–319.

Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A rapid bootstrap algorithm for the RAXML Web servers. Syst. Biol. 57: 758–771.

Taanman, J. W. 1999. The mitochondrial genome: structure, transcription, translation and replication. Biochim. Biophys. Acta. 1410: 103–123.

Talbot, G., and R. B. Sewell. 1939. The fauna of British India including Ceylon and Burma: Butterflies. Vol. 1. Today and Tomorrow’s Printers and Publishers, New Delhi, India.

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28: 2731–2739.

Tang, M., M. Tan, G. Meng, S. Yang, X. Su, S. Liu, W. Song, Y. Li, Q. Wu, A. Zhang, et al. 2014. Multiplex sequencing of pooled mitochondrial genomes—a crucial step toward biodiversity analysis using mto-metagenomics. Nucleic Acids Res. 42: e166.

Timmermans, M. J., D. C. Lees, and T. J. Simonsen. 2014. Towards a mitogenomic phylogeny of Lepidoptera. Mol. Phylogenet. Evol. 79: 169–178.

Wang, C. S. 2001. Fauna Sinica, Insecta vol. 25: Lepidoptera; Papilionidae; Papilioninae, Zerynthiinae, Parnassiinae. Science Press, Beijing, China.

Wolstenholme, D. R. 1992. Animal mitochondrial DNA: structure and evolution. Int. Rev. Cytol. 141: 173–216.

Yagi, T., G. Sasaki, and H. Takebe. 1999. Phylogeny of Japanese papilionid butterflies inferred from nucleotide sequences of the mitochondrial ND5 gene. J. Mol. Evol. 48: 42–48.

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.