昆 虫 学 报 *Acta Entomologica Sinica*, January 2012, 55(1): 91 - 100 ISSN 0454-6296

The complete mitochondrial genome of the dragon swallowtail, *Sericinus montela* Gray (Lepidoptera: Papilionidae) and its phylogenetic implication

JI Liu-Wei[1], HAO Jia-Sheng[1,*], WANG Ying[1], HUANG Dun-Yuan[1,2], ZHAO Jin-Liang[2], ZHU Chao-Dong[2,*]

(1. Laboratory of Molecular Evolution and Biodiversity, College of Life Sciences, Anhui Normal University, Wuhu, Anhui 241000, China; 2. Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China)

Abstract: The phylogenetic relationship between the zerynthiids and the other groups of Papilionidae has long been a controversial issue. In this study, the complete nucleotide sequences of the mitochondrial genome (mitogenome) of *Sericinus montela* belonging to the zerynthiids was determined using long PCR and conserved primer walking approaches. Meanwhile, the phylogenetic analysis of this species with other representative papilionid species was conducted to clarify their phylogenetic relationships on the mitogenomic level. The results showed that the entire mitochondrial DNA (mtDNA) molecule is 15 242 bp in length with the content of A, T, G and C of 40.1%, 40.8%, 7.4% and 11.7%, respectively, and its nucleotide composition of the genome is highly A + T biased (80.8%); all the 13 protein coding genes (PCGs) use standard initiation codons ATN, and all the PCGs use common stop codon (TAA), except for the ND4 and ND4L genes, which terminate with a single T; all tRNA genes form a typical clover-leaf secondary structure, except for the $tRNA^{Ser}$ (AGN), whose DHU arm forms a simple loop; there are twelve intergenic spacer regions ranging from 2 to 65 bp in size, and fifteen overlaps ranging from 1 to 8 bp in size in the mitogenome sequence. The neighbor joining and maximum parsimony phylogenetic analyses based on the 13 PCG sequences showed that *Sericins montela* and *Luehdorfia chinensis* form a clade that is sister to the *Parnassius bremeri*, suggesting that they should be designated as a taxon of tribal level within the subfamily Parnassiinae in the family of Papilionidae.

Key words: Lepidoptera; Papilionidae; *Sericinus montela*; mitochondrial genome; phylogenetic analysis

1 INTRODUCTION

Mitochondrial genomes of metazoan animals consist of a circular molecular (15 - 20 kb), which is separated from the nuclear genome. It has a remarkably conserved set of 37 genes: 13 PCGs (cox1 - 3, atp6, atp8, nadx1 - 6, nad4L and cytb), 22 transfer RNA genes (tRNAs), 2 ribosomal RNA genes (rRNAs), and a non-coding control element regulating the transcription and replication of the mitogenome (Boore, 1999; Taanman, 1999). Although gene content of animal mitogenome is well conserved, evidence has been provided that mitogenomes with gene content of more than 37, are mainly due to the emergence of tRNA gene extra copies. For example, the lepidopteran *Coreana raphaelis* has an extra copy of $tRNA^{ser}$ (AGN) (Kim *et al.*, 2006), *Acraea issoria* has an extra copy of $tRNA^{Ile}$ (AUR) (Hu *et al.*, 2010), the dipteran *Chrysomya chloropyga* has an extra $tRNA^{Ile}$ (Junqueira *et al.*, 2004), and some species of thysanopterans and hemipterans contain variable number of tRNA genes (Shao and Barker, 2003; Thao *et al.*, 2004).

The phylogeny and scientific classification of the family Papilionidae has long been a controversial and complicated issue (Su *et al.*, 2007). It has become popular to use mtDNA for the studies of phylogenetics and population genetics on the genomic level, with the development of PCR and applicability of universal primers (Kocher *et al.*, 1989; Yamauchi *et al.*, 2004). So far, only seven complete or nearly complete mitochondrial DNA

基金项目: 国家自然科学基金项目(41172004); 安徽省高校省级自然科学研究重点项目(KJ2010A142); 安徽省优秀青年科技基金(08040106811); 中国科学院动物研究所动物进化与系统学院级重点实验室专项基金(0529YX5105)

作者简介: 季刘伟, 女, 1978 年生, 安徽临泉人, 硕士研究生, 主要从事昆虫分子进化与分子系统学研究, E-mail: jiliuwei1899@163.com

通讯作者 Corresponding authors, E-mail: jshaonigpas@sina.com; zhucd@ioz.ac.cn

收稿日期 Received: 2011-04-29; 接受日期 Accepted: 2011-10-23

sequences of Papilionidae have been reported or determined: *Papilio xuthus* (Feng *et al.*, 2010), *Parnassius bremeri* (Kim *et al.*, 2009), *Papilio maraho*, *Luehdorfia chinensis*, *Troides aeacus*, *Lamproptera curius* and *Teinopalpus aureus*, and among these species, the later five were only available directly from our determination (*Troides aeacus* and *Lamproptera curius*, unpublished) or download from the GenBank (the GenBank accession numbers: *Papilio maraho*: NC_014055, *Luehdorfia chinensis*: EU622524, *Teinopalpus aureus*: NC_014398). And thus, more mitogenomes of the papilionid butterfly species are needed to clarify this problem with more details.

The dragon swallowtail, *Sericins montela* Gray, is a representative species of the tribe Zerynthiinii, subfamily Parnassiinae [sometimes subfamily Zerynthiinae by some scholars such as Chou (1998) and Li and Zhu (1992)] in the family Papilionidae. It is commonly found in Russian, Far East, Korea, China and Japan (Ackery, 1975) and may be the only multivoltine species in the Parnassiinae (Igarashi, 2003). In this paper, we report the complete mitochondrial genome of *Sericins montela*, providing the sequence information for the studies of butterfly mitogenomic evolution and phylogenetics.

2 MATERIALS AND METHODS

2.1 Specimen collection

Adult individuals of *S. montela* were collected at the Slender West Lake (Yangzhou, Jiangsu Province, China) in September, 2008. The fresh materials were preserved in 100% alcohol and stored at −20℃ until used for DNA extraction.

2.2 PCR amplification and sequencing

Total genomic DNA was extracted from thoracic muscle tissue using glass bead method after Hao *et al.* (2005). The entire genome of *S. montela* was amplified using standard PCR method and a set of mitochondrial universal primers (Simon *et al.*, 1994; Caterino and Sperling, 1999; Simmons and Weller, 2001). Universal primers were used to amplify CO Ⅰ, CO Ⅱ, Cytb and lrRNA (16S) genes. The primers for the short fragment (ND5) and the five long fragments [CO Ⅰ-CO Ⅱ, CO Ⅱ-ND5, ND5-cytb, Cytb-srRNA (12S), 12S-CO Ⅰ] were designed by the multiple sequence alignments of the complete mitogenomes of the other lepidopterans available, using the ClustalX1.8 software (Thompson *et al.*, 1997) and Primer Premier 5.0 software. The long PCR amplification was performed using TaKaRa LA Taq polymerase with the following cycling parameters: 95℃ for 5 min; 30 cycles of 95℃ for 50 s, 50℃ for 50 s, 68℃ for 2 min and 30 s; and a final extension step of 68°C for 10 min. The PCR products were detected via electrophoresis on 1.2% agarose gel, purified using the 3S Spin PCR Product Purification Kit and sequenced directly with ABI-377 automatic DNA sequencer. All the sequences of the primers for each long PCR are listed in Table 1.

Table 1 Primers for the long PCR amplification and their sequences used in this study

Primers	Upper primer sequence (5′−3′)	Lower primer sequence (5′−3′)
CO Ⅰ-CO Ⅱ	GATTGTTACCACCTTCTT	ATCTATGATTTGCTCCAC
CO Ⅱ-ND5	TTTTTGGACAATGTTCAG	TAGCTTTGTGTGGTATTC
ND5-Cytb	GGGAATACCACACAAAGC	CTGTTTGATGGAGGAATAAT
Cytb-12S	ATTATTCCTCCATCAAACAG	TTAATTCAGATCAAGATGC
12S-CO Ⅰ	AATCTGCATCTTGATCT	ATTAGAAGAAAGAGGGG

2.3 Sequnce analysis and annotation

The raw sequence was run in the NCBI BLAST program for the identification and rectification of sequence homology, and after the sequence concatenation, the complete nucleotide sequences of the *S. montela* mitogenome was obtained. PCGs and rRNA genes were identified using the ClustalX1.8 software and the NCBI Internet BLAST search function. The tRNA gene analysis was conducted using tRNAscan-SE software v. 1.21 (http://lowelab.ucsc.edu/tRNA Scan-SE), the putative tRNAs not found by tRNAscan-SE were confirmed by sequence comparison with other lepidopterans. The nucleotide composition and codon usage were calculated in MEGA 4.0 software (Tamura *et al.*, 2007). The mitogenome sequence data have been deposited into the GenBank database under the accession number HQ259122.

2.4 Sequence variation and phylogenetic analysis

The multiple alignments of the concatenated amino acid sequences of the 13 PCGs were done by the ClustalX software. The genetic distances of *S.*

montela and the 7 representative papilionid butterfly species as mentioned above in the introduction were calculated, meanwhile, the neighbor joining (NJ) (Saitou and Nei, 1987) and maximum parsimony (MP) (Fitch, 1971) phylogenetic trees of the papilionid species were reconstructed based on the PCG sequences using the JTT evolutionary model (Jones *et al.*, 1992) by the MEGA 4.0 software. In the phylogenetic analysis, a nymphalid species *Argreus hyperbius* was selected as the outgroup, and the bootstrap values with internal branch tests were obtained by 1 000 replicates to estimate the support levels for the nodes in the resultant topologies of both NJ and MP trees.

3 RESULTS

Mitogenome of *S. montela* is 15 242 bp in length, slightly longer than that of *Artogeia melete* (15 140 bp) (GenBank accession no. NC_010568) (Hong *et al.*, 2009), but much shorter than that of *Papilio maraho* (16 094 bp) (GenBank accession no. FJ810212). It contains 37 genes that are found in typical metazoan animals: 13 PCGs, 2 rRNAs and 22 tRNAs. The gene arrangement of the *S. montela* mitogenome (Fig. 1) was similar to that found in the inferred ancestral insect (Boore, 2001). Nucleotide compositions of A (40.1%), T (40.8%), G (7.4%), C (11.7%) are highly A + T biased (80.8%), which corresponds well to the AT bias generally observed in other insect mitochondrial genomes, ranging from 69.5% to 84.9% (Crozier and Crozier, 1993; Dotson and Beard, 2001).

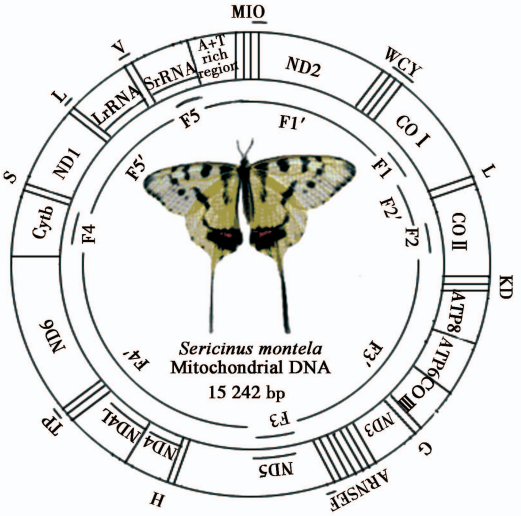

Fig. 1 Circular map of the mitogenome of *Sericunus montela*
The abbreviations for the genes are as follows: CO Ⅰ, CO Ⅱ and CO Ⅲ refer to the cytochrome oxidase subunits, CytB refers to cytochrome b, and ND1 -6 refer to NADH dehydrogenase components. tRNAs are denoted as one-letter symbol according to the IUPAC-IUB single letter amino acid codes. Gene name without underline indicates the direction of transcription from left to right and that with underline indicates right to left.

3.1 Protein-coding genes

The sequence of the 13 PCGs is 11 144 bp in length. Analysis of base composition at each codon position of the 13 PCGs of *S. montela* showed that the A + T content at the 3rd codon position (83.4%) was higher than at the first (77%) and the second (80.2%) codon position.

Table 2 shows all of the start and stop codons of the 13 PCGs. All 13 PCGs of *S. montela* mitogenome use standard ATN as start codons (ND2, ND3, ND5, ND4, ND6, ATPase8 and CO Ⅰ genes use ATT; ND1 uses ATA; ATPase6, CO Ⅱ, CO Ⅲ, ND4L and Cytb use ATG). For the stop codons, the typical TAA was used for the termination of the most putative protein sequences (Table 2), whereas, for the ND4 and ND4L genes, they terminated with a single T, which is adjacent directly to $tRNA^{His}$ downstream and to $tRNA^{Thr}$ upstream, respectively.

The number of codons (CDs) for the 13 PCGs is 3 691 in total, and all the codon usage of PCGs is listed in Table 3. Some codons have much more numbers and occur with relatively higher frequencies, such as the ATT(I), TTT(F), AAT(N) and AAA(K) are 288, 275, 274 and 265 in number, accounting for 7.8%, 7.45%, 7.4% and 7.2% of all the codons, respectively. In addition, some synonymous codons occur randomly in the genome, and thus the RSCU (relative synonymous codon usage) is often used to detect this bias (Table 3). These codons and codon families exhibit a very similar behavior with those of other lepidopteran species (Salvato *et al.*, 2008).

3.2 Transfer RNAs and ribosomal RNAs

Like all other metazoan mtDNAs sequenced, *S. montela* mtDNA contains 22 tRNAs, which range from 62 to 71 bp in size. Twenty-one tRNA sequences can be folded into perfect cloverleaf secondary structure, with the exception of $tRNA^{Ser}$ (AGN) gene, which does not form a stable stem loop structure in the DHU arm (Fig. 2). This case was seen in many other metazoan, including insect mtDNA $tRNA^{Ser}$ (AGN) (Wolstenholme, 1992). Additionally, total 8 bp overlaps exist in all the 22 tRNA sequences, the longest (8 bp) is located betwen $tRNA^{Trp}$ and $tRNA^{Cys}$ as reported in a dipteran species *Cochliomyia hominivorax* (Lessinger *et al.*, 2000).

The amino acid stems of all 22 tRNAs are invariant, all of them are 7 nt. Variability occurs at the DHU loops (3 - 11 nt) and TΨC loops (2 - 8 nt) (Fig. 2).

The 1 338-bp lrRNA (16S) and 760-bp srRNA

Table 2 Organization of the *Sericunus montela* mitochondrial genome

Gene	Direction	Nucleotide position	Size (bp)	Intergenic nucleotides	Anticodon	Start codon	Stop codon
$tRNA^{Met}$	F	1 – 69	69		CAT		
$tRNA^{Ile}$	F	70 – 134	65	0	GAT		
$tRNA^{Gln}$	R	132 – 200	69	– 3	TTG		
ND2	F	266 – 1 279	1 014	65		ATT	TAA
$tRNA^{Trp}$	F	1 278 – 1 342	65	– 2	TCA		
$tRNA^{Cys}$	R	1 335 – 1 398	64	– 8	GCA		
$tRNA^{Tyr}$	R	1 401 – 1 466	66	2	GTA		
COI	F	1 464 – 3 005	1 542	– 3		ATT	TAA
$tRNA^{Leu(UUR)}$	F	3 001 – 3 067	67	– 5	TAA		
COII	F	3 068 – 3 724	657	0		ATG	TAA
$tRNA^{Lys}$	F	3 749 – 3 819	71	24	CTT		
$tRNA^{Asp}$	F	3 819 – 3 884	66	– 1	GTC		
ATP8	F	3 885 – 4 052	168	1		ATT	TAA
ATP6	F	4 046 – 4 723	678	– 7		ATG	TAA
COIII	F	4 723 – 5 511	789	– 1		ATG	TAA
$tRNA^{Gly}$	F	5 515 – 5 579	65	3	TCC		
ND3	F	5 579 – 5 933	362	– 1		ATT	TAG
$tRNA^{Ala}$	F	5 932 – 5 996	65	– 2	TGC		
$tRNA^{Arg}$	F	5 996 – 6 059	63	– 1	TCG		
$tRNA^{Asn}$	F	6 060 – 6 125	66	0	GTT		
$tRNA^{Ser(AGN)}$	F	6 129 – 6 190	62	4	GCT		
$tRNA^{Glu}$	F	6 191 – 6 261	71	0	TTC		
$tRNA^{Phe}$	R	6 260 – 6 323	64	– 2	GAA		
ND5	R	6 323 – 8 038	1 716	– 1		ATT	TAA
$tRNA^{His}$	R	8 054 – 8 119	66	15	GTG		
ND4	R	8 117 – 9 449	1 333	– 3		ATT	T--tRNA
ND4L	R	9 459 – 9 733	274	9		ATG	TA---
$tRNA^{Thr}$	F	9 747 – 9 811	65	13	TGT		
$tRNA^{Pro}$	R	9 812 – 9 877	66	0	TGG		
ND6	F	9 880 – 10 410	5 231	2		ATT	TAA
Cytb	F	10 407 – 11 570	1 164	– 4		ATG	TAA
$tRNA^{Ser(UCN)}$	F	11 577 – 11 644	68	6	TGA		
ND1	R	11 663 – 12 598	9 536	18		ATA	TAA
$tRNA^{Leu(CUN)}$	R	12 603 – 12 672	70	4	TAG		
lrRNA	R	12 673 – 14 010	1 338	0			
$tRNA^{Val}$	R	14 011 – 14 074	64	0	TAC		
srRNA	R	14 075 – 14 833	760	0			
D-loop		14 834 – 15 241	408				

The tRNA abbreviations follow the IU-PAC-IUB three-letter code. For other abbreviations see Fig. 1 legend.

Table 3　Codon usage of PCGs in the *Sericunus montela* mitogenome

Codon (aa)	N (RSCU)	Codon (aa)	N(RSCU)	Codon (aa)	N(RSCU)	Codon (aa)	N(RSCU)
UUU(F)	275.0(1.54)	UCU(S)	53.0(2.15)	UAU(Y)	187.0(1.65)	UGU(C)	16.0(1.33)
UUC(F)	83.0(0.46)	UCC(S)	22.0(0.89)	UAC(Y)	39.0(0.35)	UGC(C)	8.0(0.67)
UUA(L)	192.0(2.32)	UCA(S)	47.0(1.91)	UAA(*)	259.0(1.61)	UGA(W)	38.0(1.58)
UUG(L)	63.0(0.76)	UCG(S)	9.0(0.37)	UAG(*)	62.0(0.39)	UGG(W)	10.0(0.42)
CUU(L)	105.0(1.27)	CCU(P)	47.0(1.72)	CAU(H)	64.0(1.64)	CGU(R)	6.0(1.26)
CUC(L)	31.0(0.38)	CCC(P)	26.0(0.95)	CAC(H)	14.0(0.36)	CGC(R)	2.0(0.42)
CUA(L)	78.0(0.94)	CCA(P)	32.0(1.17)	CAA(Q)	68.0(1.36)	CGA(R)	10.0(2.11)
CUG(L)	27.0(0.33)	CCG(P)	4.0(0.15)	CAG(Q)	32.0(0.64)	CGG(R)	1.0(0.21)
AUU(I)	288.0(1.57)	ACU(T)	47.0(1.68)	AAU(N)	274.0(1.80)	AGU(S)	12.0(0.49)
AUC(I)	78.0(0.43)	ACC(T)	24.0(0.86)	AAC(N)	31.0(0.20)	AGC(S)	12.0(0.49)
AUA(M)	240.0(1.73)	ACA(T)	31.0(1.11)	AAA(K)	265.0(1.83)	AGA(S)	26.0(1.06)
AUG(M)	37.0(0.27)	ACG(T)	10.0(0.36)	AAG(K)	24.0(0.17)	AGG(S)	16.0(0.65)
GUU(V)	26.0(1.63)	GCU(A)	26.0(2.42)	GAU(D)	78.0(1.61)	GGU(G)	12.0(1.50)
GUC(V)	6.0(0.38)	GCC(A)	4.0(0.37)	GAC(D)	19.0(0.39)	GGC(G)	3.0(0.38)
GUA(V)	26.0(1.63)	GCA(A)	12.0(1.12)	GAA(E)	91.0(1.40)	GGA(G)	14.0(1.75)
GUG(V)	6.0(0.38)	GCG(A)	1.0(0.09)	GAG(E)	39.0(0.60)	GGG(G)	3.0(0.38)

A total of 3 691 codons are analyzed, excluding the initiation and termination codons. N: Number of each codon. RSCU: Relative synonymous codon usage. * Stop codon.

genes are located between $tRNA^{Leu}$ and $tRNA^{Val}$, and between $tRNA^{Val}$ and A + T-rich region, respectively (Table 2). They are well within the size of other insects, such as *Parnassius bremeri* (1 344 bp and 773 bp), and *Papilio maraho* (1 332 bp and 778 bp). The AT contents of both rRNAs are 83.6% and 84.6%, respectively.

3.3 Non-coding regions

For *S. montela* in this study, the A + T content of AT-rich region is up to 94.1% and spans 408 bp in length, which is a little longer than that of *Coreana raphaelis* (375 bp), but much shorter than *Papilio maraho* (1 269 bp) among the known butterfly species. The *S. montela* AT-rich region does not contain any conspicuous macro-repeat units, which are commonly found in other insect species but absent from all the lepidopteran species sequenced to date (Zhang and Hewitt, 1997; Hu *et al.*, 2010; Liao *et al.*, 2010). However, it includes $(TA)_8$ and $(AT)_7$ microsatellite-like regions, and the former is preceded by a conserved motif ATTTA. The second poly-T region is close to the upstream of the $tRNA^{Met}$, spanning 8 bp long.

Besides the AT-rich region, there are another 13 non-coding regions (intergenic spacer) in the complete nucleotide sequences of *S. montela*. They are 1 – 65 bp in size (five of them are over 10 bp): the longest (65 bp) (spacer 1) is located between $tRNA^{Gln}$ and ND2 genes, and this case is usually observed in many other lepidopteran insects (Lee *et al.*, 2006; Hu *et al.*, 2010), but is not found in other insect species till now. The 24 bp long intergenic spacer (spacer 2) existed between COII gene and $tRNA^{Lys}$ gene, and this phenomenon was not found in any other lepidopteran species, which maybe derived from the duplication of one of these tRNAs (Lavrov *et al.*, 2002). The spacer of 18 bp (spacer 3) locating between $tRNA^{Ser}$ gene and ND1 gene is similar to most insect mitochondrial genomes, varying remarkably in size among lepidopteran species. Another two spacers, which are larger than 10 bp in size are located between ND5 and $tRNA^{His}$ (15 bp) and between ND4L and $tRNA^{His}$ (13 bp), respectively. Additionally, there is a 7 bp motif (ATACTAA) located between the ND1 gene and $tRNA^{Ser\,(UCR)}$ gene. In addition, there are total fifteen overlapping sequences ranging from 1 bp to 8 bp in size scattered in the whole mitogenome, and among which, the longest overlap (8 bp) is located between the $tRNA^{Trp}$ gene and $tRNA^{Cys}$ gene (Table 2).

3.4 Phylogenetic analyses

In this study, corresponding sequence variation and phylogenetic analyses were conducted to further clarify the phylogenetic relationships among the main lineages of papilionid butterflies as done in dipterans

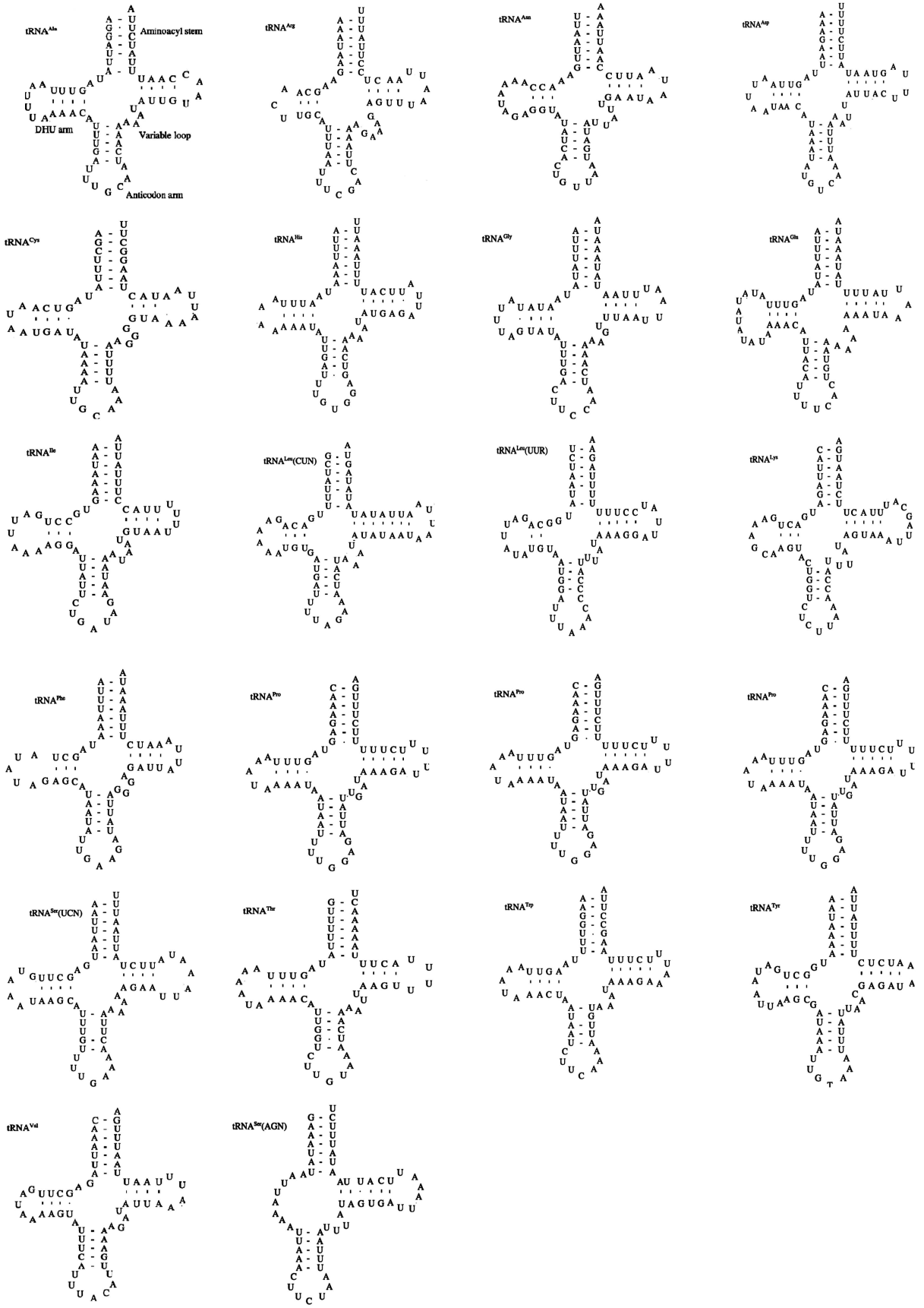

Fig. 2 Predicated secondary clover-leaf structure for the 22 tRNAs of *Sericunus montela*

The tRNAs are labelled with the abbreviations of their corresponding amino acids. Nucleotide sequences from 5′ to 3′ as indicated for $tRNA^{Ala}$. Dash (–) indicates Watson-Crick basepairing, and centred dot (·) indicates G-U base-pairing. Arms of tRNAs (clockwise from top) are the amino acid acceptor (AA) arm, TΨC (T) arm, the anticodon (AC) arm, and dihydrouridine (DHU or D) arm.

by Cameron *et al.* (2007). The 13 concatenated PCG sequences of the 8 papilionid species were multiple aligned using the ClustalX1.8 software, and the results show that the aligned sequences covers 10 360 nucleotides in total, including 6 480 conservative, 3 871 variable and 2 148 parsimony-informative sites, respectively. The results of phylogenetic analysis show that the resultant NJ and MP trees share the same topology, in which one branch is the grouping of the *S. montela*, *Luehdorfia chinensis* and *Parnassius bremeri*, and *Lamproptera curius* with the relations of (((*S. montela*, *Luehdorfia chinensis*), *Parnassius bremeri*), *Lamproptera curius*); the second branch is the clustering of *Papilio xuthus*, *Papilio maraho* and *Teinopalpus aureus*, and the two Palilio species is sister to *Teinopalpus aureus*; the third branch is the *Troides aeacus*, which is sister to the grouping of the above two branches (Fig. 3). These results are remarkably congruent with those obtained by the combined nuclear and mitochondrial genes analyses by Caterino *et al.* (2001).

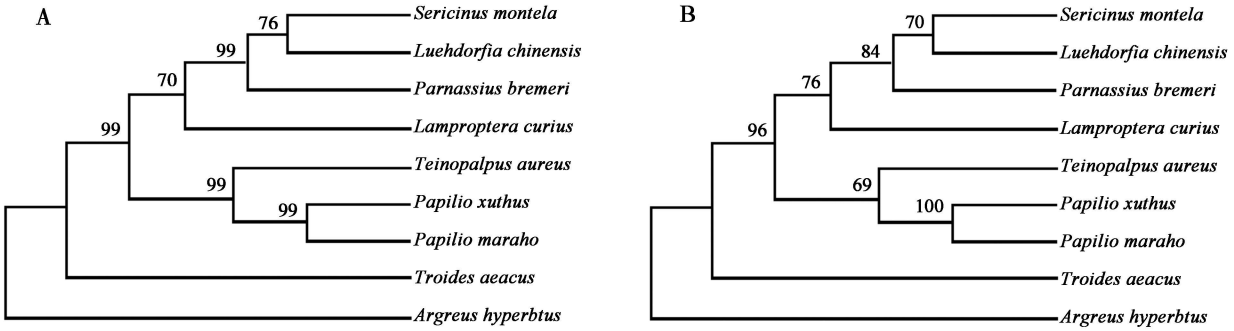

Fig. 3　NJ (A) and MP (B) phylogenetic trees based on the mitogenome PCG amino acid sequences with the JTT model
Numbers above branches indicate bootstrapping values with 1 000 replications.

4 DISCUSSION

4.1 Protein-coding genes

The facts of the PCG AT bias and the higher AT content of the third codon position compared with the first and second may be correlated with the lower pressure of the third codon site, and this case corresponds with the GC→AT trends of substitutions in the animal mitochondrial genes (Ye *et al.*, 2008).

The start codon for CO Ⅰ of *S. montela* was standard ATT, other than those non-canonical codons commonly found in other lepidopteran species examined to date (Fenn *et al.*, 2007, Liao *et al.*, 2010), such as in the cases of *Adoxophyes orana*, the codon CGA, typically special for Arginine was tentatively assumed as the CO Ⅰ start codon (Lee *et al.*, 2006), and *Coreana raphaelis*, the tetranucleotide TTAG assumed as the initiation site for CO Ⅰ (Kim *et al.*, 2006); in addition, Clary and Wolstenholme (1985) and de Bruijn (1983) also considered the start codon of CO Ⅰ were tetranucleotides ATAA, TTAA or ATTA in some insect species. Moreover, some hexanucleotides, such as TATTAG in *Ostrinia furnicalis* and *Ostrinia nubilalis* (Coates *et al.*, 2005), ATTTAA in *Anopheles gambiae* (Beard *et al.*, 1993), TATCTA in *Penaeus monodon* (Wilson *et al.*, 2000), and *Anopheles quadrimaculatus* (Mitchell *et al.*, 1993), were recognized as the start codon for CO Ⅰ, *etc.* Thus, the issue about the CO Ⅰ start codon is still far from reaching consensus, and more studies for mRNA transcripts would be needed to clarify the prediction on the CO Ⅰ initiation position.

The number and frequency of codon usage of *S. montela* are similar with those reported in all other insect species (De Brujin, 1983; Wilson *et al.*, 2000; Coates *et al.*, 2005; Kim *et al.*, 2006).

4.2 Transfer RNAs and ribosomal RNAs

In general, the sizes and structures of 22 tRNAs and 2 rRNAs, as well as the AT contents of both tRNAs and rRNAs are also well within the range of the other sequenced insects (Beard *et al.*, 1993; Kim *et al.*, 2006, 2009; Fenn *et al.*, 2007; Feng *et al.*, 2010; Hu *et al.*, 2010).

4.3 Non-coding regions

The AT-rich region known for the initiation of replication in vertebrates and invertebrate animals is located between the srRNA gene and $tRNA^{Met}$ gene (Fig. 1), and this mitochondrial control region is considerably more conservative in structure than most other insect groups in that it usually is consisted of five readily identifiable regions: the putative origin of minority or light strand replication (O_N); the microsatellite containing region; a polythymidine (poly-T) stretch; a highly variable region (150 – 375 bp in size); and a second poly-T stretch

immediately upstream of $tRNA^{Met}$ (Cameron and Whiting, 2008).

To our knowledge, the absence of the macrosatellilite-like units in *S. montela* AT-rich region are commonly found in other insect species but absent from all the lepidopteran species sequenced to date (Zhang and Hweitt, 1997; Cameron and Whiting, 2008; Hu *et al.*, 2010; Liao *et al.*, 2010); however, the microsatellite-like regions are presented as distinguishing features, such as the $(TATTA)_{31}$ between $tRNA^{Glu}$ and $tRNA^{Phe}$ in *Adoxophyes orana* (Lee *et al.*, 2006), $(TA)_{22}$ and $(TA)_{11}$ between ND3 and $tRNA^{Ala}$ in *Bombyx mori* (Yukuhiro *et al.*, 2002), $(TAA)_7$ between $tRNA^{phe}$ and ND5 in *Papilio maraho*, *etc.* A poly-T stretch is found close to the srRNA, and this polythymidine stretch has been postulated to be a transcription control and/or the initiation of replication (Zhang and Hweitt, 1997).

From the available lepidopterean mitogenome data, it is easy to find that the longest intergenic spacer shows limited sequence conservation even between closely related lepidopteran species (*Ostrinia furnicalis* and *O. nubilalis* (Coates *et al.*, 2005); *Bombyx mori* and *B. mandarina* (Pan *et al.*, 2008)). Additionally, the 7 bp motif ATACTAA is relatively conserved in location across the order Lepidoptera, and to a lesser extent across insect groups (Beard *et al.*, 1993; Dotson and Beard, 2001; Yukuhiro *et al.*, 2002; Lee *et al.*, 2006; Cameron *et al.*, 2007; Kim *et al.*, 2006, 2009; Ye *et al.*, 2008; Pan *et al.*, 2008; Hu *et al.*, 2010; Feng *et al.*, 2010; Liao *et al.*, 2010).

4.4 Phylogenetic analyses

The phylogenetic relationships among the subfamilies Zerynthiinae, Parnassiinae, and Papilioninae, especially the phylogenetic position of the Zerynthiinae have long been controversial. Up to date, there were many different opinions about this problem as mentioned below: 1) Wu (2001) proposed that the Papilionidae should be divided into three subfamilies: Papilioninae, Parnassiinae and Zerynthiinae which is further divided into three genera: *Bhutanitis* (Atkinson), *Luehdorfia* (Cruger) and *Sericinus*; 2) Chou (1998) insisted that the parnassians should be lifted up as a family parallel to the family Papilionidae which is divided into two subfamilies of Papilioninae and Zerynthiinae; 3) Li and Zhu (1992) suggested that the parnassids should be categoried as the taxon of the familial level (Parnassiidae) parallel to the family Papilionidae, and the family Parnassiidae was devided into two subfamilies of Parnassiinae and Papilioninae; 4) whereas, some European and American experts are inclined to regard them either as the taxonomic level of tribes, that is the Parnassiini and Zerynthiini, or as the subfamilial level (Parnassiinae and Zerynthiinae) (Munroe, 1961; Hancock, 1983; Igarashi, 1984; Liu, 1999; Ackery *et al.*, 1999). According to the results of this study on the grounds of mitogenomic data, it is reasonable to propose that the zerynthiids (including *S. montela* and *Luehdorfia chinensis*) should be categorized as a taxon of tribal level (the tribe Zerynthini) which was commonly accepted by international scholars at present, and constitute the subfamily Parnassiinae together with the tribe Parnassiinii.

References

Ackery PR, 1975. A guide to the genera and species of Parnassiinae (Lepidoptera: Papilionidae). *Bull. Br. Mus. Nat. Hist.* (*Ent.*), 31: 71 – 105.

Ackery PR, de Jong R, Vane-Wright RI, 1999. The butterflies: Hedyloidea, Hesperioidea and Papilionoidea. In: Kristensen NP ed. Lepidoptera, Moths and Butterflies. 1. Evolution, Systematics and Biogeography. Handbook of Zoology, 4(35), Lepidoptera. de Gruyter, Berlin. 263 – 300.

Beard CB, Hamm DMF, Collins H, 1993. The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. *Insect Mol. Biol.*, 2: 103 – 124.

Boore JL, 1999. Animal mitochondrial genomes. *Nucleic Acids Res.*, 27: 1767 – 1780.

Boore JL, 2001. Complete mitochondrial genome sequence of the polychaete annelid *Platynereis dumerilii*. *Mol. Biol. Evol.*, 18(7): 1413 – 1416.

Cameron SL, Lambkin CL, Barker SC, Whiting MF, 2007. A mitochondrial genome phylogeny of Diptera: whole genome sequences resolve relationships over broad timescales with high precision. *Syst. Entomol.*, 32: 40 – 59.

Cameron SL, Whiting MF, 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.

Caterino MS, Reed RD, Kuo MM, Sperling FAH, 2001. A partitioned likelihood analysis of swallowtail butterfly phylogeny (Lepidoptera: Papilionidae). *Syst. Biol.*, 50(1): 106 – 127.

Caterino MS, Sperling FA, 1999. *Papilio* phylogeny based on mitochondrial cytochrome oxidase I and II genes. *Mol. Phylogenet. Evol.*, 11(1): 122 – 137.

Chou I, 1998. Classification and Identification of Chinese butterflies. Henan Scientific and Technological Publishing House, Zhengzhou.

Clary DO, Wolstenholme DR, 1985. The mitochondrial DNA molecular of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *J. Mol. Evol.*, 22: 252 – 271.

Coates BS, Sumerford DV, Hellmich RL, Lewis LC, 2005. Partial mitochondrial genome sequences of *Ostrinia nubilalis* and *Ostrinia furnicalis*. *Int. Biol. Sci.*, 1: 13 – 18.

Crozier RH, Crozier YC, 1993. The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics*, 113: 97 – 117.

De Bruijn MHL, 1983. *Drosophila melanogaster* mitochondrial DNA, a novel organization and genetic code. *Nature*, 304: 234 – 241.

Dotson EM, Beard CB, 2001. Sequence and organization of the mitochondrial genome of the Chagas disease vector, *Triatoma dimidiata*. *Insect Mol. Biol.*, 10(3): 205 – 215.

Feng X, Liu DF, Wang NX, Zhu CD, Jiang GF, 2010. The mitochondrial genome of the butterfly *Papilio xuthus* (Lepidoptera: Papilionidae) and related phylogenetic analyses. *Mol. Biol. Rep.*, 37(8): 3777 - 3788.

Fenn JD, Cameron SL, Whiting MF, 2007. The complete mitochondrial genome of the Mormon cricket (*Anabrus simplex*: Tettigoniidae: Orthoptera) and an analysis of control region variability. *Insect Mol. Biol.*, 16: 239 - 252.

Fitch WM, 1971. Towards defining the course of evolution: minimum change for a special tree topology. *Syst. Zool.*, 20: 406 - 416.

Hancock DL, 1983. Classification of the Papilionidae (Lepidoptera): a phylogenetic approach. *Smithersia*, 2: 1 - 48.

Hao JS, Li CX, Sun XY, Yang Q, 2005. Phylogeny and divergence time estimation of cheilostome bryozoans based on mitochodrial 16S rRNA gene sequences. *Chin. Sci. Bull.*, 50(12): 1205 - 1212.

Hong GY, Jiang ST, Yu M, Yang Y, Li F, Xue FS, Wei ZJ, 2009. The complete nucleotide sequence of the mitochondrial genome of the cabbage butterfly, *Artogeia melete* (Lepidoptera: Pieridae). *Acta Biochim. Biophys. Sin.*, 41(6): 446 - 455.

Hu J, Zhang DX, Hao JS, Huang DY, Cameron S, Zhu CD, 2010. The complete mitochondrial genome of the yellow coaster, *Acraea issoria* (Lepidoptera: Nymphalidae: Heliconiinae: Acraeini): sequence, gene organization and a unique tRNA translocation event. *Mol. Biol. Rep.*, 37(7): 3431 - 3438.

Igarashi S, 1984. The classification of the Papilionidae mainly based on the morphology of their immature stages. *Tyo to Ga*, 34(2): 41 - 96.

Igarashi S, 2003. Life history of *Bhutanitis mansfieldi* in comparison with those of related species. *Butterflies* (*Publication of the Butterfly Society of Japan*), 35: 20 - 39.

Jones DT, Taylor WR, Thornton JM, 1992. The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences*, 8: 275 - 282.

Junqueira ACM, Lessinger AC, Torres TT, Silva FR, Vettore AL, Arruda P, Espin AMA, 2004. The mitochondrial genome of the blowfly *Chrysomya chloropyga* (Diptera: Calliphoridae). *Gene*, 339: 7 - 15.

Kim I, Lee EM, Seol KY, Yun EY, Lee YB, Hwang JS, Jin BR, 2006. The mitochondrial genome of the Korean hairstreak, *Coreana raphaelis* (Lepidoptera: Lycaenidae). *Insect Mol. Biol.*, 15(2): 217 - 225.

Kim MI, Baek JY, Kim MJ, Jeong HC, Kim KG, Bae CH, Han YS, Jin BR, Kim I, 2009. Complete nucleotide sequence and organization of the mitogenome of the red-spotted Apollo butterfly, *Parnassius bremeri* (Lepidoptera: Papilionidae) and other Lepidopteran insects. *Mol. Cells*, 28: 347 - 363.

Kocher TD, Thomas WK, Meyer A, Edwards SV, Paabo S, Villablanca FX, Wilson AC, 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. *Proc. Natl. Acad. Sci. USA*, 86: 6196 - 6200.

Lavrov DV, Boore JL, Brown WM, 2002. Complete mtDNA sequences of two millipedes suggest a new model for mitochondrial gene rearrangements: duplication and non-random loss. *Mol. Biol. Evol.*, 19: 163 - 169.

Lee ES, Shin KS, Kim MS, Park H, Cho S, Kim CB, 2006. The mitochondrial genome of the smaller tea tortrix *Adoxophyes honmai* (Lepidoptera: Tortricidae). *Gene*, 373: 52 - 57.

Lessinger AC, Martins Junqueira AC, Lemos TA, Kemper EL, Da Silva FR, Vettore AL, Arruda P, Azeredo-Espin AML, 2000. The mitochondrial genome of the primary screw worm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae). *Insect Mol. Biol.*, 9: 521 - 529.

Li CL, Zhu BY, 1992. Atlas of Chinese Butterflies. Shanghai Far East Publisher, Shanghai.

Liao F, Wang L, Wu S, Li YP, Zhao L, Huang GM, Niu CJ, Liu YQ, Li MG, 2010. The complete mitochondrial genome of the fall webworm, *Hyphantria cunea* (Lepidoptera: Arctiidae). *Int. J. Biol. Sci.*, 6(2): 172 - 186.

Liu WP, 1999. A review of *Parnassius*. *Entomological Knowledge*, 36(6): 360 - 366.

Mitchell SE, Cockburn AF, Seawright JA, 1993. The mitochondrial genome of *Anopheles quadrimaculatus* species A: complete nucleotide sequence and gene organization. *Genome*, 36: 1058 - 1073.

Munroe E, 1961. The classification of the Papilionidae (Lepidoptera). *Can. Entomol. Suppl.*, 17: 1 - 51.

Pan MH, Yu QY, Xia YL, Dai FY, Liu YQ, Lu C, Zhang Z, Xiang ZH, 2008. Characterization of the mitochondrial genome of the Chinese wild mulberry silkworm, *Bomyx mandarina* (Lepidoptera: Bombycidae). *Science in China*, 51: 693 - 701.

Saitou N, Nei M, 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.*, 4: 406 - 425.

Salvato P, Simonato M, Battisti A, Negrisolo E, 2008. The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae). *BMC Genomics*, 9: 331.

Shao R, Barker SC, 2003. The highly rearranged mitochondrial genome of the plague thrips, *Thrips imaginis* (Insecta: Thysanoptera): convergence of two novel gene boundaries and an extraordinary arrangement of rRNA genes. *Mol. Biol. Evol.*, 20: 362 - 370.

Simmons RB, Weller SJ, 2001. Utility and evolution of cytochrome b in insects. *Mol. Phylogenet. Evol.*, 20: 196 - 210.

Simon C, Frati F, Bekenbach A, Crespi B, Liu H, Flook P, 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.*, 87: 651 - 701.

Su CY, Zhu GP, Hao JS, Chen N, Pan HC, Zhang XP, 2007. Molecular phylogenetic relationships of swallowtail butterflies of the subfamily Papilioninae (Papilionidae, Lepidoptera) based on 16S rRNA gene. *Acta Zootaxonom. Sin.*, 32(2): 334 - 341.

Taanman JW, 1999. The mitochondrial genome: structure, transcription, translation and replication. *Biochim. Biophys. Acta*, 1410: 103 - 123.

Tamura K, Dudley J, Nei M, Kumar S, 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.*, 24: 1596 - 1599.

Thao ML, Baumann L, Baumann P, 2004. Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha). *BMC Evol. Biol.*, 4: 25.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG, 1997. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.*, 25: 4876 - 4882.

Wilson K, Cahill V, Ballment E, Benzie J, 2000. The complete sequence of the mitochondrial genome of the crustacean *Penaeus monodon*: are malacostracan crustaceans more closely related to insects than to branchiopods? *Mol. Biol. Evol.*, 17(6): 863 - 874.

Wolstenholme DR, 1992. Animal mitochondrial DNA: structure and evolution. *Int. Rev. Cytol.*, 141: 173 - 216.

Wu CS, 2001. Fauna Sinica, Insecta, Vol. 25. Science Press, Beijing.

Yamauchi MM, Miya MU, Nishida M, 2004. Use of a PCR-based approach for sequencing whole mitochondrial genomes of insects: two examples (cockroach and dragonfly) based on the method developed for decapod crustaceans. *Insect Mol. Biol.*, 13: 435 - 442.

Ye W, Dang JP, Xie LD, Huang Y, 2008. Complete mitochondrial genome of *Teleogryllus emma* (Orthoptera: Gryllidae) with a new gene order in Orthoptera. *Zool. Res.*, 29: 236 - 244.

Yukuhiro K, Sezutsu H, Itoh M, Shimizu K, Banno Y, 2002. Significant levels of sequence divergence and gene rearrangements have occurred between the mitochondrial genomes of the wild mulberry silkmoth, *Bombyx mandarina*, and its close relative, the domesticated silkmoth, *Bombyx mori*. *Mol. Biol. Evol.*, 19: 1385 - 1389.

Zhang DX, Hweitt GM, 1997. Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochem. Syst. Ecol.*, 25: 99 - 120.

丝带凤蝶线粒体基因组全序列及其系统学意义

季刘伟[1]，郝家胜[1,*]，王　莹[1]，黄敦元[1,2]，赵金良[2]，朱朝东[2,*]

（1. 安徽师范大学生命科学学院分子进化与生物多样性研究室，安徽芜湖 241000；
2. 中国科学院动物研究所动物进化与系统学院级重点实验室，北京 100101）

摘要：锯凤蝶类与凤蝶科其他类群的系统发生关系及其分类学地位一直存在争议。本研究采用 PCR 和 long PCR 技术测定了属于锯凤蝶类的丝带凤蝶 *Sericinus montelus* 线粒体基因组全序列；结合已有的其他凤蝶科物种的相应序列数据，基于 13 个蛋白质编码基因重建了凤蝶科主要类群的系统发生树，探讨了它们之间的系统发生关系。基因组分析结果表明：丝带凤蝶线粒体基因组全长 15 242 bp，包括 13 个编码蛋白基因（ATP6，ATP8，COⅠ－Ⅲ，ND1－6，ND4L 和 Cytb）、22 个 tRNA 基因、16S 和 12S rRNA 基因以及非编码的控制区；基因组 A，T，G 和 C 含量分别为 40.1%，40.8%，7.4% 和 11.7%，表现出明显的 AT 偏倚。所有的蛋白质编码基因都使用标准的起始密码子（ATN）；除 ND4 和 ND4L 基因使用单个的 T 作为终止密码子外，其余蛋白编码基因都使用了标准的终止密码子（TAA）。除丝氨酸 tRNA 的二氢尿苷突环缺失外，所有 tRNA 基因都形成典型的三叶草型结构。基因组中共存在 12 个大小介于 2～65 bp 之间的基因间隔区以及 15 个大小介于 1～8 bp 之间的基因重叠区，其中，存在于 COⅡ和 $tRNA^{Lys}$ 之间的 24 bp 的间隔区在其他鳞翅目昆虫中未曾见到。以邻接法和最大简约法并基于 13 个蛋白质编码基因序列对凤蝶科进行了系统发生分析。结果显示，丝带凤蝶和中华虎凤蝶 *Luehdorfia chinensis* 先构成一个支系，再和冰清绢蝶 *Parnassius bremeri* 构成姊妹群；表明锯凤蝶类应作为族级分类单元归于凤蝶科下的绢蝶亚科。

关键词：鳞翅目；凤蝶科；丝带凤蝶；线粒体基因组；系统发育分析

中图分类号：Q969　　**文献标识码**：A　　**文章编号**：0454-6296（2012）01-0091-10

（责任编辑：袁德成）