GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Complete mitochondrial genomes of the Bright Sunbeam *Curetis bulis* and the Small Copper *Lycaena phlaeas* (Lepidoptera: Lycaenidae) and their phylogenetic implications

L.L. Zhang[1], J.S. Hao[1,2], D.Y. Huang[1], X.Y. Sun[2], J.J. Hao[1], C.M. Peng[1] and Q. Yang[2]

[1]Laboratory of Molecular Evolution and Biodiversity, College of Life Sciences, Anhui Normal University, Wuhu, China
[2]State Key Laboratory of Palaeobiology and Stratigraphy, Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences, Nanjing, China

Corresponding author: J.S. Hao / Q. Yang
E-mail: jshaonigpas@sina.com / qunyang@nigpas.ac.cn

Genet. Mol. Res. 12 (4): 4434-4445 (2013)
Received September 28, 2012
Accepted January 31, 2013
Published October 10, 2013
DOI http://dx.doi.org/10.4238/2013.October.10.9

ABSTRACT. In this study, the complete mitochondrial genomes of *Curetis bulis* and *Lycaena phlaeas* were determined and analyzed. The circular genomes are 15,162 bp long for *C. bulis* and 15,280 bp long for *L. phlaeas*, with a total A+T content of 82.6 and 83.1%, respectively. Both mitogenomes contain 37 genes, and their gene orders are similar to those of other lepidopterans. All protein-coding genes (PCGs) are initiated by ATN codons, except for *cox1*, which is started with the CGA codon; all PCGs terminate in the typical stop codon TAA, except for *cox1*, *cox2*, and *nad4*, which end with a single T. The codons TTA (Leu), ATT (Ile), TTT (Phe), ATA (Met), and AAT (Asn) appear the most frequently. Both of the mitogenome A+T-rich regions harbor the motif ATAGA, followed by a 19-bp

poly(T) stretch, with *C. bulis* containing a microsatellite-like $(AT)_5$ element next to the ATTTA motif, and *L. phlaeas* containing a microsatellite-like $(TA)_6$ (AT) element next to the ATTTA motif. The phylogenetic trees of the 17 representative butterfly species, including the two species of this study, were reconstructed with the maximum likelihood and Bayesian inference methods, based on the 13 PCG nucleotide sequence data. The results of the phylogenetic analyses strongly supported the relationships of ((((Lycaenidae + Pieridae) + Nymphalidae) + Hesperiidae) + Papilionidae), which was markedly different from the traditional morphological view of the Lycaenidae and Nymphalidae considered to be sisters of each other.

Key words: Mitochondrial genome; Lepidoptera; Lycaenidae; *Curetis bulis*; *Lycaena phlaeas*; Phylogenetic analysis

INTRODUCTION

The insect mitochondrial genome (mitogenome) has a circular structure that is about 14 to 20 kb long, and includes 22 transfer RNA (tRNA) genes, two ribosomal RNA (rRNA) genes, 13 protein coding genes (PCGs), and a noncoding region (the A+T-rich region) that regulates the transcription and replication of the mitochondrial genome (Boore, 1999). So far, the mitogenomes of more than 140 insect species, including 36 lepidopteran species, have been entirely or nearly entirely sequenced (Coates et al., 2005; Kim et al., 2006; Lee et al., 2006; Liu et al., 2008; Salvato et al., 2008; Hong et al., 2009; Kim et al., 2009; Chai et al., 2012). Owing to its unique characteristics in molecular evolutionary analysis (Cheng et al., 2000), the mitogenome has been broadly applied in phylogenetic, phylogeographic, and comparative genomic studies (Zhang and Shi, 1992).

The Lepidoptera is one of the largest insect groups, accounting for more than 200,000 species. So far, many studies have focused on their phylogeny and systematics, but there are still many unresolved phylogenetic relationships within the order (Kristensen and Skalski, 1999), such as the relationships within and between the Papilionidae and Hesperiidae (butterflies) lineages. For example, we know that the family Lycaenidae is one of the largest butterfly lineages and accounts for about 6000 species worldwide (Fiedler, 1996); however, their phylogenetic relationships with other butterfly groups are still in a state of controversy. Thus, more mitogenomic data of Lycaenidae butterfly species are especially valuable for this purpose.

Curetis bulis is a medium butterfly species, belonging to the subfamily Curetinae of Lycaenidae and distributed in Southeastern Asia (New, 1993; Chou, 1994). *Lycaena phlaeas* is a butterfly species of the subfamily Lycaeninae in the family Lycaenidae, and they are commonly found in Europe, America, Africa, and Asia. In this study, we determined the complete mitogenome sequences of these two lycaenid butterfly species, described their genome features in detail, and compared these two butterfly mitogenomes with those of 14 other representative butterfly species available, in order to provide more useful information for the studies of lepidopteran comparative mitogenomics, molecular evolution, systematics, phylogeny, and so on.

MATERIAL AND METHODS

Sample collection and DNA extraction

Adult individuals of *C. bulis* and *L. phlaeas* were collected from Yandang Mountain, Zhejiang Province, China in July 2008, and from Luoyang, Henan Province, China in May 2011, respectively. After collection, samples were immediately preserved in 100% ethanol and stored at -20°C. The whole genomic DNA was extracted from thorax muscle using the Sangon Animal Genome DNA Extraction Kit in accordance with manufacturer protocols (Shanghai, China).

Primer design and polymerase chain reaction (PCR) amplification

In an effort to sequence the entire mitogenomes of the two Lycaenidae butterflies, some short fragments of the *cox1-3*, *nad4-5*, *cytb*, *rrnS*, and *rrnL* genes were initially sequenced. The primers for the *cox1* fragments for both species were adopted from those reported by Simon et al. (1994), whereas the primers for the other short fragments were designed via the alignment of several lepidopteran mitogenomes sequenced in their entirety. The short fragments were amplified under the following conditions: 95°C for 5 min; 35 cycles of 95°C for 50 s, 48 to 50°C (depending on primer combinations) for 50 s, and 72°C for 90 s; and a final extension at 72°C for 10 min, in a 50-µL reaction volume composed of 22.8 µL ddH_2O, 6 µL 10X reaction buffer, 8 µL 25 mM $MgCl_2$, 6 µL BSA, 1.5 µL dNTP mix, 1.8 µL primers (both direction), 1.5 µL template DNA, and 0.6 µL TaKaRa *Taq* polymerase (TaKaRa). For sequence information of the short fragments, some long PCR primers were designed using the software Primer Premier version 5.0 (Singh et al., 1998) to amplify overlapping long fragments. The long PCR amplification was performed using TaKaRa LA *Taq* (Dalian, China) with the following cycling parameters: 95°C for 5 min; 15 cycles of 95°C for 50 s, 50 to 56°C for 50 s, and 68°C for 150 s; another 15 cycles of 95°C for 50 s, 50 to 56°C for 50 s, and 68°C for 150 s; and a final extension at 68°C for 10 min. The PCR products were detected via electrophoresis on a 1.0% agarose gel and purified using a DNA gel extraction kit (TaKaRa) with the QIAquick PCR Purification Kit (Qiagen). All fragments were sequenced from both strands. The long PCR products were sequenced with a primer walking approach by using nested amplifications.

Sequence analysis

For the sequence analysis, we included the complete mitochondrial sequences for 17 species of lepidopterans, as well as the two that were newly sequenced in this study. Raw sequence files were proofread and assembled in BioEdit version 7.0 (Hall, 1999). PCGs and rRNA genes were identified using the ClustalX software (Thompson et al., 1997) and the NCBI online BLAST search function. The tRNA genes were identified using the tRNAscan-SE Search Server version 1.21 (Lowe and Eddy, 1997) with manual editing. The putative tRNAs not found by tRNAscan-SE were identified by alignment with other lepidopterans that had been sequenced in their entirety. The nucleotide composition and codon usage were calculated using the MEGA version 5.0 software (Tamura et al., 2011). Sequence data of the

two mitogenomes have been deposited into the GenBank database under the accession Nos. JX262888 (*C. bulis*) and JX262887 (*L. phlaeas*).

Phylogenetic analysis

Phylogenetic relationships of the 17 butterfly representative species were analyzed with Bayesian inference (BI) and maximum likelihood (ML) methods based on their 13 PCG nucleotide sequences, using the moth species *Adoxophyes honmai* (GenBank accession No. NC_008141) as the outgroup. The Bayesian analysis was performed using the MrBayes version 3.1.2 software (Huelsenbeck and Ronquist, 2001). In this process, the best-fitting nucleotide substitution model (GTR + I + Γ) was selected via Modeltest version 3.06 (Posa and Krandall, 1998); the Markov chain Monte Carlo (MCMC) was run with four chains (one cold chain and three hot chains) for 1,000,000 generations until the average standard deviation of split frequencies reached a value less than 0.01, with Bayesian posterior probabilities calculated from the sample points after the MCMC algorithm had started to converge. The ML analyses were conducted in PAUP version 4.0b10 (Swofford, 2002) with the tree bisection and reconnection branch swapping (10 random addition sequences) search method, with its best-fitting substitution model selected as in the BI analysis. The bootstrap values of the ML tree were evaluated via a bootstrapping test with 1000 iterations.

RESULTS AND DISCUSSION

Genome structure, organization, and composition

The entire mitochondrial sequence of the two lycaenid species, *C. bulis* and *L. phlaeas*, are 15,162 and 15,280 bp in length, respectively (Figure 1). Both of the mitogenomes contain a typical set of genes: 13 protein-coding genes (*cox1-3*, *nad1-6*, *nad4L*, *atp6*, *atp8*, and *cytb*), a small (*rrnS*) and a large (*rrnL*) subunit ribosomal RNA gene, 22 transfer RNA genes, and one non-coding A+T-rich region (control region, CR) (Figure 1 and Table 1). Their genome sizes are well within the size range detected in the completely sequenced lepidopteran insects, ranging from 15,140 bp in *Artogeia melete* (Hong et al., 2009) to 15,928 bp in *Agehana maraho* (Wu et al., 2010).

The gene order and orientation of the two mitogenomes are also identical to the completely sequenced lepidopteran mitogenomes, and the arrangement in lycaenid species (CR-M-I-Q-*nad*2) was a derived one. As is the case with other insect mitogenome sequences, the mitogenome nucleotide compositions are also A+T biased, as 81.4% (39 A, 42.4 T, 7.7 G, 10.9% C) for *C. bulis* (Table 2) and 82.4% (39.8 A, 42.5 T, 7.4 G, 10.3% C) for *L. phlaeas* (Table 3). Like the other published values for lepidopterans, these values also fall within the range from 77.8% in *Ochrogaster lunifer* (Salvato et al., 2008) to 82.7% in *Coreana raphaelis* (Kim et al., 2006). The highest A+T compositions of *C. bulis* and *L. phlaeas* occur in their CR (92.9, 93.2%) rather than in the *rrnS* (85.7, 85.5%), *rrnL* (84.9, 85.3%), or PCGs (Table 2 and Table 3), and this phenomenon is frequently detected in insect mitogenomes (Kim et al., 2005). In *C. bulis*, the 12 PCGs had an AT skew, ranging from -0.097 to -0.227, whereas in *L. phlaeas*, the AT skew was between -0.065 and -0.180 (Tables 2 and 3). Comparing the GC skew values of 13 PCGs, *cytb*, *rrnS*, *rrnL*, and the control region between *C. bulis* and *L. phlaeas*, the H-

strand genes of *C. bulis* had a strong GC skew, which was between -0.019 and -0.619, except for *cox1* (GC skew was 0) and the *nad1*, *nad4*, *nad4L*, and *nad5* genes, which when coded on the L-strand had a strong GC skew (GC skew between 0.265 and 0.469) (Tables 2 and 3). The same exception in *L. phlaeas* was the GC skew of *cox1*, which was 0.023. This phenomenon was incompletely quite fit in the pattern of skew values in other insects (Wei et al., 2010).

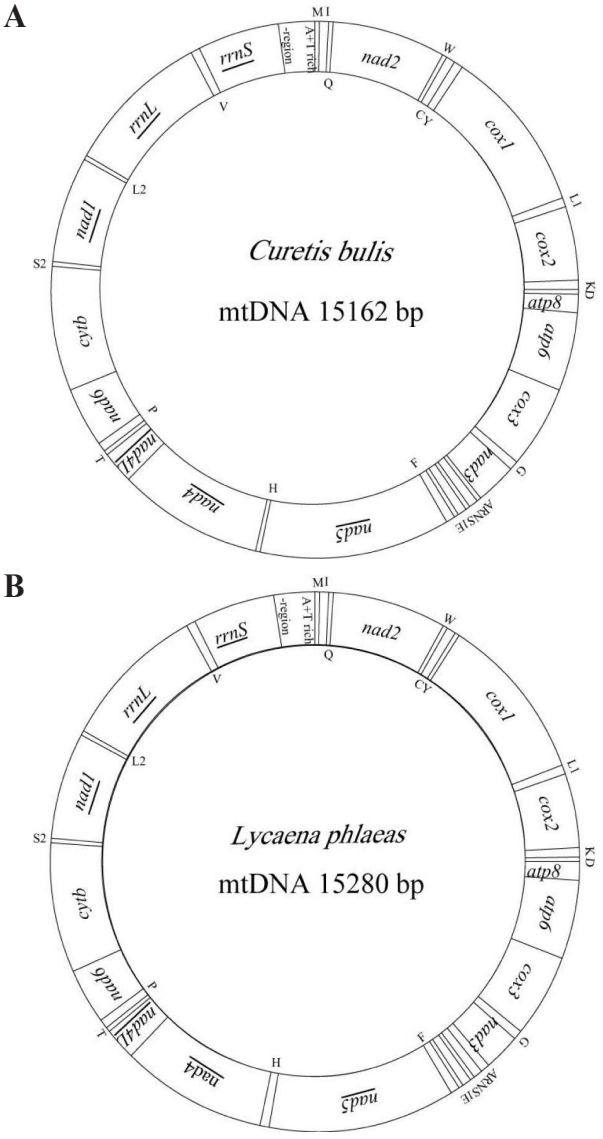

Figure 1. Circular map of the mitogenome of *Curetis bulis* (**A**) and *Lycaena phlaeas* (**B**). Gene name without underline indicates the direction of transcription from left to right and that with underline indicates right to left. Transfer RNA genes encoded by H and L strands are shown outside and inside the circular gene map, respectively. Transfer RNA genes are indicated by the IUPAC-IUB single letter amino acid codes, while L1, L2, S_1, S_2 represent *tRNA*-Leu$^{(UUR)}$, *tRNA*-Leu$^{(CUN)}$, *tRNA*-Ser$^{(AGN)}$, and *tRNA*-Ser$^{(UCN)}$, respectively.

Table 1. Summary of *Curetis bulis* (Cb) and *Lycaena phlaeas* (Lp) mitogenomes.

Gene	Direction	Nucleotide number		Size		Start codon		Stop codon	
		Cb	Lp	Cb	Lp	Cb	Lp	Cb	Lp
$tRNA^{Met}$	F	1-67	1-68	67	68	-	-	-	-
$tRNA^{Ile}$	F	68-134	71-137	67	67	-	-	-	-
$tRNA^{Gln}$	R	131-199	138-206	69	69	-	-	-	-
nad2	F	248-1273	259-1272	1026	1014	ATT	ATT	TAA	TAA
$tRNA^{Trp}$	F	1272-1338	1272-1338	67	67	-	-	-	-
$tRNA^{Cys}$	R	1331-1393	1331-1395	63	65	-	-	-	-
$tRNA^{Tyr}$	R	1394-1458	1396-1461	65	66	-	-	-	-
cox1	F	1460-2991	1462-2969	1532	1508	CGA	CGA	T-tRNA	T-tRNA
$tRNA^{Leu}$ (UUR)	F	2992-3057	2970-3036	66	67	-	-	-	-
cox2	F	3058-3736	3039-3720	679	682	ATG	ATG	T-tRNA	T-tRNA
$tRNA^{Lys}$	F	3737-3807	3721-3791	71	71	-	-	-	-
$tRNA^{Asp}$	F	3819-3886	3856-3923	68	68	-	-	-	-
atp8	F	3887-4057	3924-4091	171	168	ATC	ATC	TAA	TAA
atp6	F	4051-4728	4085-4762	678	678	ATG	ATG	TAA	TAA
cox3	F	4728-5516	4766-5554	789	789	ATG	ATG	TAA	TAA
$tRNA^{Gly}$	F	5519-5584	5572-5638	66	67	-	-	-	-
nad3	F	5585-5938	5639-5992	354	354	ATT	ATT	TAA	TAA
$tRNA^{Ala}$	F	5946-6011	5996-6062	66	67	-	-	-	-
$tRNA^{Arg}$	F	6009-6076	6060-6125	68	66	-	-	-	-
$tRNA^{Asn}$	F	6103-6168	6124-6190	66	67	-	-	-	-
$tRNA^{Ser}$ (AGN)	F	6169-6229	6207-6267	61	61	-	-	-	-
$tRNA^{Glu}$	F	6235-6302	6277-6342	68	66	-	-	-	-
$tRNA^{Phe}$	R	6301-6367	6341-6406	67	66	-	-	-	-
nad5	R	6368-8107	6406-8142	1740	1737	ATT	ATA	TAA	TAA
$tRNA^{His}$	R	8108-8173	8143-8208	66	66	-	-	-	-
nad4	R	8174-9512	8209-9547	1339	1339	ATG	ATG	T-tRNA	T-tRNA
nad4L	R	9512-9793	9547-9834	282	288	ATT	ATG	TAA	TAA
$tRNA^{Thr}$	F	9802-9867	9853-9917	66	65	-	-	-	-
$tRNA^{Pro}$	R	9868-9932	9918-9984	65	67	-	-	-	-
nad6	F	9935-10465	9985-1051	531	530	ATT	ATA	TAA	TAA
cytb	F	10465-11592	10514-11659	1128	1146	ATG	ATG	TAA	TAA
$tRNA^{Ser}$ (UCN)	F	11591-11656	11658-11723	66	66	-	-	-	-
nad1	R	11679-12617	11755-12691	939	937	ATG	ATG	TAA	TAA
$tRNA^{Leu}$ (CUN)	R	12621-12687	12693-12760	67	68	-	-	-	-
rrnL	R	12688-14008	12761-14089	1321	1329	-	-	-	-
$tRNA^{Val}$	R	14009-14075	14090-14155	67	66	-	-	-	-
*rrn*S	R	14076-14838	14156-14913	763	758	-	-	-	-
A+T–rich region	R	14839-15162	14914-15280	324	367	-	-	-	-

tRNA abbreviations follow the IU-PAC-IUB three-letter code.

Table 2. Nucleotide composition and skews of *C. bulis* mitochondrial by regions.

Gene	Nucleotide frequency				%A+T	AT-skew	GC-skew
	A	T	G	C			
atp6	0.326	0.474	0.081	0.119	80.0	-0.185	-0.190
atp8	0.346	0.549	0.020	0.085	89.5	-0.227	-0.619
cox1	0.312	0.418	0.135	0.135	73.0	-0.145	0
cox2	0.354	0.430	0.106	0.110	78.4	-0.097	-0.019
cox3	0.310	0.433	0.119	0.138	74.3	-0.166	-0.074
nad1	0.332	0.472	0.129	0.067	80.4	-0.174	0.316
nad2	0.347	0.521	0.057	0.075	86.8	-0.200	-0.136
nad3	0.339	0.483	0.071	0.107	82.2	-0.175	-0.202
nad4	0.358	0.457	0.117	0.068	81.5	-0.121	0.265
nad4L	0.362	0.491	0.108	0.039	85.3	-0.151	0.469
nad5	0.354	0.474	0.111	0.061	82.8	-0.145	0.291
nad6	0.352	0.514	0.060	0.073	86.6	-0.187	-0.098
cytb	0.338	0.433	0.101	0.120	77.1	-0.123	-0.086
rrnS	0.418	0.439	0.047	0.096	85.7	-0.025	-0.343
rrnL	0.391	0.458	0.052	0.099	84.9	-0.079	-0.311
Control region	0.426	0.503	0.031	0.040	92.9	-0.083	-0.127
Total	0.354	0.472	0.084	0.090	82.6	-0.143	-0.054

AT skew = (A% - T%) / (A% + T%); GC skew = (G% - C%) / (G% + C%).

Table 3. Nucleotide composition and skews of *Lycaena phlaeas* mitochondrial by regions.

Gene	Nucleotide frequency				%A+T	AT-skew	GC-skew
	A	T	G	C			
atp6	0.353	0.450	0.076	0.121	80.3	-0.121	-0.228
atp8	0.425	0.484	0.026	0.065	90.9	-0.065	-0.429
cox1	0.309	0.432	0.133	0.127	74.1	-0.166	0.023
cox2	0.360	0.437	0.096	0.107	79.7	-0.097	-0.054
cox3	0.346	0.426	0.106	0.122	77.2	-0.104	-0.070
nad1	0.336	0.474	0.126	0.064	81.0	-0.170	0.326
nad2	0.359	0.500	0.063	0.078	85.9	-0.164	-0.106
nad3	0.333	0.479	0.080	0.108	81.2	-0.180	-0.149
nad4	0.359	0.465	0.109	0.067	82.4	-0.129	0.239
nad4L	0.369	0.468	0.117	0.046	83.7	-0.118	0.436
nad5	0.353	0.491	0.097	0.059	84.4	-0.164	0.244
nad6	0.368	0.509	0.053	0.070	87.7	-0.161	-0.138
cytb	0.341	0.434	0.100	0.125	77.5	-0.120	-0.111
rrnS	0.402	0.453	0.050	0.095	85.5	-0.060	-0.310
rrnL	0.388	0.465	0.050	0.097	85.3	-0.090	-0.320
Control region	0.414	0.518	0.024	0.044	93.2	-0.112	-0.294
Total	0.363	0.468	0.082	0.087	83.1	-0.126	-0.059

AT skew = (A% - T%) / (A% + T%); GC skew = (G% - C%) / (G% + C%).

A+T-rich region

The A+T-rich regions of the *C. bulis* and *L. phlaeas* mitogenomes are both located between the *rrnS* and tRNA-^{Met} genes, and their sizes are 324 and 367 bp with an AT content of 92.9 and 93.2%, respectively. This region includes the ON (origin of minority or light strand replication) site, identifiable by the motif ATAGA, followed by a 19-bp poly(T) (Figure 2), which has been observed in several insect species as the recognition site for the replication origin of both strands (Saito et al., 2005) and may perform some regulatory functions together with the ATAGA motif (Kim et al., 2009). Additionally, the A+T-rich regions include multiple short microsatellite-like repeat regions. The microsatellite-like $(AT)_7$ element is observed in *C. bulis*, whereas *L. phlaeas* contains a microsatellite-like $(TA)_6$ (AT) element next to the ATTTA motif. Both of the CRs of the two lycaenid species have a 9-bp poly(A) adjacent to their tRNAs, respectively (Figure 2). Moreover, a triplicated 23-bp repeat element was found in *C. bulis*, which was similar to that in *L. phlaeas* (a duplicated 21-bp repeat element) (Figure 2). The presence of a tandem repeat in the mitochondrial A+T-rich region has been reported frequently in other insects (Yukuhiro et al., 2002; Cameron and Whiting, 2007; Liu et al., 2008; Kim et al., 2010a).

Protein-coding genes and codon usage patterns

The sequences of the 13 PCGs of *C. bulis* and *L. phlaeas* are 11,188 and 11,170 bp in length, respectively, which accounted respectively for 73.79 and 73.10% of the two lycaenid mitochondrial genomes. All protein-coding genes, with the exception of the *cox1* gene, begin with the typical ATN codons, and all of them use TAA as a termination codon, except for *cox1*, *cox2*, and *nad4*, which terminate with a single T residue, by which the complete termination signal will be produced after polyadenylation of the transcription (Ojala et al., 1981) (Table 1). Most lepidopteran species sequenced to date use CGA as the initial site for *cox1*, which is commonly detected in insects (Fenn et al., 2007). In this study, as might be expected, the *cox1* genes of the two mitogenomes start with the CGA codon.

Figure 2. Structures of the A+T-rich regions of *Curetis bulis* (**A**) and *Lycaena phlaeas* (**B**) mitogenomes. The consensus pattern repeat unit is highlighted in colors.

Among the PCGs, the *atp8* gene has the highest A+T content (89.5%), whereas the *cox1* gene has the lowest (73.0%) in *C. bulis* (Table 2). Similarly, *atp8* has the highest (90.9%) and *cox1* the lowest (74.1%) A+T content in *L. phlaeas* (Table 3). The codon usage analysis shows that a total of 3699 and 3685 amino acids are encoded by *C. bulis* and *L. phlaeas*, respectively. Some codons occur in much more numbers and higher frequencies. The most frequently occurring amino acids in *C. bulis* are TTA (L, 13.19%), ATT (I, 11.76%), TTT (F, 9.49%), ATA (M, 7.22%), and AAT (N, 6.60%) (Table S1). The amino acids frequency and order in *L. phlaeas* are identical (Table S2). Given all the above facts, we can draw a conclusion that the codons TTA (Leu), ATT (Ile), TTT (Phe), ATA (Met), and AAT (Asn) are the five most frequently occurring codons in the two lycaenid species. These codons are all composed of A or T nucleotides, thus indicating the biased usage of A and T nucleotides. As far as we know, these cases are also detected in the majority of insect PCGs (Lessinger et al., 2000).

Transfer RNA

The *C. bulis* and *L. phlaeas* mitogenomes similarly contain 22 tRNA genes (Figure 3), which is typical of animal mitogenomes (Boore, 1999), interspersed throughout the mitogenome (Table 1 and Figure 1) and ranging in size from 61 to 71 nucleotides. The two lycaenid species tRNAs have the typical cloverleaf structure of mitochondrial tRNAs, which are all identical to their counterpart tRNAs of the sequenced lepidopterans (Yukuhiro et al., 2002;

Coates et al., 2005), except for the tRNA$^{Ser(AGN)}$, which lacks the dihydrouridine (DHU) stem (Figure 3), as seen in many insect mitogenomes (Lee et al., 2006; Kim et al., 2006; Liu et al., 2008; Hong et al., 2008). In general, the lack of a DHU stem in tRNA^{Ser} is a common feature in metazoan mtDNAs (Lavrov et al., 2000).

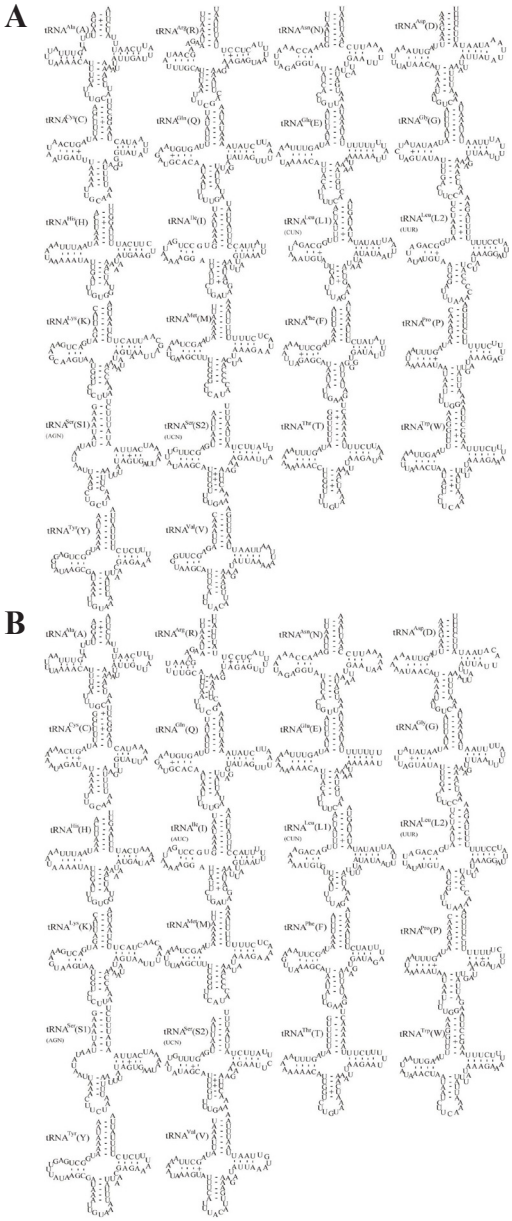

Figure 3. Predicted secondary cloverleaf structures for the 22 tRNA genes of *Curetis bulis* (**A**) and *Lycaena phlaeas* (**B**). The tRNAs are labeled with the abbreviations of their corresponding amino acids. Dashes (-) indicate Watson-Crick base pairing and centered asterisks (+) indicate unmatched base pairing.

A total of 21 mismatched base pairs occur in the tRNAs of *C. bulis*, including 8 G-U pairs, 8 U-G pairs, 4 U-U pairs, and 1 G-A pair, whereas a total of 19 mismatched base pairs exist in *L. phlaeas*, with 8 G-U pairs, 6 U-G pairs and 5 U-U pairs (Figure 3). The mismatched bases are mainly G-U, U-G, or U-U. These mismatches are corrected through RNA-editing mechanisms that are well known for arthropod mtDNAs (Lavrov et al., 2000).

Phylogenetic analysis

The results of the phylogenetic analyses show that both the BI and ML trees are the same in their topologies, with a relatively high support value for each node (Figure 4). Both trees reveal that Lycaenidae and Pieridae are sister-related with each other, to which the Nymphalidae stands as its sister; the Hesperiidae stands as a sister to the above three-family grouping, whereas the Papilionidae is a separate family diverging early from the above butterfly lineages. These above results are congruent with the conclusions of the former relevant phylogenetic analyses conducted by Kim et al. (2010a), Chai et al. (2012), and Hao et al. (2012), but contradict the suggestions of the relationship structure ((((Lycaenidae + Nymphalidae) + Pieridae) + Papilionidae) + Hesperiidae) reported by others (Wahlberg et al., 2005; Kim et al., 2010b, 2011). With respect to morphological evidences, some characteristics, such as the succinct and fully developed forelegs shared by Pieridae and Lycaenidae (or suspended pupa and degraded forelegs for the Nymphalidae), strongly hint to a close relatedness between Lycaenidae and Pieridae (Chou, 1994).

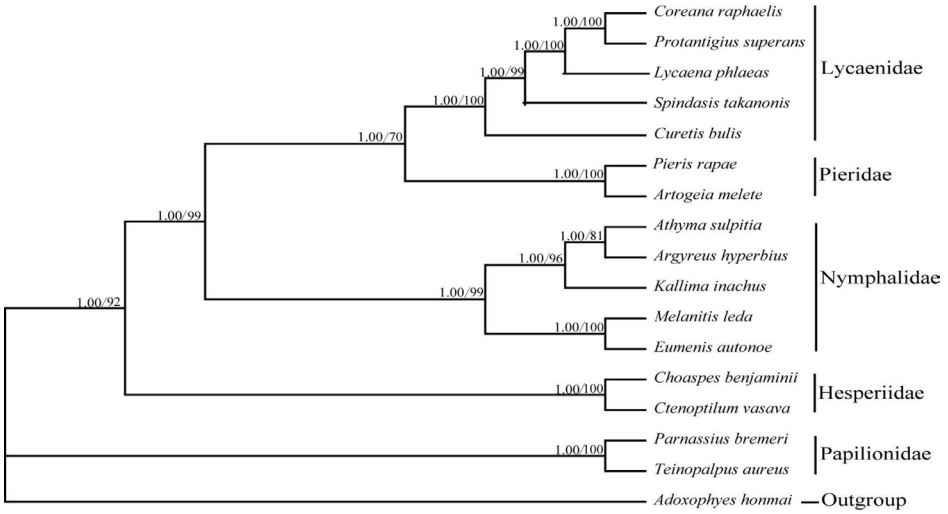

Figure 4. Inferred phylogenetic relationships among Lepidoptera based on nucleotide sequences of mitochondrial 13 PCGs using Bayesian inference (BI) and maximum likelihood (ML). Numbers at each node indicate percentages of Bayesian posterior probabilities (first value) and ML bootstrap support values (second value), respectively.

In conclusion, the phylogenetic analyses of this study at the mitogenomic level strongly supported the relationships of ((((Lycaenidae + Pieridae) + Nymphalidae) + Hesperiidae) + Papilionidae), which were strengthened by some former molecular studies, but

were not congruent with those proposed by most morphologically based studies. However, it is necessary for us to keep it in mind that more mitogenomes as well as other nuclear gene sequence data would be needed in the future to reconstruct the robust and well-accepted butterfly phylogenies.

ACKNOWLEDGMENTS

Research supported by the National Science Foundation of China (Grant #41172004), the CAS/SAFEA International Partnership Program for Creative Research Teams, the Chinese Academy of Sciences (Grant #KZCX22YW2JC104), the Provincial Key Projects of Natural Science Foundation in the Colleges of Anhui Province (Grant #KJ2010A142), and the Opening Funds from the State Key Laboratory of Palaeobiology and Stratigraphy, Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences (Grant #104143).

Supplementary material

REFERENCES

Boore JL (1999). Animal mitochondrial genomes. *Nucleic Acids Res.* 27: 1767-1780.

Cameron SL and Whiting MF (2007). Mitochondrial genomic comparisons of the subterranean termites from the genus *Reticulitermes* (Insecta: Isoptera: Rhinotermitidae). *Genome* 50: 188-202.

Chai HN, Du YZ and Zhai BP (2012). Characterization of the complete mitochondrial genomes of *Cnaphalocrocis medinalis* and *Chilo suppressalis* (Lepidoptera: Pyralidae). *Int. J. Biol. Sci.* 8: 561-579.

Cheng XY, Zhou HZ and Zhang GX (2000). Perspective of molecular biological techniques applied in insect systematics. *Acta Zootaxonomica Sin.* 25: 121-133.

Chou I (1994). Monographia Rhopalocerorum Sinensium (Monograph of Chinese Butterflies). Henan Scientific and Technological. Publishing House, Zhengzhou.

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

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

Fiedler K (1996). Host-plant relationships of lycaenid butterflies: large-scale patterns, interactions with plant chemistry, and mutualism with ants. *Entomol. Exp. Appl.* 80: 259-267.

Hall TA (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* 41: 95-98.

Hao J, Sun Q, Zhao H, Sun X, et al. (2012). The complete mitochondrial genome of *Ctenoptilum vasava* (Lepidoptera: Hesperiidae: Pyrginae) and its phylogenetic implication. *Comp. Funct. Genom.* 2012: ID 328049.

Hong GY, Jiang ST, Yu M, Yang Y, et al. (2009). The complete nucleotide sequence of the mitochondrial genome of the cabbage butterfly, *Artogeia melete* (Lepidoptera: Pieridae). *Acta Biochim. Biophis. Sin.* 41: 446-455.

Hong MY, Lee EM, Jo YH, Park HC, et al. (2008). Complete nucleotide sequence and organization of the mitogenome of the silk moth *Caligula boisduvalii* (Lepidoptera: Saturniidae) and comparison with other lepidopteran insects. *Gene* 413: 49-57.

Huelsenbeck JP and Ronquist F (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17: 754-755.

Kim I, Cha SY, Yoon MH, Hwang JS, et al. (2005). The complete nucleotide sequence and gene organization of the mitochondrial genome of the oriental mole cricket, *Gryllotalpa orientalis* (Orthoptera: Gryllotalpidae). *Gene* 353: 155-168.

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

Kim MJ, Wan X, Kim KG, Hwang JS, et al. (2010a). Complete nucleotide sequence and organization of the mitogenome of endangered *Eumenis autonoe* (Lepidoptera: Nymphalidae). *Afr. J. Biotechnol.* 9: 735-754.

Kim MI, Wan X, Kim MJ, Jeong HC, et al. (2010b). Phylogenetic relationships of true butterflies (Lepidoptera:

Papilionoidea) inferred from COI, 16S rRNA and EF-1α sequences. *Mol. Cells* 30: 409-425.
Kim MJ, Kang AR, Jeong HC, Kim KG, et al. (2011). Reconstructing intraordinal relationships in Lepidoptera using mitochondrial genome data with the description of two newly sequenced lycaenids, *Spindasis takanonis* and *Protantigius superans* (Lepidoptera: Lycaenidae). *Mol. Phylogenet. Evol.* 61: 436-445.
Kim SR, Kim MI, Hong MY, Kim KY, et al. (2009). The complete mitogenome sequence of the Japanese oak silkmoth, *Antheraea yamamai* (Lepidoptera: Saturniidae). *Mol. Biol. Rep.* 36: 1871-1880.
Kristensen NP and Skalski AW (1999). Phylogeny and Palaeontology. In: Handbuch der Zoologie, a Natural History of the Phyla of the Animal Kingdom, Vol. IV, Arthropoda: Insecta, Part 35, Lepidoptera, Moths and Butterflies, Vol. 1. Evolution, Systematics, and Biogeography. (Kristensen NP, ed.). Walter de Gruyter, Berlin and New York.
Lavrov DV, Brown WM and Boore JL (2000). A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede *Lithobius forficatus*. *Proc. Natl. Acad. Sci. U. S. A.* 97: 13738-13742.
Lee ES, Shin KS, Kim MS, Park H, et al. (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, et al. (2000). The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae). *Insect Mol. Biol.* 9: 521-529.
Liu Y, Li Y, Pan M, Dai F, et al. (2008). The complete mitochondrial genome of the Chinese oak silkmoth, *Antheraea pernyi* (Lepidoptera: Saturniidae). *Acta Biochim. Biophys. Sin.* 40: 693-703.
Lowe TM and Eddy SR (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* 25: 955-964.
New T (1993). Conservation Biology of Lycaenidae. IUCN, Switzerland.
Ojala D, Montoya J and Attardi G (1981). tRNA punctuation model of RNA processing in human mitochondria. *Nature* 290: 470-474.
Posada D and Crandall KA (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14: 817-818.
Saito S, Tamura K and Aotsuka T (2005). Replication origin of mitochondrial DNA in insects. *Genetics* 171: 1695-1705.
Salvato P, Simonato M, Battisti A and Negrisolo E (2008). The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera, Notodontidae). *BMC Genomics* 9: 331.
Simon C, Frati F, Bekenbach A, Crespi B, et al. (1994). Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chainreaction primers. *Ann. Entomol. Soc. Am.* 87: 651-701.
Singh VK, Mangalam AK, Dwivedi S and Naik S (1998). Primer premier: program for design of degenerate primers from a protein sequence. *Biotechniques* 24: 318-319.
Swofford DL (2002). PAUP* Phylogenetic Analysis Using Parsimony (*and other Methods). Version 4.0b10, Sinauer Associates. Mass, Sunderland.
Tamura K, Peterson D, Peterson N, Stecher G, et al. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28: 2731-2739.
Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, et al. (1997). The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25: 4876-4882.
Wahlberg N, Braby MF, Brower AV, de Jong R, et al. (2005). Synergistic effects of combining morphological and molecular data in resolving the phylogeny of butterflies and skippers. *Proc. Biol. Sci.* 272: 1577-1586.
Wei SJ, Shi M, Chen XX, Sharkey MJ, et al. (2010). New views on strand asymmetry in insect mitochondrial genomes. *PLoS One* 5: e12708.
Wu LW, Lees DC, Yen SH, Lu CC, et al. (2010). The complete mitochondrial genome of the near-threatened swallowtail, *Agehana maraho* (Lepidoptera: Papilionidae): evaluating sequence variability and suitable markers for conservation genetic studies. *Entomol. News* 121: 267-280.
Yukuhiro K, Sezutsu H, Itoh M, Shimizu K, et al. (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 YP and Shi LM (1992). Mitochondrial DNA polymorphisms in animals: a review. *Zool. Res.* 13: 289-298.