Complete mitogenome of the Painted Jezebel, *Delias hyparete* Linnaeus (Lepidoptera: Pieridae) and its comparison with other butterfly species

Qinghui SHI [1], Jing XIA [1], Xiaoyan SUN [2], Jiasheng HAO [1,*], Qun YANG [2,*]

1. *College of Life Sciences*, *Anhui Normal University, Wuhu 241000*, *China*;
2. *Nanjing Institute of Paeleontology and Stratigraphy*, *the Chinese Academy of Sciences*, *Nanjing 210008*, *China*

Abstract: In the present study, we report the first complete mitochondrial genome (mitogenome) of the Painted Jezebel, *Delias hyparete*. The mitogenome of *Delias hyparete* is 15186 bp in length, and has typical sets of 37 genes: 13 protein-coding genes (PCGs), 2 ribosomal RNAs, 22 transfer RNAs and a non-coding A+T-rich region. All protein-coding genes are initiated by ATN codons, except for *COI*, which is tentatively designated by the CGA codon, as observed in other butterfly species. A total of 10 PCGs harbored the complete termination codon TAA or TAG, while the *COI*, *COII* and *ND5* genes ended at a single T residue. All 22 tRNA genes show typical clover structures, with the exception of the $tRNA^{Ser(AGN)}$ which lacks the dihydrouridine (DHU) stem and is instead replaced by a simple loop. Thirteen intergenic spacers totaling 153 bp, and 13 overlapping regions totaling 46 bp are scattered throughout the whole genome. The 377 bp long of *D. hyparete* A+T-rich region is not comprised of large repetitive sequences, but harbors several features characteristic of the lepidopteran insects, including the motif ATAGA followed by an 18 bp poly-T stretch, a microsatellite-like $(AT)_5$ element preceded by the ATTTA motif, an 10 bp polyA-like stretch (AAAAATAAAA) present immediately upstream $tRNA^{Met}$.

Keywords: Lepidoptera; Pieridae; *Delias hyparete*; Mitochondrial genome

The insect mitogenome is a circular and double-stranded molecule, approximately 14–20 kb in size, typically comprised of 37 genes (including 13 PCGs, 22 tRNA, and 2 rRNA genes) and a non-coding A+T-rich region harboring the initiation sites for transcription and replication (Wolstenholme, 1992; Boore, 1999; Taanman, 1999). In recent decades, the mitogenome has been widely used in studies of phylogenetics, comparative and evolutionary genomics, population genetics and molecular evolution, etc. (Ballard & Whitlock, 2004; Simonsen et al, 2006), because of its smaller size, faster evolutionary rate, maternal inheritance and little recombination rather than the nuclear genomics (Vigilan et al, 1991; Stoneking & Soodyall, 1996). As of August 2012, the complete or nearly complete mitochondrial genomes have been determined for 376 insect species. Among these insect mitogenomes, 25 are from butterfly species, and only 3 from Peridae species (*Artogeia melete*, *Pieris rapae* and *Aporia crataegi*).

Found commonly in tropical areas of south China, India, and south-east Asia (Chou, 1998), the Painted Jezebel, *Delias hyparete* Linnaeus, is an attractive common garden butterfly species of genus *Delias*, subfamily Pierinae in the family Pieridae. Like many other Pieridae species, *Delias hyparete* is medium sized and is not easily mistaken due to its bright wing colours and graceful flight pattern. Here, we report the complete mitogenome of *Delias hyparete*, analysis of its nucleotide organization and other major characteristics

Received: 28 June 2012; Accepted: 24 October 2012
Foundation items: This work was supported by the National Science Foundation of China (Grant No.41172004), the CAS/SAFEA International Partnership Program for Creative Research Teams, Chinese Academy of Sciences (Grant No. KZCX22YW2JC104), and the Opening Funds from the State Key Laboratory of Palaeobiology and Stratigraphy, Nanjing Institute of Geology and Palaeontology, Chinese Academy of Sciences (Grant No. 104143)
*Corresponding authors, E-mails: jshaonigpas@sina.com; qunyang@nigpas.ac.cn

with those of other butterfly species available, with the aim of providing more molecular data for studies of this organisms molecular identification, population genetics, conservation biology, comparative genomics with other butterfly species and other relevant areas.

MATERIALS AND METHODS

Sample collection and DNA extraction

Adult individuals of *D. hyparete* were collected from Jinghong, Yunnan, China in July 2006. After sample collection, the fresh tissues were preserved in 100% ethanol immediately and stored at −20 C° until DNA extraction. Total genomic DNA was isolated from the thoracic muscle of an adult individual using the proteinase K-SiO_2 method, as described by Hao et al (2007).

Primer design, PCR amplification and DNA sequencing

Five short fragment sequences (500−700 bp) of *COI*, *COIII*, *ND4*, *Cytb* and 16S rRNA genes were amplified using insect universal primers (Simon et al, 1994; Caterino & Sperling, 1999; Simmons & Weller, 2001). Then, the long PCR primers were designed by Primer Premier 5.0 according to the conserved regions based on the newly acquired sequences of those short gene fragments (Table 1). All primers were synthesized by Sangon Biotechnology, Shanghai, China.

Short fragments were amplified under the following condition: 5 min of initial denaturation at 95 C°, 35 cycles of denaturation at 95 C° for 50 s, annealing at 45−51 C° (depending on primer pairs) for 50 s, extension at 72 C° for 1 min and 30 s, and a final extension at 72 C° for 10 min. Long PCR techniques were performed using TaKaRa LA Taq polymerase with the cycling parameters: an initial denaturation at 95 C° for 5 min, followed by 30 cycles of denaturation at 95 C° for 50 s, annealing at 46−55 C° (depending on primers) for 50 s, extension at 68 C° for 2 min and 30 s during the first 15 cycles, and then an additional 5 s per cycle during the last 15 cycles, and a final extension at 68 C° for 10 min. PCR products were visualized by electrophoresis on 1.2% agarose gel, then purified using a 3S Spin PCR Product Purification Kit and sequenced directly with an ABI-377 automatic DNA sequencer. For each long PCR product, the full, double-stranded sequence was determined by primer walking.

Table 1 PCR and long-PCR primers used in this study

Primers	Upper primer sequence (5'→3')	Lower primer sequence (5'→3')	TM (℃)	Size (kb)
*COI**	GGTCAACAAATCATAAAGATATTG	TAAACTTCAGGGTGACCAAAAAAT	50.0	0.67
COI-COIII	GAACTGAACTGGGAAACC	AAGCCTGAAGGATAGAAA	50.5	2.96
*COIII**	GTTCTGAGATTTCAGGTAA	TTACTAATAAATCATTTGC	49.6	0.65
COIII-ND4	ATTCCCTTTAATCCTTTCC	CGTTTAGGGAGACGAAGA	55.0	3.0
*ND4**	TTATGAAAGAAATTCTTT	CAGATATAAGGGTTAATT	45.5	0.5
ND4-Cytb	ATTACTCGCAATAAACCG	TACAGCAAATCCTCCTCA	46.5	2.6
*Cytb**	TATGTACTACCATGAGGACAAATAT	ATTACACCTCCTAATTTATTAGGAAT	47.0	0.5
Cytb-16S	ACTGAATCTGAGGAGGAT	CTTAGGGATAACAGCGTA	50.0	1.93
*16S**	CTGTACAAAGGTAGCATA	GCCAAAACTTTAGTCTAG	49.0	0.5
16S-COI	ATTACGCTGTTATCCCTAT	TTCGTGGAAAGGCTATGTC	49.8	3.0

"*" universal primer

Sequence analysis and annotation

The determined sequences were checked firstly with the NCBI Internet BLAST search function. The raw sequence files were proofread and assembled in BioEdit 7.0 (Hall, 1999) as well as ClustalX 1.8 (Thompson et al, 1997). Individual *D. hyparete* genes and the A+T-rich region were identified by aligning the sequences with homologous regions of other lepidopteran mitogenome sequences using Sequin 5.35. Sequence annotation was performed using the DNAStar package (DNAStar Inc., Madison, WI, USA) and the online blast tools available through the NCBI web site. The concatenated amino acid sequences of the 13 PCGs were obtained and analysed by ClustalX 1.8 (Thompson et al, 1997) and MEGA 5.0 (Tamura et al, 2011). Identification of tRNA genes was verified using the program tRNAscan-SE 1.21 (Lowe & Eddy, 1997). Putative tRNAs that could not be found by tRNAscan-SE were identified by sequence comparisons with other lepidopteran tRNA genes. The tandem repeats in the A+T-rich region were predicted using the Tandem Repeats Finder available online (http://tandem.bu.edu/trf/trf.html) (Benson, 1999). All mitogenome sequence data has been deposited into GenBank under the accession number JX094279.

RESULTS

Genome organization and structure

The complete mitogenome of *D. hyparete* is 15186 bp long, containing 13 protein-coding genes (*ND1-6*, *ND4L*, *COI-III*, *Cytb*, *ATP6*, *ATP8*), 2 ribosomal RNA genes (*lrRNA* and *srRNA*), 22 putative tRNA genes and one major non-coding A+T-rich region (control region) (Figure 1, Table 2). Like many other insect mitogenomes, its major strand codes for 23 genes (9 PCGs and 14 tRNAs) and the A+T-rich region, while the minor strand

codes for the remaining 14 genes (4 PCGs, 8 tRNAs and 2 rRNA genes). Besides the non-coding A+T-rich region, 13 intergenic spacer sequences ranging from 1 to 46 bp (153 bp in total) and 13 overlapping regions from 1 to 8 bp (46 bp in total) are spread over the whole mitogenome (Table 2). The 7 overlapping nucleotides (ATGATAA) are located between *ATP8* and *ATP6*. Moreover, the overlapped nucleotide sequences were also detected in all lepidopteran mitogenomes sequenced to date, and function in forming the structure of the hairpin loops for posttranslational modifications (Kim et al, 2006; Fenn et al, 2007).

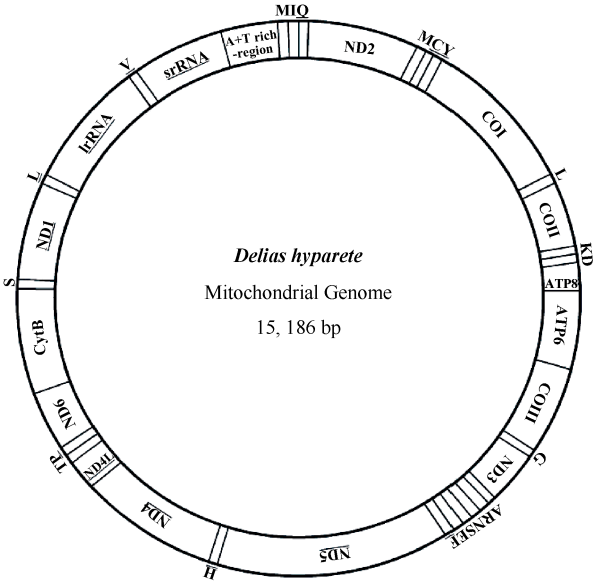

Figure 1 Circular map of *Delias hyparete* mitochondrial genome

COI, *COII* and *COIII* refer to the cytochrome oxidase subunits; *Cytb* refers to cytochrome B; *ATP6* and *ATP8* refer to subunits 6 and 8 of F0 ATPase; *ND1- 6* refer to components of NADH dehydrogenase. tRNAs are denoted as one-letter symbols consistent with the IUPAC-IUB single letter amino acid codes. Gene names that are not underlined indicate a clockwise transcriptional direction, whereas underlined indicate a counter-clockwise transcriptional direction.

Similar to other available butterfly mitogenomes, the nucleotide composition of the whole *D. hyparete* mitogenome is A+T biased (79.8%), whose value is identical to that of *Artogeia melete* (79.8%), and slightly higher than *Pieris rapae* (79.7%), and well within the range of butterfly species known to date, from 79.1% in *Eumenis autonoe* to 82.7% in *Coreana raphaelis*. Furthermore, the A+T content varies profoundly among genes and regions of the *D. hyparete* mitogenome; for example, 92.0% for the A+T-rich region, 85.1% for *srRNA* gene, 80.4% for all the tRNA genes, 83.7% for *lrRNA* gene and 78.4% for all the PCGs (Table 3, Table 4). The A+T- and G+C-skew of the whole *D. hyparete* mitogenome is −0.018 and −0.228, respectively (Table 4), indicating that more Ts and Cs than As and Gs are used. Both of the A+T- and G+C-skew values are within the corresponding values of other lepidopterans, ranging from −0.048 (*Parathyma sulpitia*) to 0.059 (*Bombyx mori*) and from −0.318 (*Ochrogaster lunifer*) to −0.158 (*Coreana raphaelis*), respectively.

Table 2 Organization of the *Delias hyparete* mitochondrial genome

Gene	Direc-tion	Position	Size (bp)	Intergenic length*	Start codon	Stop codon
tRNA^{Met}	F	1−65	65	0		
tRNA^{Ile}	F	63−127	65	−3		
tRNA^{Gln}	R	125−193	69	−3		
ND2	F	240−1253	1 014	46	ATT	TAA
tRNA^{Trp}	F	1252−1 317	66	−2		
tRNA^{Cys}	R	1 310−1 375	66	−8		
tRNA^{Tyr}	R	1 375−1 439	65	−1		
COI	F	1 442−2 972	1 531	2	CGA	T
tRNA$^{Leu(UUR)}$	F	2 973−3 038	66	0		
COII	F	3 040−3 718	679	1	ATG	T
tRNA^{Lys}	F	3 719−3 789	71	0		
tRNA^{Asp}	F	3 789−3 854	66	−1		
ATP8	F	3 855−4 016	162	0	ATT	TAA
ATP6	F	4 010−4 687	678	−7	ATG	TAA
COIII	F	4 691−-5 482	792	3	ATG	TAA
tRNA^{Gly}	F	5 485−5 550	66	2		
ND3	F	5 551−5 904	354	0	ATT	TAG
tRNA^{Ala}	F	5 903−5 968	66	−2		
tRNA^{Arg}	F	5 968−6 033	66	−1		
tRNA^{Asn}	F	6 038−6 102	65	4		
tRNA$^{Ser(AGN)}$	F	6 103−6 163	61	0		
tRNA^{Glu}	F	6 164−6 231	68	0		
tRNA^{Phe}	R	6 240−6 303	64	8		
ND5	R	6 306−8 026	1 721	2	ATT	T
tRNA^{His}	R	8 051−8 115	65	24		
ND4	R	8 155−9 461	1 307	39	ATG	T
ND4L	R	9 455−9 748	294	−7	ATA	TAA
tRNA^{Thr}	F	9 741−9 804	64	−8		
tRNA^{Pro}	R	9 805−9 869	65	0		
ND6	F	9 872−10 396	525	2	ATT	TAA
Cytb	F	10396−11544	1 149	−1	ATG	TAA
tRNA$^{Ser(UCN)}$	F	11 546−11 610	65	−2		
ND1	R	11627−12 562	936	16	ATA	TAA
tRNA$^{Leu(CUN)}$	R	12567−12 634	68	4		
lrRNA	R	12635−13 970	1 336	0		
tRNA^{Val}	R	13971−14 035	65	0		
srRNA	R	14036−14 809	774	0		
A+T-rich region		14810−15 186	377	0		

*In the column intergenic length, the positive number indicates interval base pairs between genes, while the negative number indicates the overlapping base pairs between genes.

Protein-coding genes (PCGs)

The 13 PCGs are 11 142 bp in length. Their nucleotide frequency of the majority (J strand) is T

Table 3 Characteristics of butterfly species mitogenomes

Taxon	Whole genome		PCG[b]		lrRNA		srRNA		A+T-rich region		GenBank accession no.	References
	Size(bp)	A+T (%)	No. codons[a]	A+T(%)	Size (bp)	A+T%	Size (bp)	A+T%	Size (bp)	A+T%		
Papilionidae *Parnassius bremeri*	15 389	81.3	3 734	80.1	1 344	83.9	773	85.1	504	93.6	NC_014053	Kim et al, 2009
Agehana maraho	16 094	80.5	3 718	78.2	1 333	83.7	779	85.5	1 258	95.2	NC_014055	Wu et al, 2010
Sericinus montela	15 242	80.8	3 691	79.9	1 338	83.6	760	84.6	408	94.1	HQ259122	Ji et al, 2012
Teinopalpus aureus	15 242	79.8	3 720	78.3	1 320	82.4	781	85.6	395	93.1	NC_014398	Unpublished
Nymphalidae *Acraea issoria*	15 245	79.7	3 717	78.1	1 331	83.9	788	83.7	430	96.0	NC_013604	Hu et al, 2009
Eumenis autonoe	15 489	79.1	3 728	76.8	1 335	83.7	775	85.2	678	94.6	GQ868707	Kim et al, 2010
Argyreus hyperbius	15 156	80.8	3 718	79.4	1 330	84.5	778	85.2	349	95.4	NC_015988	Wang et al, 2011
Calinaga davidis	15 267	80.4	3 737	78.8	1 337	83.8	773	85.9	389	92.0	HQ658143	Xia et al, 2011
Argynnis nerippe	15 140	80.9	3 729	79.7	1 321	84.5	773	84.9	329	95.7	NC_016419	Kim et al, 2011b
Apatura ilia	15 242	80.5	3 711	78.9	1 333	86.0	776	84.9	403	92.5	JF437925	Chen et al, 2012
Parathyma sulpitia	15 268	81.9	3 729	80.6	1 319	84.7	779	85.7	349	94.6	JQ347260	Tian et al, 2012
Sasakia charonda	15 244	79.9	3 695	78.2	1 323	84.4	775	85.0	380	91.8	NC_014224	Unpublished
Euploea mulciber	15 166	81.5	3 713	80.2	1 314	84.6	776	85.3	399	93.5	NC_016720	Unpublished
Libythea celtis	15 164	81.2	3 732	80.1	1 335	84.7	774	85.4	328	96.3	NC_016724	Unpublished
Lycaenidae *Coreana raphaelis*	15 314	82.7	3 708	81.5	1 330	85.3	777	85.8	375	94.2	DQ102703	Kim et al, 2006
Protantigius superans	15 248	81.7	3 720	80.4	1 331	85.1	739	85.6	361	93.6	NC_016016	Kim et al, 2011a
Spindasis takanonis	15 349	82.3	3 728	81.1	1 333	85.6	777	84.7	371	94.6	NC_016018	Kim et al, 2011a
Pieridae *Artogeia melete*	15 140	79.8	3 715	78.4	1 319	83.4	777	85.5	351	89.2	NC_010568	Hong et al, 2009
Pieris rapae	15 157	79.7	3 721	78.2	1 320	84.0	764	85.0	393	91.6	HM156697	Mao et al, 2010
Aporia crataegi	15 140	81.3	3 708	79.9	-	85.4	-	85.5	354	95.2	JN796473	Park et al, 2012
Delias hyparete	15 186	79.8	3 703	78.4	1 336	83.7	774	85.1	377	92.0	JX094279	This study
Hesperoidea *Ctenoptilum vasava*	15 468	80.5	3 698	78.9	1 343	84.1	774	86.4	429	88.1	JF713818	Hao et al, 2012

a: Termination codons were excluded in total codon count; b: Protein coding genes.

Table 4 Nucleotide composition and skewness in different regions of the *Delias hyparete* mitogenome

Region	Size (bp)	Nucleotide composition (%)					AT-skew	GC-skew
		T	C	A	G	A+T		
All gene	15 186	40.6	12.4	39.2	7.8	79.8	−0.018	−0.228
Genes on J-strand	6 884	43.7	13.4	33.2	9.7	76.9	−0.137	−0.160
Genes on N-strand	4 258	33.4	12.8	47.4	6.4	80.8	0.173	−0.333
PCGs	11 142	39.8	13.1	38.6	8.4	78.4	−0.015	−0.219
First codon positions	3 703	38.0	10.0	41.5	10.0	79.5	0.044	0.000
Second codon positions	3 703	37.0	17.7	33.5	12.1	70.5	−0.050	−0.188
Third codon positions	3 703	44.0	11.8	40.8	3.3	84.8	−0.038	−0.563
tRNA	1 447	39.1	11.3	41.3	8.3	80.4	0.027	−0.153
lrRNA	1 336	42.8	10.9	40.9	5.4	83.7	−0.023	−0.337
srRNA	774	46.5	10.2	38.6	4.7	85.1	−0.093	−0.369
A+T-rich region	377	49.3	4.5	42.7	3.4	92.0	−0.072	−0.139

(43.7%), A (33.2%), C (13.4%) and G (9.7%), whereas, the minority strand (N strand) is A (47.4%), T (33.4%), C (12.8%) and G (6.4%) (Table 4).

All PCGs of the *D. hyparete* mitogenome are initiated by typical ATN codons (*ND2*, *ATP8*, *ND3*, *ND5* and *ND6* with ATT; *COII*, *ATP6*, *COIII*, *ND4* and *Cytb* with ATG; *ND4L* and *ND1* with ATA), while the COI gene is tentatively designated by the CGA codon. Nine PCGs harbored the complete termination codon TAN (8 genes with TAA, *ND3* with TAG) and the remaining 4 genes (*COI*, *COII*, *ND5* and *ND4*) ended with a single T right ahead of tRNA genes (Table 2).

D. hyparete harbors 3 703 codons, excluding termination codons, and this number is slightly lower than those of *Aporia crataegi* (3 708) and *Coreana raphaelis* (3 708) (Table 3). The base composition at each codon position of the concatenated 13 PCGs showed that the A+T content at the third codon position (84.8%) was higher than those of the first (79.5%) and second (70.5%), consistent with other sequenced lepidopteran species. The relative synonymous codon usage (RSCU) revealed that NNU and NNA codons were greater than 1, indicating that the U and A are more frequently used for the third codon positions. The codon usage bias and A+T bias of the third codon position in PCGs are positively correlated with each other. Furthermore, ATT (Ile), AAT (Asn), TTT (Phe) and TTA (Leu) are the most frequently used codons, accounting for 9.2%, 7.2%, 6.6% and 6.4% of all the codons, respectively.

Transfer RNA genes and ribosomal RNA genes

In total, 22 transfer RNA genes are interspersed throughout the *D. hyparete* mitogenome, ranging in length from 61 bp ($tRNA^{Ser(AGN)}$) to 71 bp ($tRNA^{Lys}$) (Table 2). All of them are shown to be folded into the expected clover-leaf secondary structures except that the $tRNA^{Ser(AGN)}$ lacks the dihydrouridine (DHU) stem, which is replaced by a simple loop (Figure 2). The nucleotide composition of the 22 tRNA genes (1 447 bp in total size) is also A+T biased (80.4%). Sixteen tRNA genes possess 24 pair mismatches in their stems, including 5 pairs in the amino acid acceptor stems, 8 pairs in the DHU stems, 9 pairs in the anticodon stems and 2 pairs in the TΨC stems. These mismatched bases are mainly G•U and U•U, with the exception of $tRNA^{Arg}$ and $tRNA^{Leu(UUR)}$ that exhibit A•C mismatches.

Like all other insect mitogenome sequences, two ribosomal RNA genes (1 336 bp *lrRNA* and 774 bp *srRNA*) are presented in *D. hyparete* mitogenome. These two rRNA genes are composed of 83.7% and 85.1% A+T content, and located between $tRNA^{Leu(CUN)}$ and $tRNA^{Val}$, and between $tRNA^{Val}$ and the A+T-rich region, respectively (Table 2, Table 3).

A+T-rich region

The 377 bp A+T-rich region of *D. hyparete* mitogenome is located between the *srRNA* and $tRNA^{Met}$ (Figure 1, Table 2) and shows a relatively higher level of A+T content (92.0%) than those of other *D. hyparete* mitogenome regions (Table 4). This region does not contain any conspicuous macro-repeat units, but harbors some structures typical of other butterfly mitogenomes: the ON (origin of minority or light strand replication) located 18 bp upstream of the 5'-end of the *srRNA* gene, which contains the motif ATAGA followed by an 18 bp poly-T stretch; the microsatellite-like elements $(TA)_5$ and $(AT)_7$ located 196 bp and 273 bp upstream of *srRNA* gene (Figure 3); the microsatellite-like repeat $(AT)_5$ preceded by the ATTTA motif; and an 10 bp polyA-like stretch (AAAAATAAAA) present immediately upstream of $tRNA^{Met}$ (Figure 3).

DISCUSSION

General features of *D. hyparete* mitogenome

The 15 186 bp *D. hyparete* mitogenome is the largest among the three available Pieridae species, however its size is well within the range detected in completely sequenced butterfly species, ranging from 15 140 bp in *Artogeia melete* (Hong et al, 2009), *Aporia crataegi* (Park et al, 2012) and *Argynnis nerippe* (Kim et al, 2011b) to 16 094 bp in *Agehana maraho* (Wu et al, 2010). Likewise, the gene content, orientation and order are identical to the majority of other lepidopterans. This type of gene arrangement has been considered to be a synapomorphy for the order Lepidoptera (Kim et al, 2011a), differing from those ancestral insects due to the movement of $tRNA^{Met}$ to a position 5'-upstream of $tRNA^{Ile}$, resulting in the order $tRNA^{Met}$, $tRNA^{Ile}$, and $tRNA^{Gln}$, rather than the order $tRNA^{Ile}$, $tRNA^{Gln}$, and $tRNA^{Met}$ (Boore et al, 1998).

Although the gene content of the insect mitogenome is highly conserved, specific characteristics vary considerably among its different groups, e.g. variable length of A+T-rich region, different number of tRNA genes, and variable organization of genes (Hong et al, 2009). For instance, *Coreana rapaelis* has an extra copy of $tRNA^{Ser}$(AGN) (Kim et al, 2006), *Acraea issoria* harbors an extra copy of $tRNA^{Ile(AUR)b}$ (Hu et al, 2010), *Parnassius bremeri* harbors two $tRNA^{Trp}$-like and $tRNA^{Leu(UUR)}$-like sequences (Kim et al, 2009), *Argynnis nerippe* possesses two extra $tRNA^{Met}$-like and $tRNA^{Leu(UUR)}$-like sequences (Kim et al, 2011b), and *Ctenoptilum vasava* includes an extra copy of the trnS (AGN) gene and a tRNA-like insertion trnL (UUR) (Hao et al, 2012).

PCGs

To date, no agreement has been reached concerning

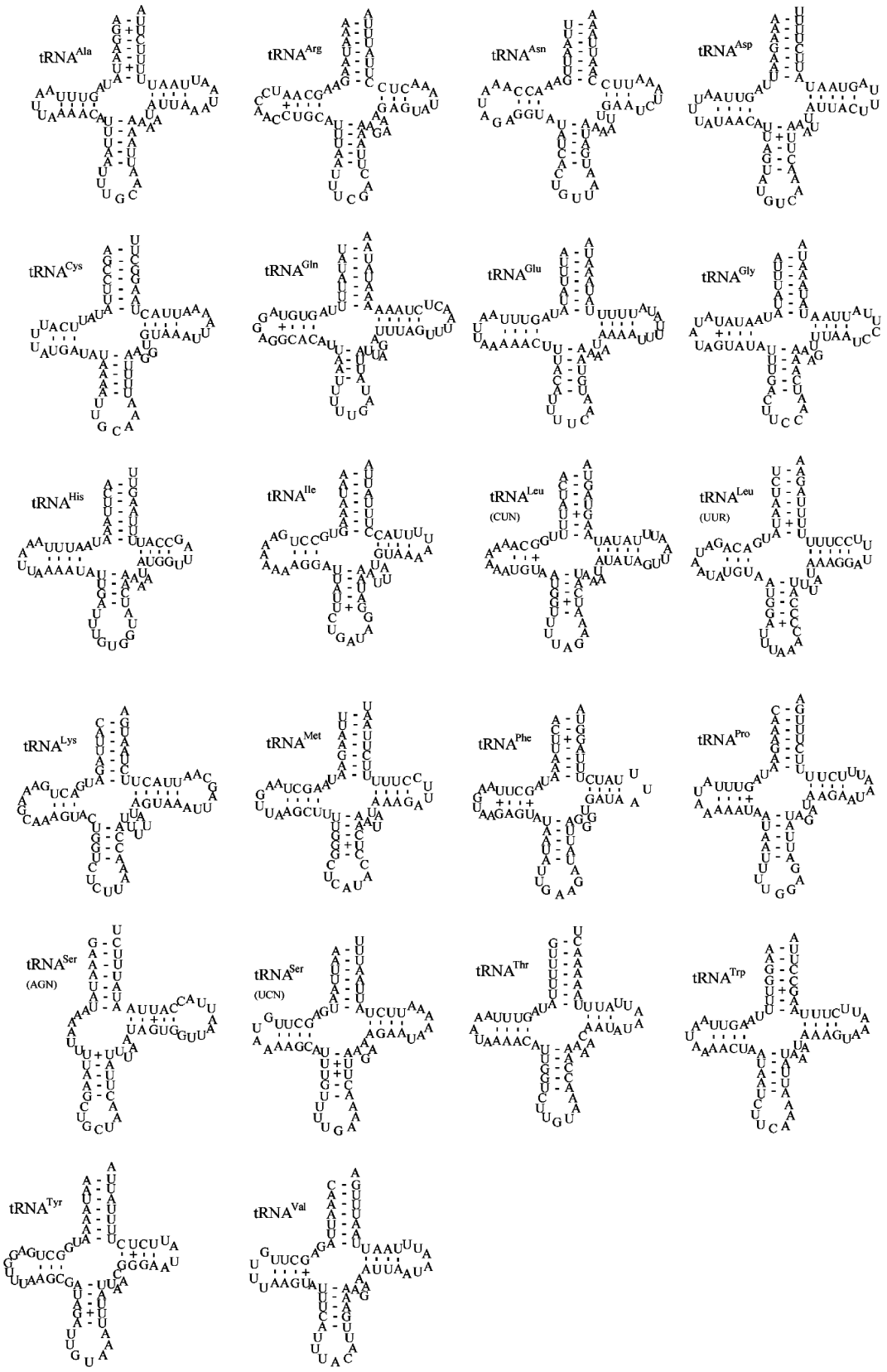

Figure 2 Predicted secondary clover-leaf structure of the *Delias hyparete* 22 tRNA genes

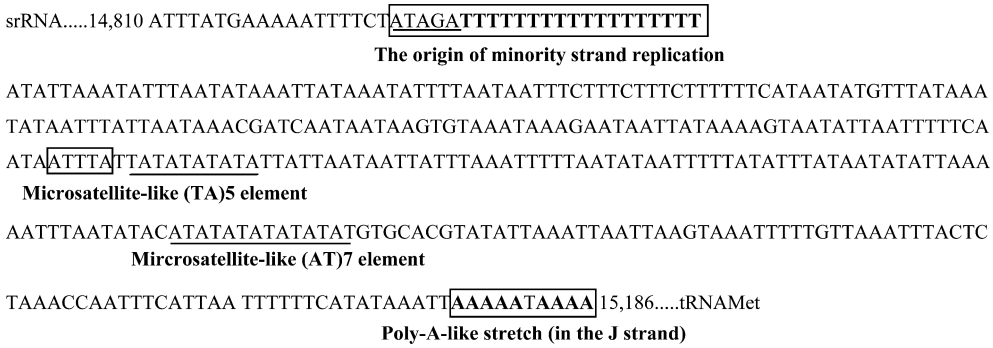

Figure 3 The structure of the A+T-rich region of the *Delias hyparete* mitochondrial genome

the insect *COI* start codons. For example, the trinucleotide TTG was supposed to be the initiation codon for *COI* gene in the *Caligula boisduvalii* (Hong et al, 2008), *Acraea issoria* (Hu et al, 2010) and *Calinaga davidis* (Xia et al, 2011); ATA for *Sericinus montela* (Ji et al, 2012), ATT for *Ctenoptilum vasava* (Hao et al, 2012) and *Aporia crataegi* (Park et al, 2012); the tetranucleotide TTAG for *Antheraea pernyi* (Liu et al, 2008) and *Coreana raphaelis* (Kim et al, 2006); and the hexanucleotides TATTAG for *Ostrinia nubilalis* and *O. furnicalis* (Coates et al, 2005), ATTACG for *Papilio xuthus* (Feng et al, 2010), TTAAAG for *Pieris rapae* (Mao et al, 2010). Here, by sequence homology comparison with other available lepidopteran species, we tentatively presumed CGA as the start codon for *COI* gene (Lee et al, 2006; Jiang et al, 2009; Hong et al, 2009; Kim et al, 2009; Kim et al, 2010; Wang et al, 2011; Tian et al, 2012, Chen et al, 2012).

Transfer and ribosomal RNA genes

The *D. hyparete* mitogenome harbors 22 tRNA genes scattered throughout the whole genome. Unlike some other butterfly species, no extra tRNA gene has been detectedin *D. hyparete* (Kim et al, 2006; Hu et al, 2010; Kim et al, 2009; Kim et al, 2011b). The *D. hyparete* tRNAs harbor a total of 24 unmatched base pairs within the stem region: 17 are G•U pairs, the remaining 7 are atypical pairs, including 5 U•U pairs and 2 A•C pairs, all of which form a weak bond in the tRNAs and could be corrected through RNA-editing mechanisms (Yokobori & Pääbo, 1995; Lavrov et al, 2000). All tRNAs possess an invariable 7 bp in the aminoacyl stem, as commonly found in other butterfly species (Kim et al, 2009; Kim et al, 2010; Chen et al, 2012), 21 tRNAs harbor 7 bp in the anticodon loop (except 9 bp in $tRNA^{His}$), and 19 tRNAs harbor 5 bp in the anticodon stem (except 4 base pairs in $tRNA^{Ile}$, $tRNA^{Gln}$ and $tRNA^{Lys}$). Additionally, most tRNA size variations result from length variations in the DHU and TΨC arms, within which loop sizes (3−9 bp) are slightly more variable than the stem sizes (3−5 bp).

In *D. hyparete,* the two rRNA locations are the same as other lepidopterans. Their sizes are likewise well within the range of other known butterfly species, from 1 314 bp of *Euploea mulciber* (Unpublished, NC_016720) to 1344 bp of *Parnassius bremeri* (Kim et al, 2009) for lrRNA, and from 739 bp of *Protantigius superans* (Kim et al, 2011a) to 788 bp of *Acraea issoria* (Hu et al, 2010) for srRNA, respectively. The A+T content value of the two rRNA genes is consistent with those observed in other available butterfly species (Table 3).

Non-coding regions

Intergenic spacer sequences The four spacers longer than 10 bp are located respectively between $tRNA^{Gln}$ and *ND2* (46 bp), $tRNA^{His}$ and *ND4* (39 bp), *ND5* and $tRNA^{His}$ (24 bp) and $tRNA^{Ser(UCN)}$ and *ND1*(16 bp). The remaining 9 spacers shorter than 10 bp are dispersed throughout the whole genome, accounting for 18.3% of the total length of entire intergenic spacers.

The first longer intergenic spacer located between $tRNA^{Gln}$ and *ND2* is not only found in *D. hyparete*, but has also been detected in other sequenced lepidopterans (Lee et al, 2006; Kim et al, 2009; Hu et al, 2010; Kim et al, 2010), ranging in size from 40 bp in *Parnassius bremeri* to 87 bp in *Sasakia charonda*. This intergenic spacer is usually considered a constitutive synapomorphic feature and a molecular signature of lepidopteran mitogenomes due to its absence in non-lepidopteran insects (Junqueira et al, 2004; Cha et al, 2007; Cameron & Whiting, 2008). Additionally, the 67.4% homology of this intergenic spacer with its neighboring *ND2* gene reveals its *ND2* duplication origin (Kim et al, 2009; Kim et al, 2010), similar to the cases presented in other lepidopterans, including *Artogeia melete* (68.8%) (Hong et al, 2009), *Pieris rapae* (71.7%) (Mao et al, 2010), *Coreana raphaelis* (62.5%) (Kim et al, 2006), *Eumenis autonoe* (74%) (Kim et al, 2010), *Parnassius bremeri* (70%) (Kim et al, 2009), *Bombyx mandarina* (70.2%) (Pan et al, 2008).

The second largest spacer is inserted between $tRNA^{His}$ and *ND4* (39 bp), which is only correspondingly detected in the *Acraea issoria* mitogenome (4 bp) (Hu et al, 2010), while absent in other sequenced butterfly

species (e.g., Kim et al, 2006; Kim et al, 2009; Kim et al, 2010; Wang et al, 2011; Hong et al, 2009; Mao et al, 2010; Ji et al, 2012; Hao et al, 2012). The sequence alignment of this spacer with the neighboring *ND4* gene revealed a sequence homology of 82.1%, also suggesting its origin of *ND4* duplication (Figure 4). The third largest spacer, located between *ND5* and $tRNA^{His}$, is 24 bp long and this region is highly variable in length among the butterfly species (8−24 bp) (Hu et al, 2010; Kim et al, 2006; Wang et al, 2011; Ji et al, 2012) and sometimes even overlaps between the two genes, such as in *Parnassius bremeri* (Kim et al, 2009) and *Eumenis autonoe* (Kim et al, 2010).

***Delias hyparete* (82.1%)**

spacer(8116-8154) TAATAAATTTAAAATCAAA - - - - TTCTATAATCAAATTT - - - TAAA

ND4(8904-8949) TCTTAAATATAAAATTAAAGAATTTCTTTCATAAAATTTATATAAA

* ***** ****** *** **** * ** ****** ****

Figure 4 Sequence alignment of the intergenic spacer located between $tRNA^{His}$ and *ND4* gene with its neighboring partial *ND4* gene

In some butterfly species, such as *Acraea issoria*, *Argyreus hyperbius*, *Calinaga davidis*, *Argynnis nerippe* and *Parathyma sulpitia* there exists 2 bp overlapping sequence resided between the $tRNA^{Ser(UCN)}$ and *ND1* genes. However, for *D. hyparete,* which we used in this study and other butterfly (moth) species, intergenic spacer sequences are detected, such as 9 bp spacer in *Diatraea saccharalis* (Unpublished, NC_013274) and 38 bp spacer in *Ostrinia nubilalis* (Coates et al, 2005). All of these intergenic spacers harbor the ATACTAA motif, which is conserved in all sequenced lepidopteran species and has been suggested as a possible recognition site for the transcription termination peptide (mtTERM protein) (Cameron & Whiting, 2008; Taanman, 1999) (Figure 5).

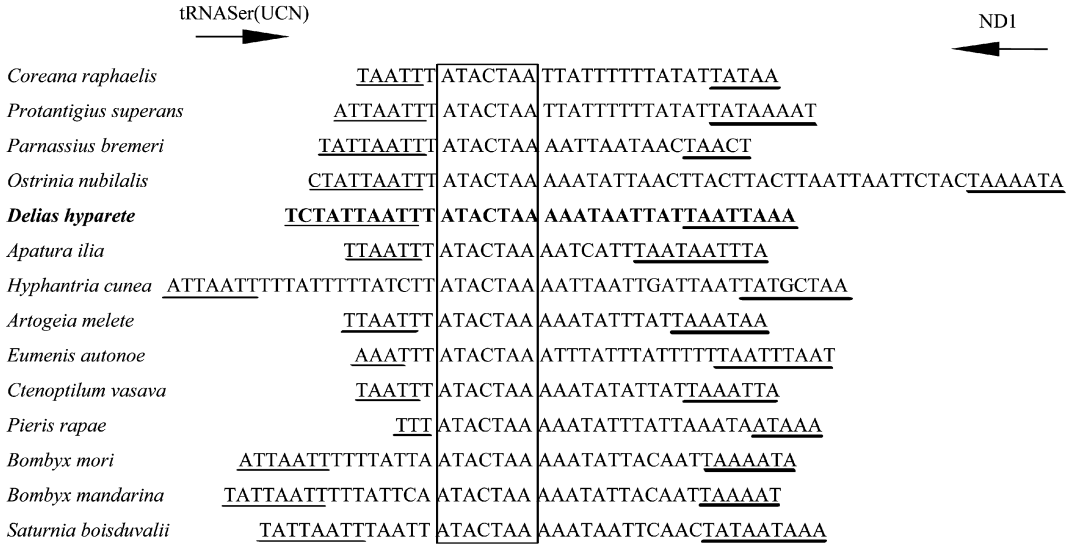

Figure 5 Sequence alignment of the internal spacer regions located between $tRNA^{Ser(UCN)}$ and *ND1*

Boxed nucleotides indicate the conserved heptanucleotide region (ATACTAA) detected in all sequenced lepidopteran insects. Underlined nucleotides indicate the adjacent partial sequences of the $tRNA^{Ser(UCN)}$ gene and *ND1* gene. Arrows indicate the transcriptional direction.

A+T-rich region As in other lepidopteran insects, the A+T-rich region of *D. hyparete* mitogenome includes the origin sites for transcription and replication (Taanman, 1999). Its 377 bp size is well within the range observed in the completely sequenced butterfly species, ranging from 328 bp in *Libythea celtis* (Unpublished, NC_016724) to 1270 bp in *Papilio maraho* (Wu et al, 2010). The A+T content of *D. hyparete* A+T-rich region (92%) is slightlylower than the corresponding regions of *Aporia crataegi* (95.2%) (Park et al, 2012) within the determined Pieridae species, but this value is also within the range from 88.1% in *Ctenoptilum vasava* (Hao et al, 2012) to 96.3% in *Libythea celtis* (Unpublished, NC_016724).

The presence of varying copy numbers of tandem repeated elements in the mitochondrial A+T-rich region has frequently been reported in insects (Zhang et al, 1995; Cameron & Whiting, 2007), including some lepidopteran insects. For example, the *Antheraea pernyi* A+T-rich region harbors a repeat element of 38 bp tandemly repeated 6 times (Liu et al, 2008), the *A. proylei* has 5 repeat elements (Arunkumar et al, 2006; Liu et al, 2008), Japanese *Bombyx mandarina* harbors a tandem triplication of a 126 bp repeat unit, while only one of the repeat elements is found in the A+T-rich region of *B. mori* and Chinese *B. mandarina* (Yukuhiro et al, 2002; Pan et al, 2008). Within Papilionoidea, the nymphalid

Eumenis autonoe is the only species reported to have a tandem repeat sequence, which harbors 10 identical 27 bp long tandem repeats and one 13 bp long incomplete final repeat (Kim et al, 2010). In this study, no large sequence repeats were detected in the *D. hyparete* A+T-rich region, which harbors several features common to the other lepidopteran insects, such as the motif ATAGA followed by 18 bp long poly-T stretches close to srRNA gene, and a microsatellite-like $(AT)_5$ preceded by the ATTTA motif and a $(AT)_7$ element (Fig. 3). The microsatellite-like repeat preceded by the ATTTA motif is common in insect mitogenomes, and has been found in all other lepidopteran species. For example, ATTTA $(TA)_8$ in *Hyphantria cunea* (Liao et al, 2010), ATTTA $(AT)_7$ in *Coreana raphaelis* (Kim et al, 2006), ATTTA $(TA)_9$ in *Pieris rapae* (Mao et al, 2010), ATTTA $(TA)_9$ in *Argyreus hyperbius* (Wang et al, 2011), ATTTA $(TA)_{10}$ in *Apatura ilia* (Chen et al, 2012) and so on. The function of these sequences is still not known and requires more data to come to any conclusive decisions (Kim et al, 2011b). The 10 bp polyA-like structure upstream $tRNA^{Met}$ is absent in other two sequenced Pieridae species *Artogeia melete* (Hong et al, 2009) and *Pieris rapae* (Mao et al, 2010). However, a similar case was found in other butterfly species, for example, the polyA structure of *Acraea issoria* and *Apatura ilia* are also inserted by a single G and T base, respectively (Hu et al, 2010; Chen et al, 2012).

References

Arunkumar KP, Metta M, Nagaraju J. 2006. Molecular phylogeny of silkmoths reveals the origin of domesticated silkmoth, *Bombyx mori* from Chinese *Bombyx mandarina* and paternal inheritance of *Antheraea proylei* mitochondrial DNA. *Mol Phylogenet Evol*, **40**(2): 419-427.

Ballard JWO, Whitlock MC. 2004. The incomplete natural history of mitochondrial. *Mol Ecol*, **13**(4): 729-744.

Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*, **27**(2): 573-580.

Boore JL, Lavrov DV, Brown WM. 1998. Gene translocation links insects and crustaceans. *Nature*, **392**(6677): 667-668.

Boore JL. 1999. Animal mitochondrial genomes. *Nucleic Acids Res*, **27**(8): 1767-1780.

Cameron SL, Whiting MF. 2007. Mitochondrial genomic comparisons of the subterranean termites from the Genus *Reticulitermes* (Insecta: Isoptera: Rhinotermitidae). *Genome*, **50**(2): 188-202.

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**(1-2): 112-123.

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

Cha SY, Yoon HJ, Lee EM, Yoon MH, Hwang JS, Jin BR, Han YS, Kim I. 2007. The complete nucleotide sequence and gene organization of the mitochondrial genome of the bumblebee, *Bombus ignitus* (Hymenoptera: Apidae). *Gene*, **392**(1-2): 206-220.

Chen M, Tian LL, Shi QH, Cao TW, Hao JS. 2012. Complete mitogenome of the Lesser Purple Emperor *Apatura ilia* (Lepidoptera: Nymphalidae: Apaturinae) and comparison with other nymphalid butterflies. *Zool Res*, **33**(2): 191-201.

Chou I. 1998. Classification and Identification of Chinese Butterflies[M]. Zhengzhou: Henan Scientific and Technological Publishing House. (in Chinese)

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

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): 3877-3888.

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**(2): 239-252.

Hao JS, Su CY, Zhu GP, Chen N, Wu DX, Zhang XP. 2007. The molecular morphologies of mitochondrial 16S rDNA of the main butterfly lineages and their phylogenetic significances. *Genet Mol Biol*, **18**(2): 111-123.

Hao JS, Sun QQ, Zhao HB, Sun XY, Gai YH, Yang Q. 2012. The complete mitochondrial genome of *Ctenoptilum vasava* (Lepidoptera: Hesperiidae: Pyrginae) and its phylogenetic implication. *Comp Funct Genomics*, **2012**: 328049, doi: 10.1155/2012/328049.

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser*, **41**: 95-98.

Hong MY, Lee EM, Jo YH, Park HC, Kim SR, Hwang JS, Jin BR, Kang PD, Kim KG, Han YS, Kim I. 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**(1-2): 49-57.

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.

Ji LW, Hao JS, Wang Y, Huang DY, Zhao JL, Zhu CD. 2012. The complete mitochondrial genome of the dragon swallowtail, *Sericinus montela* Gray (Lepidoptera: Papilionidae) and its phylogenetic implication. *Acta Entomol Sin*, **55**(1): 91-100.

Jiang ST, Hong GY, Yu M, Li N, Yang Y, Liu YQ, Wei ZJ. 2009. Characterization of the complete mitochondrial genome of the giant silkworm moth, *Eriogyna pyretorum* (Lepidoptera: Saturniidae). *Int J*

Biol Sci, **5**(4): 351-365.

Junqueira ACM, Lessinger AC, Torres TT, Da Silva FR, Vettore AL, Arruda P, Azeredo-Espin AML. 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 MII, 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 comparison with other lepidopteran insects. *Mol Cells*, **28**(4): 347-363.

Kim MJ, Wan XL, Kim KG, Hwang JS, Kim I. 2010. Complete nucleotide sequence and organization of the mitogenome of endangered *Eumenis autonoe* (Lepidoptera: Nymphalidae). *Afri J Biotech*, **9**(5): 735-754.

Kim MJ, Kang AR, Jeong HC, Kim KG, Kim I. 2011a. 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**(2): 436-445.

Kim MJ, Jeong HC, Kim SR, Kim I. 2011b. Complete mitochondrial genome of the nerippe fritillary butterfly, *Argynnis nerippe* (Lepidoptera: Nymphalidae). *Mitochondrial DNA*, **22**(4): 86-88.

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

Lavrov DV, Brown WM, Boore JL. 2000. A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede Lithobius forficatus. *Proc Natl Acad Sci USA*, **97**(25): 13738-13742.

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 YQ, Li YP, Pan MH, Dai FY, Zhu XW, Lu C, Xiang ZH. 2008. The complete mitochondrial genome of the Chinese oak silkmoth, *Antheraea pernyi* (Lepidoptera: Saturniidae). *Acta Biochim Biophys Sin*, **40**(8): 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**(5): 955-964.

Mao ZH, Hao JS, Zhu GP, Hu J, Si MM, Zhu CD. 2010. Sequencing and analysis of the complete mitochondrial genome of *Pieris rapae* Linnaeus (Lepidoptera: Pieridae). *Acta Entomol Sin*, **53**(11): 1295-1304. (in Chinese)

Pan MH, Yu QY, Xia YL, Dai FY, Liu YQ, Lu C, Zhang Z, Xiang ZH. 2008. Characterization of mitochondrial genome of Chinese wild mulberry silkworm, *Bomyx mandarina* (Lepidoptera: Bombycidae). *Sci Chn*, **51**(8): 693-701.

Park JS, Cho Y, Kim MJ, Nam SH, Kim I. 2012. Description of complete mitochondrial genome of the black-veined white, *Aporia crataegi* (Lepidoptera: Papilionoidea), and comparison to papilionoid species. *J Asia Pac Entomol*, **15**(3): 331-341.

Simmons RB, Weller SJ. 2001. Utility and evolution of cytochrome *b* in insects. *Mol Phylogenet Evol*, **20**(2): 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**(6): 651-701.

Simonsen TJ, Wahlberg N, Brower AVZ, Jong R. 2006. Morphology, molecules and fritillaries: approaching a stable phylogeny for Argynnini (Lepidoptera: Nymphalidae). *Insect Syst Evol*, **37**(4): 405-418.

Stoneking M, Soodyall H. 1996. Human evolution and the mitochondrial genome. *Curr Opin Genet Dev*, **6**(6): 731-736.

Taanman JW. 1999. The mitochondrial genome: structure, transcription, translation and replication. *Biochim Biophys Acta*, **1410**(2): 103-123.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Bio Evol*, **28**(10): 2731-2739.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The Clustal X windows interface: flexible strategies for multiple sequences alignment aided by quality analysis tools. *Nucleic Acids Res*, **25**(24): 4876-4882.

Tian LL, Sun XY, Chen M, Gai YH, Hao JS, Yang Q. 2012. Complete mitochondrial genome of the Five-dot Sergeant *Parathyma sulpitia* (Nymphalidae: Limenitidinae) and its phylogenetic implications. *Zool Res*, **33**(2): 133-143.

Vigilan L, Stoneking M, Harpending H, Hawkes K, Wilson A. 1991. African populations and the evolution of human mitochondrial. *Science*, **253**(5027): 1503-1507.

Wang XC, Sun XY, Sun QQ, Zhang DX, Hu J, Yang Q, Hao JS. 2011. Complete mitochondrial genome of the laced fritillary *Argyreus hyperbius* (Lepidoptera: Nymphalidae). *Zool Res*, **32**(5): 465-475.

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

Wu LW, Lees DC, Yen SH, Lu CC, Hsu YF. 2010. The complete mitochondrial genome of the near-threatened swallowtail, *Agehana maraho* (Lepidoptera: Papilionidae): evaluating sequence variability and suitable markers for conservation genetic studies. *Etomol News*, **121**(3): 267-280.

Xia J, Hu J, Zhu GP, Zhu CD, Hao JS. 2011. Sequence and analysis of the complete mitochondrial genome of *Calinaga davidis* Oberthür (Lepidoptera: Nymphalidae). *Acta Entomol Sin*, **54**(5): 555-565. (in Chinese)

Yokobori S, Pääbo S. 1995. Transfer RNA editing in land snail mitochondrial. *Proc Natl Acad Sci USA*, **92**(22): 10432-10435.

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 domes-ticated silkmoth, *Bombyx mori*. *Mol Biol Evol*, **19**(8): 1385-1389.

Zhang DX, Szymura JM, Hewitt GM. 1995. Evolution and structural conservation of the control region of insect mitochondrial DNA. *J Mol Evol*, **40**(4): 382-391.