Mol. Cells *28*, 347-363, October 31, 2009
DOI/10.1007/s10059-009-0129-5

Molecules and Cells

Complete Nucleotide Sequence and Organization of the Mitogenome of the Red-Spotted Apollo Butterfly, *Parnassius bremeri* (Lepidoptera: Papilionidae) and Comparison with Other Lepidopteran Insects

Man Il Kim, Jee Yeon Baek, Min Jee Kim, Heon Cheon Jeong[1], Ki-Gyoung Kim[2], Chang Hwan Bae[2], Yeon Soo Han, Byung Rae Jin[3], and Iksoo Kim*

The 15,389-bp long complete mitogenome of the endangered red-spotted apollo butterfly, *Parnassius bremeri* (Lepidoptera: Papilionidae) was determined in this study. The start codon for the COI gene in insects has been extensively discussed, and has long remained a matter of some controversy. Herein, we propose that the CGA (arginine) sequence functions as the start codon for the COI gene in lepidopteran insects, on the basis of complete mitogenome sequences of lepidopteran insects, including *P. bremeri,* as well as additional sequences of the COI start region from a diverse taxonomic range of lepidopteran species (a total of 53 species from 15 families). In our extensive search for a tRNA-like structure in the A+T-rich region, one $tRNA^{Trp}$-like sequence and one $tRNA^{Leu}$(UUR)-like sequence were detected in the *P. bremeri* A+T-rich region, and one or more tRNA-like structures were detected in the A+T-rich region of the majority of other sequenced lepidopteran insects, thereby indicating that such features occur frequently in the lepidopteran mitogenomes. Phylogenetic analysis using the concatenated 13 amino acid sequences and nucleotide sequences of PCGs of the four macrolepidopteran superfamilies together with the Tortricoidea and Pyraloidea resulted in the successful recovery of a monophyly of Papilionoidea and a monophyly of Bombycoidea. However, the Geometroidea were unexpectedly identified as a sister group of the Bombycoidea, rather than the Papilionoidea.

INTRODUCTION

The red-spotted apollo butterfly, *Parnassius bremeri*, is a member of the snow apollo genus of the swallowtail family, Papilionidae, within the insect order Lepidoptera. This species is distributed throughout Russia, Korea and China, and is detected in open landscapes on forest steppes, as well as slopes with woodlands reaching up to the alpine zone (Kim, 2005). In Korea, adults of the species are seen once per year between May and June in host plants including *Sedum kamtschaticum* and *Cirsium japonicum* (Ko et al., 2004). This species has historically been detected in a broad swath of the Korean peninsula, with the exception of some lowlands along the coastline of the Yellow Sea. However, the species has been observed only very rarely in recent years (Ko et al., 2004) and is thus listed as a second-degree endangered wild animal in Korea (Kim, 2005). Although the reasons for the disappearance of the species have been speculated to be habitat diminishment (including host plant) and/or global warming, no proper investigations have yet been conducted in this regard.

Animal mitogenomes are usually circular, approximately 16-20 kb in size, and composed of 13 PCGs, 22 tRNA genes, two rRNA genes (lrRNA and srRNA), and one large non-coding, region, referred to in insects as the A+T-rich region (Wolstenholme 1992). The A+T-rich region has been determined to harbor the origin of heavy-strand mtDNA replication in vertebrates (Tapper and Clayton, 1984), the replication origin for both mtDNA strands in *Drosophila* species (Brehm et al., 2001; Clary and Wolstenholme, 1987; Fauron and Wolstenholme, 1980; Saito et al., 2005), and the replication origin for minor-strand mtDNA in lepidopteran *Bombyx mori*, coleopteran *Tribolium castaneum*, and orthopteran *Locusta migratoria* (Saito et al., 2005).

Thus far, mitogenome sequences from more than 120 species have been determined from a variety of insects. Among them, the mitogenomes of 14 species belonging to six superfamilies in the lepidopteran insects have been entirely or nearly entirely sequenced. Considering the diversity of the lepidopterans, which contain approximately 160,000 species, the information currently

College of Agriculture and Life Sciences, Chonnam National University, Gwangju 500-757, Korea, [1]Insect Research Institute of Hampyeong, Chonnam 525-811, Korea, [2]Biological Resources Research Department, National Institute of Biological Resources, Incheon 404-708, Korea, [3]College of Natural Resources and Life Science, Dong-A University, Busan 604-714, Korea
*Correspondence: ikkim81@chonnam.ac.kr

Received May 6, 2009; accepted August 4, 2009; published online October 9, 2009

Keywords: complete mitogenome, lepidopteran phylogeny, macrolepidoptera, mitochondrial DNA, *Parnassius bremeri*, red-spotted apollo butterfly, structural element, tRNA-like structure

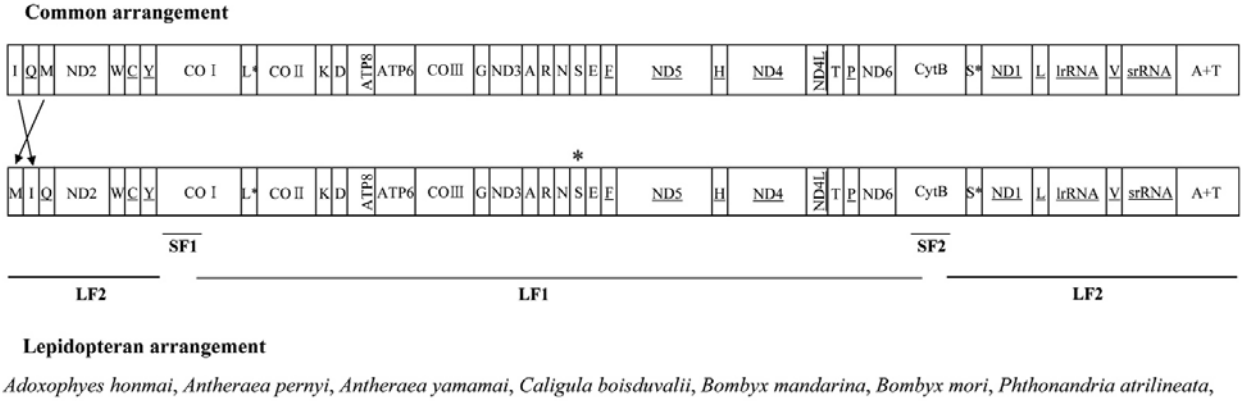


Fig. 1. Linear arrangement of the mitogenomes of lepidopteran insects including *Parnassius bremeri* and the most common type found in insects. Translocated tRNAs from the common arrangement are indicated by arrows. tRNAs are denoted as one-letter symbols in accordance with the IUPAC-IUB single letter amino acid codes. The one-letter symbols L, L*, S, and S* denote codon $tRNA^{Leu}$(CUN), $tRNA^{Leu}$(UUR), $tRNA^{Ser}$(AGN), and $tRNA^{Ser}$(UCN), respectively. Gene names that are not underlined indicate a clockwise direction of transcription, whereas the underline indicates a counter-clockwise transcriptional direction. The lines under the linear map indicate two short (SF1 and SF2) and two long overlapping fragments (LF1 and LF2) amplified for the sequencing of the whole *Parnassius bremeri* mitogenome.

existing on lepidopteran mitogenomes remains quite limited. Particularly, considering that butterfly species are vulnerable to local extinction as the result of global warming (Wettstein and Schmid, 1999), an accumulation of minimal genetic information from the endangered butterfly species is considered crucial for the long-term objective of conservation.

In this study, we have determined the complete mitogenome sequence of *P. bremeri*, and its sequence was compared to other insect mitogenomes, particularly those of lepidopteran species, with regard to arrangement, composition, and structural elements in the A+T-rich region. Furthermore, the concatenated amino acid and nucleotide sequences of 13 PCGs were employed for the reconstruction of phylogenetic relationships among six lepidopteran superfamilies. These included four macrolepidopteans (Papilionoidea, Bombycoidea, Geometroidea, and Noctuoidea) and two microlepidopterans belonging to the Obtectomera (Tortricoidea and Pyraloidea).

MATERIALS AND METHODS

DNA extraction

An adult specimen of *P. bremeri* was collected from Mt. Jiri, Korea in May, 2008. DNA was extracted with a Wizard™ Genomic DNA Purification Kit in accordance with the manufacturer's instructions (Promega, USA).

Primer design, PCR, and sequencing

In an effort to sequence the whole mitogenome of *P. bremeri*, 500-700 bp of *P. bremeri* COI and CytB (SF1 and SF2 in Fig. 1, respectively) were initially sequenced. The primers for the COI fragment (SF1) were adapted from those reported by Folmer et al. (1994), whereas those for the CytB fragment (SF2) were designed via the alignment of several insect mitogenomes sequenced in their entirety. The primer sequences were as follows: LCO1490, 5′-GGTCAACAAATCATAAAGATATTGG-3′ and HCO2198, 5′-TAAACTTCAGGGTGACCAAAAAATCA-3′ for SF1 and LEP-CytBF, 5′-TTCATATTGGACGAGGAATTTA-3′ and LEP-CytBR, 5′-TAAAAGCAATAACTCCTCCTAA-3′ for SF2. These sequences correspond to nucleotide positions 1460-1484 for LCO1490 and 2143-2168 for HCO2198, with respect to the mitogenome of *P. bremeri*, respectively. These short fragments were amplified with AccuPower® PCR PreMix (Bioneer, Korea) under the following conditions: initial denaturation for 7 min at 94°C, followed by 35 cycles of 60 s at 94°C, 60 s at 54°C, and 2 min at 72°C, with a subsequent final 7-min extension at 72°C. Based on the sequence information, two pairs of primers were designed to amplify two overlapping long fragments (LF1 and LF2 in Fig. 1) using LA Taq™ (Takara Biomedical, Japan) under the following conditions: initial dena-turation for 1 min at 94°C, followed by 30 cycles of 10 s at 94°C and 15 min at 60-61°C and a subsequent final 10-min extension at 72°C. The primer sequences for LF1 and LF2 were as follows: PB-COIF1, 5′-GTCGAAAATGGAGCAGGAACTG-3 and PB-CytBR, 5′-GATT-AGTAATAACTGTAGCTCC-3′ for LF1 and PB-CytBF, 5′-TTAT-TCCAGCTAATCCTTTAGTAAC-3′ and PB-COIR, 5′-GCTATAT-CTGGAGCTCCTAG-3′ for LF2. These se-quences corresponded to nucleotide positions 1789-1810 and 10897-10918 for PB-COIF1 and PB-CytBR in LF1, respectively, and 11240-11264 and 1196-1215 for PB-CytBF and PB-COIR in LF2, respectively, with respect to the *P. bremeri* mitogenome. These PCR products were then employed in the construction of a shotgun library. In brief, the DNAs were sheared into 1-5 kb fragments using Hydroshear (Gene Machine, USA) and the DNA fraction was collected using a Chromaspin TE 1000 column. The DNA fraction was then cloned into the pUC118 vector (Takara Biomedical, Japan), and each of the resultant plasmid DNAs was isolated using a Wizard *Plus* SV Minipreps DNA Purification System (Promega, USA). DNA sequencing was conducted using the ABI PRISM® BigDye® Terminator v3.1 Cycle Sequencing Kit and the ABI PRISM™ 3100 Genetic Analyzer (PE Applied Biosystems, USA). All fragments were sequenced from both strands.

In an effort to verify the abnormal translational starts of the COI sequence in lepidopteran insects, the start region of the COI gene and the neighboring $tRNA^{Tyr}$ gene were additionally sequenced from 39 species of lepidopteran insects belonging to eight lepidopteran families. The primers for the amplification of the region were designed via the alignment of several lepidopteran mitogenomes sequenced in their entirety. The primer

sequences were as follows: Lep-nd2F1, 5′-CCMTTTTTAGG-ATTTTTTCCTAAATGA-3′ and Lep-co1R2, 5′-GAAAGCTAT-ATCAGGGGCTCC-3′. PCR was conducted using Accu Power® PCR PreMix (Bioneer, Korea) under the following conditions: initial denaturation for 7 min at 94°C, followed by 35 cycles of 60 s at 94°C, 60 s at 50-60°C, and 2 min at 72°C, with a subsequent final 7-min extension at 72°C.

Sequence analysis

Sequences from overlapping fragments were assembled via the alignment of neighboring fragments using CLUSTAL X software (Thompson et al., 1997). Individual *P. bremeri* mt genes and the A+T-rich region were determined via the alignment of the sequences with homologous regions of known full-length insect mt sequences using CLUSTAL X software (Thompson et al., 1997). The nucleotide sequences of the PCGs were translated on the basis of the invertebrate mtDNA genetic code. The secondary structures of most of the tRNA genes were predicted with tRNAscan-SE 1.21 (Lowe and Eddy, 1997) using invertebrate codon predictors and a cove score cut-off of 1, but some [e.g., $tRNA^{Ser(AGN)}$] were drawn by hand, on the basis of the nucleotide sequences of the tRNA genes of other insects sequenced in their entirety and visually edited, with careful consideration given to the anticodon sequences. The sequence data were deposited into the GenBank database under accession no. FJ871125.

Phylogenetic analysis

In order to gain insight into the phylogenetic relationships existing among the apoditrysian superfamilies in Lepidoptera, we utilized each of the amino acid sequences and nucleotide sequences of the concatenated 13 PCGs, including those of *P. bremeri* belonging to the Papilionoidea superfamily.

The alignment of the amino acids and sequences of the individual 13 PCGs were conducted using CLUSTAL W software with the default gap opening/gap extension scheme (Thompson et al., 1994) within BioEdit (Hall, 1999). On the other hand, the alignment of the nucleotide sequences of 13 individual PCGs was constructed using RevTrans ver. 1.4, which aligns the coding sequences on the basis of the protein alignment (Wernersson and Pedersen, 2003). Then, the well-aligned blocks from the amino acid sequences and nucleotide sequences were selected with GBlocks 0.91b (Castresana, 2000), with the maximum number of contiguous non-conserved positions set to four. These were subsequently concatenated into an amino acid (3,498 sites, which is 91% of original sites) and a nucleotide (10,779 sites, which is 93% of original sites) sequence alignment. This alignment is available upon request.

The degree of compositional heterogeneity of each sequence was determined using TREE-PUZZLE ver. 5.2 (Schmidt et al., 2002). This analysis revealed statistically significant compositional deviations in both the amino acid and nucleotide sequences of *Coreana raphaelis* (superfamily Papilionoidea) and in the nucleotide sequences of *Antheraea pernyi* (superfamily Bombycoidea), *Manduca sexta* (superfamily Bombycoidea), and *Ochrogaster lunifer* (superfamily Noctuoidea) in the 5% chi-square test. For simplicity's sake, these sequences were excluded from all subsequent phylogenetic analyses, leaving ten species: *Bombyx mori* and *B. mandarina* belonging to the family Bombycidae (superfamily Bombycoidea), *A. yamamai* and *Caligula boisduvalii* belonging to the family Saturniidae (superfamily Bombycoidea), *Artogeia melete* belonging to the Pieridae (superfamily Papilionoidea), *P. bremeri* belonging to the Papilionidae (superfamily Papilionoidea), *Phthonandria atrilineata* belonging to the Geometridae (superfamily Geometroidea), *Adoxophyes honmai* belonging to the Tortricidae (superfamily Torraicoidea), and *Ostrinia furnacalis* and *O. nubilalis* belonging to the family Crambidae (superfamily Pyraloidea).

Substitution model selection was conducted via a comparison of Akaike Information Criterion (AIC) scores (Akaike, 1974), calculated using the programs ProTest ver. 1.4 (Abascal et al., 2005) for amino acid sequence alignment and Modeltest ver. 3. 7 (Posada and Crandall, 1998) for nucleotide sequence alignment. The mtRev-24 (Adachi and Hasegawa, 1996) + I + G + F model was selected as a model for Bayesian inference (BI) and Maximum likelihood (ML) analyses for amino acid sequences in the absence of the recently developed mtArt model (Abascal et al., 2007) in the MrBayes package. On the other hand, the GTR (Lanave et al., 1984) + I + G was selected as the best-fitting model for nucleotide sequences for BI and ML analyses. The BI analyses for both amino acid and nucleotide sequences were conducted using MrBayes ver. 3.1 (Huelsenbeck and Ronquist, 2001) under the following conditions: 1,000,000 generations, four chains (one hot chain and three cold chains), and a burn-in step of the first 10,000. The confidence values of the BI tree are presented as the Bayesian posterior probabilities in percent (BPP). The ML analyses for both amino acid and nucleotide sequences were conducted using PHYML (Guindon et al., 2005) under the following conditions: the proportion of invariable sites as "estimated", number of substitution rate categories as six, gamma distribution parameter as "estimated", and starting tree as a BIONJ distance-based tree. The confidence values of the ML tree were evaluated via the bootstrap test with 100 iterations. The dipteran, *Drosophilla yakuba* (Clary and Wolstenholme, 1985), *Anopheles gambiae* (Beard et al., 1993), and *Bactrocera oleae* (Nardi et al., 2003) were employed as a co-outgroup.

RESULTS AND DISCUSSION

Genome structure, organization, and composition

The *P. bremeri* mitogenome contains a typical set of genes, such as 13 PCGs, 22 tRNA genes, 2 rRNA genes, and one non-coding A+T-rich region (Table 1). The genome size of 15,389 bp is well within the range detected in the completely sequenced lepidopteran insects, with sizes ranging from 15,140 in *Artogeia melete* to 15,928 in *B. mandarina* (Table 2). *P. bremeri* harbors 3,734 codons, excluding termination codons, and this number is identical to that of *C. boisduvalii* (Table 2). All individual *P. bremeri* mt genes were well within the range detected in the respective genes of other lepidopteran insects (Supplementary Table 1).

The orientation and gene order of the lepidopteran mitogenomes are identical, including those of *P. bremeri* (Fig. 1), but differ from the most common type detected in a variety of insect orders. The difference between the two involves the movement of $tRNA^{Met}$ to a position 5′-upstream of $tRNA^{Ile}$, which results in the following order of $tRNA^{Met}$, $tRNA^{Ile}$, and $tRNA^{Gln}$, rather than the common-type order $tRNA^{Ile}$, $tRNA^{Gln}$, and $tRNA^{Met}$. This most common type was inferred to be ancestral for insects (Boore et al., 1998).

As is the case in other insect mitogenome sequences, the nucleotide composition of the *P. bremeri* mitogenome is also biased toward A+T content, as 81.3% (Table 2). The analysis of the base composition at each codon position of the concatenated 13 PCGs of *P. bremeri* demonstrated that the third codon position (94.8%) harbored an A+T content higher than that of the first (74.7%) and second (70.6%) codon positions, as has also been noted with other sequenced lepdopteran species (Table 3). The genome-wise A+T bias is also reflected in the codon usage of the *P. bremeri* mitogenome (Supplemantary Table 2). The relative

Table 1. Summary of *Parnassius bremeri* mitogenome

Gene	Direction	Nucleotide number	Size	Anticodon	Start codon	Stop codon
$tRNA^{Met}$	F	1-69	69	33-35 CAT	-	-
$tRNA^{Ile}$	F	70-133	64	99-101 GAT	-	-
$tRNA^{Gln}$	R	131-199	69	167-169 TTG	-	-
ND2	F	240-1253	1014	-	ATT	TAA
$tRNA^{Trp}$	F	1253-1318	66	1283-1285 TCA	-	-
$tRNA^{Cys}$	R	1311-1376	66	1345-1347 GCA	-	-
$tRNA^{Tyr}$	R	1381-1444	64	1412-1414 GTA	-	-
COI	F	1447-2982	1536	-	CGA	TAA
$tRNA^{Leu}$(UUR)	F	2978-3044	67	3008-3010 TAA	-	-
COII	F	3045-3726	682	-	ATG	T-tRNA
$tRNA^{Lys}$	F	3727-3797	71	3757-3759 CTT	-	-
$tRNA^{Asp}$	F	3797-3863	67	3828-3830 GTC	-	-
ATP8	F	3864-4028	165	-	ATA	TAA
ATP6	F	4022-4699	678	-	ATG	TAA
COIII	F	4699-5487	789	-	ATG	TAA
$tRNA^{Gly}$	F	5491-5557	67	5521-5523 TCC	-	-
ND3	F	5555-5911	357	-	ATA	TAA
$tRNA^{Ala}$	F	5911-5976	66	5941-5943 TGC	-	-
$tRNA^{Arg}$	F	5976-6041	66	6005-6007 TCG	-	-
$tRNA^{Asn}$	F	6042-6108	67	6072-6074 GTT	-	-
$tRNA^{Ser}$(AGN)	F	6112-6172	61	6133-6135 GCT	-	-
$tRNA^{Glu}$	F	6216-6281	66	6247-6249 TTC	-	-
$tRNA^{Phe}$	R	6280-6349	70	6311-6313 GAA	-	-
ND5	R	6351-8087	1737	-	ATT	TAA
$tRNA^{His}$	R	8085-8148	64	8116-8118 GTG	-	-
ND4	R	8148-9488	1341	-	ATG	TAA
ND4L	R	9488-9778	291	-	ATG	TAA
$tRNA^{Thr}$	F	9781-9846	66	9812-9814 TGT	-	-
$tRNA^{Pro}$	R	9847-9911	65	9879-9881 TGG	-	-
ND6	F	9914-10444	531	-	ATC	TAA
CytB	F	10462-11610	1149	-	ATA	TAA
$tRNA^{Ser}$(UCN)	F	11612-11678	67	11641-11643 TGA	-	-
ND1	R	11695-12633	939	-	ATG	TAG
$tRNA^{Leu}$(CUN)	R	12635-12703	69	12671-12673 TAG	-	-
lrRNA	R	12704-14047	1344	-	-	-
$tRNA^{Val}$	R	14048-14112	65	14080-14082 TAC	-	-
srRNA	R	14113-14885	773	-	-	-
A+T-rich region	R	14886-15389	504	-	-	-

*tRNA abbreviations follow the IUPAC-IUB three letter code.

synonymous codon frequencies (RSCU) revealed that codons harboring A or T in the third position were always overused as compared to other synonymous codons in the *P. bremeri* mitogenome. For example, the codon TTC (Phe) was utilized only 28 times for phenylalanine, corresponding to an RSCU of 0.15, but the synonymous codon TTT for phenylalanine was profoundly overused, corresponding to an RSCU of 9.13 (Supplemantary Table 2). This trend has also been noted in other sequenced lepidopteran insects (Supplemantary Table 2). Furthermore, the codons TTT (Phe), TTA (Leu), ATT (Ile), and ATA (Ile) are the four most frequently utilized codons in the *P. bremeri* PCGs, accounting for 38.72% (Supplemantary Table 2). In other sequenced lepidopteran insects, the values range from 35.09% in *Ochrogaster lunifer* (Paola et al., 2008) to 54.89% in *Coreana raphaelis* (Kim et al., 2006) (Supplemantary Table 2). These codons are all composed of A or T nucleotides, thus indicating the biased usage of A and T nucleotides in the PCGs of lepidopteran insects.

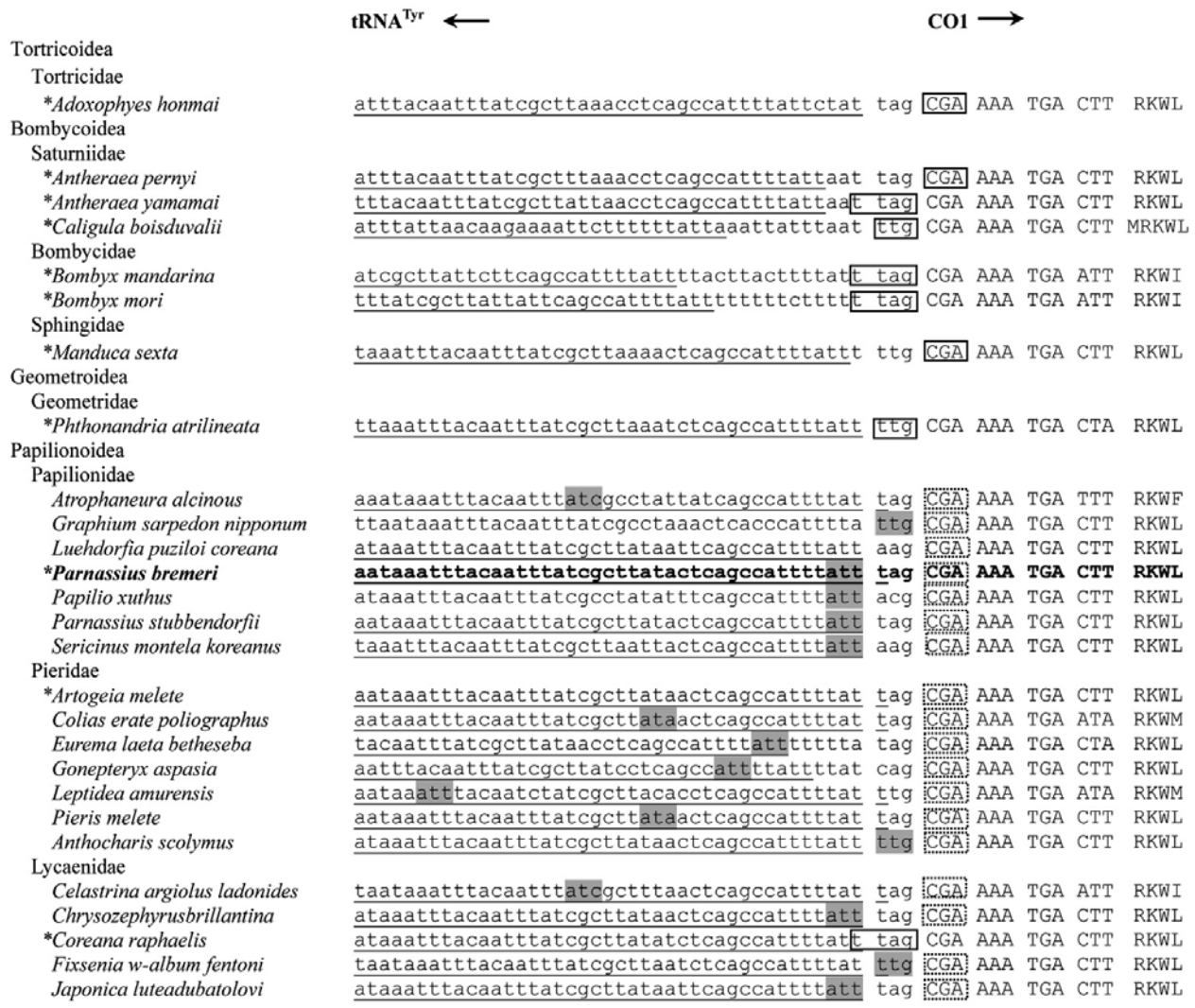
PCGs

All protein-coding sequences with the exception of the COI gene began with the typical ATN codons (isoleucine) in the *P. bremeri*

Table 2. Characteristics of the lepidopteran mitogenomes

Taxon	Size (bp)	A+T content (%)	No. codons[a]	PCG[b] A+T content (%)	lrRNA		srRNA		A+T-rich region		GenBank accession no.	References
					Size (bp)	A+T %	Size (bp)	A+T %	Size (bp)	A+T %		
Tortricoidea												
Tortricidae												
Adoxophyes honmai	15,680	80.4	3,748	78.5	1,387	83.6	779	85.4	489	94.3	DQ073916	Lee et al. (2006)
Bombycoidea												
Saturniidae												
Antheraea pernyi	15,575	80.2	3,732	78.5	1,369	83.9	775	84.1	552	90.4	AY242996	Liu et al. (2008)
Antheraea yamamai	15,338	80.2	3,729	79.3	1,380	83.5	776	85.9	334	90.4	EU726630	Kim et al. (2009)
Caligula boisduvalii	15,360	80.6	3,734	79.1	1,391	84.8	774	84.1	330	91.5	EF622227	Hong et al. (2008)
Bombycidae												
Bombyx mandarina	15,928	81.7	3,722	79.6	1,377	84.7	783	86.0	747	95.2	AB070263	Yukuhiro et al. (2002)
Bombyx mori	15,656	81.4	3,720	79.5	1,378	84.4	783	85.6	494	95.5	AB070264	Yukuhiro et al. (2002)
Sphingidae												
Manduca sexta	15,516	81.8	3,718	80.2	1,391	84.7	777	86.8	324	95.1	EU286785	Cameron and Whiting (2008)
Geometroidea												
Geometridae												
Phthonandria atrilineata	15,499	81.1	3,724	79.0	1,400	85.1	803	87.5	457	98.5	NC010522	Yang et al. (2009)
Papilionoidea												
Papilionidae												
Parnassius bremeri	**15,389**	**81.3**	**3,734**	**80.2**	**1,344**	**83.8**	**773**	**85.1**	**504**	**93.6**	**FJ871125**	**This study**
Pieridae												
Artogeia melete	15,140	79.8	3,715	78.4	1,319	83.4	777	86.9	351	88.0	NC010568	Hong et al. (2009a)
Lycaenidae												
Coreana raphaelis	15,314	82.7	3,708	81.5	1,330	85.3	777	85.8	375	94.1	DQ102703	Kim et al. (2006)
Noctuoidea												
Notodontidae												
Ochrogaster lunifer	15,593	77.9	3,746	75.7	1,351	81.4	806	84.3	319	92.7	AM946601	Salvato et al. (2008)
Pyraloidea												
Crambidae												
Ostrinia furnacalis	14,536	80.4	3,719	79.4	1.339	85.0	435	82.8	-	-	AF467260	Coates et al. (2005)
Ostrinia nubilalis	14,535	80.2	3,718	79.1	1,339	84.9	434	82.0	-	-	AF442957	Coates et al. (2005)

[a]Termination codons were excluded in total codon count
[b]Protein coding genes
Bar (-) indicates lack of sequence information on the A+T rich region in the genome



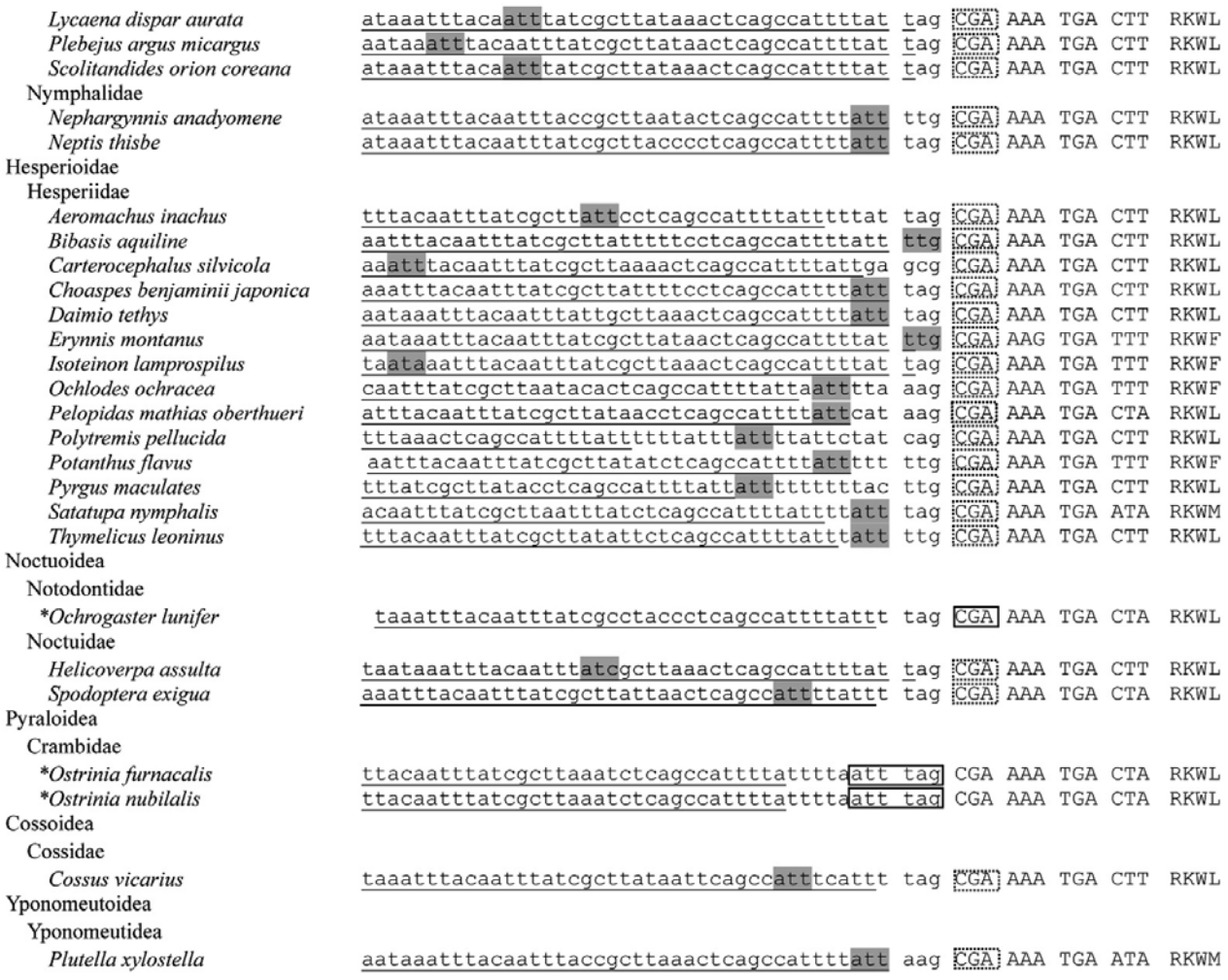
(continued)

Fig. 2. Alignment of the initiation site for the COI genes of 47 lepidopteran species belonging to 15 families. Asterisks on the species names indicate the 14 lepidopteran insects, sequenced in their entire or near entire mitogenomes, including that of *Parnassius bremeri*, whereas no asterisks indicate 33 newly sequenced species in their COI regions. The first four or five codons and their amino acids are provided on the right-hand side of the figure. The boxed nucleotides indicate previously known translation initiators for the lepidopteran COI gene. Dotted boxes indicate the newly-proposed CGA codon found in all sequenced lepidopteran insects. Shaded boxes indicate the in-frame typical start codon that is detected within $tRNA^{Tyr}$. Underlined nucleotides indicate the adjacent partial sequence of $tRNA^{Tyr}$. Arrows indicate the transcriptional direction.

mitogenome (six with ATG, three with ATA, two with ATT, and one with ATC; Table 1). In the COI gene, however, no canonical ATN initiator for mt PCGs was detected in the start region of the COI gene, nor in the immediately neighboring sequence of the precedent $tRNA^{Tyr}$ (Fig. 2). The only plausible traditional start codon for COI gene is ATA, located 24 bp upstream of the COI gene within the $tRNA^{Tyr}$ gene. This ATA sequence requires seven-eight additional amino acids, resulting in a peculiar alignment as compared with other lepidopteran insects. Furthermore, a TAG-stop codon is present at the beginning region of the COI gene. Thus, this ATA sequence cannot be the start codon for the COI gene. With the exception of this triplet, no canonical start codon for the *P. bremeri* COI gene is currently available. Similarly, the majority of sequenced lepidopteran insects harbor an in-frame ATT sequence within the $tRNA^{Tyr}$ gene, but also harbor a TAG-stop codon at the beginning region of the COI gene (Fig. 2).

The uncertainty of the start codon for the COI gene is also manifest in several other arthropod species (Wilson et al., 2000; Woo et al., 2007) including the Lepidoptera. Currently, the start codon for the COI gene in the lepidopteran insects of the completely (or nearly completely) sequenced mitogenomes was theorized to be composed of TAG (*A. melete*), TTAG (*C. raphaelis*, *B. mandarina*, *B. mori*, and *A. yamamai*), TTG (*P. atrilineata* and *C. boisduvalii*), CGA (*A. pernyi*, *A. honmai*, *O. lunifer*, and *M. sexta*), or ATTTAG (*O. furnacalis* and *O. nubilalis*) (Fig. 2). Among these options, TTG is the only canonical start codon that is acceptable in terms of the current invertebrate mt code, but is available from the limited number of lepidopteran species.

In the current situation, in which no mRNA expression data for *P. bremeri* are available, this would be one strategy for the



selection of a start codon that is universally present in the lepidopteran insects with the consideration of a probable logic that would minimize intergenic space and gene overlaps in the evolutionary economic perspective. According to these criteria, the first non-overlapping codon in the COI gene is the CGA designating arginine (Fig. 2). This codon exists in a region that is highly conserved throughout all sequenced lepidopteran insects. Furthermore, our additional sequences of 39 lepidopteran insect species, encompassing eight families, demonstrate that the CGA is well conserved in lepidopteran insects (Fig. 2), thus suggesting that the sequence may be functionally constrained. This allowed us to hypothesize that the CGA (arginine) may be a synapomorphic characteristic of the Lepidoptera.

The 12 of the 13 PCGs harbor the complete termination codon TAA or TAG, but the COII gene harbors the incomplete termination codon T (Table 1). The partial termination codons (i.e., T or TA) are observed in all sequenced lepidopteran insects (Supplementary Table 3) and this phenomenon has been interpreted in terms of post-transcriptional polyadenylation, by which "A" residue(s) are added to create TAA termini (Anderson et al., 1981; Ojala et al., 1981).

Intergenic spacer sequences

The *P. bremeri* mt genes are interleaved with a total of 138 bp of intergenic spacer sequences, which are spread over 14 regions, ranging in size from 1 to 45 bp, with the relatively longer ones located between $tRNA^{Gln}$ and ND2 (40 bp), between $tRNA^{Ser}$(AGN) and $tRNA^{Glu}$ (43 bp), between ND6 and CytB (17 bp), and between $tRNA^{Ser}$(UCN) and ND1 (16 bp) (Table 4). The first relatively long intergenic spacer sequences located between $tRNA^{Gln}$ and ND2 are always detected in the sequenced lepidopteran insects with somewhat larger size (47-72 bp). The sequence alignment of this intergenic spacer sequence from the *P. bremeri* mitogenome to the neighboring ND2 gene revealed a sequence homology of 70% (Fig. 3A), and this may suggest that this spacer sequence originated from a partial duplication of the ND2 gene. In a similar fashion, the intergenic spacer sequences of *A. melete* (70%), *C. raphaelis* (62%), *C. boisduvalii* (62%), *B. mori* (63%), and *P. atrilineata* (70%) also evidence sequence homology substantially higher to their respective ND2 gene (Fig. 3A). On the other hand, the remaining sequenced lepidopteran species evidenced relatively lower sequence homology with the neighboring ND2 (< 40%). These species may very well have undergone further rapid sequence divergence due to the non-coding nature of the intergenic spacer sequence. As more sequence information from lepidopteran species is accumulated, further informative conclusions may be drawn.

The second large intergenic spacer sequence located between $tRNA^{Ser}$ (AGN) and $tRNA^{Glu}$ (43 bp) of the mitogenome of *P.*

Table 3. Summary of base composition at each codon position of the concatenated 13 PCGs in the lepidopteran mitogenomes

Species	1st codon position				2nd codon position				3rd codon position				Overall			
	A	T	C	G	A	T	C	G	A	T	C	G	A	T	C	G
Adoxophyes honmai	36.4	36.6	10.5	16.5	22.4	48.0	16.3	13.4	41.2	50.7	4.8	3.3	33.3	45.1	10.5	11.1
Antheraea pernyi	35.4	37.5	10.5	16.6	21.7	48.5	16.6	13.3	40.9	51.4	4.9	2.8	32.7	45.8	10.6	10.9
Antheraea yamamai	35.6	37.2	10.7	16.5	21.6	48.7	16.5	13.2	41.4	51.9	4.1	2.6	32.9	45.9	10.4	10.8
Caligula boisduvalii	35.9	37.9	10.2	16.0	22.1	48.7	16.1	13.2	41.0	51.7	4.2	3.1	33.0	46.1	10.1	10.8
Bombyx mandarina	37.5	37.5	9.3	15.7	22.3	48.7	15.9	13.1	43.4	49.4	4.3	2.9	34.4	45.2	9.9	10.6
Bombyx mori	37.2	37.5	9.5	15.7	22.2	48.7	15.9	13.2	43.7	49.2	4.3	2.7	34.4	45.2	9.9	10.5
Manduca sexta	37.1	37.7	9.5	15.7	22.2	48.8	16.0	13.1	43.7	51.3	2.6	2.5	34.3	45.9	9.3	10.4
Phthonandria atrilineata	36.8	37.5	9.4	16.3	22.4	48.2	16.1	13.4	42.8	49.2	4.8	3.2	34.0	45.0	10.1	10.9
Parnassius bremeri	**36.9**	**37.8**	**9.5**	**15.8**	**22.0**	**48.6**	**16.4**	**13.1**	**44.0**	**50.8**	**3.3**	**1.8**	**34.3**	**45.7**	**9.7**	**10.2**
Artogeia melete	36.9	37.8	10.4	16.2	21.8	48.0	16.8	13.4	42.2	49.8	4.5	3.5	33.6	44.8	10.6	11.0
Coreana raphaelis	38.5	37.9	8.8	14.8	22.0	49.2	15.7	13.1	45.7	51.1	2.0	1.2	35.4	46.1	8.8	9.7
Ochrogaster lunifer	36.3	35.6	11.3	16.8	21.7	48.1	16.5	13.6	38.8	46.2	9.0	5.9	32.3	43.3	12.3	12.1
Ostrina furnacalis	37.7	36.8	9.5	16.1	21.6	48.7	16.3	13.3	43.7	49.6	4.0	2.7	34.3	45.0	10.0	10.7
Ostrina nubilalis	35.8	38.3	10.8	15.1	26.0	48.6	14.2	11.2	41.0	47.6	5.1	6.3	34.3	44.8	10.0	10.9

Stop codon was excluded in the count.

bremeri is unique in that other sequenced lepidopteran insects harbor very short spacer sequences within this region (1-10 bp) (Table 4). The sequence of *P. bremeri* appears to be the result of the duplication of the TTTCCTTTT segment or the triplication of the TTTCC segment within the 43 bp spacer sequence (Fig. 3B). The third intergenic spacer sequence located between ND6 and CytB (17 bp) of the *P. bremeri* mitogenome is highly variable in length among lepidopteran species (1-55 bp) and even overlaps in *A. yamamai* and *A. melete* (Table 4). Sequence homology with the neighboring ND6 and CytB genes is quite low in *P. bremeri* (data not shown).

The fourth intergenic spacer sequence, which is located between the $tRNA^{Ser}$(UCN) and ND1 (16 bp) of the *P. bremeri* mitogenome are always detected in the sequenced lepidopteran insects, with somewhat larger size (16-38 bp). This intergenic spacer sequence harbors the ATACTAA motif, which is conserved in all sequenced lepidopteran species (Fig. 4). This 7-bp consensus sequence was suggested to be the possible binding site for mtTERM, the transcription termination peptide, with the consideration that the intergenic spacer sequence can be detected at the end site of the major-strand coding region in the circular mtDNA (Cameron and Whiting, 2008; Taanman, 1999). Similarly, all sequenced coleopteran insects also harbor the 5-bp motif, which is shortened by one-bp at both ends from the 7-bp motif of lepidopteran insects (Sheffield et al., 2008). Excluding these, the size of the intergenic spacer in *P. bremeri* is generally far less than 10 bp.

Transfer RNA and ribosomal RNA genes

The *P. bremeri* mitogenome harbors 22 tRNA genes (Fig. 1). All *P. bremeri* tRNAs evidence the typical clover-leaf structure of mt tRNAs, with the exception of $tRNA^{Ser}$(AGN), the dihydrouridine (DHU) arm of which forms a simple loop (Fig. 5). The $tRNA^{Ser}$(AGN) with the simple loop in the DHU arm has also been detected in many insects (Wolstenholme, 1992) including Lepidoptera (e.g., Hong et al., 2008; Kim et al., 2006; Salvato et al., 2008). Nuclear magnetic resonance analysis of the tertiary structure of nematode $tRNA^{Ser}$ (AGN) indicated that such aberrant tRNA lacking the DHU arm also fits the ribosome by adjusting its structural conformation in a fashion similar to that of the tRNAs usually detected in the ribosome (Ohtsuki, 2002). The *P. bremeri* tRNAs harbor a total of three mismatches within the stem region (Table 5): one U-U mismatch each in the amino-acyl stem region of $tRNA^{Leu}$(CUN) and $tRNA^{Leu}$(UUR), and one C-U mismatch in the anticodon stem of $tRNA^{Lys}$. The *P. bremeri* tRNAs range in size between 61 ($tRNA^{Asp}$) and 71 bp ($tRNA^{Lys}$) (Table 5). All *P. bremeri* tRNAs invariably harbor 7 bp sequences in the amino-acyl stem, 5 bp in the anticodon stem, and 7 bp in the anticodon loop, but other portions of tRNAs are variable, particularly within the DHU and TΨC loops (4-11 bp and 4-10 bp, respectively). As is the case with all other insect mitogenome sequences, two rRNAs genes were detected in *P. bremeri*. These were located between $tRNA^{Leu}$(CUN) and $tRNA^{Val}$, and between $tRNA^{Val}$ and the A+T-rich region, respectively (Fig. 1). The lengths of the *P. bremeri* lrRNA and srRNA were determined to be 1,344 bp and 773 bp, respectively. The lengths of both the lrRNA and srRNA were well within the size range reported for other lepidopteran insects (Table 2). The A+T contents of the lrRNA and srRNA genes were 83.8% and 85.1%. These values are also well within the range reported for other lepidopteran insects (Table 2).

A+T-rich region

The 504-bp *P. bremeri* A+T-rich region is located between srRNA and $tRNA^{Met}$ (Fig. 1; Table 1). The A+T-rich region harbors the highest A+T content (93.6%) of any region of the *P. bremeri* mitogenome. The majority of the *P. bremeri* A+T-rich-region is composed of non-repetitive sequences, but harbors a poly-T stretch (18 thymine nucleotides located 21 bp upstream from srRNA), a microsatellite-like AT repeat (9 repeats located 412 bp upstream from srRNA), and a poly-A stretch (12 adenine nucleotides abutting to the 5′ end of $tRNA^{Met}$). The microsatellite-like $(AT)_9$ repeat is preceded by the ATTTA motif (data not shown) detected in the majority of other sequenced lepidopteran A+T-rich regions (Cameron and Whiting, 2008; Salvato et al., 2008). However, no large sequence repeats detected frequently in other insect orders, including the lepidopterans *A. pernyi* (Liu et al., 2008) and *B. mandarina* (Yukuhiro

Table 4. Overlapping and intergenic-space sequences of the lepidopteran mitogenomes

	Ahon	Aper	Ayam	Cboi	Bman	Bmor	Msex	Patr	**Pbre**	Amel	Crap	Olun	Ofur	Onub
$tRNA^{Met}$ - $tRNA^{Ile}$	(I)1	(I)9	(I)9	(I)1	(O)2		(I)10					(O)3	(I)1	(I)1
$tRNA^{Ile}$- $tRNA^{Gln}$	(I)1	(O)3	(O)3	(O)3	(O)3	(O)3	(O)3	(O)4	**(O)3**	(O)3	(O)3			
$tRNA^{Gln}$ - ND2	(I)64	(I)56	(I)53	(I)53	(I)47	(I)47	(I)54	(I)63	**(I)40**	(I)48	(I)56	(I)72	(I)62	(I)62
ND2 - $tRNA^{Trp}$	(I)1	(I)8	(I)7	(I)8	(I)5	(I)5		(I)11	**(O)1**	(O)2	(I)2	(I)1	(I)7	(I)7
$tRNA^{Trp}$- $tRNA^{Cys}$	(O)6	(O)8	(O)8	(O)8	(O)8	(O)7	(O)8	(O)8	**(O)8**	(O)8	(O)8	(O)8	(O)7	(O)7
$tRNA^{Cys}$ - $tRNA^{Tyr}$	(I)1	(I)19	(I)9		(I)6	(I)9	(I)10	(I)11	**(I)4**			(I)19	(I)3	(I)3
$tRNA^{Tyr}$ - COI	(I)7	(I)6	(I)2	(I)11	(I)17	(I)17	(I)4	(I)3	**(I)2**	(I)2	(I)3	(I)9	(I)11	(I)11
COI - $tRNA^{Leu}$(UUR)			(O)5	(O)6					**(O)5**		(I)1		(O)5	(O)5
$tRNA^{Leu}$(UUR) - COII	(I)1													
COII - $tRNA^{Lys}$														
$tRNA^{Lys}$ - $tRNA^{Asp}$	(I)14	(I)23	(I)16	(I)22	(O)1		(I)21		**(O)1**	(O)1	(I)13	(I)13	(O)1	(O)1
$tRNA^{Asp}$ - ATP8	(I)1	(I)36								(I)1			(O)1	(O)1
ATP8 - ATP6	(O)7	(O)7	(O)7	(O)7	(O)7	(O)7	(O)7	(O)8	**(O)7**	(O)7	(O)7	(O)7	(O)7	(O)7
ATP6 - COIII	(I)3	(O)1	(O)1		(I)12	(I)14	(I)18	(O)1	**(O)1**	(I) 8		(I)3	(O)1	(O)1
COIII - $tRNA^{Gly}$	(I)2	(I)2	(I)2	(I)2	(I)2	(I)2	(I)6	(I)12	**(I)3**	(O)1	(I)14	(I)2	(I)2	(I)2
$tRNA^{Gly}$ - ND3			(O)3		(I)3	(I)3			**(I)3**					(I)3
ND3 - $tRNA^{Ala}$	(O)1	(O)2	(O)2		(I)53	(I)31	(I)20	(I)3	**(O)1**	(O)2	(I)10	(I)74	(I)12	(I)10
$tRNA^{Ala}$ - $tRNA^{Arg}$	(I)1	(O)1	(O)1	(I)12	(I)50	(I)4	(I)1	(O)1	**(O)1**	(O)1	(I)1	(I)7	(O)1	(O)1
$tRNA^{Arg}$ - $tRNA^{Asn}$	(I)1				(I)1	(I)1	(I)50	(I)5		(I)1	(I)2	(I)12	(I)1	(O)1
$tRNA^{Asn}$ - $tRNA^{Ser}$(AGN)	(I)3	(I)1	(I)2	(O)1	(O)1		(I)3	(O)1	**(I)3**		(I)2	(I)1	(I)1	(I)1
$tRNA^{Ser}$(AGN) - $tRNA^{Ser}$(AGN)†											(I)11			
$tRNA^{Ser}$(AGN) - $tRNA^{Glu}$		(I)1	(I)1	(I)6	(I)10	(I)7		(O)1	**(I)43**	(I)15	(I)10	(I)6		
$tRNA^{Glu}$ - $tRNA^{Phe}$	(I)222	(I)10	(I)6	(I)2	(O)1		(O)2	(O)2	**(O)2**	(I)1	(O)2	(I)70		
$tRNA^{Phe}$ - ND5	(O)29	(O)15		(O)17	(I)4	(I)5		(O)15	**(I)1**	(I)2			(O)15	(O)16
ND5 - $tRNA^{His}$	(I)23		(O)3		(I)18	(I)21	(I)12	(I)18	**(O)3**	(I)18	(I)16		(I)15	(I)15
$tRNA^{His}$ - ND4	(O)61	(I)9	(I)12	(I)1	(I)59	(I)47			**(O)1**	(O)1			(O)6	(O)6
ND4 - ND4L	(O)1	(I)5	(I)3		(O)1	(O)1	(I)83	(I)36	**(O)1**	(O)4		(I)2	(I)7	(I)7
ND4L - $tRNA^{Thr}$	(I)1	(I)7	(I)5	(I)6	(I)4	(I)4	(I)13	(I)2	**(I)2**	(O)8	(I)14	(I)2	(I)9	(I)9
$tRNA^{Thr}$ - $tRNA^{Pro}$	(I)2				(O)1		(O)1	(I)1			(I)2			(O)1
$tRNA^{Pro}$ - ND6	(I)2	(I)2	(I)2	(I)2	(I)2	(I)2	(I)1	(I)2	**(I)2**	(I)2		(I)7	(I)2	
ND6 - CytB	(I)18	(I)2	(O)1	(I)4	(I)55	(I)50	(I)17	(I)1	**(I)17**	(O)1			(I)2	(I)2
CytB - $tRNA^{Ser}$(UCN)	(O)1	(I)15	(I)24	(I)41	(O)1	(I)3	(I)21	(O)1	**(I)1**			(I)9	(O)1	(O)1
$tRNA^{Ser}$(UCN) - ND1	(I)25	(I)21	(I)24	(I)22	(I)25	(I)25	(I)31	(I)18	**(I)16**	(I)16	(I)19	(I)17	(I)34	(I)38
ND1 - $tRNA^{Leu}$(CUN)	(I)2	(I)2	(I)1	(I)1	(O)6	(I)6		(I)6	**(I)1**	(I)3	(I)1	(I)1	(I)10	(I)10
$tRNA^{Leu}$(CUN) - lrRNA	(I)2											(I)48	(O)1	(O)1
lrRNA - $tRNA^{Val}$					(O)1									
$tRNA^{Val}$ - srRNA	(I)2													
Total nucleotides	(O)106	(O)37	(O)36	(O)42	(O)33	(O)18	(O)21	(O)42	**(O)35**	(O)39	(O)20	(O)18	(O)46	(O)49
	(I)400	(I)234	(I)178	(I)194	(I)373	(I)321	(I)375	(I)192	**(I)138**	(I)117	(I)177	(I)375	(I)179	(I)181

The 22 tRNAs are denoted by one-letter symbol and L*, L, S*, and S denote $tRNA^{Leu}$(UUR), $tRNA^{Leu}$(CUN), $tRNA^{Ser}$(AGN), and $tRNA^{Ser}$(UCN), respectively. N2, C1, C2, A8, A6, C3, N3, N5, N4, 4L, N6, CB, and N1 represent the ND2, COI, COII, ATP8, ATP6, COIII, ND3, ND5, ND4, ND4L, ND6, CytB, and ND1, respectively. Species names are abbreviated by using one alphabet from genus name and three alphabets from species name. Full name of the species are presented in Table 2. (O), Overlapping sequences; and (I), Intergenic space sequences. †An extra copy found only in *C. raphaelis*. Empty column means neighboring genes are abutting to each other.

Table 5. Size of each region of *Parnassius bremeri* tRNAs

Region	M	I	Q	W	C	Y	L*	K	D	G	A	R	N	S	E	F	H	T	P	S*	L	V
	69	64	69	66	66	64	67	71	67	67	66	66	67	61	66	70	64	66	65	67	69	65
Amino-acyl stem	7	7	7	7	7	7	7^1	7	7	7	7	7	7	7	7	7	7	7	7	7	7^1	7
DHU stem	4	3	3	4	4	3	3	3	4	4	4	4	3	-	4	4	4	4	4	4	3	4
Anticodon stem	5	5	5	5	5	5	5^1	5^1	5	5	5	5	5	5	5	5	5	5	5	5	5	5
TΨC stem	5	4	5	4	3	7	5	4	5	4	5	5	4	4	4	4	4	4	4	6	5	4
DHU loop	7	6	5	5	4	7	7	7	6	5	5	4	7	11	6	11	5	6	5	4	7	5
Anticodon loop	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7	7
Variable loop	4	5	4	4	4	4	4	6	4	4	4	5	5	7	4	4	4	4	4	4	4	4
TΨC loop	5	4	7	6	9	4	5	10	4	7	4	4	6	8	5	4	4	5	5	5	7	5

Transfer RNA genes are labeled by one-letter symbol according to the IUPAC-IUB single little amino acid codes.
L, L*, S and S* indicate $tRNA^{Leu}$(CUN), $tRNA^{Leu}$(UUR), $tRNA^{Ser}$(AGN), and $tRNA^{Ser}$(UCN), respectively.
Superscripts indicate number of mismatch in each region

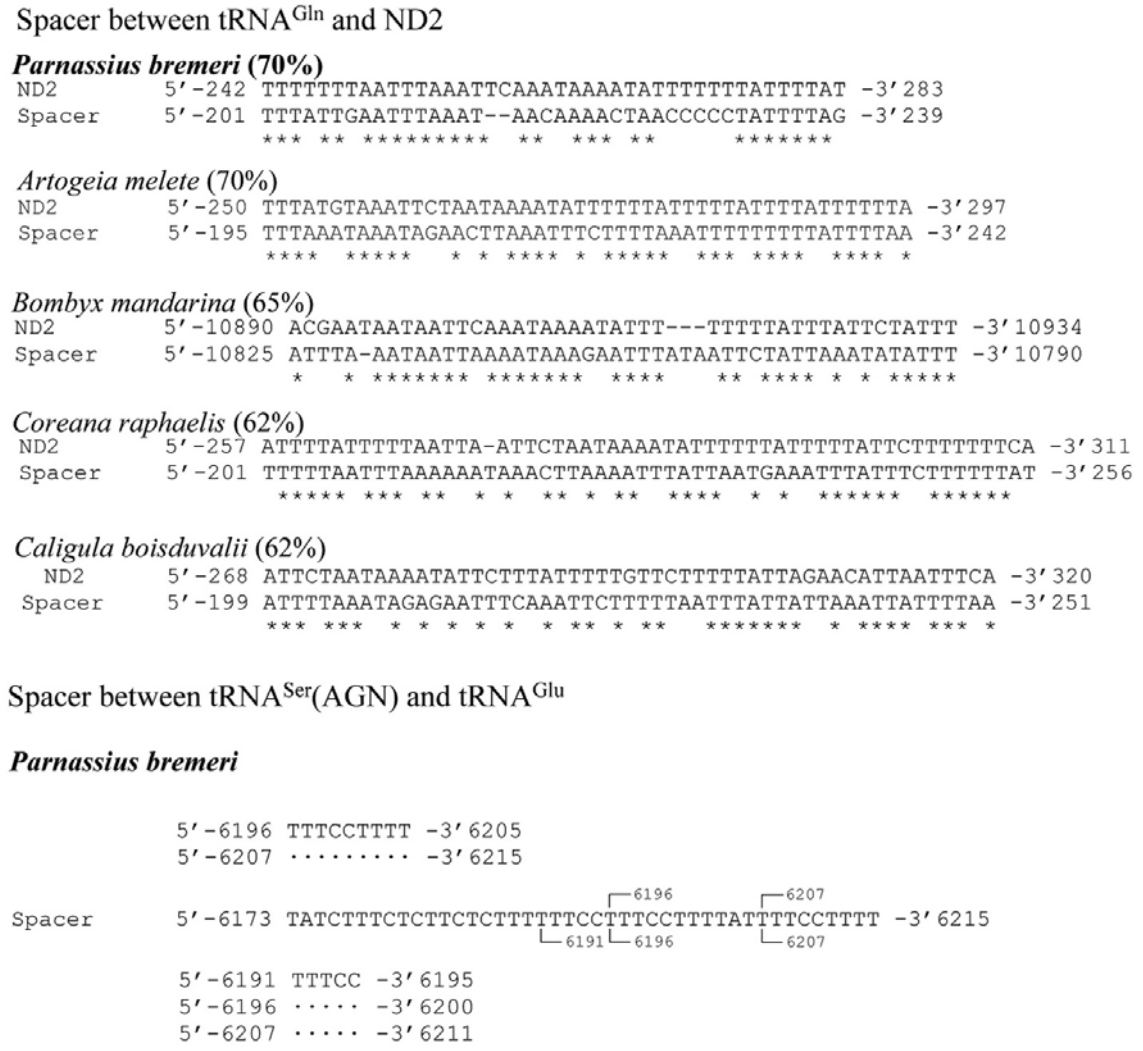


Fig. 3. Sequences of two relatively large intergenic spacer sequences. (A) Alignment of the intergenic space sequence located between $tRNA^{Gln}$ and ND2 gene and neighboring partial ND2 gene from some lepidopteran insects, including *Parnassius bremeri*. Asterisks indicate consensus sequences in the alignment between the intergenic spacer sequence and the ND2 gene. Sequence homology between the spacer and the ND2 gene is shown in the parenthesis next to the species name. The nucleotide position is indicated at the beginning and end sites of the sequence. (B) The intergenic space sequence detected between the $tRNA^{Ser}$(AGN) and $tRNA^{Glu}$ of *P. bremeri* (43 bp), and the alignment of repeat sequences detected within the intergenic spacer sequence. The nucleotide position is indicated at the beginning and end sites of the sequence.

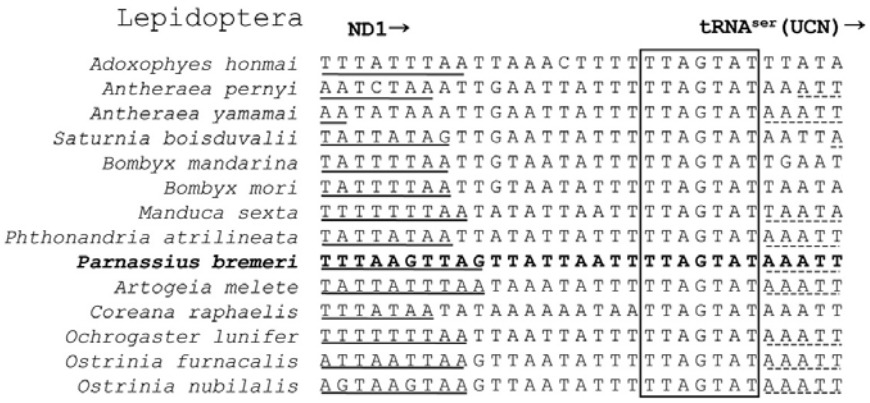


Fig. 4. Alignment of the internal spacer region located between ND1 and tRNA^{Ser} (UCN), in which complete or nearly complete mitogenomes are available. The boxed nucleotides indicate the conserved heptanucleotide region (TTAGTAT) detected in all sequenced lepidopteran insects. Underlined and dotted nucleotides, respectively, indicate the adjacent partial sequences of ND1 and tRNA^{Ser}(UCN). Arrows indicate the transcriptional direction.

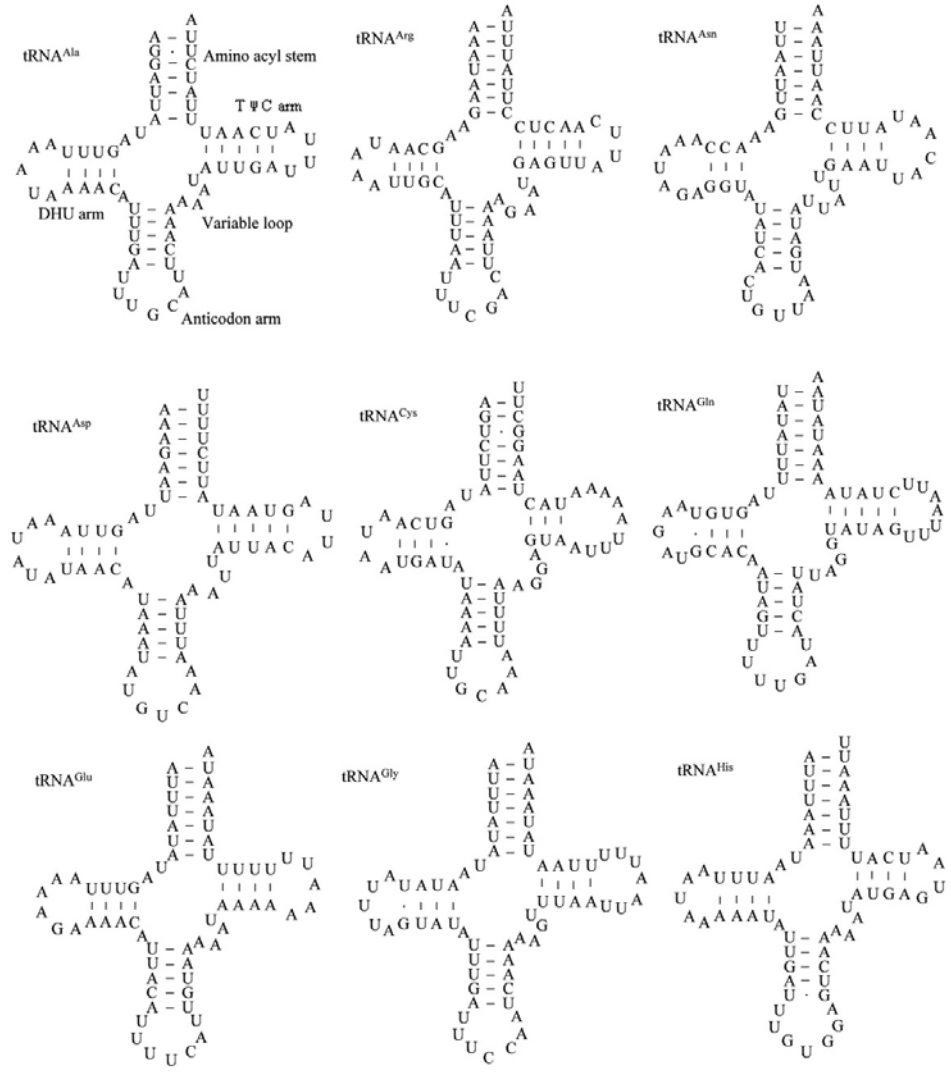
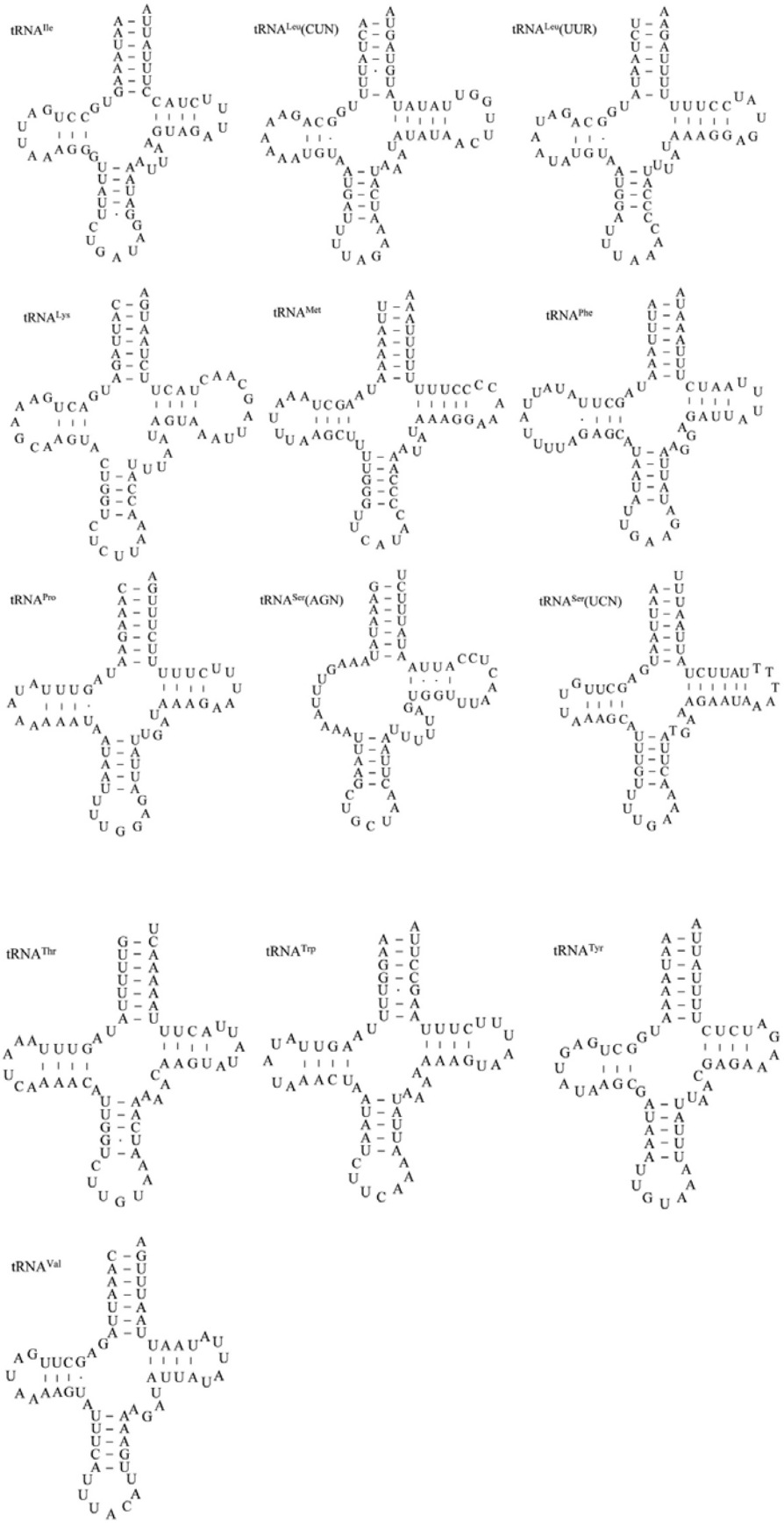


Fig. 5. Predicted secondary cloverleaf structures for the 22 tRNA genes of *Parnassius bremeri*. The tRNAs are labeled with the abbreviations of their corresponding amino acids. Nucleotide sequences from 5′ to 3′ are indicated for tRNA^{Ala}. Dashes (-) indicate Watson-Crick base-pairing, and centered asterisks (*) indicate 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 the dihydrouridine (DHU) arm.

(continued)

et al., 2002) were detected in the *P. bremeri* A+T-rich region.

The control region of the vertebrate mitogenome, which is equivalent to the A+T-rich region of the insect mitogenome, has been shown to harbor the origin of heavy-strand mtDNA replication (Tapper and Clayton, 1984). In insects, the replication origin for both mtDNA strands has been detected in the A+T-rich regions of *Drosophila* species (Clary and Wolstenholme, 1987; Fauron and Wolstenholme, 1980). More recently, Saito et al. (2005) determined the precise position of the replication origin for minor-strand mtDNA from species belonging to Diptera, Lepidop-


tRNA^Ile
tRNA^Leu(CUN)
tRNA^Leu(UUR)
tRNA^Lys
tRNA^Met
tRNA^Phe
tRNA^Pro
tRNA^Ser(AGN)
tRNA^Ser(UCN)
tRNA^Thr
tRNA^Trp
tRNA^Tyr
tRNA^Val

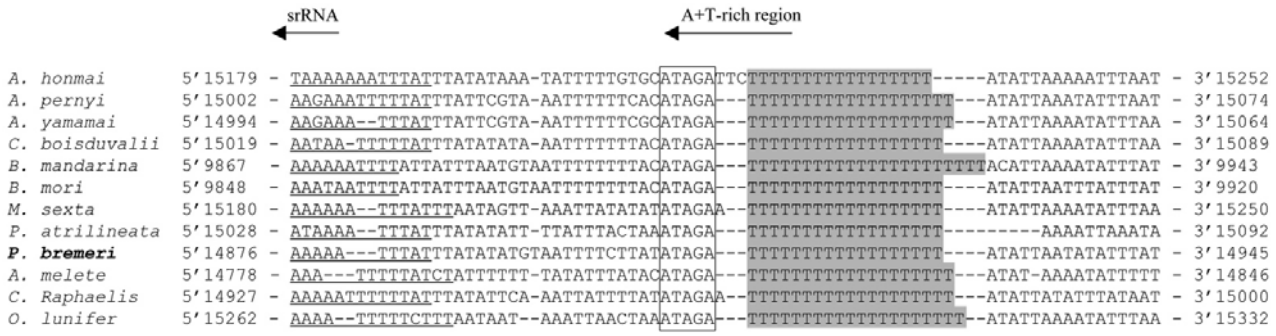


Fig. 6. Alignment of partial A+T-rich region (no underline) and srRNA (underline). The shaded nucleotides indicate the poly-T stretch and the boxed nucleotides indicate the conserved motif, ATAGA. The direction of replication is indicated by arrows. The nucleotide position is indicated at the beginning and end sites of the sequence with respect to each mitogenome.

tera, Coleoptera, and Orthoptera, and also reported that the lepidopteran *Bombyx mori* harbors this origin immediately downstream of a poly-T stretch located at the 3′-end of the *B. mori* A+T-rich region (upstream of the 5′-end of the srRNA gene). Thus, this poly-T stretch has been suggested to function as a structural signal for the recognition of proteins in the initiation of replication of minor-strand mtDNA. The *P. bremeri* also harbors the 18-bp long poly-T stretch at the similar position 26 bp upstream of the 3′-end of the A+T-rich region (Fig. 6). In fact, other sequenced lepidopteran mitogenomes all harbor the poly-T stretch ranging in size from 17-22 bp at 22-28 bp upstream of the 3′-end of the A+T-rich region (Fig. 6). With regard to positioning, other insect orders, including the Diptera, Coeloptera, and Orthoptera, harbor this regulatory sequence in the middle of the A+T-rich region (Saito et al., 2005), but those of lepidopteran species are present near the end of the A+T-rich region (Fig. 6).

Immediately downstream of the poly-T stretch another conserved motif, ATAGA, located 22 bp upstream of srRNA in the case of *P. bremeri*, is uniquely detected in all sequenced lepidopteran insects (Fig. 6). Thus, this sequence motif may also perform some regulatory function together with the poly-T stretch. Considering the taxonomic diversity of the sequenced lepidopteran insects, which encompass six superfamilies, the conservancy of the motif in the lepidopteran insects is worth noting, particularly considering that no such conserved sequence is detected immediately downstream of the A+T-rich region of the 12 *Drosophila* species belonging to a dipteran genus (Saito et al., 2005). Thus, the lepidopteran insects may have obtained such unique regulatory features independently after splitting from an ancestral species.

By way of contrast with the origin of minor-strand replication, no experiments have yet ascertained precisely the location of the structural signals for major-strand replication in insects, with the exception of *Drosophila* species, wherein a poly-T stretch located 16-21 bp upstream of $tRNA^{Ile}$ within the A+T-rich region was identified as the regulatory sequence for the recognition of the origin of major-strand replication (Saito et al., 2005). In 12 *Drosophila* species, this poly-T stretch ranges in size from 13-25 bp (Saito et al., 2005), but that of the sequenced lepidopteran species ranges from 1 bp (*A. yamamai*; Kim et al., 2009) ~12 bp (*P. bremeri*; This study), or is interrupted by non-AT nucleotides in certain species (data not shown). Thus, this poly-T stretch is less apparent than in *Drosophila* and less apparent than minor-strand replication in Lepidoptera, thus casting doubt on the notion that such a short T-stretch in some lepidopteran species may function either as a recognition site for protein interaction, or not. This poly-T stretch or T nucleotide in sequenced lepidopteran insects abuts the neighboring $tRNA^{Met}$. Thus, the origin of major-strand replication in the lepidopteran insects must exist within $tRNA^{Met}$, if this poly-T stretch functions as a regulatory signal.

The presence of tRNA-like sequences within the A+T-rich region has been reported in Hymenoptera (Cha et al., 2007; Coizier and Crozier, 1993; Hong et al., 2008), Lepidoptera (*A. yamamai*; Kim et al., 2009), and Coleoptera (Hong et al., 2009b). The *P. bremeri* A+T-rich region also harbors one $tRNA^{Trp}$-like sequence and one $tRNA^{Leu}$(UUR)-like sequence, both encoded in the major strand (Fig. 7). The possession of the proper anticodon and the formation of a cloverleaf structure allowed us to recognize these tRNA-like sequences. The sequence homology between the regular and tRNA-like structure was 54% for $tRNA^{Trp}$ and 46% for $tRNA^{Leu}$(UUR), thereby indicating the presence of fairly substantial sequence divergence between the regular tRNAs and tRNA-like sequences. Furthermore, the stem regions of both tRNA-like sequences evidence many mismatches (Fig. 7). Thus, the functionality of these tRNA-like sequences remains unknown. Our effort to detect the tRNA-like sequences from other completely sequenced lepidoteran insects yielded the finding that all, with the exceptions of *A. honmai* and *M. sexta*, harbor at least one tRNA-like structure (Fig. 7), thus suggesting that such a feature appears to occur frequently in the A+T-rich region of insect mitogenomes, at least in the lepidopteran A+T-rich region. It has been previously demonstrated in studies of the replication origin for the heavy strand of mammalian mtDNA that the nucleotide sequence--which can be folded into tRNA-like structures--probably functions as a primer that may be necessary for DNA synthesis (Brown et al., 1986). To extend this by analogy, tRNAs themselves may function as primers for mtDNA synthesis, and the subsequent failure to cleave the tRNA primer from the nascent DNA strand may possibly induce the incorporation of a tRNA gene into the mitogenome (Cantatore et al., 1987). Once incorporated, the tRNA-like sequences would begin to accumulate sequence divergence by following the evolutionary ratio of the non-coding nature of the A+T-rich region. Interestingly, the $tRNA^{Leu}$(UUR)-like sequence was most frequently observed as five among 16 tRNA-like structures found in the 10 completely sequenced lepidopteran mitogenomes, whereas others range from one to three (Fig. 7). The highest frequency of the $tRNA^{Leu}$(UUR)-like sequence in the lepidopteran mitogenomes may be explained in terms of the highest usage of the TTA codon encoding for leucine (Supplemantary Table 2). It seems that the $tRNA^{Leu}$(UUR)-like sequence may have the highest turnover ratio in cells, and this may have increased the chance to incorporate into the

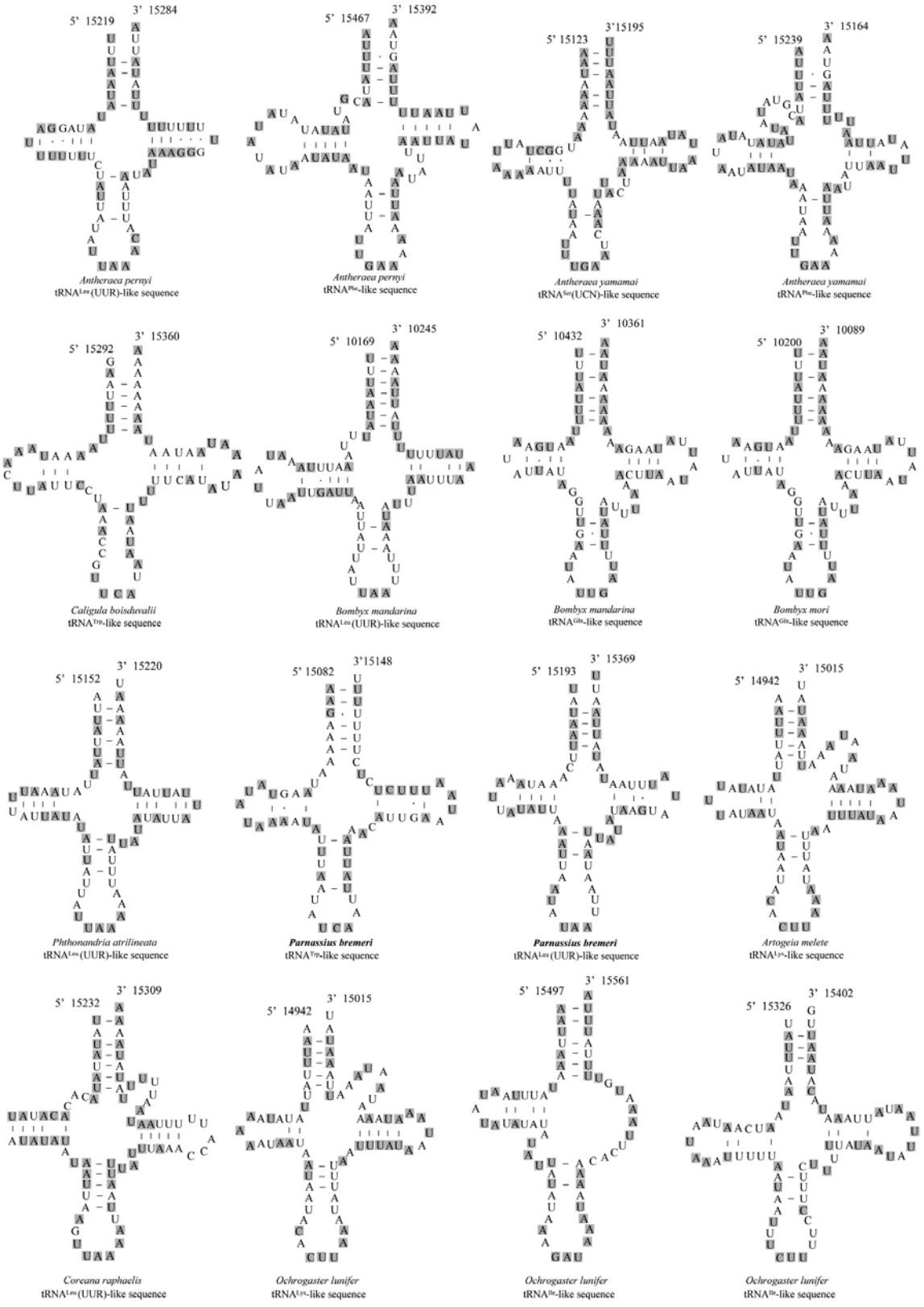


Fig. 7. Predicted secondary clover-leaf structure of tRNA-like sequences detected in the A+T-rich region of the sequenced lepidopteran species, including the tRNA^{Trp}-like sequence and tRNA^{Leu}(UUR)-like sequence from the *Parnassius bremeri* A+T-rich region. The nucleotide position of the clover-leaf structure is indicated at the beginning and end sites of the structure with regard to each mitogenome. The shaded nucleotides indicate the consensus sequences as compared with each corresponding regular tRNA.

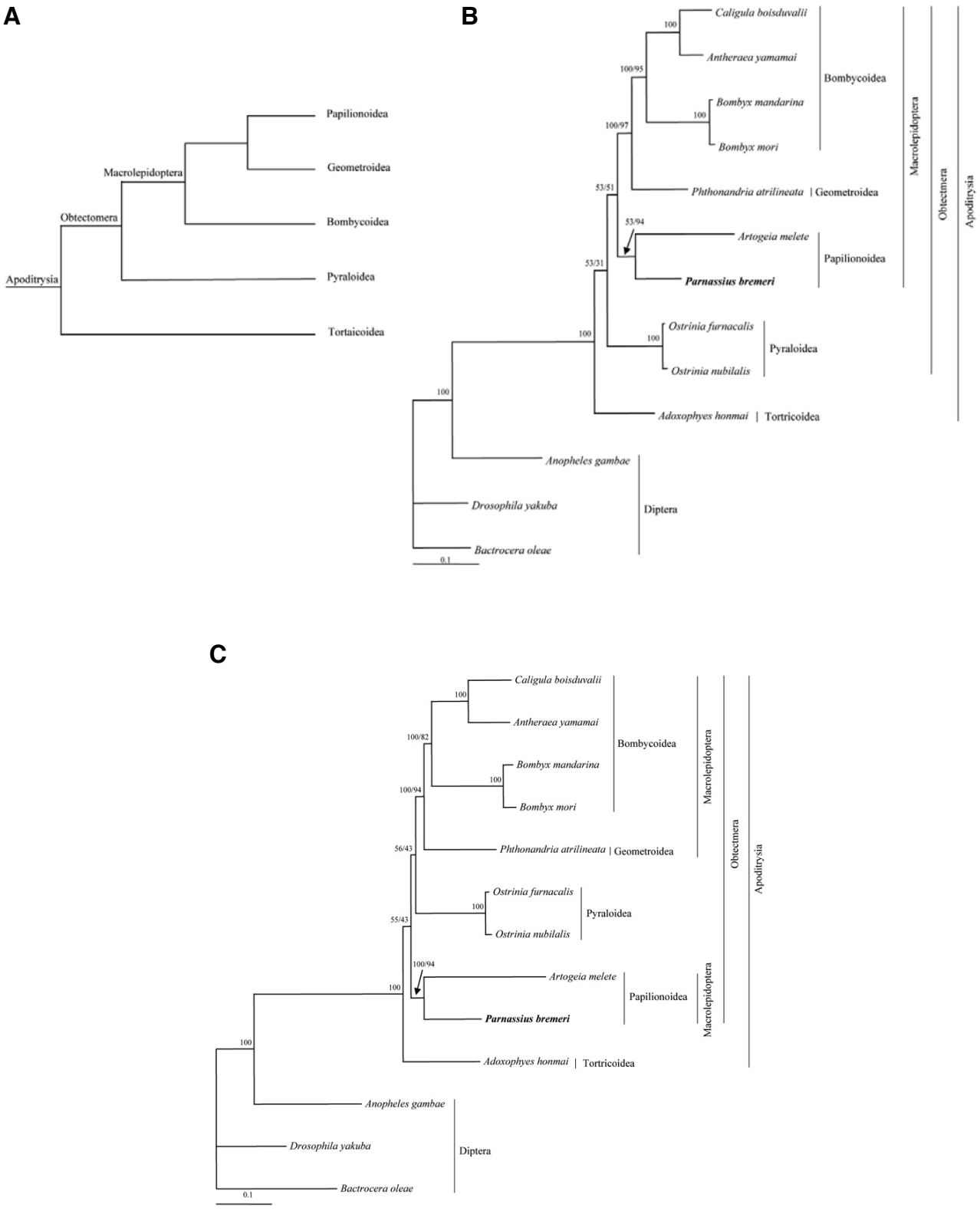


Fig. 8. Phylogeny of apoditrisian superfamilies. (A) The current hypothesis of apoditrisian superfamily relationships (Minet, 1991; Nielsen, 1989). (B) Bayesian Inference phylogram of apoditrisian superfamilies obtained with the amino acid dataset. The dipterans, *Drosophilla yakuba* (Clary and Wolstenholme, 1985), *Anopheles gambiae* (Beard et al., 1993), and *Bactrocera oleae* (Nardi et al., 2003) were employed as the co-outgroup. The numbers at each node specify BPP by BI analysis (first value) and bootstrap percentages of 100 pseudoreplicates by ML analysis (second value). Single value at each node indicates identical one by both analyses. The scale bar indicates the number of substitutions per site. (C) Bayesian Inference phylogram of apoditrisian superfamilies obtained with the nucleotide dataset. The detailed information of the tree is the same as in (B).

mitogenome. Considering that the non-identical type of tRNA-like structures are detected in the within-generic species *A. pernyi* [tRNA^{Leu}(UUR)-like sequence and tRNA^{Phe}-like sequence] and *A. yamamai* [tRNA^{Ser}(UCN)-like sequence and tRNA^{Phe}-like sequence], this incorporation process appears to be evolutionarily random, as opposed to having taxonomic implications (Fig. 7). As more complete mitogenome sequences become available, it may become possible to acquire greater insight into this matter.

Phylogenetic relationships

The 10 lepidopteran mitogenomes that passed heterogeneity test represent five lepidopteran superfamilies. Among them, the superfamilies Papilionoidea, Bombycoidea, and Geometroidea are collectively referred to as the Macrolepidoptera. There has been substantial controversy surrounding the phylogenetic relationships among macrolepidopteran superfamilies (Minet, 1991; 1994; Scott, 1986) and several lines of evidence indicate a closer relationship between the Papilionoidea and the Geometroidea, leaving the Bombycoidea as a basal lineage (Fig. 8A; after Minet, 1991; Nielsen, 1989). Our phylogenetic analysis of these Macrolepidoptera together with other members of the lepidopteran insects belonging to the Apoditrysia showed an unexpected clustering of the Geometroidea and Bombycoidea, in both the BI and ML trees, leaving Papilionoidea represented by *A. melete* belonging to Pieridae and our *P. bremeri* belonging to Papilionidae as a basal lineage (Figs. 8B and 8C). The node support for the Geometroidea and Bombycoidea group was found to be high, at 100% and 97%, by the BI and ML analyses, respectively, using amino acid sequences (Fig. 8B) and 100% and 94% by the BI and ML analyses, respectively, using the nucleotide sequences (Fig. 8C). No previous phylogenetic analysis, whether molecular analysis, morphological analysis, or combined analyses has thus far suggested such a possibility (e.g., Minet, 1994; Regier et al., 2008; Weller and Pashely, 1995). Nevertheless, a recent phylogenetic analysis using the nucleotide and amino acid sequences of the concatenated 13 mt PCGs also supported the notion that the Geometroidea were closely related to the Bombycoidea (Yang et al., 2009). In that study, the same ML method was performed as in this study, as well as additional MP and neighbor-joining methods with a different outgroup scheme (one Dipteran and one Orthoptera), and strong support was detected for a close relationship between the Geometroidea and the Bombycoidea. These results indicate that macrolepidopteran evolution is more complex than we understand, and justify the necessity for further research into lepidopteran mitogenomes.

Within Obtectomera, a monophylely of Macrolepidoptera was either weakly supported by BI (53%) and ML (51%) using amino acid sequences (Fig. 8B) or not supported by both BI and ML analyses using nucleotide sequences (Fig. 8C). Previous phylogenetic analysis using concatenated 13 mt PCGs also revealed a similar pattern when the nucleotide sequences were utilized (Yang et al., 2009). Nevertheless, this result is quite unusual in that the Macrolepidoptera are generally well-grouped on the basis of the observed improved flying ability conferred via changes in wing shape, reduction in the adult mandibles, and changes in the arrangement of the crochets (hooks) on the larval prolegs (Grimaldi and Engel, 2005), all of which demonstrate the existence of a robust monophyletic group (Kristensen et al., 2007; Minet, 1991; Nielsen, 1989). In future studies, the mitogenomic information from more non-macrolepidopteran species may be required.

Collectively, the information currently available supports a clustering of the Geometroidea and Bombycoidea, and weak support for a monophyly of Macrolepidoptera, whereas it supports a monophyly of Papilionoidea and a monophyly of Bombycoidea. In order to further evaluate the phylogenetic relationships among macrolepidopteran and obtectomeran insects, a larger number of complete mitogenome sequences that encompass more of the taxonomic diversity will be required, as our analysis covered only a small portion of the taxonomic diversity, due to limited availability.

Note: Supplementary information is available on the Molecules and Cells website (www.molcells.org).

ACKNOWLEDGMENTS

This work was supported by research grants from National Institute of Biological Resources, "The Genetic Evaluation of Important Biological Resources" and Origin of Biological Diversity of Korea: Molecular Phylogenetic Analyses of Major Korean Taxa" awarded to Iksoo Kim.

REFERENCES

Abascal, F., Zardoya, R., and Posada, D. (2005). ProTest: selection of best-fit models of protein evolution. Bioinformatics *21*, 2104-2105.

Abascal, F., Posada, D., and Zardoya, R. (2007). MtArt: a new model of amino acid replacement for arthropoda. Mol. Biol. Evol. *24*, 1-5.

Adachi, J., and Hasegawa, M. (1996). Model of amino acid substitution in proteins encoded by mitochondrial DNA. J. Mol. Evol. *42*, 459-468.

Akaike, H. (1974). A new look at the statistical model identification. IEEE Trans. Autom. Contr. *19*, 716-723.

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijin, M.H.L., Droujn, A.R.J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., et al. (1981). Sequence and organization of the human mitochondrial genome. Nature *290*, 457-465.

Beard, C.B., Mills, D., and Collins, F.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, J.L., lavrov, D., and Brown,W.M. (1998). Gene translocation links insects and crustaceans. Nature *393*, 667-668.

Brehm, A., Harris, D.J., Hernández, M., Cabrera, V.M., Larruga, J.M., Pinto, F.M., and González, A.M. (2001). Structure and evolution of the mitochondrial DNA complete control region in the *Drosophila subobscura* subgroup. Insect Mol. Biol. *10*, 573-578.

Brown, G.G. (1986). Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. J. Mol. Biol. *192*, 503-511.

Cameron, S.L., and Whiting, M.F. (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.

Cantatore, P., Gadaleta, M.N., Roberti, M., Saccone, C., and Wilson, A.C. (1987). Duplication and remodeling of tRNA genes during the evolutionary rearrangement of mitochondrial genomes. Nature *329*, 853-855.

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic tool. Curr. Opin. Genet. Dev. *8*, 668-674.

Cha, S.Y., Yoon, H.J., Lee, E.M., Yoon, M.H., Hwang, J.S., Jin, B.R., Han, Y.S., and Kim, I. (2007). The complete nucleotide sequence and gene organization of the mitochondrial genome of the bumblebee, Bombus ignitus (Hymenoptera: Apidae). Gene *392*, 206-220.

Clary, D.O., and Wolstenholme, D.R. (1985). The mitochondrial DNA molecule of Drosophila yakuba: nucleotide sequence, gene organization, and genetic code. J. Mol. Evol. *22*, 252-271.

Clary, D.O., and Wolstenholme, D.R. (1987). *Drosophila* mitochondrial DNA: Conserved sequences in the A+T-rich region and supporting evidence for a secondary structure model of the small ribosomal RNA. J. Mol. Evol. *25*, 116-125.

Coates, B.S., Sumerford, D.V., Hellmich, R.L., and Lewis, L.C. (2005). Partial mitochondrial genome sequences of *Ostrinia nubilalis* and

Ostrinia furnicalis. Int. J. Biol. Sci. *1*, 13-18.
Crozier, R.H., and Crozier, Y.C. (1993). The mitochondrial genome of the honeybee Apis mellifera: complete sequence and genome organization. Genetics *133*, 97-117.
Fauron, C.M.R., and Wolstenholmn, D.R. (1980). Extensive diver-sity among *Drosophila* species with respect to nucleotide sequences within the adenine+thymine-rich region of mitochondrial DNA molecules. Nucleic Acids Res. *8*, 2439-2452.
Folmer, O., Black, M., Hoeh, W., Lutz, R., and Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol. Mar. Biol. Biotechnol. *3*, 294-299.
Grimaldi, D., and Engel, M.S. (2005). Evolution of the insects. (New York: Cambridge University Press, Cambridge, U.K.).
Guindon, S., Lethiec, F., Duroux, P., and Gascuel, O. (2005). PHYML Online - a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res. *33*, W557-W559.
Hall, T.A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acids Symp. Ser. *41*, 95-98.
Hong, M.Y., Lee, E.M., Jo, Y.H., Park, H.C., Kim, S.R., Hwang, J.S., Jin, B.R., Kang, P.D., Kim, K.-G., Han, Y.S., et al. (2008). Complete nucleotide sequence and organization of the mito-genome of the silk moth *Caligula boisduvalii* (Lepidoptera: Saturniidae) and comparison with other lepidopteran insects. Gene *413*, 49-57.
Hong, G., Jiang, S., Yu, M., Yang Y., Li, F., Xue, F., and Wei, Z. (2009a). The complete nucleotide sequence of the mitochondrial genome of the cabbage butterfly, Artogeia melete (Lepidoptera: Pieridae). Acta. Biochim. Biophys. Sin. *41*, 446-455.
Hong, M.Y., Jeong, H.C., Kim, M.J., Jeong, H.U., Lee, S.H., and Kim, I. (2009b). Complete mitogenome sequence of the jewel beetle, *Chrysochroa fulgidissima* (Coleoptera: Buprestidae). Mitochondrial DNA. *20*, 46-60.
Huelsenbeck, J.P., and Ronquist, F. (2001). MrBayes: Bayesian inference of phylogeny. Bioinformatics *17*, 754-755.
Kim, Y.S. (2005). Illustrated book of Korean butterflies in color (Seoul, Korea: Kyo-Hak Pub. Co.).
Kim, I., Lee, E.M., Seol, K.Y., Yun, E.Y., Lee, Y.B., Hwang, J.S., and Jin, B.R. (2006). The mitochondrial genome of the Korean hairstreak, Coreana raphaelis (Lepidoptera: Lycaenidae). Insect Mol. Biol. *15*, 217-225.
Kim, S.R., Kim, M.I., Hong, M.Y., Kim, K.Y., Kang, P.D., Hwang, J.S., Han, Y.S., Jin, B.R., and Kim, I. (2009). The complete mitogenome sequence of the Japanese oak silkmoth, *Antheraea yamamai* (Lepidoptera: Saturniidae). Mol. Biol. Rep. *36*, 1871-1880.
Ko, M.S., Lee, J.S., Kim, C.H., Kim, S.S., and Park, K.T. (2004). Distributional data and ecological characteristics of *Parnassius bremeri* Bremer in Korea. Kor. J. App. Entomol. *43*, 7-14.
Kristensen, N.P., Scoble, M.J., and Karsholt, O. (2007). Lepidoptera phylogeny and systematic: the state of inventorying moth and butterfly diversity. Zootaxa *1668*, 699-747.
Lanave, C., Preparata, G., Saccone, C., and Serio, G. (1984). A new method for calculating evolutionary substitution rates. J. Mol. Evol. *20*, 86-93.
Lee, E.S., Shin, K.S., Kim, M.S., Park, H., Cho, S., and Kim, C.B. (2006). The mitochondrial genome of the smaller tea tortrix *Adoxophyes honmai* (Lepidoptera: Tortricidae). Gene *373*, 52-57.
Liu, Y., Li, Y., Pan, M., Dai, F., Zhu, X., Lu, C., and Xiang, Z. (2008). The complete mitochondrial genome of the Chinese oak silkmoth, Antheraea pernyi (Lepidoptera: Saturniidae). Acta. Biochim. Biophys. Sin. *40*, 693-703.
Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. *25*, 955-964.
Minet, J. (1991). Tentative reconstruction of the ditrysian phylogeny (Lepidoptera: Glossata). Entomol. Scand. *22*, 69-95.
Minet, J. (1994). The Bombycoidea: phylogeny and higher classification (Lepidoptera: Glossata). Entomol. Scand. *25*, 63-88.
Nardi, F., Caeapelli, A., Dallai, R., and Frati, F. (2003). The mitochondrial genome of the olive fly *Bactrocera oleae*: two haplotypes from distant geographic locations. Insect Mol. Biol. *12*, 605-611.
Nielsen, E.S. (1989). Phylogeny of major lepidopteran groups. In the Hierarchy of Life, B. Fernholm, K. Bremer, and H. Jörnvall, eds., (Amsterdam: Elsevier), pp. 281-294.
Ohtsuki, T., Kawai, G.., and Watanabe, K. (2002). The minimal tRNA: unique structure of *Ascaris suum* mitochondrial $tRNA^{Ser}$-UCU having a short T arm and lacking the entire D arm. FEBS Lett. *514*, 37-43.
Ojala, D., Montoya, J., and Attardi, G. (1981). tRNA punctuation model of RNA processing in human mitochondria. Nature *290*, 470-474.
Posada, D., and Crandal, K.A. (1998). Modeltest: testing the model of DNA substitution. Bioinformatics *14*, 817-818.
Saito, S., Tamuea, K., and Aotsuka, T. (2005). Replicatiion origin of mitochondrial DNA in insects. Genetics *171*, 433-448.
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.
Schmidt, H.A., Strimmer, K., and von Haeseler, A. (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. Bioinformatics *18*, 502-504.
Scott, J. (1986). On the monophyly of the Macrolepidoptera, including a reassessment of their relationship to Cossoidea and Castnioidea, and a reaasignment of Mimallonidae to Pyraloidea. J. Res. Lepid. *25*, 30-38.
Sheffield, N.C., Song, H., Cameron, S.L., and Whiting, M.F. (2008). A comparative analysis of mitochondrial genomes in Coleoptera (Arthropoda: Insecta) and genome descriptions of six new beetles. Mol. Biol. Evol. *25*, 2499-2509.
Taanman, J.W. (1999). The mitochondrial genome: structure, transcription, translation and replication. Biochim. Biophys. Acta *1410*, 103-123.
Tapper, D.A., and Clayton, D.A. (1981). Mechanism of replication of human mitochondrial DNA: localization of the 5′ ends of nacent daughter strands. J. Biol. Chem. *256*, 5109-5115.
Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). Clustal-W - improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. *22*, 4673-4680.
Thompson, J.D., Gibson, T.J, Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997) The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. *24*, 173-216.
Weller, S.J., and Pashely, D.P. (1995). In search of butterfly origins. Mol. Phylogenet. Evol. *4*, 235-246.
Wernersson, R., and Pedersen, A.G. (2003). Multiple alignment of coding DNA from aligned amino acid sequences. Nucleic Acids Res. *31*, 3537-3539.
Wettstein, W., and Schmid, B. (1999). Conservation of arthropod diversity in montane wetlands: effect of altitude, habitat quality and habitat fragmentation on butterflies and grasshoppers. J. Appl. Ecol. *36*, 363-373.
Wilson, K., Cahill, V., Ballment, E., and 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*, 863-874.
Wolstenholme, D.R. (1992). Animal mitochondrial DNA: structure and evolution. Int. Rev. Cytol. *141*, 173-216.
Woo, H.J., Lee, Y.S., Park, S.J., Lim, J.T., Jang, K.H., Choi, E.H., Choi, Y.G., and Hwang, U.W. (2007). Complete mitochondrial genome of a troglobite millipede *Antrokoreana gracilipes* (Diplopoda, Juliformia, Julida), and juliformian phylogeny. Mol. Cells, *23*, 182-191.
Yang, L., Wei, Z.J., Hong, G.Y., Jiang, S.T., and Wen, L.P. (2009). The complete nucleotide sequence of the mitochondrial genome of *Phthonandria atrilineata* (Lepidoptera: Geometridae). Mol. Biol. Rep. *36*, 1441-1449.
Yukuhiro, K., Sezutsu, H., Itoh, M., Shimizu, K., and Banno, Y. (2002). Significant levels of sequence divergence and gene rearrangements have occurred between the mitochondrial genomes of the wild mulberry silk moth, *Bombyx mandarina*, and its close relative, the domesticated silk moth, *Bombyx mori*. Mol. Biol. Evol. *19*, 1385-1389.