*Full Length Research Paper*

# Complete nucleotide sequence and organization of the mitogenome of endangered *Eumenis autonoe* (Lepidoptera: Nymphalidae)

## Min Jee Kim[1], Xinlong Wan[1], Ki-Gyoung Kim[2], Jae Sam Hwang[3] and Iksoo Kim[1]*

[1]College of Agriculture and Life Sciences, Chonnam National University, Gwangju 500-757, Republic of Korea.
[2]Biological Resources Research Department, National Institute of Biological Resources, Incheon, 404-708, Republic of Korea.
[3]National Academy of Agricultural Science RDA, Suwon 441-707, Republic of Korea.

*Eumenis autonoe,* a member of the lepidopteran family, Nymphalidae (superfamily Papilionoidea) is an endangered species and is found only on one isolated remote island Jeju in South Korea, on Halla Mt, at altitudes higher than 1,400 m. In this study, the complete mitochondrial genome (mitogenome) of *E. autonoe* was reported. The 15,489-bp long *E. autonoe* genome evidenced the typical gene content found in animal mitogenomes, and harbors the gene arrangement identical to all other sequenced lepidopteran insects, which differs from the most common type found in insects, due to the movement of tRNA$^{Met}$ to a position 5'-upstream of tRNA$^{Ile}$. As has been observed in many other lepidopteran insects, no typical ATN codon for the COI gene is available. Thus, we also designated the CGA (arginine) found at the beginning of the COI gene as a lepidopteran COI starter, in accordance with previous suggestions. The 678 bp long A + T-rich region, which is second longest in sequenced lepidopteran insects, harbored 10 identical 27 bp long tandem repeats plus one 13 - bp long incomplete final repeat. Such a repeat sequence has been, thus far, only rarely detected in lepidopteran mitogenomes. The *E. autonoe* A + T-rich region harbored a poly-T stretch of 19 bp and a conserved ATAGA motif located at the end of the region, which have been suggested to function as structural signals for minor-strand mtDNA replication. Phylogenetic reconstruction using the concatenated 13 amino acid and nucleotide sequences of the protein-coding genes (PCGs) consistently supported a close relationship between Bombycoidea and Geometroidea among six available lepidopteran superfamilies (Tortricoidea, Pyraloidea, Papilionoidea, Bombycoidea, Geometroidea and Noctuoidea). Among the true butterflies (Pieridae, Nymphalidae, Lycaenidae and Papilionidae), a closer relationship between Lycaenidae and Pieridae, excluding Nymphalidae was consistently concluded to exist, although this result deviated from the traditional view.

**Key words:** Mitochondrial DNA, complete mitogenome, *Eumenis autonoe*, the A + T-rich region, papilionoidea, macrolepidoptera, lepidopteran phylogeny.

## INTRODUCTION

*Eumenis autonoe,* a member of the lepidopteran family Nymphalidae (superfamily Papilionoidea) is listed as a first-degree endangered wild animal in Korea (Kim, 2005). Historically, this species has been distributed throughout the Northern hemisphere, including Europe,

Korean Peninsula (Joo and Kim, 2002). Currently, this Russia and North-eastern China, as well as on the species is detected only in two isolated regions of the Korean peninsula: In South, on the remote Island of Jeju, on Halla Mt. at altitudes higher than 1,400 m and in North in Hamkyeumg-do in the far North of the Korean Peninsula. The Halla Mt. population has been generally regarded as a remnant of the Pleistocene glaciations, the result of migration to the cooler higher mountain range

---

*Corresponding author. E-mail: ikkim81@chonnam.ac.kr.

regions (Joo and Kim, 2002). Thus, this species is a very rare example that provided us with an opportunity to evaluate the historical adaptation of butterfly species to the cooler temperatures prevailing on the Korean Peninsula. Nevertheless, the on-going global warming is predicted to further diminish the current population, which is already quite small. This may be particularly true in that true butterfly species are vulnerable to local extinction as a result of global warming (Wettstein and Schmid, 1999). Therefore, it seems essential to accumulate a minimal, but significant amount of genetic species information, which is largely unknown from the genetic perspective. Such information is expected to be helpful for long-term conservation objectives, as well as future utility. In this regard, the full-length mitochondrial genome (mitogenome) information is an important component of this knowledge. For example, the sequence information can be utilized in future population level work, including genetic diversity estimates, population relationships between donor and donee populations, primer design and so on.

Mitogenome sequences have already been determined in a variety of insects. This chart includes 16 lepidopteran species, belonging to six super families. However, only three of these species are true butterflies (Papilionoidea) belonging to the Papilionidae, Pieridae and Lycaenidae, respectively, but lacks another major true butterfly family, namely the Nymphalidae. In this study, thus, the complete mitogenome sequence of *E. autonoe* belonging to the Nymphalidae was determined. The mitogenome sequence was described via comparison to other insect mitogenomes, in particular to those of 16 sequenced lepidopteran species in terms of whole genome organization, arrangement, the major characteristics of individual genes and the composition of non-coding regions, including the A+T-rich region. Thus, the newly sequenced *E. autonoe* mitogenome is anticipated to enrich our understanding of the comparative biology of insect mitogenomes, particularly those of lepidopteran species.

Furthermore, the concatenated amino acid and nucleotide sequences of 13 protein-coding genes (PCGs) of the *E. autonoe* mitogenome were utilized in order to reconstruct the phylogenetic relationships among available lepidopteran superfamilies (Papilionoidea, Bombycoidea, Geometroidea, Noctuoidea, Tortricoidea and Pyraloidea) as well as among the true butterfly families within Papilionoidea (Pieridae, Nymphalida, Lycaenidae and Papilionidae).

## MATERIALS AND METHODS

### Genomic DNA extraction

An adult specimen of *E. autonoe* was collected from Halla Mt., Korea on May, 2008. Total genomic DNA was extracted with the Wizard™ Genomic DNA Purification Kit, in accordance with the manufacturer's instructions (Promega, USA).

### Mitochondrial DNA amplification by long PCR

In order to sequence the complete mitogenome of *E. autonoe*, 500 - 700 bp of *E. autonoe* ND5 and srRNA genes (SF1 and SF2 in Figure 1, respectively) were initially sequenced. The primers for SF1 and SF2 were designed via the alignment of several insect mitogenomes sequenced in their entirety. Based on the sequence information, two pairs of primers specific to *E. autonoe* were designed to amplify two overlapping long fragments (LF1 and LF2 in Figure 1) using LA Taq™ (Takara Biomedical, Japan) under the following conditions: initial denaturation 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 10 min final extension at 72°C. The primer sequences for these short and long fragments are listed in Table 1. These PCR products were then utilized in the construction of a shotgun library. In brief, the DNAs were sheared into 1 - 5 kb fragments with a 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 with an ABI PRISM® Big-Dye® Terminator v 3.1 Cycle Sequencing Kit and an ABI PRISM™ 3100 Genetic Analyzer (PE Applied Biosystems, USA).

### Gene identification and tRNA structures

Sequences from overlapping fragments were assembled with the neighboring fragments using CLUSTAL X software (Thompson et al., 1997). Via comparison of the DNA or amino acid sequences with homologous regions of known full-length insect mitogenome sequences, 13 PCGs, two rRNA genes and the A + T-rich region were determined via the alignment of the 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 22 tRNA genes were identified by their proposed cloverleaf secondary structure and anticodon sequences with the aid of tRNAscan-SE 1.21 using invertebrate codon predictors and a cove score cut-off of 1 (Lowe and Eddy, 1997) the folding of the predicted tRNA sequences was further confirmed by visual inspection. The sequence data were deposited into the GenBank database under accession no. GQ868707.

### Comparative mitochondrial gene analyses

The A + T-content of each gene and the whole genome were calculated via DNA frequency analysis (http://kd.lab.nig.ac.jp/mishima/nucfreq1.html). Nucleotide composition at each codon position of the PCGs was calculated using PAUP ver. 4.0b10 (Swofford, 2002) software. Gene overlap and intergenic-space sequences were hand-counted. Nucleotide composition, termed "compositional skew" was calculated for the PCGs between two strands and the whole genome with the EditSeq program included in the Lasergene software package (www.dnastar.com) using the following formula proposed by Perna and Kocher (1995) GC-skew = (G-C)/ (G+C) and AT-skew = (A-T)/(A+T), where C, G, A and T are the frequencies of the four bases.

### Phylogenetic analysis

A phylogenetic reconstruction of relationships within the Lepidoptera was conducted on the basis of complete mitogenomes using Bayesian Inference (BI) and maximum likelihood (ML) algorithms. The alignment of the amino acid sequences of 13 individual PCGs from 17 sequenced lepidopteran PCGs with three dipteran out
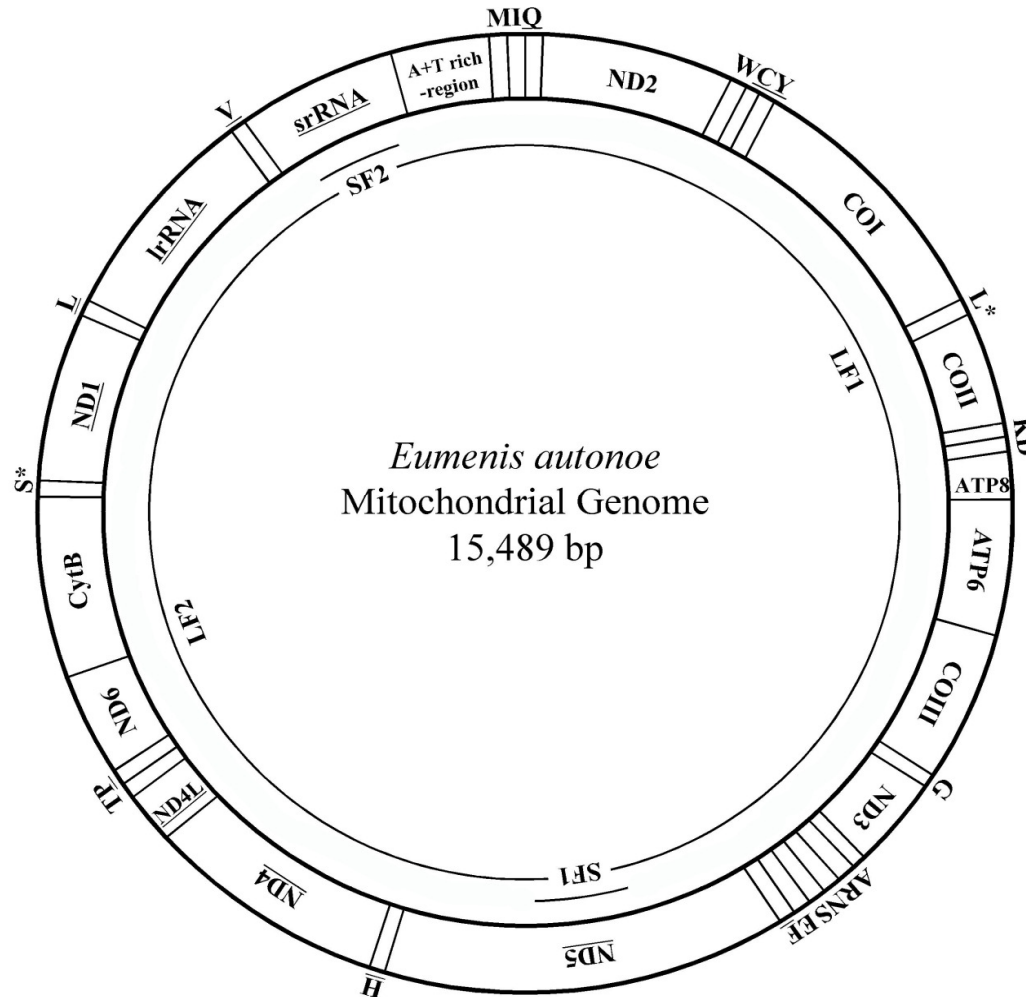
**Figure 1.** Circular map of the mitochondrial genome of *Eumenis autonoe*. COI, COII and COIII refer to the cytochrome oxidase subunits; CytB refers to cytochrome B; ATP6 and ATP8 refer to subunits 6 and 8 of F$_0$ ATPase; ND1 - 6 refer to components of NADH dehydrogenase. tRNAs are denoted as one-letter symbols consistent with the IUPAC-IUB single letter amino acid codes. Gene names that are not underlined indicate a clockwise transcriptional direction, whereas underlines indicate a counter-clockwise transcriptional direction. The *E. autonoe* mitogenome was sequenced by four overlapping fragments (SF1, SF2, LF1 and LF2), shown as single lines within a circle.

groups, including *Bactrocera oleae* (Nardi et al., 2003), *Drosophila yakuba* (Clary and Wolstenholme, 1985) and *Anopheles gambiae* (Beard et al., 1993) was constructed using CLUSTAL X software (Thompson et al., 1997) within BioEdit (Hall, 1999), under default conditions. On the other hand, the alignment of nucleotide sequences of 13 individual PCGs was constructed using RevTrans ver. 1.4 (Wernersson and Pedersen, 2003), which aligns coding sequences on the basis of protein alignments.

The well-aligned blocks from the amino acid and nucleotide sequences were selected with GBlocks 0.91b (Castresana, 2000) with the maximum number of contiguous non-conserved positions set to eight. These were subsequently concatenated into amino acid (3,482 sites, which is 90% of original sites) and nucleotide (10,608 sites, which is 91% of original sites) sequence alignments. This alignment is available upon request.

Substitution model selection was conducted via a comparison of Akaike Information Criterion (AIC) scores (Akaike, 1974), calculated using the ProTest program ver. 1.4 (Abascal et al., 2005) for amino

acid sequence alignment and Modeltest software 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 BI and ML analyses for amino acid sequence in the absence of a recently developed mtArt model (Abascal et al., 2007) in the MrBayes package. On the other hand, the GTR (Lanave et al., 1984) + I + G were selected as the best-fitting model for nucleotide sequences to be utilized in BI and ML analyses. The BI analyses were conducted using MrBayes ver. 3.1 (Huelsenbeck and Ronquist, 2001) under the following conditions: 1,000,000 generations, four chains (one cold chain and three hot chains) and a burn-in step of the first 10,000. The confidence values of the BI tree were expressed as the Bayesian posterior probabilities in percentages (BPP). The ML analysis was conducted using PHYML (Guindon et al., 2005) under the following conditions: the proportion of invariable sites as "estimated", the number of substitution rate categories as four, the gamma distribution parameter as "estimated" and the starting tree as a BIONJ distance-based tree. The

**Table 1.** List of primers used to amplify and sequence the mitogenome of *Eumenis autonoe*.

| Fragment name | Primer name | Direction | Location[a] | Sequence (5' → 3') |
|---|---|---|---|---|
| **Short fragments** | | | | |
| SF1 | Lep-ND5-F2 | F | 6962 | AAAGGAATTCCACATAAAGCTA |
| | Lep-ND5-R2 | R | 7583 | CCAATACGATTAGATAAAGCAG |
| SF2 | Lep-srRNA-F2 | F | 14063 | CTCTACTTTGTTACGACTTATT |
| | Lep-srRNA-R2 | R | 14635 | CACTGCTGGCACAAAATTTGTT |
| **Long fragments** | | | | |
| LF1 | EA-srRNA-F1 | F | 14389 | CTAAATGAACTAAAATACCGCC |
| | EA-ND5-R1 | R | 7118 | GCATGAGTTAATAAATGAAAATAAG |
| LF2 | EA-ND5-F2 | F | 7394 | GGAGCAGCTATAGCTGCAGGT |
| | EA-srRNA-R1 | R | 14250 | CTACATCTTGATCTGAATTAATTTCTTA |

[a] Locations are with respect to *Eumenis autonoe* mitogenome.

confidence values of the ML tree were evaluated via a bootstrap test with 100 iterations.

## RESULTS AND DISCUSSION

### General mitogenome features of *E. autonoe*

The complete mtDNA sequence of *E. autonoe* was 15,489 bp in length (Table 2). The complete mitogenome consists of 2 rRNAs, 22 tRNAs, 13 PCGs and one major non-coding A+T-rich region. As is the case in many insect mitogenomes, the major strand coded for a somewhat higher number of genes (9 PCGs and 14 tRNAs), whereas somewhat fewer genes were coded in the minor strand (4 PCGs, 8 tRNAs and 2 rRNA genes), as is shown in Figure 1. The genome size of 15,489 bp was well within the range detected in the completely sequenced lepidopteran insects, with sizes ranging from 15,140 in *Artogeia melete* to 15,928 in *Bombyx mandarina* (Table 3). *E. autonoe* harbors 3,728 codons, excluding termination codons and this number is most similar to those of *Antheraea pernyi* (3,732) and *Phthonandria atrilineata* (3,724) (Table 3). All individual *E. autonoe* mitochondrial genes were well within the range detected in the respective genes of other lepidopteran insects (data not shown).

Thus far, the complete mitogenome sequences of all lepidopteran insects, including *E. autonoe,* evidence identical orientation and gene order (Figure 1). However, this differs from the most common type that has been suggested as ancestral for insects (Boore et al., 1998). The difference between the two involves the movement of tRNA$^{Met}$ to a position 5'-upstream of tRNA$^{Ile}$, which resulted in the following order: tRNA$^{Met}$, tRNA$^{Ile}$ and tRNA$^{Gln}$.

### Nucleotide composition and base bias

The nucleotide composition of the mitogenome of *E.*

*autonoe* is also biased toward A+T content at 79.1%, as is the case with other lepidopteran mitogenome sequences (Table 3). This value is well within the range found in the sequenced lepidopteran insects, where these values range from 77.9% in *Ochrogaster lunifer* to 82.7% in *Coreana raphaelis* (Table 3). The composition of the PCGs of *E. autonoe* was A, 32.3%; T, 44.5%; C, 11.5% and G, 11.5% (Table 4). Both guanine and cytosine were quite rare and accounted for only 23.2% (AT at 76.8%). This value is well within the range detected in the sequenced lepidopteran insects, but was the second highest of these values, next to *O. lunifer* (24.4%) among the sequenced lepidopteran insects (Table 4). The analysis of the base composition at each codon position of the concatenated 13 PCGs of *E. autonoe* showed that the third codon position (88.1%) harbored a higher A+T content than that of the first (72.2%) and second (69.9%) codon positions and a similar pattern was also detected in other sequenced lepidopteran species (Table 4).

To evaluate the degree of the base bias, the base-skew was measured and it was determined that AT-skew and GC-skew in the whole genome of *E. autonoe* (measured from the major strand) were -0.016 and -0.243, respectively (Table 5), thereby indicating that more Ts and more Cs are encoded in the entire mitogenome of *E. autonoe*. In other lepidopteran insects, the AT-skew values ranged between -0.047 (*C. raphaelis*) and 0.059 (*Bombyx mori*). Thus, the frequency of adenine varied only slightly from species to species in the Lepidoptera, in that the values varied within ± 0.1 (Table 5). In most other insects, a slight A-skew has been reported and indeed, the only exceptional cases that can be referenced were *Reticulitermes* (Isoptera) at 0.30 and *Locusta migratoria* (Orthoptera) at 0.18 (Cameron and Whiting, 2007; Flook et al., 1995). Thus, lepidopteran insects are typical in terms of AT-skew in the whole genome among insects. Unlike the AT-skew estimates, GC-skew estimates, including those of *E. autonoe* (-0.243) in sequenced lipidopteran species were all negative with somewhat large

**Table 2.** Summary of mitogenome of *Eumenis autonoe*.

| Gene | Direction | Nucleotide number | Size | Anticodon | Start codon | Stop codon |
|---|---|---|---|---|---|---|
| tRNA$^{Met}$ | F | 1 - 67 | 67 | 32 - 34 CAT | - | - |
| tRNA$^{Ile}$ | F | 68 - 132 | 65 | 98 - 100 GAT | - | - |
| tRNA$^{Gln}$ | R | 130 - 198 | 69 | 166 - 168 TTG | - | - |
| ND2 | F | 249 - 1262 | 1014 | - | ATT | TAA |
| tRNA$^{Trp}$ | F | 1261 - 1327 | 67 | 1292 - 1294 TCA | | |
| tRNA$^{Cys}$ | R | 1320 - 1384 | 65 | 1352 - 1354 GCA | | |
| tRNA$^{Tyr}$ | R | 1385 - 1448 | 64 | 1415 - 1417 GTA | | |
| CO I | F | 1457 - 2992 | 1536 | - | CGA | TAA |
| tRNA$^{Leu}$(UUR) | F | 2988 - 3054 | 67 | 3018 - 3020 TAA | | |
| CO II | F | 3055 - 3730 | 676 | - | ATG | T (tRNA) |
| tRNA$^{Lys}$ | F | 3731 - 3801 | 71 | 3761 - 3763 CTT | | |
| tRNA$^{Asp}$ | F | 3803 - 3868 | 66 | 3833 - 3834 GTC | | |
| ATP8 | F | 3869 - 4033 | 165 | - | ATC | TAA |
| ATP6 | F | 4027 - 4704 | 678 | - | ATG | TAA |
| CO III | F | 4704 - 5492 | 789 | - | ATG | TAA |
| tRNA$^{Gly}$ | F | 5495 - 5560 | 66 | 5525 - 5527 TCC | | |
| ND3 | F | 5561 - 5914 | 354 | - | ATC | TAA |
| tRNA$^{Ala}$ | F | 5915 - 5980 | 66 | 5945 - 5947 TGC | - | - |
| tRNA$^{Arg}$ | F | 5981 - 6042 | 62 | 6008 - 6010 TCG | - | - |
| tRNA$^{Asn}$ | F | 6043 - 6109 | 67 | 6074 - 6076 GTT | - | - |
| tRNA$^{Ser}$(AGN) | F | 6107 - 6166 | 60 | 6127 - 6129 GCT | - | - |
| tRNA$^{Glu}$ | F | 6169 - 6234 | 66 | 6199 - 6201 TTC | - | - |
| tRNA$^{Phe}$ | R | 6233 - 6296 | 64 | 6264 - 6266 GAA | - | - |
| ND5 | R | 6271 - 8034 | 1764 | - | ATC | T (tRNA) |
| tRNA$^{His}$ | R | 8032 - 8097 | 66 | 8065 - 8067 GTG | - | - |
| ND4 | R | 8098 - 9436 | 1339 | - | ATG | T (tRNA) |
| ND4L | R | 9436 - 9726 | 291 | - | ATG | TAA |
| tRNA$^{Thr}$ | F | 9729 - 9792 | 64 | 9759 - 9761 TGT | | |
| tRNA$^{Pro}$ | R | 9793 - 9857 | 65 | 9825 - 9827 TGG | | |
| ND6 | F | 9860 - 10387 | 528 | - | ATT | TAA |
| CytB | F | 10387 - 11538 | 1152 | - | ATG | TAA |
| tRNA$^{Ser}$(UCN) | F | 11547 - 11624 | 68 | 11588 - 11590 TGA | - | - |
| ND1 | R | 11641 - 12569 | 939 | - | ATG | TAA |
| tRNA$^{Leu}$(CUN) | R | 12571 - 12637 | 67 | 12606 - 12608 TAG | - | - |
| lrRNA | R | 12638 - 13972 | 1335 | - | - | - |
| tRNA$^{Val}$ | R | 13973 - 14036 | 64 | 14005 - 14007 TAC | - | - |
| srRNA | R | 14037 - 14811 | 775 | - | - | - |
| A+T-rich region | R | 14812 - 15489 | 678 | - | - | - |

tRNA abbreviations follow the IUPAC-IUB three letter code.

values, ranging from -0.318 (*O. lunifer*) to -0.158 (*C. raphaelis*), thereby indicating that more Cs were clearly present in lepidopteran species (Table 5). Such a trend has also been noted in previous studies on Lepidoptera (Jiang et al., 2009) and Diptera (Junqueria et al., 2004), although the reason for this observed skewness remains to be clearly elucidated.

With regard to PCGs, the degree of base bias was calculated in different strands of the *E. autonoe* mito-genome (Table 5). The major strand that encodes for 9

PCGs (ND2, ND3, ND6, COI, COII, COIII, ATP6, ATP 8and CytB) exhibited T-skew at -0.136, whereas the minor strand which encodes for only 4 PCGs (ND1, ND4, ND4Land ND5) evidenced T-skew at - 0.191 (Table 5). Thus, the major strand which encodes for more PCGs is considered to be less biased for T-skew. Similarly, all other lepidopteran species evidenced profound T-skew in the minor strand as compared to the major strand. With regard to GC-skew, the major strand PCGs of *E. autonoe* evidenced a value of -0.184, whereas those of the minor

**Table 3.** Characteristics of the lepidopteran mitogenomes.

| Taxon | Size (bp) | A+T content (%) | No. codons[a] | PCG[b] A+T content (%) | lrRNA Size (bp) | lrRNA A+T % | srRNA Size (bp) | srRNA A+T % | A+T-rich region Size (bp) | A+T-rich region A+T % | GenBank accession no. | References |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **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 | 3,714 | 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) |
| *Eriogyna pyretorum* | 15,327 | 80.8 | 3,711 | 79.3 | 1,338 | 84.6 | 762 | 84.5 | 374 | 91.7 | FJ685653 | Jiang et al. (2009) |
| **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 | Kim et al. (2009) |
| **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. (2009) |
| **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) |
| **Nymphalidae** | | | | | | | | | | | | |
| *Eumenis autonoe* | 15,489 | 79.1 | 3,728 | 76.8 | 1,335 | 83.7 | 775 | 85.3 | 678 | 94.5 | GQ868707 | This study |
| **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) |
| **Lymantriidae** | | | | | | | | | | | | |
| *Lymantria dispar* | 15,569 | 79.9 | 3,742 | 77.8 | 1,351 | 84.2 | 799 | 85.2 | 435 | 96.1 | FJ617240 | Unpublished |

**Table 3.** Contd…

| 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.

**Table 4.** 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 |
| *Eriogyna pyretorum* | 36.1 | 37.1 | 10.4 | 16.3 | 21.7 | 48.9 | 16.2 | 13.0 | 41.4 | 52.8 | 3.5 | 2.1 | 33.0 | 46.3 | 10.5 | 10.0 |
| *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 |
| *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 |
| *Eumenis autonoe* | 36.0 | 36.2 | 11.2 | 16.3 | 21.4 | 48.5 | 16.3 | 13.6 | 39.5 | 48.6 | 7.1 | 4.5 | 32.3 | 44.5 | 11.5 | 11.5 |
| *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 |
| *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 |
| *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 |
| *Lymantria dispar* | 36.5 | 36.4 | 10.3 | 16.6 | 22.0 | 48.1 | 16.2 | 13.5 | 40.8 | 49.2 | 6.2 | 3.6 | 33.1 | 44.6 | 10.9 | 11.3 |
| *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.

minor strand exhibited a value of 0.342. Thus, the frequency of Gs and Cs differ profoundly depending on the strand. This trend is also identical in other lepidopteran species (Table 5). Thus, mutational pressures that favor Gs and Ts are more severe on the minor-strand PCGs, although such strand-based inequalities base frequencies are yet to be clearly understood.

## PCGs

Table 2 shows the start and stop codons of the 13 PCGs in the mitogenome of *E. autonoe*. Among 13 PCGs, 7 PCGs begin with ATG, 2 with ATT and 3 with ATC. However, in the COI gene, no canonical ATN initiator was detected in the start region of the COI gene (Figure 2). The only plausible traditional start codon for the COI gene is ATC, which was located 19 bp inside the 5' end of the tRNA[Tyr] gene. However, this ATC sequence required eight to nine additional amino acids,

**Table 5.** Composition and skewness in the lepidopteran mitogenomes.

| Species | Major- strand PCGs | | | Minor- strand PCGs | | | Whole PCGs* | | | Whole genome* | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | A+T % | AT skew | GC skew | A+T % | AT skew | GC skew | A+T % | AT skew | GC skew | A+T % | AT skew | GC skew |
| *Adoxophyes honmai* | 77.05 | - 0.117 | - 0.123 | 80.53 | - 0.201 | 0.296 | 78.41 | - 0.150 | 0.024 | 80.39 | - 0.001 | - 0.196 |
| *Antheraea pernyi* | 76.96 | - 0.155 | - 0.146 | 80.89 | - 0.187 | 0.311 | 78.47 | - 0.168 | 0.011 | 80.16 | - 0.021 | - 0.216 |
| *Antheraea yamamai* | 77.60 | - 0.153 | - 0.152 | 80.95 | - 0.182 | 0.324 | 78.89 | - 0.164 | 0.014 | 80.29 | - 0.022 | - 0.220 |
| *Caligula boisduvalii* | 78.01 | - 0.152 | - 0.132 | 80.84 | - 0.187 | 0.328 | 79.10 | - 0.166 | 0.031 | 80.62 | - 0.024 | - 0.217 |
| *Eriogyna pyretorum* | 77.38 | - 0.163 | - 0.122 | 82.00 | - 0.171 | 0.256 | 79.39 | - 0.166 | 0.022 | 80.82 | - 0.031 | - 0.204 |
| *Bombyx mandarina* | 78.64 | 0.446 | - 0.128 | 81.09 | - 0.269 | 0.331 | 79.58 | - 0.136 | 0.035 | 81.68 | 0.055 | - 0.213 |
| *Bombyx mori* | 78.46 | - 0.045 | - 0.132 | 81.19 | - 0.271 | 0.325 | 79.51 | - 0.134 | 0.030 | 81.32 | 0.059 | - 0.216 |
| *Manduca sexta* | 79.18 | - 0.113 | - 0.085 | 81.95 | - 0.193 | 0.313 | 80.24 | - 0.144 | 0.054 | 81.79 | - 0.005 | - 0.181 |
| *Phthonandria atrilineata* | 78.05 | - 0.101 | - 0.107 | 80.66 | - 0.198 | 0.305 | 79.05 | - 0.139 | 0.039 | 81.02 | 0.007 | - 0.192 |
| *Parnassius bremeri* | 78.85 | - 0.126 | - 0.110 | 82.19 | - 0.166 | 0.284 | 80.14 | - 0.142 | 0.026 | 81.27 | - 0.011 | - 0.194 |
| *Artogeia melete* | 77.66 | - 0.098 | - 0.149 | 79.73 | - 0.211 | 0.324 | 78.46 | - 0.142 | 0.022 | 79.78 | 0.012 | - 0.222 |
| *Coreana raphaelis* | 80.24 | - 0.142 | - 0.071 | 83.42 | - 0.115 | 0.273 | 81.46 | - 0.131 | 0.047 | 82.66 | - 0.047 | - 0.158 |
| *Eumenis autonoe* | 75.36 | - 0.136 | - 0.184 | 79.27 | - 0.191 | 0.341 | 76.87 | - 0.158 | - 0.003 | 79.09 | - 0.016 | - 0.243 |
| *Ochrogaster lunifer* | 74.39 | - 0.089 | - 0.236 | 77.69 | - 0.226 | 0.418 | 75.67 | - 0.144 | - 0.005 | 77.84 | 0.030 | - 0.318 |
| *Lymantria dispar* | 76.56 | - 0.105 | - 0.170 | 79.68 | - 0.212 | 0.361 | 77.76 | - 0.147 | 0.016 | 79.88 | 0.016 | - 0.248 |
| *Ostrina furnacalis* | 78.58 | - 0.108 | - 0.064 | 81.31 | - 0.217 | 0.309 | 79.62 | - 0.150 | 0.065 | 80.37 | 0.032 | - 0.194 |
| *Ostrina nubilalis* | 77.97 | - 0.079 | - 0.107 | 80.91 | - 0.216 | 0.305 | 79.10 | - 0.133 | 0.038 | 80.17 | 0.032 | - 0.192 |

Stop codon was excluded in the count.
*The skewness of whole PCGs and whole genome was calculated from major strand.

resulting in a peculiar alignment compared with other lepidopteran insects. Thus, the ATC sequence may not be the start codon for the COI gene. Rather, GCA (arginine) was designated as a start codon for lepidopteran insects, as has been suggested previously (Kim et al., 2009). This codon is present as a highly conserved region throughout all sequenced lepidopteran insects, including *E. autonoe* (Figure 2). Furthermore, it has been previously demonstrated that the CGA is well conserved in 39 lepidopteran insect species, encompassing eight families. Thus, it has been previously suggested that the sequence may

be functionally constrained and may represent a synapomorphic characteristic for Lepidoptera (Kim et al., 2009). TAA functions as a stop codon in 10 genes, but three of the 13 PCGs harbored an incomplete stop codon consisting of a single thymine (COII, ND5 and ND4). Such an incomplete stop codon is frequently detected in insect mitogenomes (Cha et al., 2007; Hong et al., 2008; Kim et al., 2009). This phenomenon was caused by post-transcriptional modifications occurring during the mRNA maturation process, such as polyadenylation (Anderson et al., 1981; Ojala et al., 1981), but more experimental data will

be required for a further and more decisive, conclusion.


**rRNA and tRNA genes**

The lrRNA and srRNA genes of the *E. autonoe* mitogenome were 1,335 and 775 bp in length, respectively. As has been observed in other insects, including lipidopteran insects (Boore et al., 1998; Kim et al., 2009), these genes are located between tRNA[Leu] (CUN) and tRNA[Val] and between tRNA[Val] and the A+T-rich region,

← tRNA^Tyr                                           CO1 →

```
Adoxophyes honmai    aatttatcgcttattaactcagccatttttattaaaattata  tag  CGA AAA TGA CTT  RKWL
Antheraea pernyi     atttacaatttatcgctttaaacctcagccatttttattaat  tag  CGA AAA TGA ATT  RKWI
Antheraea yamamai    tttacaatttatcgcttattaacctcagccatttttattaat  tag  CGA AAA TGA CTT  RKWL
Caligula boisduvalii atttattaacaagaaaattctttttttattaaattatttaat  ttg  CGA AAA TGA CTA  MRKWL
Eriogyna pyretorum   taaatttacaatttatcgcttattaactcagccattttatt   tag  CGA AAA TGA CTA  RKWL
Bombyx mandarina     tatcgcttattcttcagccatttttatttacttactttttat  tag  CGA AAA TGA ATT  RKWI
Bombyx mori          aaaatcgcttattattcagccatttttatttttttttcttttt tag  CGA AAA TGA ATT  RKWI
Manduca sexta        taaatttacaatttatcgcttaaaactcagccattttattt   ttg  CGA AAA TGA CTT  RKWL
Phthonandria atrilineata ataaatttacaatttatcgcttaaatctcagccattttatt ttg  CGA AAA TGA CTA  MRKWL
Parnassius bremeri   aataaatttacaatttatcgcttatactcagccattttatt   tag  CGA AAA TGA CTA  RKWL
Artogeia melete      aataaatttacaatttatcgcttataactcagccattttat   tag  CGA AAA TGA CTT  RKWL
Coreana raphaelis    ataaatttacaatttatcgcttatatctcagccattttatt   tag  CGA AAA TGA CTT  RKWL
Eumenis autonoe      tttacaatttatcgcttatccctcagccatttttactatttt  aag  CGA AAA TGA CTT  RKWL
Ochrogaster lunifer  tttacaatttatcgcctaccctcagccatttttattattttt  tag  CGA AAA TGA CTA  RKWL
Lymantria dispar     ttacaatttaccgcttaaactcagccattttattaaaaaat   aag  CGA AAA TGA TTA  RKWL
Ostrinia furnacalis  ttacaatttatcgcttaaatctcagccattttatttttaatt  tag  CGA AAA TGA CTA  RKWL
Ostrinia nubilalis   ttacaatttatcgcttaaatctcagccattttattttaatt   tag  CGA AAA TGA CTA  RKWL
```

**Figure 2.** Alignment of the initiation context of the COI genes of lepidopteran insects, including that of *Eumenis autonoe*. The first four or five codons for COI and their amino acids are shown on the right-hand side of the figure. Underlined nucleotides indicate the adjacent partial sequence of tRNA^Tyr. Arrows indicate the transcriptional direction. Boxed nucleotides indicate the currently proposed translation initiators for the COI gene of lepidopteran insects. The start codon for *E. autonoe* was designated as CGA.

respectively. The A+T content of the lrRNA and srRNA genes is 83.7 and 85.3%, respectively, which is consistent with the A+T content observed for these genes in other lepidopteran insects (Table 3).

A total of 22 tRNA genes (one specific for each amino acid and two for leucine and serine) were identified within the mitogenome. The tRNAs were interspersed throughout the mitogenome and ranged in length from 60 to 71 bp (Table 2). All tRNAs but tRNA^Ser(AGN) were shown to be folded into the expected cloverleaf secondary structures (Figure 3). The unusual tRNA^Ser (AGN) lacked the DHU loop (Figure 3). This incomplete tRNA^Ser (60 bp) structure has been detected in the mitogenomes of other animals, including insects (Wolstenholme, 1992).

A total of 26 unmatched base pairs were detected in the *E. autonoe* mt tRNAs, but 20 of them were G-U pairs, which form a weak bond in the tRNAs. The remaining six were atypical pairings: one mismatch in the tRNA^Ala (U-U), 2 in the tRNA^Leu (UUR) (one U-U and one A-C), one in the tRNA^Lys (C-U) and 2 -tRNA^Ser (UCN) (2 U-U) (Figure 3). This number of mismatches in the *E. autonoe* tRNAs is well within the number reported for other lepidopteran insect tRNAs: 11 in *O. lunifer* (Salvato et al., 2008), 8 in *C. raphaelis* (Kim et al., 2006), 2 in *Eriogyna pyretorum* (Jiang et al., 2009) and one in *P. atrilineata* (Yang et al., 2009). The postulated tRNA cloverleaf structure harbored an invariable 7 bp in the aminoacyl stem, 5 bp in the anticodon stem and 7 bp in the anticodon loop, but also contained a variable length of the didydrouridine (DHU) arm and the TΨC arm, particularly within the loops.

**Intergenic spacer sequences and overlapping sequences**

The genes of *E. autonoe* were interleaved with a total of 139 bp, which were spread over ten regions, which ranged in size between 1 and 50 bp. The majority of intergenic spacer sequences were short (1 - 2 bp), but two locations have relatively long intergenic spacer sequences. The longest one, which is located between tRNA^Gln and ND2 in the *E. autonoe* mitogenome (50 bp) is consistently detected with similar sizes in the sequenced lepidopteran insects, ranging in size from 40 bp in *Parnassius bremeri* to 72 bp in *Ochrogaster lunifer* (Table 6). Nevertheless, other holometabolous insects such as Coleoptera, Diptera and Neuroptera, do not harbor such a long intergenic spacer sequence between the ND2 gene and the neighboring tRNA^Met gene. Instead, only a very short intergenic spacer sequence is present in a few insects, such as the coleopterans *Tribolium castaneum* (Friedrich and Muqim, 2003) and *Hydroscapha granulum* (Unpublished, GenBank accession number AM493667) and the dipterans *Mayetiola destructor* (Unpublished, GenBank accession number NC013066), *Culicoides arakawae* (Unpublished, GenBank acces-
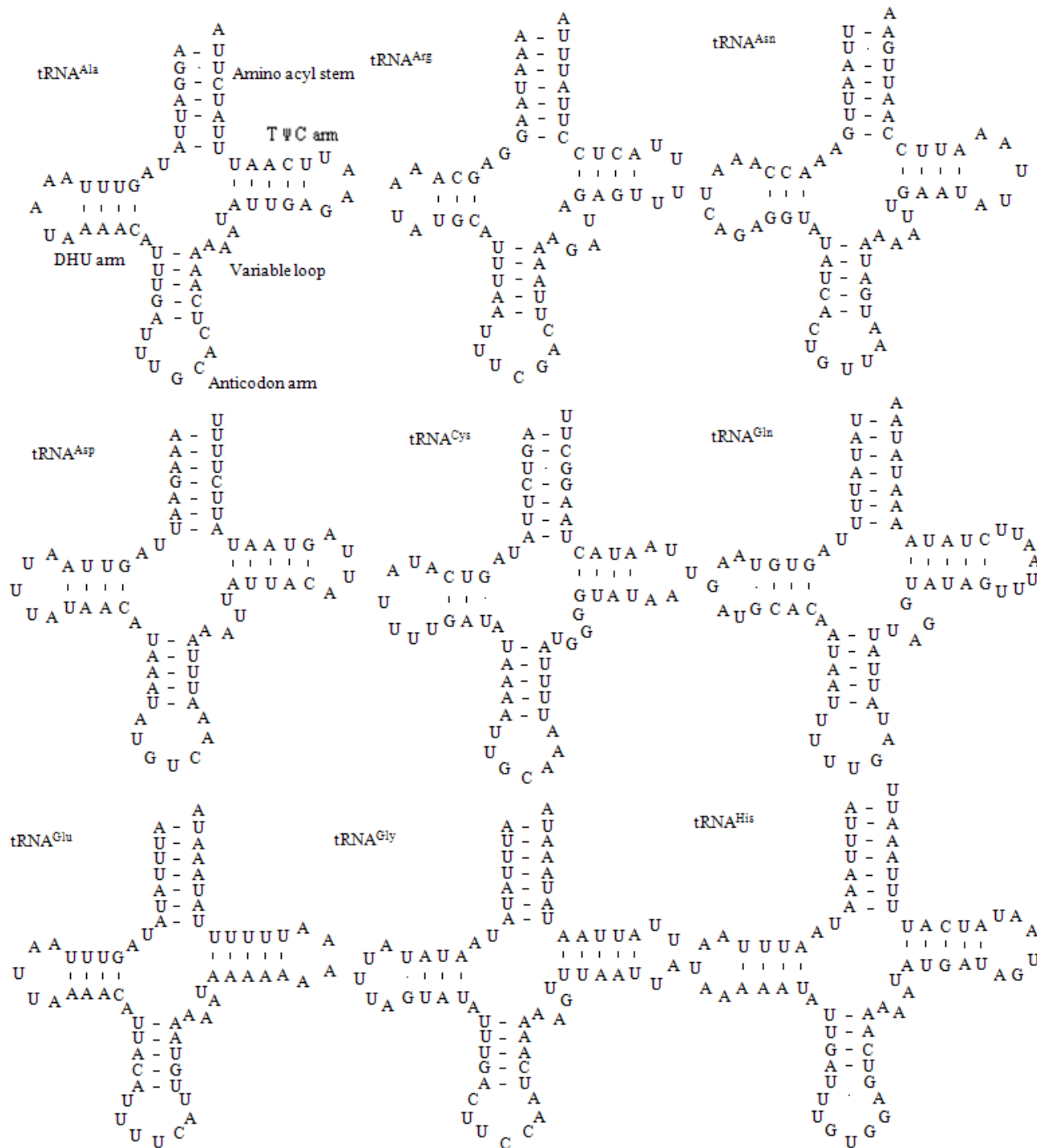
**Figure 3.** Predicted secondary cloverleaf structures for the 22 tRNA genes of *Eumenis autonoe*. The tRNAs are labeled with the abbreviations of their corresponding amino acids. Nucleotide sequences from 5' to 3' are indicated for tRNA^Ala. Dashes (-) indicate Watson-Crick base-pairing and centered 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.

sion number AB361004) and *Ceratitis capitata* (Spanos et al., 2000) ranging in size from 1 to 5 bp. Thus, this intergenic spacer sequence appeared to be synapomorphic in Lepidoptera. More interestingly, the sequence alignment of this intergenic spacer sequence to the neighboring ND2 gene revealed a sequence homology of 74% in *E. autonoe* (Figure 4). Previously, Kim et al. (2009) also detected substantially high sequence homo-
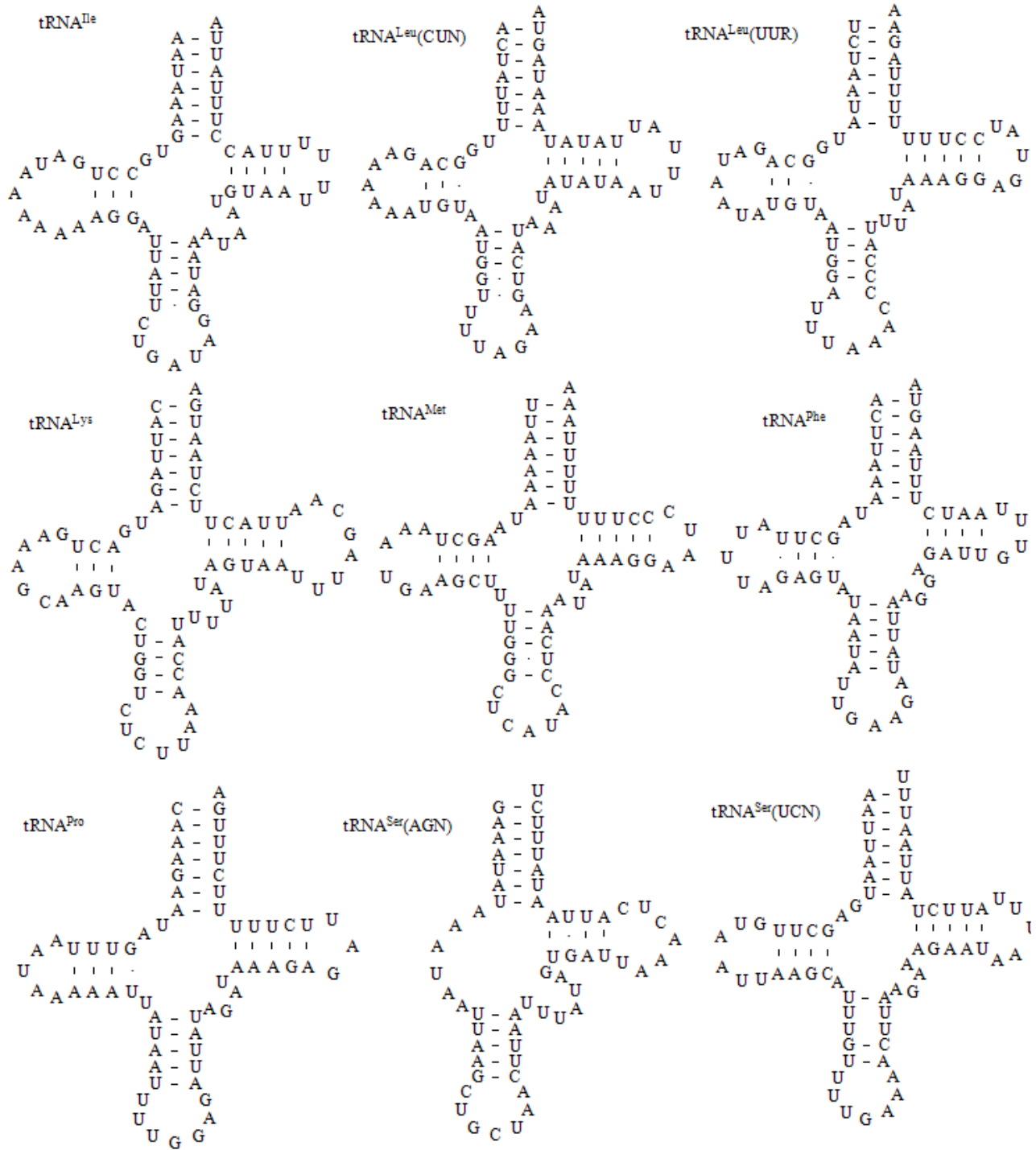
**Figure 3.** Contd…

logy between the intergenic spacer sequence and the neighboring ND2 from several lepidopteran insects, including *A. melete* (70%), *C. raphaelis* (62%), *Coligula boisduvalii* (62%), *B. mori* (63%) and *P. atrilineata* (70%). Careful analysis of this region from the newly sequenced lepidopteran insect, *Lymantria dispar* (Unpublished, Gen Bank accession number FJ617240) also evidenced substantially high sequence homology at 60% (Figure 4). This indicated that the spacer sequence may have originated from a partial duplication of the ND2 gene, but the non-coding nature of this region may have allowed for a rapid sequence divergence from the original ND2 gene.

Other longer intergenic spacer sequences included the 16-bp long sequence located between the tRNA[Ser] (UCN)
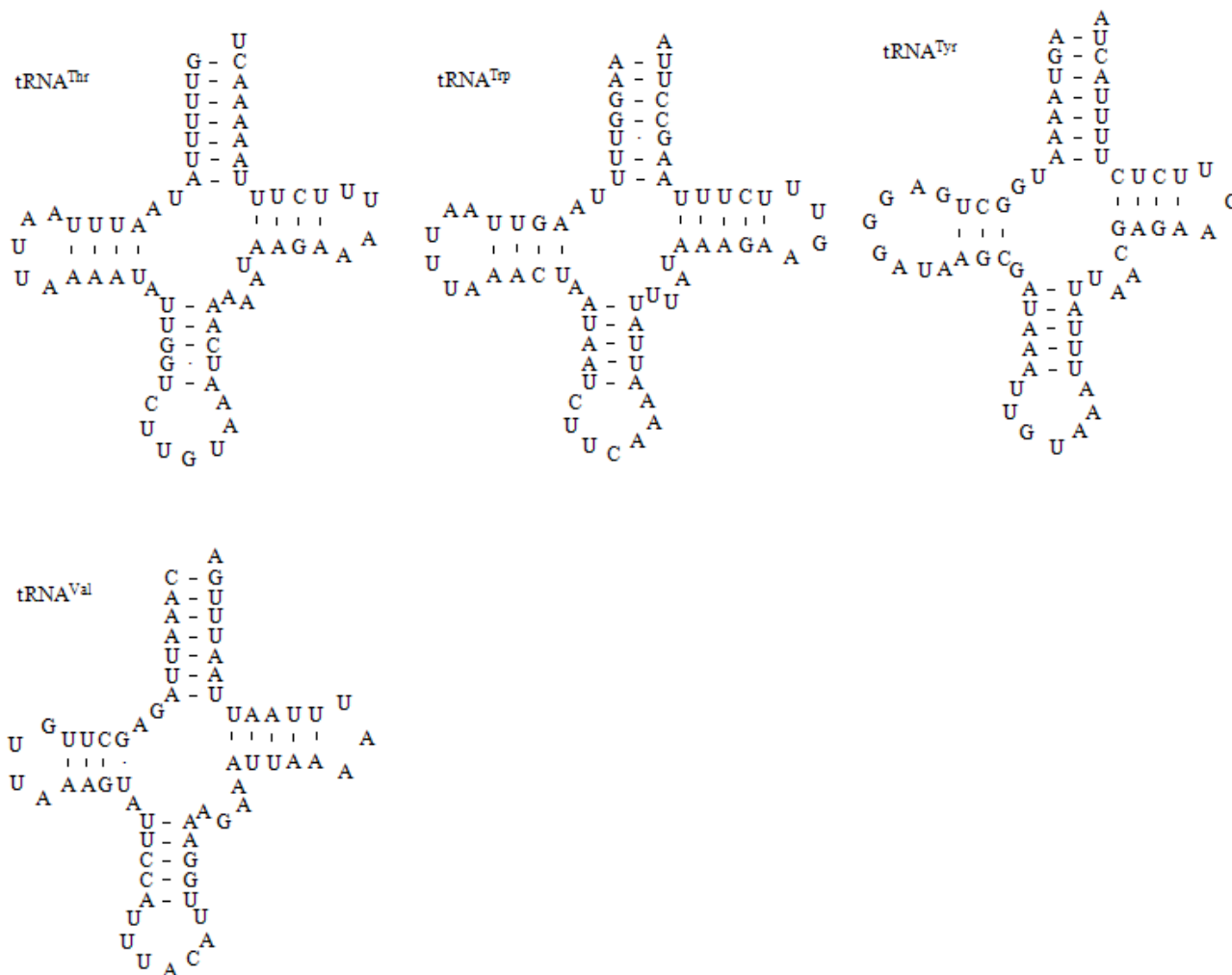
**Figure 3.** Contd

and ND1 of the mitogenome of *E. autonoe* (Table 6). This intergenic spacer sequence was also detected in all sequenced lepidopteran insects, ranging in size from 16 bp in *P. bremeri*, *A. melete* and the current *E. autonoe* to 38 bp in *Ostrinia nubilalis* (Table 6). Within this 16-bp long intergenic spacer sequence exists the 7-bp long ATACTAA motif, which is conserved in all lepidopteran species thus far sequenced (Figure 5). This 7-bp long motif has been suggested to be a possible mitochondrial transcription termination peptide binding site, in that the intergenic spacer sequence is located just past the final PCG, CytB, within the major strand (Cameron and Whiting, 2008; Taanman, 1999).

The *E. autonoe* mitochondrial genes overlap in a total of 51 bp at 11 locations, with the longest overlap measuring 16 bp, which was located between tRNA[Phe] and ND5. Similarly-sized overlapping sequences were also detected between tRNA[Phe] and ND5 in *A. pernyi*, *C. boisduvalii*, *E. pyretorum*, *P. bremeri*, *Ostrinia furnacalis* and *O. nubilalis* and a somewhat large one, 29 bp, was

detected in *Adoxophyes honmai* (Table 6).

## A+T-rich region

The A+T-rich region of the *E. autonoe* mitogenome was located between the srRNA and tRNA[Met] (Figure 1 and Table 2) and exhibits the highest A + T contents (94.5%) of any region of the *E. autonoe* mitogenome (Table 3). The 678 -bp long A + T-rich region of the *E. autonoe* mitogenome is the second longest among the completely sequenced lepidopteran insects, after the 747 bp long *B. mandarina* (Yukuhiro et al., 2002). This region comprised of a tandem repeat consisting of 10 duplicated and identical 27 bp copies and one partial copy of 13 bp, lacking 14 bp of the end portion of the 27 bp copy (Figure 6A). This repeat sequence consists of 26 A + T nucleotides and one C nucleotide, providing a very high A+T content (96.3%).

The presence of a tandem repeat in the mitochondrial

**Table 6.** Overlapping and intergenic-space sequences of the lepidopteran mitogenomes.

| Region | Ahon | Aper | Ayam | Cboi | Epyr | Bman | Bmor | Msex | Patr | Pbre | Amel | Crap | Eaut | Olun | Ldia | 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)3 | (O)4 | (O)3 | (O)3 | (O)3 | (O)3 | | (I)4 | | |
| tRNA$^{Gln}$ - ND2 | (I)64 | (I)56 | (I)53 | (I)53 | (I)63 | (I)47 | (I)47 | (I)54 | (I)63 | (I)40 | (I)48 | (I)56 | (I)50 | (I)72 | (I)56 | (I)62 | (I)62 |
| ND2 - tRNA$^{Trp}$ | (I)1 | (I)8 | (I)7 | (I)8 | (I)8 | (I)5 | (I)5 | | (I)11 | (O)1 | (O)2 | (I)2 | (O)2 | (I)1 | | (I)7 | (I)7 |
| tRNA$^{Trp}$ - tRNA$^{Cys}$ | (O)6 | (O)8 | (O)8 | (O)8 | (O)8 | (O)8 | (O)7 | (O)8 | (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)6 | (I)9 | (I)10 | (I)11 | (I)4 | | | | (I)19 | (I)12 | (I)3 | (I)3 |
| tRNA$^{Tyr}$ - CO I | (I)7 | (I)6 | (I)2 | (I)11 | (I)12 | (I)17 | (I)17 | (I)4 | (I)3 | (I)2 | (I)2 | (I)3 | (I)8 | (I)9 | (I)10 | (I)11 | (I)11 |
| COI - tRNA$^{Leu}$(UUR) | | | (O)5 | (O)6 | | | | | | (O)5 | | (I)1 | (O)5 | | | (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 | (I)23 | (O)1 | | | (I)21 | | (O)1 | (O)1 | (I)13 | (I)1 | (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)7 | (O)8 | (O)7 | (O)7 | (O)7 | (O)7 | (O)7 | (O)7 | (O)7 | (O)7 |
| ATP6 – COIII | (I)3 | (O)1 | (O)1 | | (I)10 | (I)12 | (I)14 | (I)18 | (O)1 | (O)1 | (I) 8 | | (O)1 | (I)3 | (I)11 | (O)1 | (O)1 |
| COIII - tRNA$^{Gly}$ | (I)2 | (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 | (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)4 | (I)12 | (I)10 |
| tRNA$^{Ala}$ - tRNA$^{Arg}$ | (I)1 | (O)1 | (O)1 | (I)12 | (I)7 | (I)50 | (I)4 | (I)1 | (O)1 | (O)1 | (O)1 | (I)1 | | (I)7 | (I)18 | (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)13 | (I)1 | (O)1 |
| tRNA$^{Asn}$ - tRNA$^{Ser}$(AGN) | (I)3 | (I)1 | (I)2 | (O)1 | (O)1 | (O)1 | | (I)3 | (O)1 | (I)3 | | (I)2 | (O)3 | (I)1 | (O)1 | (I)1 | (I)1 |
| tRNA$^{Ser}$(AGN) - tRNA$^{Ser}$(AGN)$^{†}$ | | | | | | | | | | | | (I)11 | | | | | |
| tRNA$^{Ser}$(AGN) - tRNA$^{Glu}$ | | (I)1 | (I)1 | (I)6 | (O)1 | (I)10 | (I)7 | | (O)1 | (I)43 | (I)15 | (I)10 | | (I)6 | (I)20 | | |
| tRNA$^{Glu}$ - tRNA$^{Phe}$ | (I)222 | (I)10 | (I)6 | (I)2 | (I)13 | (O)1 | | (O)2 | (O)2 | (O)2 | (I)1 | (O)2 | (O)2 | (I)70 | (I)2 | | |
| tRNA$^{Phe}$ - ND5 | (O)29 | (O)15 | | (O)17 | (O)17 | (I)4 | (I)5 | | (O)15 | (I)1 | (I)2 | | (O)16 | | (I)21 | (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 | (O)3 | | | (I)15 | (I)15 |
| tRNA$^{His}$ - ND4 | (O)61 | (I)9 | (I)12 | (I)1 | (I)1 | (I)59 | (I)47 | | | (O)1 | (O)1 | | | | | (O)6 | (O)6 |
| ND4 - ND4L | (O)1 | (I)5 | (I)3 | | (I)2 | (O)1 | (O)1 | (I)83 | (I)36 | (O)1 | (O)4 | | (O)1 | (I)2 | (I)22 | (I)7 | (I)7 |
| ND4L - tRNA$^{Thr}$ | (I)1 | (I)7 | (I)5 | (I)6 | (I)8 | (I)4 | (I)4 | (I)13 | (I)2 | (I)2 | (O)8 | (I)14 | (I)2 | (I)2 | (I)12 | (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)2 | (I)1 | (I)2 | (I)2 | (I)2 | | (I)2 | (I)7 | (I)10 | (I)2 | |
| ND6 - CytB | (I)18 | (I)2 | (O)1 | (I)4 | (O)1 | (I)55 | (I)50 | (I)17 | (I)1 | (I)17 | (O)1 | | (I)1 | | (I)49 | (I)2 | (I)2 |
| CytB - tRNA$^{Ser}$(UCN) | (O)1 | (I)15 | (I)24 | (I)41 | (I)2 | (O)1 | (I)3 | (I)21 | (O)1 | (I)1 | | | (I)8 | (I)9 | (I)2 | (O)1 | (O)1 |
| tRNA$^{Ser}$(UCN) - ND1 | (I)25 | (I)21 | (I)24 | (I)22 | (I)25 | (I)25 | (I)25 | (I)31 | (I)18 | (I)16 | | (I)16 | (I)19 | (I)16 | (I)17 | (I)34 | (I)38 |
| ND1 - tRNA$^{Leu}$(CUN) | (I)2 | (I)2 | (I)1 | (I)1 | (I)1 | (O)6 | (I)6 | | (I)6 | (I)1 | (I)3 | (I)1 | (I)1 | (I)1 | | (I)10 | (I)10 |
| tRNA$^{Leu}$(CUN) - lrRNA | (I)2 | | | | (I)19 | | | | | | | | | (I)48 | | (O)1 | (O)1 |

**Table 6.** Contd..

| lrRNA - tRNA$^{Val}$ | | | | | | (O)1 | | | | | | | | | (I)15 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| tRNA$^{Val}$ - srRNA | (I)2 | | | | | | | | | | | | | | | | |
| Total nucleotides | (O)106 (I)400 | (O)37 (I)234 | (O)36 (I)178 | (O)42 (I)194 | (O)38 (I)204 | (O)33 (I)373 | (O)18 (I)321 | (O)21 (I)375 | (O)42 (I)192 | (O)35 (I)138 | (O)39 (I)117 | (O)20 (I)177 | (O)51 (I)139 | (O)18 (I)375 | (O)17 (I)319 | (O)46 (I)179 | (O)49 (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 3 (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.

```
Caligula boisduvalii (62%)

ND2       5' – 268 ATTCTAATAAAATATTCTTTATTTTTGTTCTTTTTATTAGAACATTAATTTCA – 3' 320
Spacer    5' – 199 ATTTTAAATAGAGAATTTCAAATTCTTTTTAATTTATTATTAAATTATTTTAA – 3' 251
                   *** ***  * * * * * ** * **    *******  * **** *** *

Bombyx mandarina (65%)

ND2       5' – 10890 ACGAATAATAATTCAAATAAAATATTT---TTTTTATTTATTCTATTT – 3' 10934
Spacer    5' – 10825 ATTTA-AATAATTAAAATAAAGAATTTATAATTCTATTAAATATATTT – 3' 10790
                     *   * ******* ******* ****   ** **** * * *****

Parnassius bremeri (70%)

ND2       5' – 242 TTTTTTTAATTTAAATTCAAATAAAATATTTTTTTATTTTAT – 3' 283
Spacer    5' – 201 TTTATTGAATTTAAAT--AACAAAACTAACCCCCTATTTTAG – 3' 239
                   *** ** ********* **  *** **    *******

Artogeia melete (70%)

ND2       5' – 250 TTTATGTAAATTCTAATAAAATATTTTTTATTTTTTATTTTATTTTTTA – 3' 297
Spacer    5' – 195 TTTAAATAAATAGAACTTAAATTTCTTTTAAATTTTTTTTTTATTTTAA – 3' 242
                   ****  ***** * * ***** * *****  *** ****  **** *

Coreana raphaelis (62%)

ND2       5' – 257 ATTTTATTTTTAATTA-ATTCTAATAAAATATTTTTTATTTTTATTCTTTTTTTCA – 3' 311
Spacer    5' – 201 TTTTTAATTTAAAAAATAAACTTAAAATTTATTAATGAAATTTATTTCTTTTTTAT – 3' 256
                   *****  *** **  * * ** * **  ****  * * ******  ******

Eumenis autonoe (74%)

ND2       5' – 900 TTTTAAATTTAAATGATTTAAAATTTTTATTAAAAATAATTTTTTAATTTTAA – 3' 952
Spacer    5' – 199 CTTAAAAAAAAATGATATTAAATCATC--TAAAGAAATTTCTTA-TTTTAA – 3' 248
                   ** ***  ******* * **** *    **** * ***** *** ******

Lymantria dispar (60%)

ND2       5' – 278 AAAAATATTCTTTTTCTATATATTAATTTTTAGTACTTTAATTGCAATTTCTTCAA – 3' 333
Spacer    5' – 199 TAAAATATAATTATTTAATCAATGAATTTATAATTCACCAAATAATTTTATCTTAA – 3' 254
                   *******  ** ** ** ***** ** * *  ** *   ** * **
```
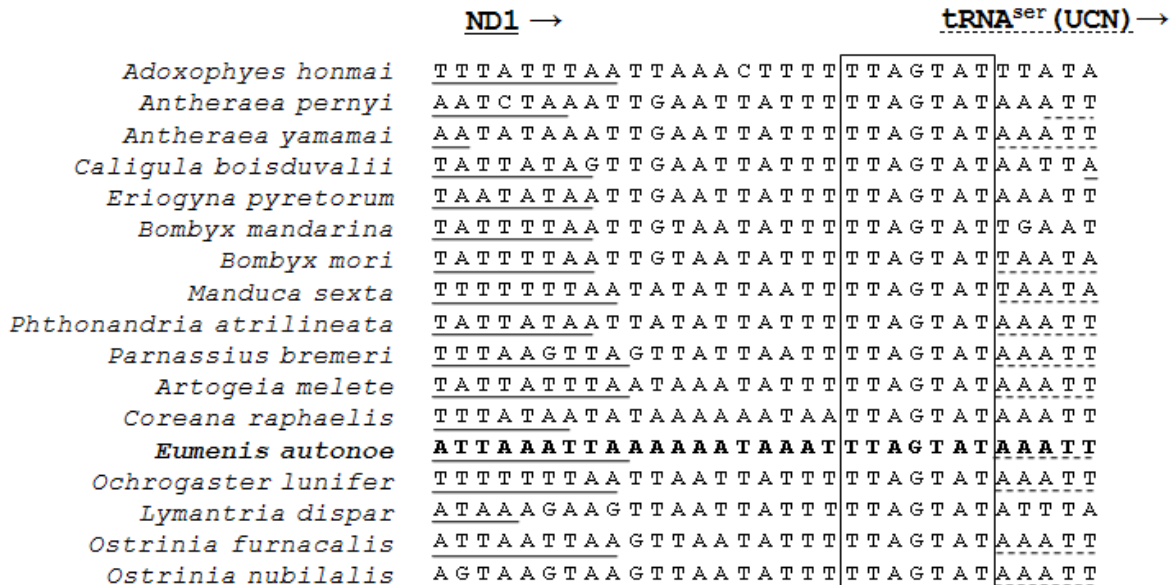
**Figure 4.** Alignment of the intergenic spacer sequence located between tRNA$^{Gln}$ and ND2 gene and neighboring partial ND2 gene from several lepidopteran insects, including *Eumenis autonoe*. Only lepidopteran species evidencing a sequence homology of more than 60% between the intergenic spacer sequence and the ND2 gene are presented. Asterisks indicate consensus sequences in the alignment. 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.

ND1 →                                                    tRNA^ser (UCN) →

```
Adoxophyes honmai        T T T A T T T A A T T A A A C T T T T  T T A G T A T  T T A T A
Antheraea pernyi         A A T C T A A A T T G A A T T A T T T  T T A G T A T  A A A T T
Antheraea yamamai        A A T A T A A A T T G A A T T A T T T  T T A G T A T  A A A T T
Caligula boisduvalii     T A T T A T A G T T G A A T T A T T T  T T A G T A T  A A T T A
Eriogyna pyretorum       T A A T A T A A T T G A A T T A T T T  T T A G T A T  A A A T T
Bombyx mandarina         T A T T T T A A T T G T A A T A T T T  T T A G T A T  T G A A T
Bombyx mori              T A T T T T A A T T G T A A T A T T T  T T A G T A T  T A A T A
Manduca sexta            T T T T T T T A A T A T A T T A A T T  T T A G T A T  T A A T A
Phthonandria atrilineata T A T T A T A A T T A T A T T A T T T  T T A G T A T  A A A T T
Parnassius bremeri       T T T A A G T T A G T T A T T A A T T  T T A G T A T  A A A T T
Artogeia melete          T A T T A T T T A A T A A A T A T T T  T T A G T A T  A A A T T
Coreana raphaelis        T T T A T A A T A T A A A A A T A A   T T A G T A T  A A A T T
Eumenis autonoe          A T T A A A T T A A A A A A T A A A T  T T A G T A T  A A A T T
Ochrogaster lunifer      T T T T T T T A A T T A A T T A T T T  T T A G T A T  A A A T T
Lymantria dispar         A T A A A G A A G T T A A T T A T T T  T T A G T A T  A T T T A
Ostrinia furnacalis      A T T A A T T A A G T T A A T A T T T  T T A G T A T  A A A T T
Ostrinia nubilalis       A G T A A G T A A G T T A A T A T T T  T T A G T A T  A A A T T
```

**Figure 5.** Alignment of the internal spacer region located between ND1 and tRNA^Ser(UCN) from all sequenced lepidopteran insects. 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 the ND1 gene and tRNA^Ser(UCN) gene. The arrows indicate the transcriptional direction.

A+T-rich region has been reported frequently in other insects (Cameron and Whiting, 2007), but is rare in sequenced lepidopteran insects. Careful analysis of the newly sequenced *L. dispar* mitogenome, which harbored a 435 bp long A+T-rich region, revealed the presence of a tandem repeat consisting of two duplicated 45 bp copies, the second copy of which harbored one substituted nucleotide (Unpublished, GenBank accession number FJ617240). Another example in the complete lepidopteran mitogenome is the 747 bp long Japanese *B. mandarina* A + T-rich region (Yukuhiro et al., 2002). It harbored a tandem triplication of a ≈ 126 bp fragment consisting of identical first and second copies, as well as one nucleotide-substituted and an AT-inserted third copy. Each of the 126 bp elements consisted of a ≈ 64 bp unit and a ≈ 62 bp repeated unit, each of which consisted of a 44 bp core sequence flanked by a 10 bp perfect inverted repeat in the case of the ≈ 64 bp unit and a 50 bp core sequence flanked by 6 bp perfect inverted repeats in the case of the ≈ 62 bp unit, respectively (Arunkumar et al., 2006). Liu et al. (2008) also reported the presence of a tandem repeat composed of six duplicated 38-bp copies, containing a ≈ 20 bp core motif flanked by 9-bp perfect inverted repeats in the 552-bp long *A. pernyi* A+T-rich region. Such repeats were considered characteristic of *Antheraea,* in that the partial mitogenome sequences of *A. roylei* and *A. proylei* also harbored highly similar repeat elements (Arunkumar et al., 2006; Liu et al., 2008). A common interpretation of the origin and persistence of repeat units within the A+T-rich region is the tandem duplication occurring via slipped-strand mispairing during replication (Moriz and Brown, 1987).

The remaining sequences of the A+T-rich region were composed of non-repetitive sequences, but harbored several poly-runs of T, A and TA. A BLAST search conducted to detect any relationship of the repeat sequence to other organisms or sequences proved only minimally successful, but the 27 bp long repeat sequence showed, interestingly, a high degree of sequence homology with a stretch of sequence located in the lrRNA gene at 74% (Figure 6B). The mechanism responsible for the location of such a similar sequence stretch in different locations within a mitogenome may be scrutinized further as more mitogenome sequence information is accumulated.

The A + T-rich region of the insect mitogenome, which is equivalent to the control region of the vertebrate mitogenome, has been demonstrated to harbor the replication origin for both strands in *Drosophila* species (Clary and Wolstenholme, 1987; Fauron and Wolstenholme, 1980) and the region located immediately downstream of a poly-T stretch at the 3'-end of the A + T-rich region has been identified as the position of the minor-strand replication origin in *B. mori* (Saito et al., 2005). Thus, the poly-T stretch has been suggested to function as a possible recognition site for the initiation of replication of the minor strand of mtDNA. The *E. autonoe* A + T-rich region harbored a 19-bp long T stretch upstream of the 5'-end of the srRNA (Figure 7). This poly-T stretch is quite well conserved in all sequenced lepidopteran insects, ranging in size from 16 bp in *L. dispar* to 22 bp in *B. mandarina* (Figure 7). Additionally, immediately downstream of the poly-T stretch in the A + T-rich region is another conserved motif ATAGA, which is

**(A)**

| | | |
|---|---|---|
| Repeat 1 | 15159 TATATAAATATTTATTATATAAACATT 15185 | |
| Repeat 2 | 15186 ......................................................... 15212 | |
| Repeat 3 | 15213 ....................................................... 15239 | |
| Repeat 4 | 15240 ...................... ........................... 15266 | |
| Repeat 5 | 15267 ....................................................... 15293 | |
| Repeat 6 | 15294 ....................................................... 15320 | |
| Repeat 7 | 15321 ....................................................... 15347 | |
| Repeat 8 | 15348 ....................................................... 15374 | |
| Repeat 9 | 15375 ...................................................... 15401 | |
| Repeat 10 | 15402 ....................................................... 15428 | |
| Repeat 11 | 15429 ....................................................... 15441 | |

**(B)**

| | | |
|---|---|---|
| Repeat | 15159 TATATAAATATTTATTATATAAACATT 15185 | |
| lrRNA | 13925 ...AT.T..A..........T...T........T............... 13951 | |

**Figure 6.** Tandem repeat units detected in the *Eumenis autonoe* A+T-rich region. (A) Alignment among the repeats and (B) Alignment between the repeat and neighboring srRNA. Nucleotide positions of the sequences are provided at each end of the sequence.



**Figure 7.** Alignment of partial A+T-rich region and srRNA. The shaded nucleotides indicate the poly-T stretch and the boxed nucleotides indicate the conserved ATAGA motif. 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.

very well-conserved in all sequenced lepidopteran insects, including *E. autonoe* (Figure 7). Previously, this motif has also been suggested to play some regulatory role together with the poly-T stretch (Kim et al., 2009). Nevertheless, this motif is conserved only in lepidopteran insects, but not in the Coleoptera and Diptera (data not shown). Thus, more experimental data are required for further conclusive decision.

**Phylogenetic relationships**

The 17 available lepidopteran mitogenomes, including that of *E. autonoe,* belong to the Obtectomera, repre-senting six lepidopteran superfamilies (Tortricoidea, Pyraloidea, Papilionoidea, Bombycoidea, Geometroidea and Noctuoidea). Among them, the superfamilies Papilionoidea, Bombycoidea, Geometroidea and Noctuo-idea are referred to as the Macrolepidoptera. The phylogenetic relationships of macrolepidopteran super-families have been the subject of substantial controversy, but no relationships have yet been clearly elucidated (Minet, 1991, 1994; Nielsen, 1989; Scott, 1986). One of the most compelling hypotheses in this regard is a closer relationship between the Papilionoidea and the Geome-troidea, with the unresolved relationships of this group to
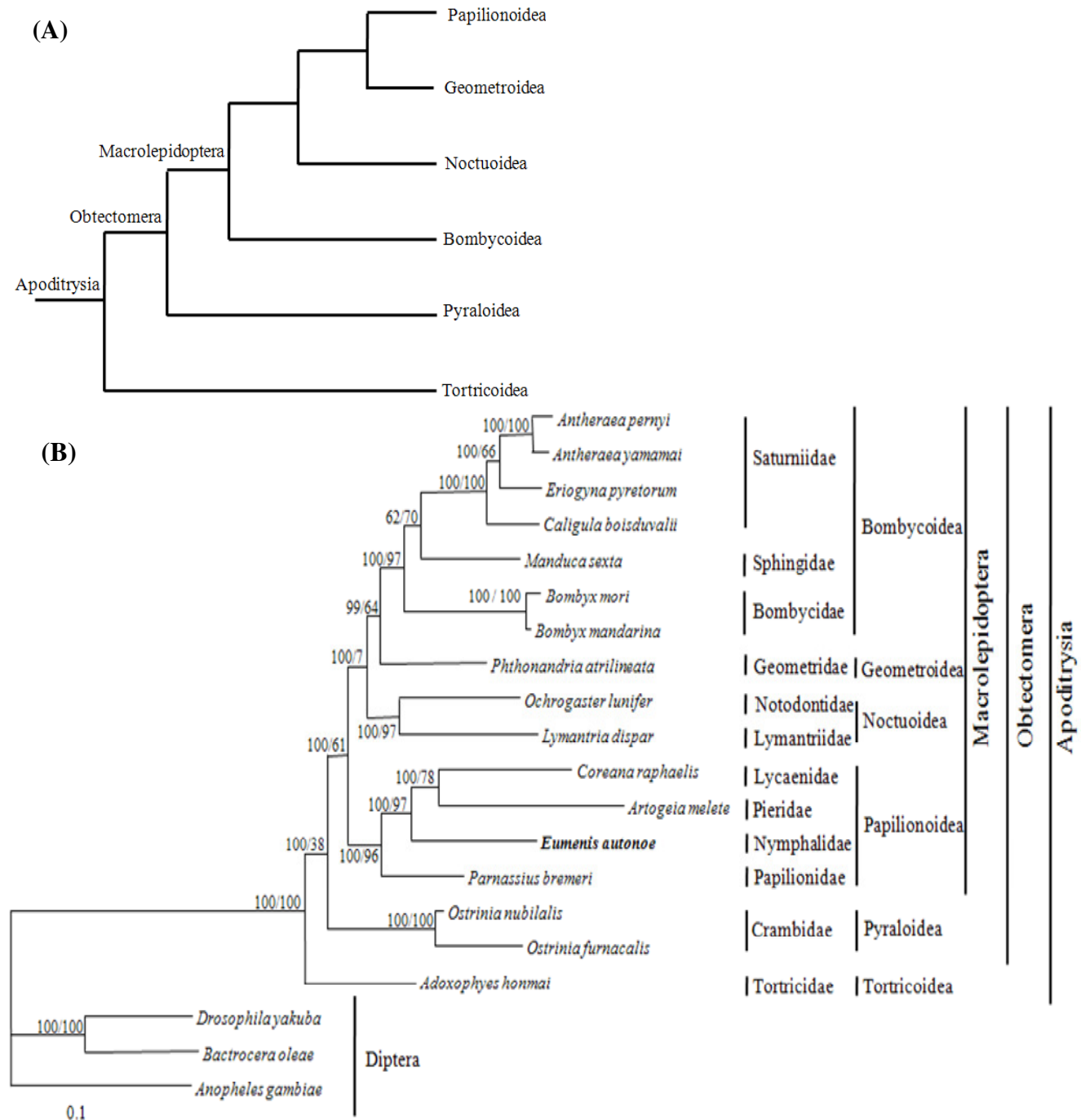
**Figure 8.** Phylogeny of apoditrisian superfamilies. (A) The current hypothesis of apoditrisian superfamily relationships (Minet, 1991; Nielsen, 1989). (B) Bayesian Inference phylogram of apoditrisian superfamilies obtained with an amino acid dataset. (C) Bayesian Inference phylogram of apoditrisian superfamilies obtained with a nucleotide dataset. Numbers at each node specify BPP by BI analysis (first value) and bootstrap percentages of 100 pseudoreplicates from ML anaslysis (second value), respectively. The dipterans, *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. The scale bar indicates the number of substitutions per site.

each Noctuoidea and Bombycoidea, resulting in tricho- tomy (Minet, 1991; Nielsen, 1989; Figure 8A).

The phylogenetic reconstruction of the relationships within Apoditrysia, both by concatenated amino acid and the nucleotide sequences of PCGs revealed an unexpected clustering of the Bombycoidea, Geome- troidea and Noctuoidea, excluding the Papilionoidea (Figures 8B and C). In particular, the Geometroidea was
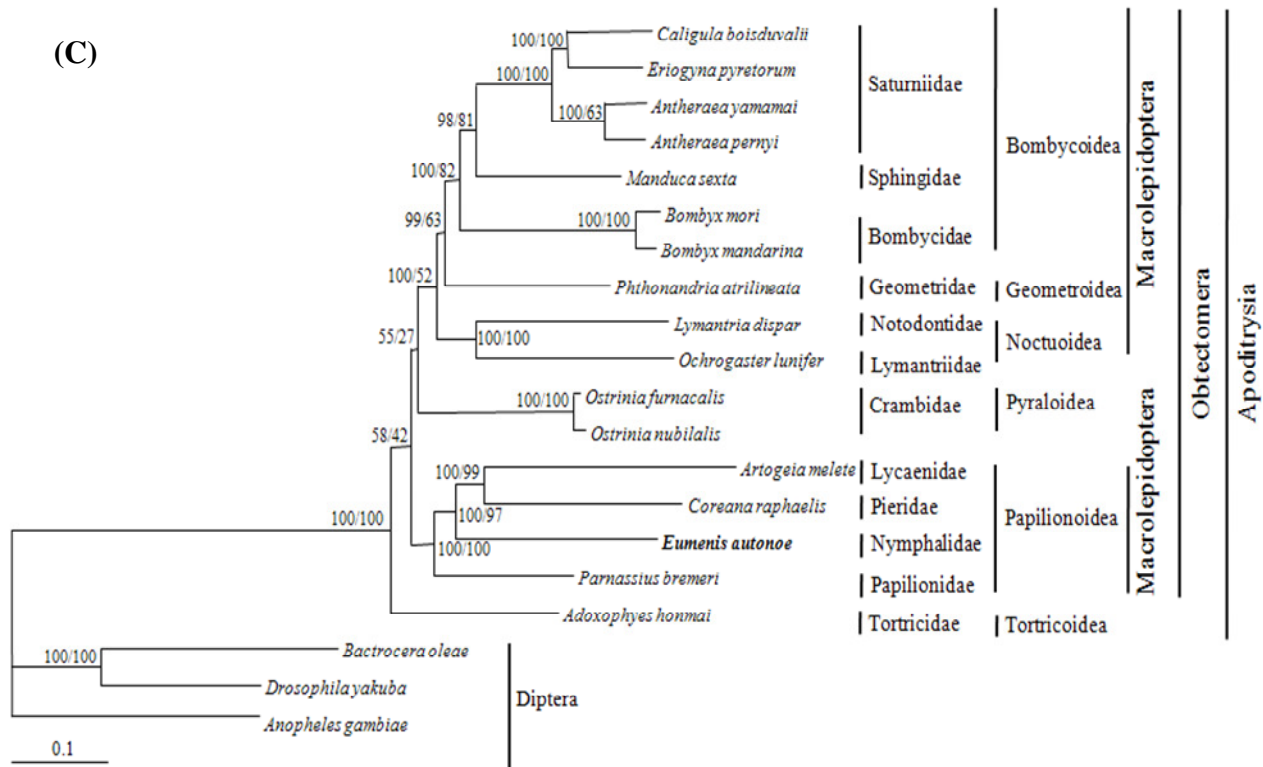
**(C)**



**Figure 8.** Contd.

identified as a sister taxon of the Bombycoidea with high nodal support on BI analyses using both amino acid and nucleotide data at 99% (Figures 8B and C), or relatively high nodal support on ML analyses using amino acid at 64% (Figure 8B) and nucleotide data at 63% (Figure 8C), respectively. On the other hand, the close relationship of Noctuoidea to the Bombycoidea + Geometroidea group is not always clear, in that this grouping was only strongly supported by BI analyses using both amino acid and nucleotide data at 100% (Figures 8B and C), whereas it was poorly supported by ML analyses using both data-sets (Figures 8B and C). Nevertheless, a close relation-ship between the Bombycoidea and the Geometroidea is worth noting, in that this result deviates from the tradi-tional view and those of other previously conducted phylogenetic studies namely, a sister group relationship between the Geometroidea and Papilionoidea, leaving Noctuoidea and Bombycoidea unresolved from the group (Mint, 1994), a close relationship between Noctuoidea and Papilionoidea, leaving Geometroidea and Bomby-coidea unresolved from the group by five nuclear PCGs (Regier et al., 2008) and a close relationship between Geometroidea and Papilionoidea and between Noctuo-Idea and Bombycoidea based on the mitochondrial ND1 gene, nuclear rRNA genes and morphological data (Weller and Pashely, 1995). Only a recent phylogenetic analysis using the complete mitogenomes of the sequenced lepidopteran insects also supported a close

relationship between the Geometroidea and Noctuoidea group and the Bombycoidea, in a fashion similar to that of this study (Jiang et al., 2009). This result suggested that macrolepidopteran evolution may be more complex than are currently understood. As more sequence information from a diverse taxonomic group becomes available, more comprehensive conclusions could be drawn.

A substantial debate has raged regarding the phylogenetic relationships existing among true butterfly families (Ehrlich and Ehrlich, 1967; Kristensen, 1976; Robbins, 1988; Scott, 1986). One of the most widely accepted relationships was that of Kristensen (1976), in which the Pieridae were identified as a sister to the Nymphalidae and Lycaenidae group, with the Papilio-nidae established as the basal lineage. This relationship was well supported by the recent work of Wahlberg et al. (2005), wherein the data matrix from substantially long DNA fragments and morphological features were collectively utilized, along with the elaborated phylo-genetic algorithms.

The phylogenetic analysis of the four families of true butterflies (Papilionoidea), each represented by a single species, showed the monophyly of Papilionoidea with very high nodal support at 100% by BI and 96% by ML using the amino acid sequence data (Figure 8B) and 100% both by BI and ML using the nucleotide sequence data (Figure 8C). With regard to the internal relationships existing among the families of Papilionoidea, all analyses

have positioned Nymphalidae as a sister to the Pieridae and Lycaenidae group, with the Papilionidae established as the basal lineage (Figures 8B and C). This relationship is consistent among all analyses, with very high respecttive nodal support (Figures 8B and C). One compromising result of our analysis to the most widely accepted relationships of true butterflies is the establishment of Papilionidae as the basal lineage (Kristensen, 1976). However, the phylogenetic relationships of true butterflies obtained by the full-length mitogenome sequences in this study overall do provide highly unconventional clustering; this has never previously been suggested in the relevant literature. One possible reason for the observed result may include a shortage in the numbers of available species, in that only a single species representing each family of the true butterflies is currently available. This may be the case, particularly because all current hypotheses contain multiple genera within the family (Wahlberg et al., 2005).

Collectively, the information currently available at least supports a strong clustering of Geometroidea and Bombycoidea, excluding Papilionoidea among the macrolepidopteran superfamilies. A monophyly of Papilionoidea and a monophyly of Bombycoidea were also well supported by the results of this study. All polygenetic analyses consistently placed Nymphalidae as a sister to the Pieridae and Lycaenidae group. Nevertheless, this topology has never been previously proposed. Thus, we are reluctant to draw a definitive conclusion regarding the phylogenetic relationships among butterflies, considering the rather limited taxonomic diversity. In order to further evaluate the phylogenetic relationships among the macrolepidopteran insects and among the true butterflies, a larger number of complete mitogenome sequences that encompass more of the taxonomic diversity will be required.

## ACKNOWLEDGMENTS

### REFERENCES

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

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

Adachi J, 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 AT, Barrell BG, de Bruijin MHL, Droujn ARJ, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981). Sequence and organization of the human mitochondrial genome. Nature 290: 457-465.

Beard CB, Mills D, Collins FH (1993). The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organizationand comparisons with mitochondrial sequences of other insects. Insect Mol. Biol. 2: 103-124.

Boore JL, Lavrov D, Brown WM (1998). Gene translocation links insects and crustaceans. Nature, 393: 667-668.

Cameron SL, Whiting MF (2007). Mitochondrial genomic comparisons of the subterranean termites from the Genus Reticulitermes (Insecta: Isoptera: Rhinotermitidae). Genome, 50: 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: 112-123.

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

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: 206-220.

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

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

Ehrlich PR, Ehrlich AH (1967). The phenetic relationships of the butterflies I. adult taxonomy and the nonspecificity hypothesis. Syst. Zool. 18: 374-385.

Fauron CMR, Wolstenholmn DR (1980). Extensive diversity among *Drosophila* species with respect to nucleotide sequences within the adenine+thymine-rich region of mitochondrial DNA molecules. Nucleic Acids Res. 8: 2439-2452.

Flook PK, Rowell CH, Gellissen G (1995). The sequence, organizationand evolution of the *Locusta migratoria* mitochondrial genome. J. Mol. Evol. 41: 928-941.

Friedrich M, Muquim N (2003). Sequence and phylogenetic analysis of the complete mitochondrial genome of the flour beetle *Tribolium castaneum*. Mol. Phylogenet. Evol. 26: 502-512.

Guindon S, Lethiec F, Duroux P, Gascuel O (2005). PHYML Online-a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res. 33: W557-W559.

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

Hong MY, Lee EM, Jo YH, Park HC, Kim SR, Hwang JS, Jin BR, Kang PD, Kim K-G, 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: 49-57.

Huelsenbeck JP, Ronquist F (2001). MrBayes: Bayesian inference of phylogeny. Bioinformatics, 17: 754-755.

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: 351-365.

Joo HJ, Kim, SS (2002). Butterflies in Jeju Island. Jeong Heng Sa, Seoul, Korea, p. 185.

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

Kim YS (2005). Illustrated Book of Korean Butterflies in Color. Kyo-Hak Pub. Co., Seoul, Korea, p. 305.

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

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: 217-225.

Kim SR, Kim MI, Hong MY, Kim KY, Kang PD, Hwang JS, Han YS, Jin BR, Kim I (2009). The complete mitogenome sequence of the Japanese oak silkmoth, *Antheraea yamamai* (Lepidoptera: Saturniidae). Mol. Biol. Rep. 36: 1871-1880.

Kristensen NP, (1976). Remarks on the family-level phylogeny of butterflies (Insecta Lepidoptera, Rhopalocera). Zeit. Zool. Syst. Evol. Forsch. 14: 25-33.

Lanave C, Preparata G, Saccone C, Serio G (1984). A new method for calculating evolutionary substitution rates. J. Mol. Evol. 20: 86-93.

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.

Liu Y, Li Y, Pan M, Dai F, Zhu X, Lu C, Xiang Z (2008). The complete mitochondrial genome of the Chinese oak silkmoth, *Antheraea pernyi* (Lepidoptera: Saturniidae). Acta. Biochim. Biophys. Sin. (Shanghai) 40: 693-703.

Lowe TM, Eddy SR (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.

Moriz C, Brown WM (1987). Tandem duplication in animal mitochondrial DNAs: variation in incidence and gene content among lizards. Proc. Natl. Acad. Sci. USA, 84: 7183-7187.

Nardi F, Caeapelli A, Dallai R, 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 ES (1989). Phylogeny of major lepidopteran groups. In: Fernholm B, Bremer K, Jörnvall H (eds) The Hierarchy of Life. Amsterdam, Elsevier. pp. 281-294.

Ojala D, Montoya J, Attardi G (1981). tRNA punctuation model of RNA processing in human mitochondria. Nature 290: 470-474.

Perna NT, Kocher TD (1995) Patterns of nucleotide composition at four fold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. 41: 353-358.

Posada D, Crandal KA (1998). Modeltest: testing the model of DNA substitution. Bioinformatics, 14: 817-818.

Regier JC, Cook C, Mitter C, Hussey A (2008). A phylogenetic study of the 'bombycoid complex' (Lepidoptera) using five protein-coding nuclear genes, with comments on the problem of macrolepidopteran phylogeny. Syst. Entomol. 33: 175-189.

Robbins RK (1988). Comparative morphology of the butterfly forleg coxa and trochanter (Lepidoptera) and its systematic implications. Proc. Entomol. Soc. Wash. 90: 133-154.

Saito S, Tamuea K, Aotsuka T (2005). Replicatiion origin of mitochondrial DNA in insects. Genetics, 171: 433-448.

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

Scott J (1986). On the monophyly of the Macrolepidoptera, including a reassessment of their relationship to Cossoidea and Castnioideaand a reaasignment of Mimallonidae to Pyraloidea. J. Res. Lepid. 25: 30-38.

Spanos L, Koutroumbas G, Kostyfakis M, Louis C (2000). The mitochondrial genome of the mediterranean fruitfly, *Ceratitis capitata*. Insect Mol. Biol. 9: 139-144.

Swofford DL (2002). PAUP*. Phylogenetic analysis using parsimony (*and other methods) ver 4.10 (Sunderland M.A.: Sinauer Associates).

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

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997). The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 24: 173-216.

Wahlberg N, Braby MF, Brower AVZ, de Jong R, Lee MM, Nylin S, Pierce N, Sperling FA, Vila R, Warren AD, Zakharov E (2005). Synergistic effects of combining morphological and molecular data in resolving the phylogeny of butterflies and skippers. Proc. R. Soc. B. 272: 1577-1586.

Weller SJ, Pashely DP (1995). In search of butterfly origins. Mol. Phylogenet. Evol. 4: 235-246.

Wernersson R, Pedersen AG (2003). Multiple alignment of coding DNA from aligned amino acid sequences. Nucleic Acids Res. 31: 3537-3539.

Wettstein W, 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.

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

Yang L, Wei ZJ, Hong GY, Jiang ST, Wen LP (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, 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.