Mol. Cells *30*, 409-425, November 30, 2010
DOI/10.1007/s10059-010-0141-9

Molecules and Cells

Phylogenetic Relationships of True Butterflies (Lepidoptera: Papilionoidea) Inferred from COI, 16S rRNA and EF-1α Sequences

Man Il Kim, Xinlong Wan, Min Jee Kim, Heon Cheon Jeong[1], Neung-Ho Ahn[2], Ki-Gyoung Kim[2], Yeon Soo Han, and Iksoo Kim*

The molecular phylogenetic relationships among true butterfly families (superfamily Papilionoidea) have been a matter of substantial controversy; this debate has led to several competing hypotheses. Two of the most compelling of those hypotheses involve the relationships of (Nymphalidae + Lycaenidae) + (Pieridae + Papilionidae) and (((Nymphalidae + Lycaenidae) + Pieridae) + Papilionidae). In this study, approximately 3,500 nucleotide sequences from cytochrome oxidase subunit I (COI), 16S ribosomal RNA (16S rRNA), and elongation factor-1 alpha (EF-1α) were sequenced from 83 species belonging to four true butterfly families, along with those of three outgroup species belonging to three lepidopteran superfamilies. These sequences were subjected to phylogenetic reconstruction via Bayesian Inference (BI), Maximum Likelihood (ML), and Maximum Parsimony (MP) algorithms. The monophyletic Pieridae and monophyletic Papilionidae evidenced good recovery in all analyses, but in some analyses, the monophylies of the Lycaenidae and Nymphalidae were hampered by the inclusion of single species of the lycaenid subfamily Miletinae and the nymphalid subfamily Danainae. Excluding those singletons, all phylogenetic analyses among the four true butterfly families clearly identified the Nymphalidae as the sister to the Lycaenidae and identified this group as a sister to the Pieridae, with the Papilionidae identified as the most basal linage to the true butterfly, thus supporting the hypothesis: (Papilionidae + (Pieridae + (Nymphalidae + Lycaenidae))).

INTRODUCTION

Butterflies are recognized to comprise somewhere between four and 14 families (Kristensen, 1976; Smart, 1989); however, the butterflies are most generally categorized into five families (Papilionidae, Pieridae, Nymphalidae, Riodinidae, and Lycaenidae). On the other hand, the family Riodinidae has occasionally been included in the Lycaenidae (Ackery, 1984; Ehrlich, 1958; Ehrlich and Ehrlich, 1967; Kristensen, 1976; Scott, 1985), but is still usually treated as a sister to the Lycaenidae (Harvey, 1987). The substantial debate raging around the phylogenetic relationships among the butterflies has led to a contention among four competing phylogenetic hypotheses regarding the true butterfly families (Fig. 1). Ehrlich and Ehrlich (1967) have reported a close relationship between the Pieridae and Papilionidae via numerical taxonomic methods (Fig. 1A). Scott (1985) generated a cladogram that categorized the Pieridae and Papilionidae into one group and the Lycaenidae (including Riodinidae as a subfamily) and Nymphalidae into another, thus positing essentially two true groups of butterflies (Fig. 1B). Kristensen (1976) placed the Pieridae as the sister to the Nymphalidae + Lycaenidae group, with Papilionidae identified as the basal lineage (Fig. 1C). This hypothesis has been supported in a number of other studies (Ackery, 1999; de Jong et al., 1996). Finally, Robbins (1988) has demonstrated some unresolved relationships among the Pieridae, Lycaenidae, and Papilionidae; the results of that study were suggestive of trichotomy (Fig. 1D).

With the advent of new molecular technologies, DNA sequence data has also been employed to evaluate phylogenetic relationships among the true butterfly families. For example, Martin and Pashley (1992) sequenced 200-300 bp of the 28S rRNA gene from 28 species encompassing five true butterfly families (including Riodinidae), which contained eight nymphalid subfamilies along with three Hesperioidea species as an outgroup. They generated four equally parsimonious trees that placed the Papilionidae as the basal lineage for the monophyletic Papilionoidea, but the relationships among the remaining families remained unresolved (Martin and Pashley, 1992). This result conflicts with previous findings supporting the sister relationship between the Pieridae and Papilionidae (Ehrlich, 1958; Ehrlich and Ehrlich, 1967; Scott, 1985). Weller et al. (1996) sequenced the mitochondrial ND1 gene (320 bp) and combined it with previously collected nuclear 28S rRNA information and 50 morphological characteristics from the relevant literature; they reported that only the combined data supported

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
*Correspondence: ikkim81@chonnam.ac.kr

Received March 9, 2010; revised June 26, 2010; accepted August 30, 2010; published online September 15, 2010

Keywords: 16S rRNA, COI, EF-1α, Lepidoptera, molecular phylogeny, true butterflies

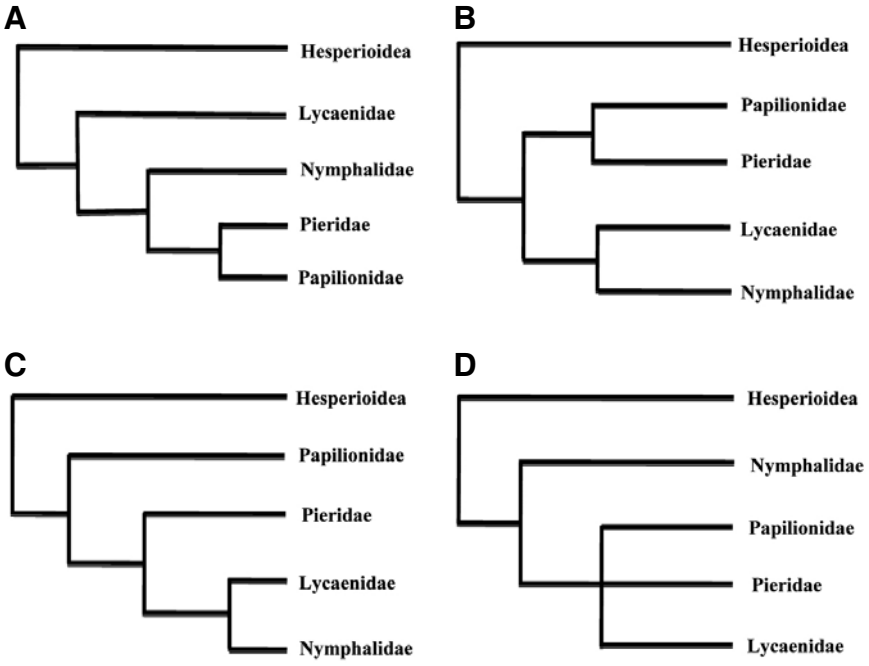


Fig. 1. Competing hypotheses regarding familial relationships of butterflies based on morphological studies. (A) Ehrlich and Ehrlich (1976). (B) Scott (1986). (C) Kristensen (1976). (D) after Robbins (1988).

Kristensen's hypothesis (1976) that the Nymphalidae, a sister to the Riodinidae, was also a sister to the Lycaenidae. The Pieridae was identified as a sister taxon to this group, thus positioning the Papilionidae as the most basal lineage to the true butterflies. More recently, Wahlberg et al. (2005a) sequenced DNA fragments of substantial length (3,258 bp) comprising three genes (COI, EF-1α, and wingless) and analyzed them with the preexisting morphological data for 57 taxa, representing all major lineages from the three putative butterfly superfamilies (Hedyloidea, Hesperioidea, and Papilionoidea). They determined that the combined molecular and morphological dataset strongly supported the relationship of (Hesperiidae + (Papilionidae + (Pieridae + (Nymphalidae + (Lycaenidae + Riodinidae))))); this is consistent with the earlier hypothesis of Kristensen (1976), with the exception of somewhat lower node support for the monophyletic Nymphalidae in some analyses (Wahlberg et al., 2005a).

Several recent studies of lepidopteran phylogeny have corroborated deep and shallow levels of divergence, detected via combined analyses of nuclear, mitochondrial, or ribosomal genes (Caterino et al., 2000; 2001; Kandul et al., 2004; Monteiro and Pierce, 2001; Wahlberg et al., 2003; 2005a; Zakharov et al., 2004). In fact, it is now common practice in the field of insect molecular systematics to combine one or more mitochondrial genes with one or more nuclear genes, as the two types of data are unlinked and evolve under different constraints (Lin et al., 2004).

In this study, therefore, we sequenced mitochondrial cytochrome oxidase I (COI), mitochondrial 16S ribosomal RNA (16S rRNA), and nuclear elongation factor-1α (EF-1α) in order to evaluate previous phylogenetic hypotheses regarding the true butterfly families (Papilionidae, Pieridae, Lycaenidae, and Nymphalidae). A total of 83 butterfly species belonging to 17 subfamilies in four true butterfly families in the butterfly superfamily Papilionoidea were employed as an ingroup.

MATERIALS AND METHODS

Taxon sampling

The Papilionoidea superfamily includes five families (Papilionidae, Pieridae, Nymphalidae, Riodinidae and Lycaenidae), but this study included four families, excluding the family Riodinidae (Harvey, 1987), because taxon sampling for the family was unsuccessful. There are an estimated 13,700 species of Papilionoidea extant in the world (Robbins, 1982), and 234 extant species currently inhabit the Korean Peninsula (Joo et al., 2005). A total of 83 butterflies belonging to four true butterfly families (Papilionidae, Pieridae, Nymphalidae, and Lycaenidae) were used as an ingroup, and three species of moths belonging to the superfamilies Zygaenoidea (*Chalcosia remota*), Copromorphoidea (*Carposina sasakii*), and Tortricoidea (*Grapholita molesta*) were used as an outgroup (Table 1). The 83 true butterflies were identified as belonging to a total of 54 genera in 17 subfamilies. Voucher specimens were deposited in the National Institute of Biological Resources, Incheon, Republic of Korea.

Molecular methods

After collection in the field, the samples were frozen at -70°C until being used for molecular analysis. Total DNA was extracted using a Wizard™ Genomic DNA Purification Kit, in accordance with the manufacturer's instructions (Promega, USA).

In order to sequence 1,099 bp of mitochondrial COI, 1,284-1,419 bp of 16S rRNA and 1,066 bp of nuclear EF-1α, each gene was amplified into two independent fragments for easy processing. Alternatively, in some cases, the whole fragment for each gene was PCR amplified and sequenced either immediately, or after cloning. The primer sets utilized for the amplification of each fragment are listed in Table 2. These primer sets were either adapted from preexisting published ones, or were newly designed from GenBank-registered sequences. DNA 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 54-58°C, and 2 min at 72°C, with a final 7-minute extension

Table 1. A list of the full taxon names, collecting localities, nucleotide frequencies, gene size, and GenBank accession no. of the species included in this study

Taxon	Collecting locality	Nucleotide frequencies & gene size (bp)															GenBank accession no.		
		COI					16S rRNA					EF-1α					COI	16S	EF-1α
		T	C	A	G	size	T	C	A	G	size	T	C	A	G	size			
Papilionoidea																			
Papilionidae																			
Papilioninae																			
Papilionini																			
Papilio bianor dehaanii	Hampyeong, Jeonnam, Korea	40.2	14.2	31.0	14.6	1099	43.0	11.3	40.5	5.1	1345	21.4	28.7	25.4	24.5	1066	GU372541	GU372450	GU372632
Papilio protenor demetrius	Hampyeong, Jeonnam, Korea	40.2	14.6	30.6	14.6	1099	43.3	11.6	39.9	5.1	1342	21.6	28.7	25.4	24.3	1066	GU372542	GU372451	GU372633
Papilio xuthus	Hampyeong, Jeonnam, Korea	39.9	15.0	30.7	14.5	1099	41.9	11.5	41.4	5.2	1346	21.6	29.0	25.1	24.3	1066	GU372543	GU372452	GU372634
Papilio maackii	Hampyeong, Jeonnam, Korea	40.0	14.4	31.2	14.4	1099	42.9	11.4	40.6	5.1	1346	21.2	29	25.4	24.4	1066	GU372544	GU372453	GU372635
Papilio macilentus	Gurye, Jeonnam, Korea	40.1	14.7	30.8	14.3	1099	43.0	11.4	40.4	5.1	1348	22.1	28.2	25.1	24.5	1066	GU372545	GU372454	GU372636
Papilio machaon hippocrates	Andeok-valley, Jeju, Korea	40.2	14.5	31.0	14.3	1099	44.8	11	38.6	5.6	1345	22.9	27.7	24.7	24.8	1066	GU372546	GU372455	GU372637
Troidini																			
Atrophaneura alcinous	Jungsun, Gangwon, Korea	41.3	14.8	29.2	14.6	1099	44.5	10.7	39.4	5.4	1359	22.4	28.3	25.5	23.7	1066	GU372547	GU372456	GU372638
Graphiini																			
Graphium sarpedon nipponum	Hampyeong, Jeonnam, Korea	39.9	15.5	30.2	14.4	1099	42.5	10.5	41.7	5.3	1366	22	28.6	24.6	24.8	1066	GU372548	GU372457	GU372639
Parnassiinae																			
Parnassiini																			
Parnassius stubbendorfii korea	Paju, Gyeonggi, Korea	40.6	13.6	31.2	14.6	1099	42.5	10.7	41.4	5.4	1355	25.9	25.0	26.0	23.2	1066	GU372549	GU372458	GU372640
Luehdorfiini																			
Luehdorfia puziloi coreana	Hampyeong, Jeonnam, Korea	40.9	13.6	31.0	14.5	1099	41.1	11.1	42.4	5.5	1375	24.8	25.6	25.9	23.7	1066	GU372550	GU372459	GU372641
Zerynthiini																			
Sericinus montela greyi	Ssangyong, Gangwon, Korea	40.8	14.1	30.6	14.6	1099	42.3	11.6	40.1	6.0	1358	25.0	24.4	26.0	24.6	1066	GU372551	GU372460	GU372642
Pieridae																			
Pierinae																			
Pierini																			
Pieris canidia kaolicola	Paju, Gyeonggi, Korea	38.1	15.7	31.6	14.6	1099	42.1	11.0	41.6	5.4	1345	25.6	25.2	26.5	22.6	1066	GU372552	GU372461	GU372643
Pieris melete minor	Paju, Gyeonggi, Korea	38.7	15.5	31.1	14.7	1099	42.3	11.0	41.5	5.2	1344	25	25.9	26.1	23	1066	GU372553	GU372462	GU372644
Pieris napi dulcinea	Mt. Odae, Gangwon, Korea	39.5	14.7	30.9	14.8	1099	42.2	11.2	41.2	5.4	1343	24.9	26.1	26.2	22.9	1066	GU372554	GU372463	GU372645
Pieris rapae orientalis	Hampyeong, Jeonnam, Korea	39.0	14.6	31.6	14.7	1099	42.1	11.1	41.4	5.4	1344	25.7	25.1	26.2	23.0	1066	GU372555	GU372464	GU372646
Anthocharidini																			
Anthocharis scolymus	Hampyeong, Jeonnam, Korea	40.3	15.2	29.8	14.6	1099	44.3	10.7	39.8	5.3	1360	24.3	25.6	27	23.1	1066	GU372556	GU372465	GU372647
Coliadinae																			
Gonepteryx Aspasia coreansis	Mt. Mudeung, Kwangju, Korea	39.7	15.1	30.6	14.6	1099	43.3	10.9	40.5	5.2	1357	22.8	28.3	24.7	24.2	1066	GU372557	GU372466	GU372648
Gonepteryx rhamni amurensis	Ssangyong, Gangwon, Korea	40.2	14.4	30.9	14.5	1099	43.9	10.6	40.3	5.3	1354	23.1	28.2	24.3	24.4	1066	GU372558	GU372467	GU372649
Eurema hecabe	Hampyeong, Jeonnam, Korea	39.8	15.4	30.1	14.7	1099	42.9	11.2	40.6	5.3	1339	23.5	27.2	25.5	23.7	1066	GU372559	GU372468	GU372650
Eurema laeta betheseba	Naju, Jeonnam, Korea	39.9	16.1	29.5	14.6	1099	43.0	11.0	40.9	5.1	1366	24.5	26.8	25.0	23.7	1066	GU372560	GU372469	GU372651
Colias erate poliographus	Seogu, Incheon, Korea	39.9	14.5	31.0	14.6	1099	42.9	10.6	41.4	5.1	1351	22.8	28.3	24.6	24.3	1066	GU372561	GU372470	GU372652
Dismorphinae											1351								
Leptidea amurensis	Yeongwol, Gangwon, Korea	39.2	15.3	30.2	15.3	1099	45.6	10.6	38.2	5.6	1355	24.7	24.8	27.9	22.7	1066	GU372562	GU372471	GU372653
Leptidea morsei morseides	Haesan-ryeong, Gangwon, Korea	39.1	15.3	31.2	14.4	1099	45.5	10.6	38.5	5.4	1356	24.8	24.6	27.6	23.1	1066	GU372563	GU372472	GU372654
Lycaenidae																			
Lycaeninae																			
Lycaenini																			
Lycaena dispar aurata	Pocheon, Gyeonggi, Korea	42.1	13.0	30.6	14.3	1099	44.0	10.4	40.4	5.2	1352	23.0	26.8	25.6	24.6	1066	GU372564	GU372473	GU372655
Lycaena phlaeas chinensis	Naju, Jeonnam, Korea	41.9	13.0	30.7	14.5	1099	45.6	9.4	39.9	5.1	1361	23.1	26.7	25.4	24.8	1066	GU372565	GU372474	GU372656
Theclinae																			
Theclini																			
Japonica lutea dubatolovi	Paju, Gyeonggi, Korea	39.9	13.9	31.8	14.3	1099	46.7	9.9	38.4	5.0	1350	24.8	25.7	26.5	23.1	1066	GU372566	GU372475	GU372657
Japonica saepestriata	Mt. Suri, Gyeonggi, Korea	40.5	13.5	31.8	14.2	1099	46.6	9.7	38.5	5.2	1351	24.8	25.8	26.4	23.1	1066	GU372567	GU372476	GU372658
Artopoetes pryeri	Mt. Odae, Gangwon, Korea	39.9	13.7	31.9	14.5	1099	45.1	10.1	39.5	5.3	1340	24.8	25.3	26.7	23.2	1066	GU372569	GU372478	GU372660
Ussuriana michaelis	Whacheon, Gangwon, Korea	40.9	12.7	32.0	14.4	1099	46.4	9.8	38.5	5.3	1344	24.4	25.9	26.5	23.3	1066	GU372570	GU372479	GU372661
Thecla betulae coreana	Jangseong, Jeonnam, Korea	39.6	13.8	32.3	14.3	1099	45.3	10.2	39.2	5.3	1359	24.5	26.2	25.9	23.5	1066	GU372599	GU372508	GU372690
Antigius attilia	Gapyeong, Gyeonggi, Korea	39.5	14.7	31.4	14.4	1099	46.4	10.1	38.4	5.1	1352	24.8	25.5	26.9	22.8	1066	GU372571	GU372480	GU372662
Antigius butleri	Gurye, Jeonnam, Korea	39.9	14.3	31.8	14.1	1099	46.3	10.0	38.4	5.3	1348	24.7	25.5	26.7	23.1	1066	GU372598	GU372507	GU372689
Favonius cognatus	Mt. Mudeung, Kwangju, Korea	40.9	13	31.2	14.9	1099	45.6	10.4	38.8	5.2	1343	23.7	26.5	26.1	23.7	1066	GU372572	GU372481	GU372663
Favonius taxila	Gapyeong, Gyeonggi, Korea	40.4	13.6	31.2	14.8	1099	45.2	10.5	39.1	5.2	1348	23.7	26.5	26.1	23.7	1066	GU372573	GU372482	GU372664
Favonius koreanus	Chuncheon, Gangwon, Korea	40.8	13.1	31.2	14.9	1099	45.7	10.3	38.8	5.3	1346	23.6	26.5	26.0	23.8	1066	GU372574	GU372483	GU372665
Favonius korshunovi	Whacheon, Gangwon, Korea	40.4	13.6	31.2	14.8	1099	45.3	10.4	39.2	5.1	1348	23.7	26.5	26.1	23.7	1066	GU372538	GU372448	GU372630
Wagimo signata	Whacheon, Gangwon, Korea	39.3	14.7	31.6	14.4	1099	45.9	10.4	38.5	5.1	1345	24.8	25.9	26.5	22.9	1066	GU372575	GU372484	GU372666
Coreana raphaelis	Gapyeong, Gyeonggi, Korea	41.1	12.6	32.0	14.2	1099	46.0	9.8	39.0	5.2	1347	24.0	25.8	26.7	23.5	1066	GU372533	GU372442	GU372624
Favonius saphirinus	Jangseong, Jeonnam, Korea	40.7	13.2	31.3	14.8	1099	45.5	10.2	39	5.3	1345	23.6	26.5	26.2	23.6	1066	GU372534	GU372444	GU372626
Favonius ultramarinus	Ssangyong, Gangwon, Korea	40.8	13.1	31.2	14.9	1099	45.7	10.1	38.9	5.3	1345	23.8	26.4	26.1	23.7	1066	GU372535	GU372445	GU372627
Chrysozephyrus smaragdinus	Gapyeong, Gyeonggi, Korea	40.7	13.2	31.6	14.6	1099	45.1	10.5	39.1	5.3	1347	23.5	26.9	26.1	23.5	1066	GU372536	GU372446	GU372628

(continued)

Taxon	Collecting locality	Nucleotide frequencies & gene size (bp)															GenBank accession no.		
		COI					16S rRNA					EF-1α					COI	16S	EF-1α
		T	C	A	G	size	T	C	A	G	size	T	C	A	G	size			
Deudorigini																			
Rapala caerulea	Mt. Odae, Gangwon, Korea	40.1	13.2	32.0	14.6	1099	45.7	9.9	39.4	5.0	1353	22.8	27.6	26.0	23.6	1066	GU372576	GU372485	GU372667
Eumaeini																			
Fixsenia prunoides	Ssangyong, Gangwon, Korea	39.3	13.8	32.1	14.7	1099	46.3	9.7	38.8	5.2	1353	23.9	25.3	27.0	23.7	1066	GU372577	GU372486	GU372668
Fixsenia herzi	Gapyeong, Gyeonggi, Korea	40.6	12.6	32.4	14.5	1099	46.1	9.7	39.2	5	1348	23.4	26.0	26.7	23.9	1066	GU372539	GU372443	GU372625
Fixsenia pruni	Gapyeong, Gyeonggi, Korea	40.7	12.9	32.4	14.0	1099	46.0	9.3	39.3	5.3	1350	23.4	26.1	26.8	23.7	1066	GU372540	GU372449	GU372631
Fixsenia eximia	Ssangyong, Gangwon, Korea	40.8	12.9	32.1	14.2	1099	46.0	9.4	39.4	5.2	1357	23.5	25.7	26.7	24.0	1066	GU372578	GU372487	GU372669
Fixsenia w-album fentoni	Ssangyong, Gangwon, Korea	39.2	13.8	32.2	14.7	1099	46.2	9.7	38.9	5.2	1354	23.9	25.3	27.0	23.7	1066	GU372568	GU372477	GU372659
Callophrys ferrea korea	Hampyeong, Jeonnam, Korea	40.9	12.3	32.7	14.2	1099	46.9	9.4	38.5	5.2	1349	23.7	26.3	26.2	23.8	1066	GU372579	GU372488	GU372670
Arhopalini																			
Arhopala japonica	Jocheon, Jeju, Korea	41.6	13.4	30.9	14.1	1099	45.0	9.2	40.7	5	1373	23.9	26.3	26.5	23.3	1066	GU372580	GU372489	GU372671
Polyommatinae																			
Polyommatini																			
Celastrina oreas mirificus	Ssangyong, Gangwon, Korea	40.7	13.2	31.9	14.2	1099	43.5	9.9	41.3	5.4	1345	20.9	29.9	24.4	24.8	1066	GU372581	GU372490	GU372672
Plebejus argus micargus	Mt. Halla, Jeju, Korea	40.1	13.6	32	14.2	1099	45.1	9.9	39.9	5.2	1345	17.1	34.1	21.6	27.3	1066	GU372582	GU372491	GU372673
Lycaeides argyronomon ussurica	Damyang, Jeonnam, Korea	40.7	13.2	31.9	14.2	1099	45.8	9.1	40.1	5.0	1353	17.1	33.8	22.0	27.2	1066	GU372583	GU372492	GU372674
Lampides boeticus	Yewol, Jeju, Korea	41.0	13.0	31.4	14.6	1099	44.3	9.7	40.8	5.2	1344	20.2	30.0	24.5	25.3	1066	GU372584	GU372493	GU372675
Chilades pandava	Yewol, Jeju, Korea	40.5	13.6	30.8	15.0	1099	45.2	9.4	40.1	5.2	1345	18.0	33.3	22.0	26.7	1066	GU372585	GU372494	GU372676
Scolitandides orion coreana	Jungsun, Gangwon, Korea	39.7	13.1	32.6	14.6	1099	44.2	10.5	40.0	5.4	1344	20.3	29.1	25.3	25.3	1066	GU372588	GU372497	GU372679
Shijimiaeoides divina	Jecheon, Chungbuk, Korea	40.6	13.3	31.8	14.4	1099	44.4	9.6	40.7	5.3	1350	20.3	29.7	24.9	25.1	1066	GU372587	GU372496	GU372678
Everes argiades hellotia	Seogu, Incheon, Korea	39.1	13.8	32.4	14.6	1099	43.4	9.8	41.4	5.3	1342	16.8	34.6	20.9	27.7	1066	GU372537	GU372447	GU372629
Niphandini																			
Niphanda fusca	Jangseong, Jeonnam, Korea	38.4	14.6	32.8	14.3	1099	43.2	10.2	41.2	5.4	1336	22.3	27.5	26.4	23.8	1066	GU372586	GU372495	GU372677
Miletinae																			
Spalgini																			
Taraka hamada	Hampyeong, Jeonnam, Korea	39.1	16.0	30.4	14.5	1099	39.6	11.5	43.5	5.4	1284	27.2	23.3	26.7	22.8	1066	GU372589	GU372498	GU372680
Nymphalidae																			
Limenitidinae																			
Limenitidini																			
Limenitis doerriesi chosensis	Yeonggwang, Jeonnam, Korea	40.9	14.6	30.1	14.4	1099	46.0	10.8	38.2	5.1	1344	23.4	27	26.4	23.3	1066	GU372509	GU372418	GU372600
Limenitis moltrechti takamukuana	Mt. Gyebang, Gangwon, Korea	39.8	15.5	30.2	14.6	1099	45.3	11.0	38.6	5.1	1346	23.7	26.9	26.3	23.1	1066	GU372510	GU372419	GU372601
Nymphalinae																			
Nymphalini																			
Polygonia c-aureum	Hampyeong, Jeonnam, Korea	39.1	14.5	31.9	14.5	1099	45.2	10.4	39.2	5.2	1355	23.4	27.4	25.9	23.4	1066	GU372511	GU372420	GU372602
Polygonia c-album hamigera	Mt. Gyebang, Gangwon, Korea	39.7	14.0	32.0	14.3	1099	45.1	10.6	39.0	5.3	1359	23.8	27.0	26.2	23.0	1066	GU372512	GU372421	GU372603
Araschnia burejana	Jungsun, Gangwon, Korea	40.4	14.1	30.8	14.6	1099	44.3	11.0	39.6	5.2	1414	23.9	26.9	26.3	22.9	1066	GU372513	GU372422	GU372604
Kaniska canace no-japonicum	Yeongcheon, Gyeongbuk, Korea	39.9	14.2	31.7	14.3	1099	45.6	10.5	38.8	5.1	1354	24.0	26.9	25.6	23.5	1066	GU372514	GU372423	GU372605
Vanessa indica	Haenam, Jeonnam, Korea	38.3	15.1	32.2	14.4	1099	44.3	10.7	39.8	5.2	1349	23.7	27.1	25.7	23.5	1066	GU372515	GU372424	GU372606
Cynthia cardui	Mt. Mudeung, Kwangju, Korea	39.6	14.1	31.6	14.7	1099	46.2	10.3	38.4	5.1	1353	23.6	27.3	25.8	23.3	1066	GU372516	GU372425	GU372607
Pseudergolinae																			
Dichorragia nesimachus nesiotes	Haenam, Jeonnam, Korea	41.5	13.2	30.7	14.6	1099	45	10.7	39.1	5.1	1360	23.6	26.4	26.9	23.1	1066	GU372517	GU372426	GU372608
Satyrinae																			
Satyrini																			
Eumenis autonoe zezutonis	Mt. Halla, Jeju, Korea	37.9	15.7	31.2	15.3	1099	45.2	11.4	38.1	5.3	1352	23.5	27.7	24.4	24.4	1066	GU372518	GU372427	GU372609
Apaturinae																			
Apatura iris peninsularis	Gurye, Jeonnam, Korea	39.5	14.6	31.0	14.8	1099	45.1	10.8	39.1	5.1	1357	23.9	26.8	26.1	23.2	1066	GU372519	GU372428	GU372610
Apatura metis heijona	Mt. Mudeung, Kwangju, Korea	39.8	14.3	31.4	14.6	1099	45.3	11.0	38.5	5.2	1353	24.2	26.5	26.0	23.3	1066	GU372520	GU372429	GU372611
Sephisa princeps	Hampyeong, Jeonnam, Korea	39.1	14.8	31.5	14.6	1099	43.4	12.2	39.1	5.3	1341	24.5	26.5	26.3	22.7	1066	GU372521	GU372430	GU372612
Mimathyma schrenckii	Whacheon, Gangwon, Korea	38.9	14.6	31.8	14.6	1099	44.5	10.9	39.2	5.3	1356	24.3	26.4	26.4	23.0	1066	GU372522	GU372431	GU372613
Athymodes nycteis furukawai	Ssangyong, Gangwon, Korea	39.1	14.9	31.5	14.5	1099	43.6	10.6	40.8	5.0	1354	24.5	26.2	26.4	23.0	1066	GU372523	GU372432	GU372614
Hestina persimilis seoki	Bukgu, Gwangju, Korea	40.2	14.7	30.9	14.1	1099	44.5	10.8	39.3	5.4	1352	22.9	27.6	26.5	23.1	1066	GU372524	GU372433	GU372615
Heliconiinae																			
Argynnis paphia geisha	Gurye, Jeonnam, Korea	40.1	14.2	31.4	14.3	1099	45.0	10.8	39.1	5.0	1347	24.7	26.0	26.5	22.9	1066	GU372525	GU372434	GU372616
Argyreus hyperbius	Hampyeong, Jeonnam, Korea	39.6	14.4	31.6	14.5	1099	45.0	10.6	38.9	5.4	1348	24.7	25.7	26.9	22.7	1066	GU372526	GU372435	GU372617
Clossiana perryi	Mt. Gyebang, Gangwon, Korea	39.2	14.9	31.7	14.2	1099	44.9	10.7	39.0	5.5	1339	24.0	26.8	25.9	23.3	1066	GU372527	GU372436	GU372618
Argyronome laodice japonica	Mt. Suri ,Gunpo, Gyeonggi, Korea	39.9	14.5	31.1	14.6	1099	45.1	11.1	38.3	5.5	1349	24.9	25.6	26.6	22.9	1066	GU372528	GU372437	GU372619
Nephargynnis anadyomene ella	Jungsun, Gangwon, Korea	39.0	15.1	31.2	14.6	1099	45.3	10.8	38.7	5.3	1348	24.5	26.1	26.6	22.8	1066	GU372529	GU372438	GU372620
Fabriciana nerippe coreana	Mt. Bukhan, Gyeonggi, Korea	38.9	14.7	31.4	14.9	1099	45.6	10.9	38.4	5.1	1343	25.2	25.5	26.9	22.3	1066	GU372530	GU372439	GU372621
Danainae																			
Danaini																			
Parantica sita niphonica	Mt. Gyebang, Gangwon, Korea	39.1	15.8	30.6	14.5	1099	45.6	10.6	38.5	5.3	1336	22.7	28.3	25.2	23.7	1066	GU372531	GU372440	GU372622
Libytheinae																			
Libythea celtis celtoides	Hampyeong, Jeonnam, Korea	40.2	14.3	30.7	14.8	1099	44.9	10.9	39.1	5.1	1353	22.5	28.1	25.3	24.0	1066	GU372532	GU372441	GU372623
Average		40.0	14.2	31.3	14.5	1099	44.6	10.5	39.7	5.2	1350.1	23.4	27.1	25.8	23.8	1066			

Table 2. Primers used for the amplification of COI, 16S rRNA, and EF-1α gene

Gene	Fragment	Name of primer (forward or reverse reading)	Primer sequences (from 5′ to 3′)	Length (bp) of sequenced product	Annealing temperature	References
COI	F1	LCO1490 (f)	GGTCAACAAATCATAAAGATATTGG	658	54°C	Folmer et al. (1994)
		HCO2198 (r)	TAAACTTCAGGGTGACCAAAAAATCA			Folmer et al. (1994)
	F1	LepF (f)	ATTCAACCAATCATAAAGATATTGG	658	54°C	Hajibabaei et al. (2006)
		LepR(r)	TAAACTTCTGGATGTCCAAAAAATCA			Hajibabaei et al. (2006)
	F2	COI2F (f)	TGTWTGAGCTGTCGGAATTACAGC	571	58°C	designed in this study
		COI2R (r)	AATWGCAAATACWGCTCCTAT			designed in this study
		LCO1490/LepF and COI2R*		1099	54-56°C	
16S rRNA	F1	16S1F (f)	AATATTTRATCCTTTCGTAC	545/605	56°C	modified from Niehuis et al. (2006)
		16S1R (r)	CTTGTTTATCAAAAACATGTC			modified from Niehuis et al. (2006)
	F2	16S2F (f)	TAARAGACTAATGATTATGC	833/916	58°C	modified from Niehuis et al. (2006)
		16S2R (r)	GTA ACAAAGTAGAGGTACTGG			modified from Niehuis et al. (2006)
		16S1F and 16S2R*		1284/1419		
EF-1α	F1	ef44 (f)	GCYGARCGYGARCGTGGTATYAC	569	54°C	Monteiro and Pierce (2001)
		ef44m (f)	GCTGAGCGYGAGCGTGGTATTAC			modified from Monteiro and Pierce (2001)
		ELF1R (r)	GTTTCAACTCTGCCTACKGGCAC			designed in this study
	F2	ELF2F (f)	AAAATGCCCTGGTTCAAGGGA	642	57°C	designed in this study
		efrcM4 (r)	ACAGCVACKGTYTGYCTCATRTC			Monteiro and Pierce (2001)
		ef44/ef44m and efrcM4*		1066	54-57°C	

F1, fragment 1; and F2, fragment 2

*The primers used to amplify whole length of sequences.

at 72°C. Detailed annealing temperatures for each amplification are shown in Table 2. In order to confirm successful DNA amplification, electrophoresis was conducted using 0.5× TAE buffer on 0.5% agarose gel. The PCR product was then purified with a PCR purification kit (QIAGEN, Germany). Cloning was conducted using pGEM-T Easy vector (Promega, USA) and XL1-Blue competent cells (Stratagene, USA). The resultant plasmid DNA was isolated using a Wizard Plus SV Minipreps DNA Purification System (Promega, USA). DNA sequencing was conducted using an ABI PRISM® BigDye® Terminator v3.1 Cycle Sequencing Kit with an ABI 377 Genetic Analyzer (PE Applied Biosystems, USA). All products were sequenced from both strands.

Sequence alignment and phylogenetic analyses

The nucleotide sequence alignment of each gene was conducted using MAFFT ver. 6 (Katoh et al., 2002) with the gap opening penalty set to 1.53 and the offset value (≈ gap extension penalty) set to 0.5. The well-aligned blocks were then selected with GBlocks 0.91b (Castresana, 2000), with the maximum number of contiguous non-conserved positions set to four for COI and EF-1α genes. On the other hand, the 16S rRNA gene was subjected to GBlocks analysis using parameters optimized for rDNA alignments (minimum length of a block as 5, allowing gaps in half position) (Massana et al., 2004). These were utilized either separately or jointly for phylogenetic reconstruction. Additionally, the concatenated first and second codon positions of COI and EF-1α genes and the 16S rRNA gene were also used for phylogenetic reconstruction. The sequences of the first and second codon positions of the COI and EF-1α genes were selected by MEGA 4 (Tamura et al., 2007). These alignments are available upon request. Sequence divergence (%) within subfamily and family for the combined and single gene sequences were calculated via pairwise comparisons using PAUP* ver. 4.0b10 (Swofford, 2002).

Substitution model selection was conducted via comparison of Akaike Information Criterion (AIC) scores (Akaike, 1974), calculated using the ModelTest ver. 3.7 (Posada and Crandall, 1998). In each of the single-gene and combined gene sequences (both including and excluding third codon positions of COI and EF-1α), the GTR (Lanave et al., 1984) + I + G model was selected, and, thus, the selected model and relevant parameters were subjected to Bayesian Inference (BI) and Maximum Likelihood (ML) methods.

The ML analyses 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 four, gamma distribution parameter as “estimated”, and the starting tree as a BIONJ distance-based tree. The confidence values of the ML tree were evaluated via the bootstrap test with 500 iterations. For the BI analysis, each gene was partitioned and the model chosen by ModelTest (Posada and Crandall, 1998) was applied to each unlinked gene. The BI analyses was conducted using MrBayes ver. 3.1 (Huelsenbeck and Ronquist, 2001). Two independent runs of four incrementally heated MCMC chains (one cold chain and three hot chains) were simultaneously run for three - nine million generations depending on the dataset, with sampling conducted every 100 generations. The convergence of MCMC, which was monitored by determining the average standard deviation of split frequencies, was achieved (< 0.01) within three - nine million generations depending on the dataset, and the first 25% of the sampled trees were discarded as burn-in. The confidence values of the BI tree are presented as the Bayesian posterior probabilities in percentages (BPP). Maximum parsimony (MP) analysis was conducted using PAUP* ver. 4.0b10 (Swofford, 2002) via heuristic search using tree-bisection-reconnection (TBR) for a branch-swapping algorithm, steepest descent option not in effect, stepwise addition option for the starting tree, number of trees held at each step during stepwise addition for

Table 3. Within-family sequence divergence (%)

Taxon	No. of species	COI	16S rRNA	EF-1α	Concatenated divergence
Lepidoptera	91				
Papilionoidea	83				
Papilionidae	11	12.67	11.08	12.68	11.43
Papilioninae	8	12.58	11.32	11.27	11.3
Parnassiinae	3	10.11	11.08	9.77	10.23
Pieridae	12	15.86	13.79	14.18	14.2
Pierinae	5	13.49	9.93	11.08	11.46
Coliadinae	5	12.85	9.68	7.70	10.06
Dismorphinae	2	4.47	1.59	0.75	2.26
Lycaenidae *(including Miletinae)	36	18.32	20.14	19.06	18.61
Lycaenidae *(excluding Miletinae)	35	11.57	9.92	17.46	11.75
Lycaeninae	2	5.47	3.35	2.01	3.66
Theclinae	24	10.03	8.05	5.20	8.41
Polyommatinae	9	10.67	8.15	5.20	10.65
Miletinae	1	NA	NA	NA	NA
Nymphalidae	24	15.31	12.64	9.22	12.65
Limenitidinae	2	12.76	3.35	6.27	3.37
Nymphalinae	6	11.21	6.71	6.15	6.85
Satyrinae	1	NA	NA	NA	NA
Apaturinae	6	10.67	8.50	5.08	7.46
Heliconiinae	6	9.66	5.87	5.20	6.82
Pseudergolina	1	NA	NA	NA	NA
Danainae	1	NA	NA	NA	NA
Libytheinae	1	NA	NA	NA	NA
Average within family		15.54	14.41	13.16	14.22

*Sequence divergence of Lycaenidae was calculated including and excluding the Miletinae, because the subfamily represented by a single species, *Taraka hamada*, revealed unusually high sequence divergence from other species of Lycaenidae.
NA, not available

one, and an initial "MaxTrees" setting of 200. The branches were collapsed if the maximum branch length was zero. Trees were evaluated via bootstrapping (Felsenstein, 1985), with 1,000 iterations. Phylogenetic analyses were conducted for each of the gene sequences and the concatenated sequences.

RESULTS AND DISCUSSION

Dataset characteristics

The sequence lengths of the 83 true butterfly and eight skipper species were 1,099 bp in COI, 1,284-1,414 bp in 16S rRNA, and 1,066 bp in EF-1α, respectively, revealing length variations only in the mitochondrial 16S rRNA (Table 1). The concatenated sequences of the three genes ranged from 3,449-3,579 bp among species, and this increased to 3,898 bp when gaps were introduced to improve alignment. However, the conserved blocks selected via GBlocks analysis (Castresana, 2000) eventually provided a total of 3,354 bp, which were subsequently utilized as a concatenated dataset for phylogenetic analysis. This 3,354 bp corresponds to 89% original sequences, composed of 1,095 from COI (99% of original sequences), 1,194 bp from 16S rRNA (68% of original sequences), and 1,065 from EF-1α (100% of original sequences). On the other hand, the concatenated first and second codon positions, excluding the third codon position of COI and EF-1α genes and the 16S rRNA gene, were a total of 2,634 bp in length.

Within-familial sequence divergence ranged from 12.67-18.32% in COI, 11.08-20.14% in 16S rRNA, and 9.22-19.06% in EF-1α (Table 3). On average, the divergence was 15.54% in COI, 14.41% in 16S rRNA, and 13.16% in EF-1α; this indicates the highest level of divergence in the COI gene (Table 3). Among the four butterfly families, the sequence divergence of Lycaeidae was highest (18.32% in COI, 20.14% in 16S rRNA, and 19.06% in EF-1α) among all other butterfly families (Table 3). The fact that family Lycaenidae contains the most butterfly species may be one reason for this (36 species vs. 11-24 species in other families), but the inclusion of a divergent species belonging to the Miletinae subfamily, *Taraka hamada,* may also account for this remarkably high divergence in the Lycaenidae. In fact, the sequence divergence of Lycaenidae dropped from 18.32% to 11.57% in COI and from 20.14% to 9.92% in 16S rRNA without *T. hamada;* these estimates then became lowest among the true butterfly families (Table 3). However, the sequence divergence of EF-1α was still highest among the true butterfly families, even when *T. hamada* was excluded (Table 3). The Miletinae subfamily, which includes *T. hamada,* has been identified as unusual due to its entirely aphytophagous feeding behavior; the species feeds primarily on the Homoptera, but sometimes on ant broods (Corbet and Pendle-bury, 1992; Eliot, 1973). Therefore, the ecological specialty of the subfamily may influence the genetic aspects observed as a profound sequence divergence.

The average A/T contents of COI, 16S rRNA, and EF-1α were, respectively, 71.3%, 84.3%, and 49.2% (Table 1). This result demonstrates A/T bias in the mitochondrial genes, but not in the nuclear EF-1α. Between the two mitochondrial genes,

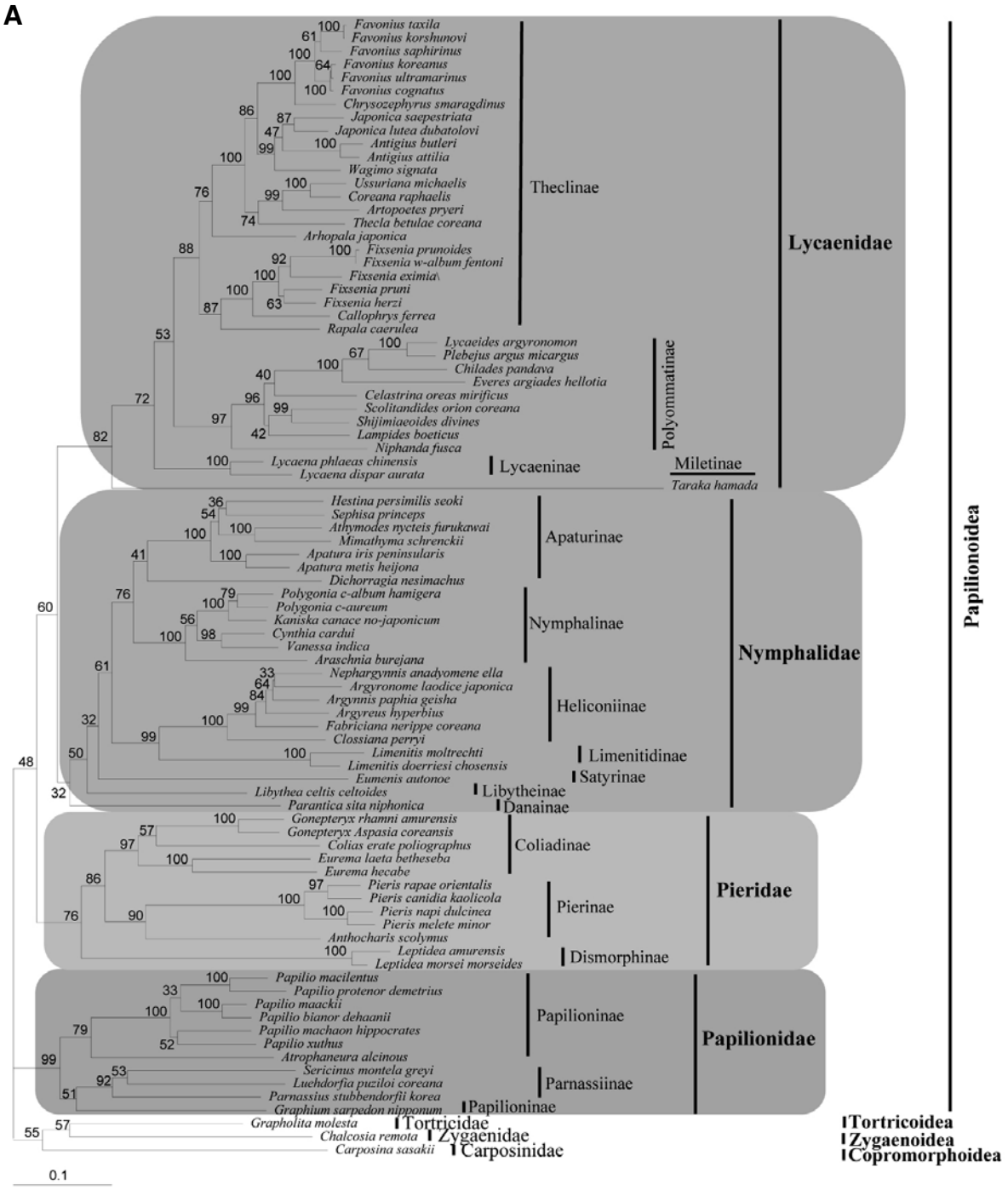
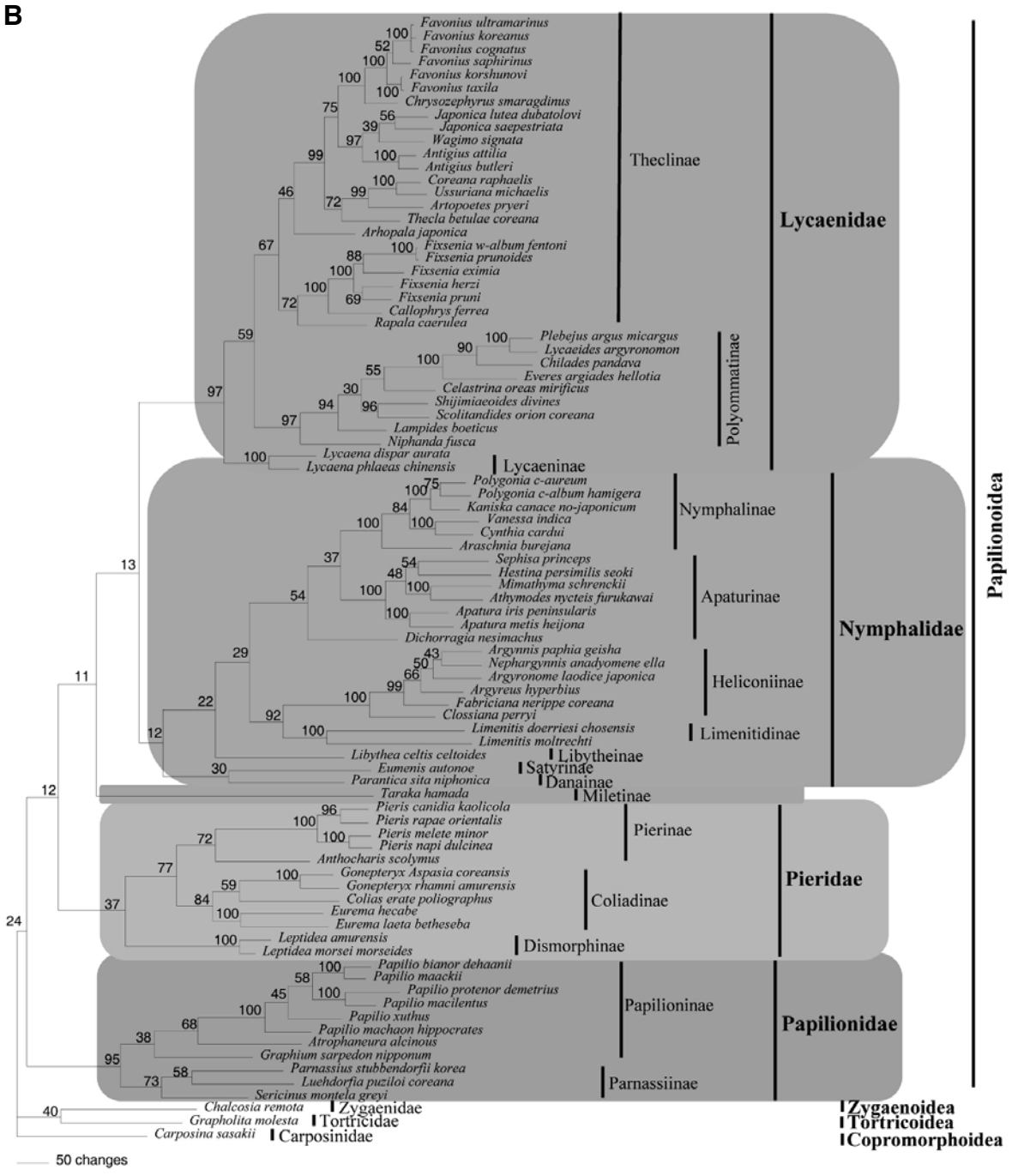


Fig. 2. Phylogenetic analyses using the concatenated mitochondrial COI, 16S rRNA, and nuclear EF-1α genes. Three species of moths were used as outgroups. (A) Tree generated via maximum likelihood analysis. The numbers at each node specify bootstrap percentage of 500 pseudoreplicates. The log likelihood value of the best tree is -54788.11. (B) Tree generated via maximum parsimonious analysis. Numbers at each node specify bootstrap percentage of 1,000 replications. The tree length is 11,560 steps, consistency index is 0.21, retention index is 0.56, and homoplasy index is 0.79. (C) Tree generated via Bayesian analysis. Average likelihood was -53629.43 and numbers at each node specify BPP.

the 16S rRNA gene evidenced a significantly higher A/T content than was detected in the COI gene. A strong A/T bias in mitochondrial DNA has also recently been reported in many mitochondrial genome studies, including studies of the Lepidoptera (Kim et al., 2006; 2009). The A/T contents of the protein coding genes were measured as approximately 76-82%, and that of 16S rRNA was approximately 81-85% in the 78-83% of A/T content in the full-length mitochondrial genomes (Kim et al.,

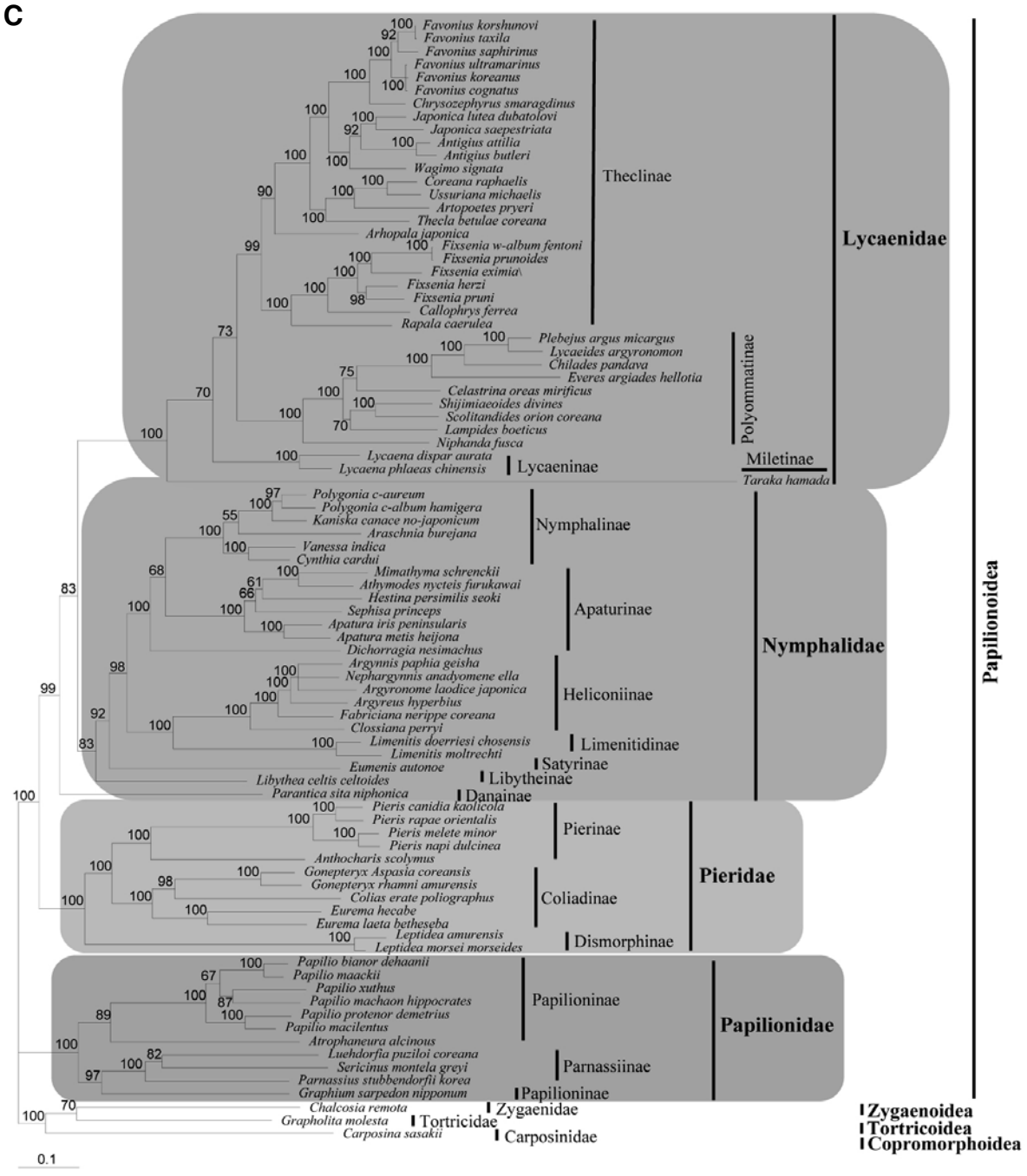


(continued)

2009), thus indicating somewhat higher A/T content in 16S rRNA.

Monophyly of true butterfly families

A number of phylogenetic analyses using individual COI, 16S rRNA, and EF-1α sequences have been conducted. However, the resolving power of the single genes appears insufficient, in that unexplainable splitting among members of the same family, extremely low node support at many branches, and/or both prevailed (Supplementary Figs. 1, 2, and 3). This may have principally been the consequence of a shortage of the characteristics required to resolve several levels of taxonomic hierarchy. A similar phenomenon has been frequently noted in other published phylogenetic studies. Thus, in the field of insect phylogeny, there is a strong trend toward the use of many more molecular markers than ever before (i.e., Savard et al., 2006). In the case of butterfly studies, the combination of both mitochondrial and nuclear genes is fairly common (Caterino et al., 2000; 2001; Kandul et al., 2004; Monteiro and Pierce, 2001; Wahlberg et al., 2003; 2005a; Zakharov et al., 2004). Therefore, we limited our data presentation and our discussion of phylogenetic reconstruction to the trees generated on the basis of the



(continued)

three concatenated gene sequences.

Via ML analysis, each monophyletic Lycaenidae, Pieridae, and Papilionidae were recovered, with node support at 82%, 76%, and 99%, respectively; however, the monophyletic Nymphalidae evidenced poor recovery, with node support at 32% (Fig. 2A). Unlike the results generated by the ML analysis, the MP tree recovered the Pieridae, Papilionidae, and Nymphalidae as monophyletic groups, with node support at 37%, 95%, and 12%, respectively (Fig. 2B). On the other hand, the monophyly of Lycaenidae was hampered due to the placement of the Miletinae, represented by the single species, *Taraka hamada*, as the basement of the group encompassing the Lycaenidae and Nymphalidae, rather than as the basal lineage of the Lycaenidae alone (Fig. 2B). Via BI analysis, the monophylies of the Lycaenidae, Pieridae, and Papilionidae were strongly supported (100%), but the monophyletic Nymphalidae was not recovered by the placement of the nymphalid subfamily Danainae, represented by the single species, *Parantica sita niphonica*, to the basement of the group encompassing the Lycaenidae and Nymphalidae (Fig. 2C). Thus, the monophylies of the Lycaenidae and Nymphalidae were not demonstrated on MP and BI analysis, respectively, although the monophylies of

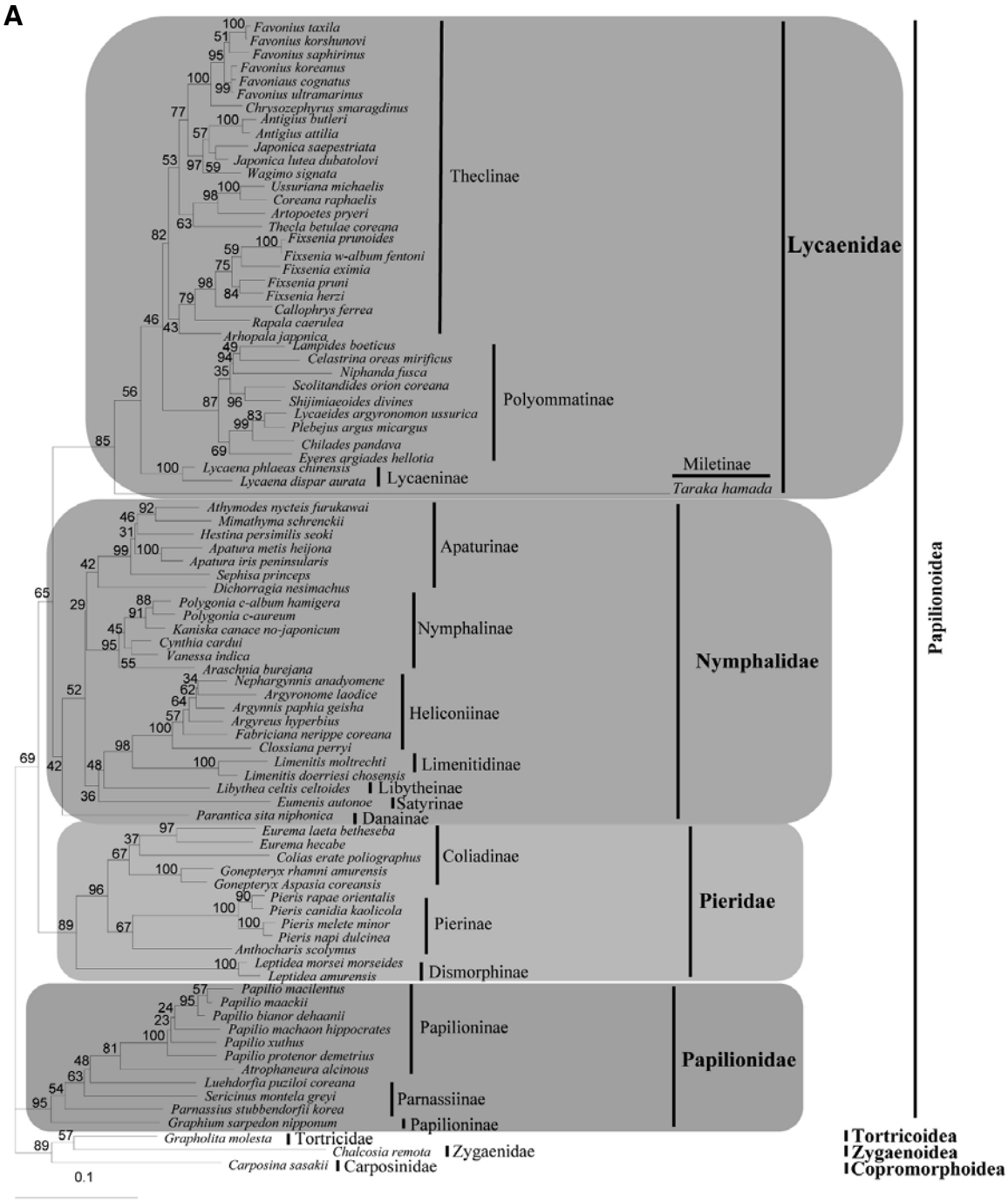
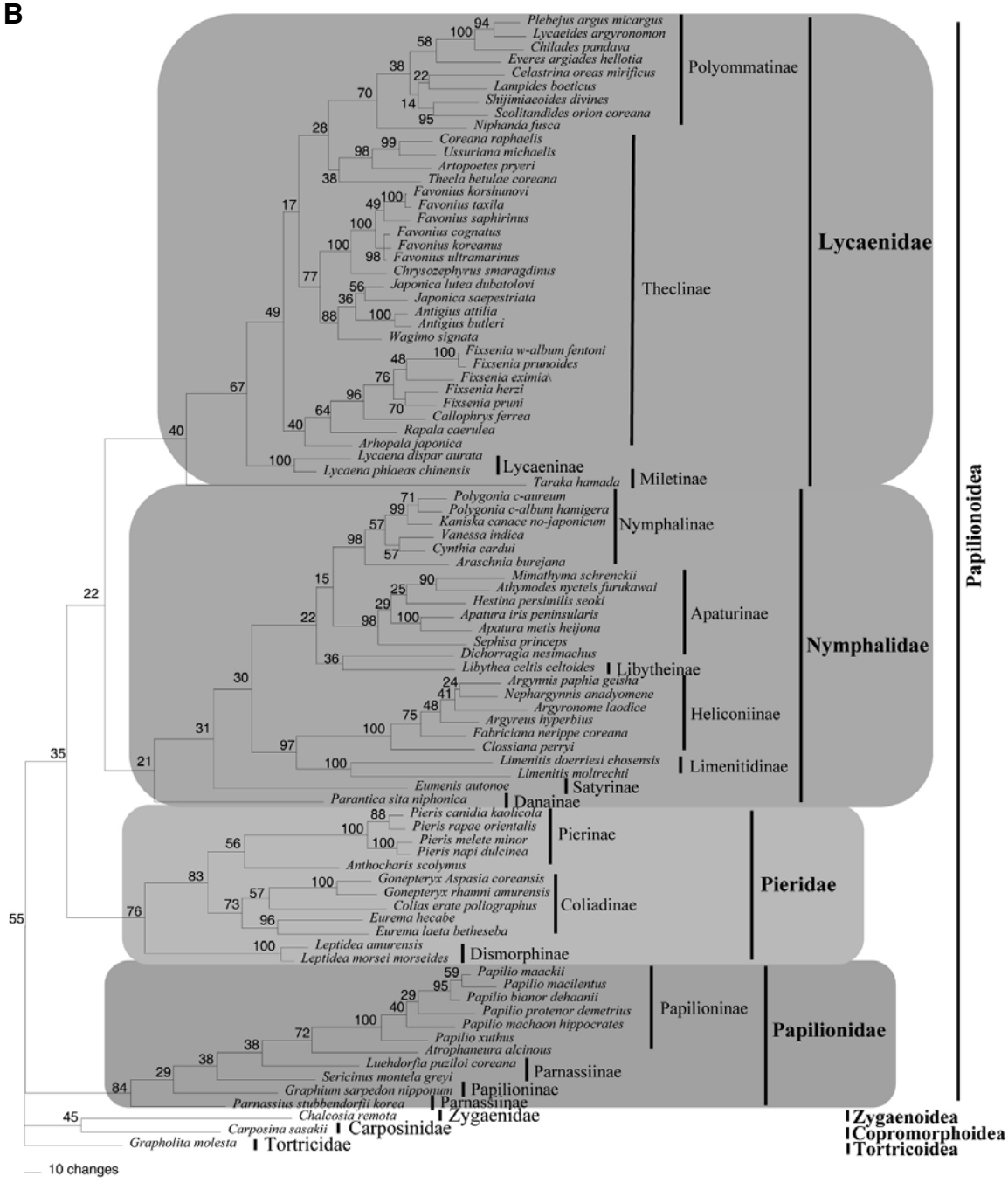


Fig. 3. The phylogenetic analyses using the concatenated first and second codon positions of mitochondrial COI and nuclear EF-1α genes, and 16S rRNA. Three spe-cies of moths were used as outgroups. (A) Tree generated via maximum likelihood analysis. The numbers at each node specify boot-strap percentages of 500 pseudo-replicates. The log likelihood value of the best tree is -25614.43. (B) Tree generated via maximum parsimony analysis. Numbers at each node specify bootstrap percentage of 1,000 replications. The tree length is 4,655 steps, consistency index is 0.27, retention index is 0.62, and homoplasy index is 0.73. (C) Tree generated via Bayesian analysis. Average likelihood was -24964.45 and numbers at each node specify BPP.

the Pieridae and the Papilionidae were consistently well supported. With regard to the lycaenid subfamily Miletinae, it should be noted that the subfamily, represented by the single species, *T. hamada*, generated the longest branch among all phylogenetic analyses (Fig. 2) and exhibited an unusually high sequence divergence from other members of the lycaenid species (Table 3). It appears likely that these aspects may have affected the placement of the subfamily at its unconventional position in the MP analysis (Fig. 2B).

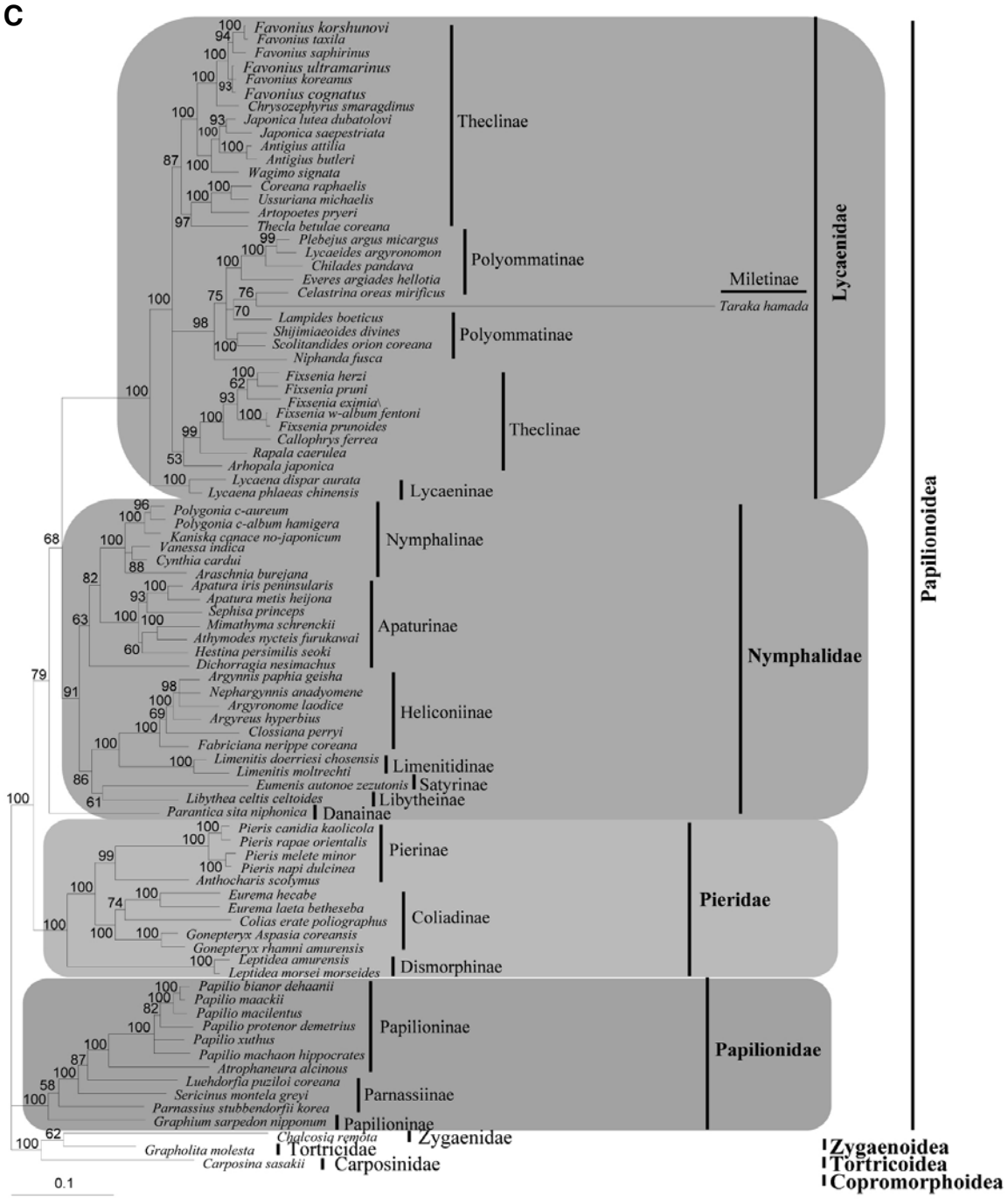
When the dataset excluded the third codon position of COI



(continued)

and EF-1α, overall node support was increased in all analyses (Fig. 3). ML analysis showed monophyly for all true butterfly families, but the node support for the monophyletic Nymphalidae was very low (42%) and the problematic lycaenid subfamily Miletinae, represented by *T. hamada*, continued to generate the longest branch (Fig. 3A). Nevertheless, this topology is basically the same as that obtained from the whole dataset via the ML method (Fig. 2A). According to the MP analysis, all true butterfly families also evidenced monophyly (Fig. 3B). The aspect of the MP tree that differs from that of the whole dataset was the recovery of monophyletic Lycaenidae in this analysis, but the node support for this family is quite low (40%). According to the results of BI analysis, the monophylies of all families but the Nymphalidae were recovered (Fig. 3C). The placement of the nymphalid subfamily Danainae to the basal position of the group encompassing the Lycaenidae and Nymphalidae was obtained as seen in the tree obtained from the whole dataset via BI analysis (Fig. 2C).

In order to clarify the phylogenetic stability of the majority bodies of each of the true butterfly families, the lycaenid *T. hamada* belonging to the Miletinae and the nymphalid *P. s. niphonica* belonging to the Danainae, which caused fluctuating



(continued)

positions in some analyses--thus hampering the monophyly of Lycaenidae and Nymphalidae--were excluded in subsequent analyses. The phylogenetic analysis using the whole dataset supported the stable identification of each Lycaenidae and Nymphalidae as monophyletic groups in all analyses, with strong node support--100% in all analyses for the Lycaenidae (Fig. 4) and 72% and 96% by BI analysis for the Nymphalidae (Figs. 4A and 4C). However, the result of MP analysis provided relatively low node support, 49%, for the monophyly of the Nymphalidae (Fig. 4B). When the dataset excluded the third codon position of COI and EF-1α, nearly identical topology and node support were obtained in all analyses to those obtained from the whole dataset (Supplementary Fig. 4).

The identification as a monophyletic group of the lycaenid butterflies, including the subfamily Miletinae, has been accepted in several recent studies. For example, Campbell et al. (2000) yielded a well-resolved topology, supporting the monophyly of the Lycaenidae with strong node support by parsimony analysis using the wingless gene. This result was consistent with the findings of the concatenated molecular and morpho-

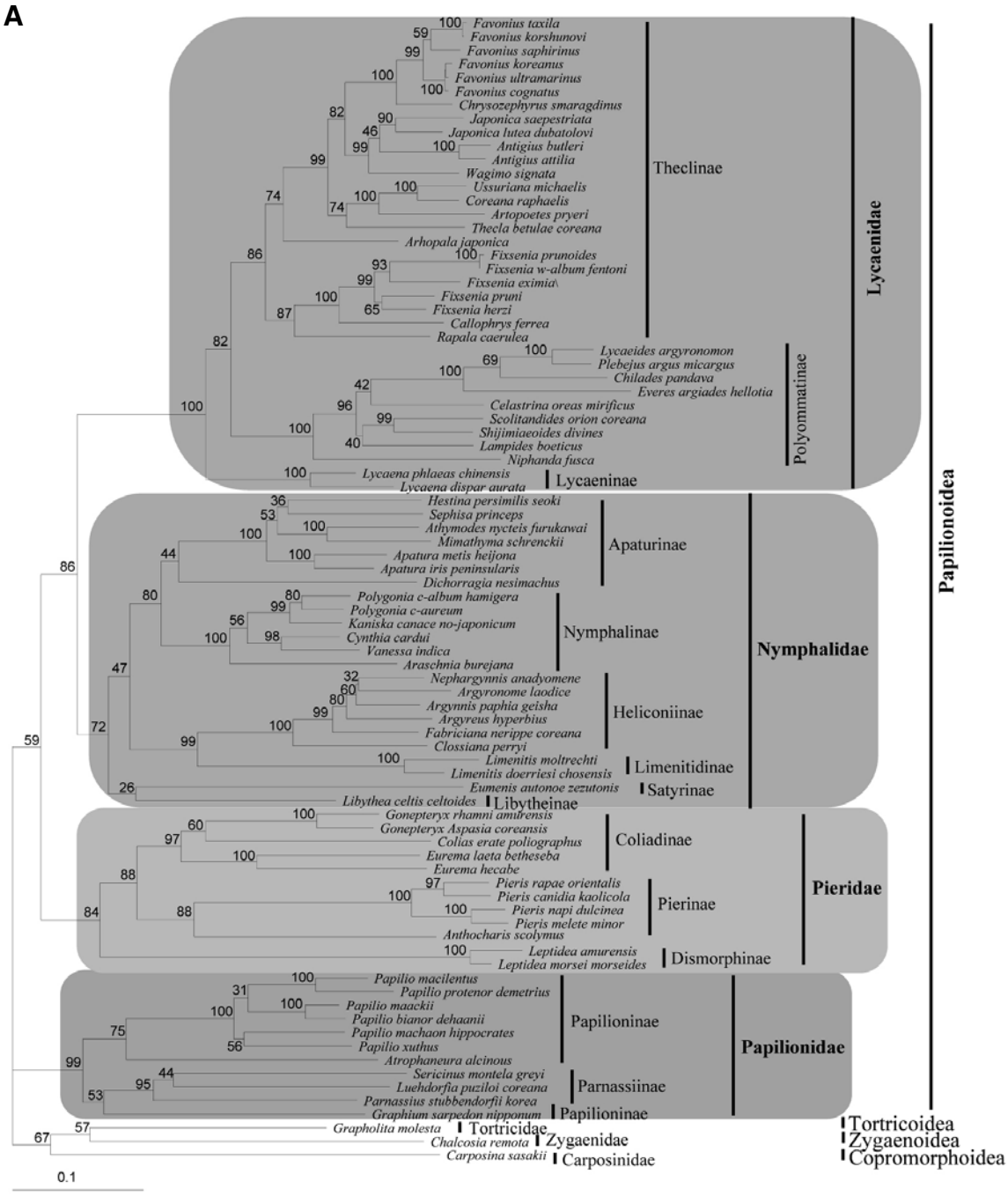
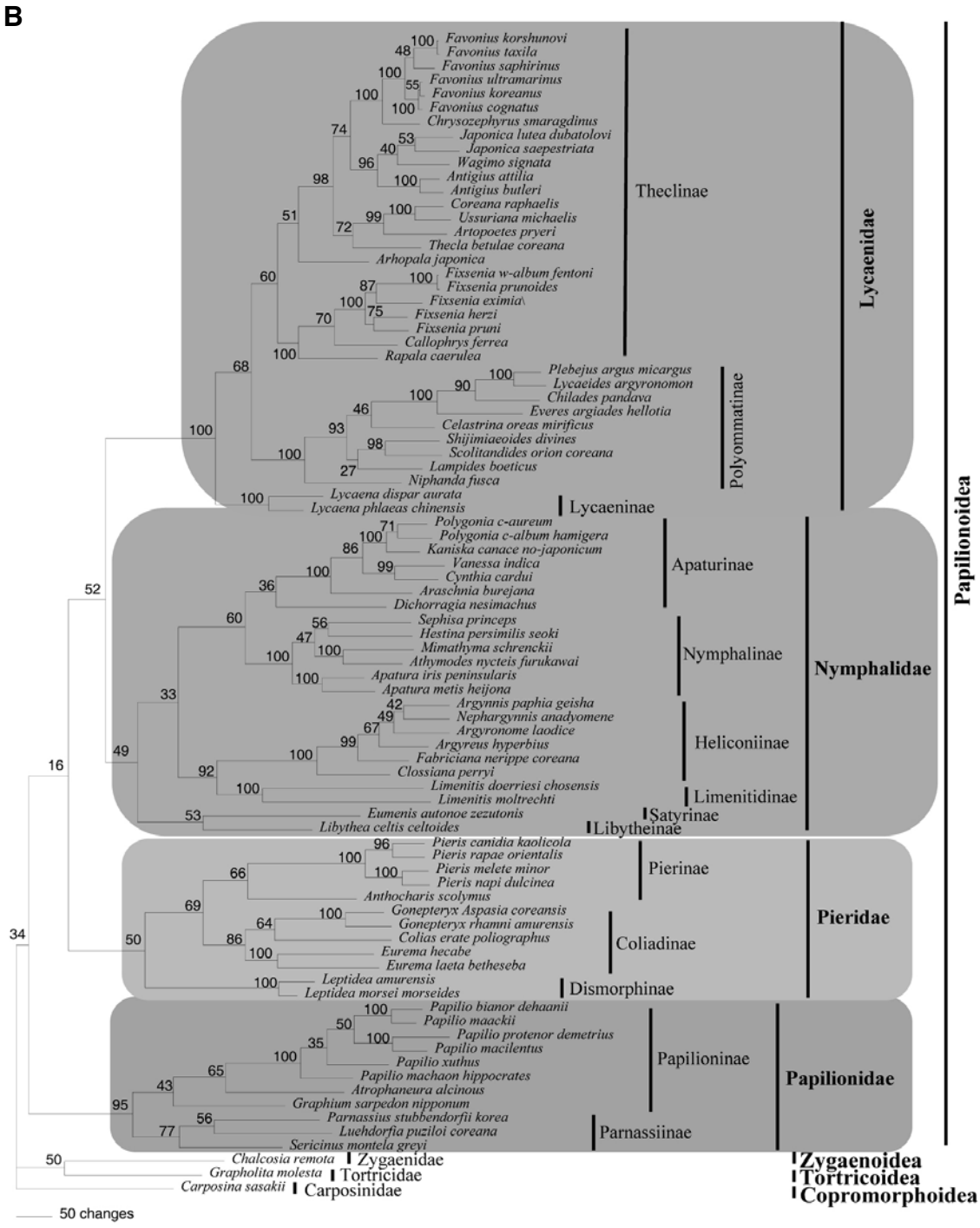


Fig. 4. The phylogenetic ana-lyses using the concatenated mitochondrial COI, 16S rRNA, and nuclear EF-1α genes, excluding the included single species from the subfamilies Danainae and Miletinae, which evidenced fluctuating posi-tions in Fig. 2. Three species of moths were used as out-groups. (A) Tree generated via maximum likelihood analy-sis. The numbers at each node specify bootstrap per-centag es of 500 pseudoreplicates. The log likelihood value of the best tree is -52481.95. (B) Tree generated via maximum parsimony analysis. Numbers at each node specify boot-strap percentage of 1,000 replications. The tree length is 10,843 steps, consistency index is 0.22, retention index is 0.58, and homoplasy index is 0.78. (C) Tree generated via Bayesian analysis. Average likelihood was -51732.64 and numbers at each node specify BPP.

logical studies (Wahlberg et al., 2005a) and with previous analyses of morphological characteristics (Hauser, 1993; Jordan, 1898). In the case of Nymphalidae, its monophyly has been supported by the identification of several morphological characteristics, including the ventromesial surface of the antennae with three longitudinal ridges, separating two continuous sulci or a pair of shallow depressions on each segment (Jordan, 1898) and the female abdomen with the von Siebold organ (Häuser, 1993). Molecular evidence further supports the notion of a monophyletic Nymphalidae (Brower, 2000; Freitas and Brown, 2004; Weller et al., 1996; Wahlberg et al., 2005b).

In summary, the traditionally recognized true butterfly families

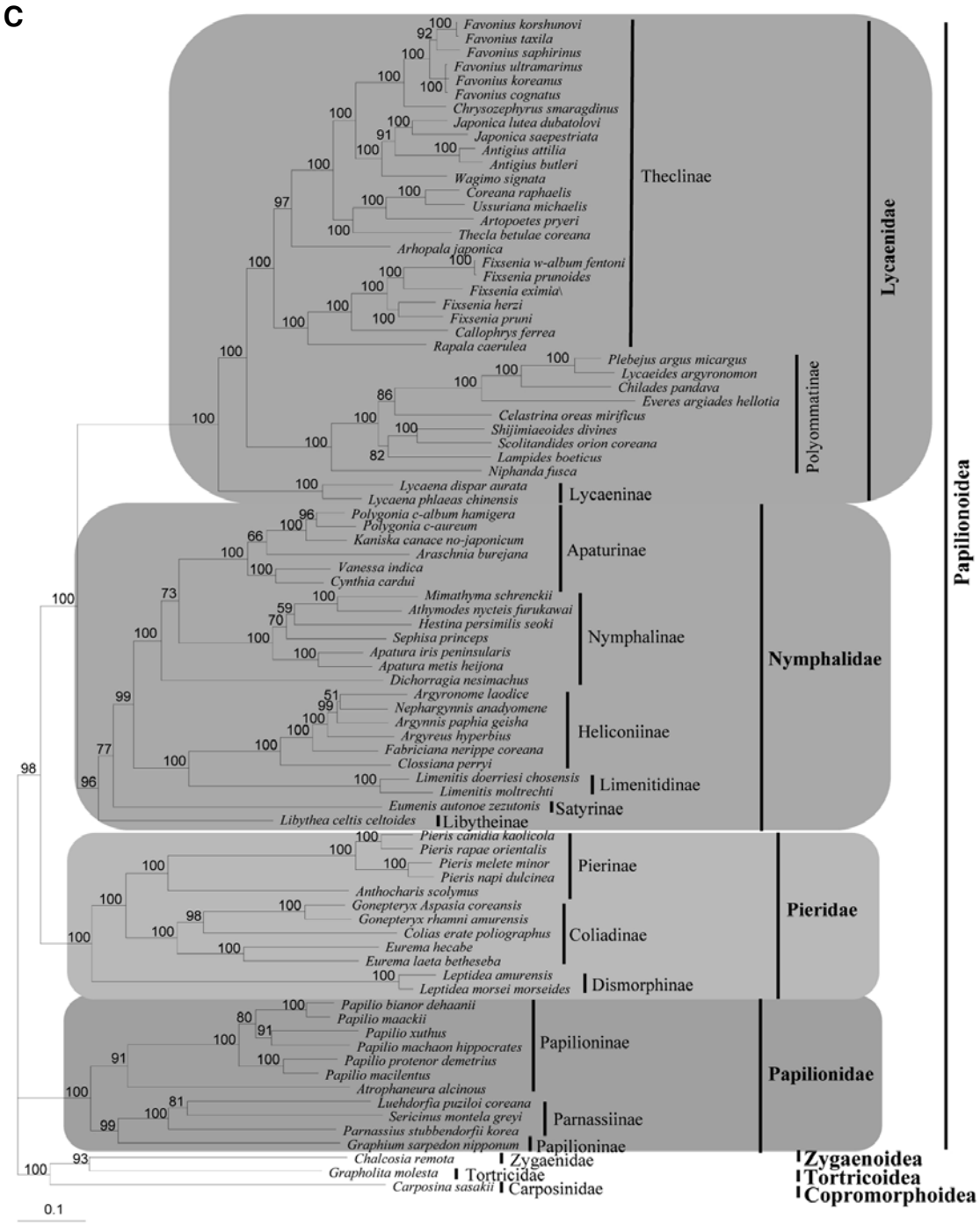


(continued)

were well recovered overall, each occurring as a monophyletic group using the three concatenated gene sequences. The lycaenid Miletinae and the nymphalid Danainae are required for further clear conclusions regarding their phylogenetic positions within each family, as well as the monophyly of the Lycaenidae and the Nymphalidae.

Phylogenetic relationships among the true butterfly families

Phylogenetic analysis conducted via ML and BI methods using total sequence data identified the Nymphalidae and Lycaenidae as sisters, with 60% bootstrap support by the ML method and 83% BPP by the BI method (Figs. 2A and 2C), whereas no sister group relationship was identified by the MP method by the placement of the lycaenid Miletinae as sister to the group composed of Nymphalidae and the remainder of the Lycaenidae (Fig. 2B). Nevertheless, in ML, MP, and BI analyses, the Lycaenidae + Nymphalidae group was consistently placed as a sister to the Pieridae, identifying the Papilionidae as the basal lineage of other true butterfly families (Fig. 2). When the dataset



(continued)

excluded the third codon position of COI and EF-1α, all methods also placed the Pieridae as sister to the Lycaenidae + Nymphalidae group, placing the Papilionidae as the basal lineage of other true butterfly families (Fig. 3). The analysis excluding the lycaenid subfamily Miletinae and the nymphalid subfamilies Danainae also consistently evidenced the relationship: (((Lycaenidae + Nymphalidae) + Pieridae) + Papilionidae) (Fig. 4); the same relationships were also recovered when the dataset excluded the third codon position of COI and EF-1α with the ingroup scheme (Suplementary Fig. 4). Thus, all our phylogenetic analyses consistently supported a topology consistent with Kristensen's hypothesis (1976): (((Lycaenidae + Nymphalidae) + Pieridae) + Papilionidae).

The sister relationship of the Nymphalidae and Lycaenidae has been well supported by the conspicuous differentiation of a unique secondary sclerite behind the metascutellum and a secondary sternopleural suture within the mesothorax (Brock, 1971; 1990) as well as by the combination of the relevant morphological and molecular characteristics (Wahlberg et al., 2005a; Weller et al., 1996); however, the relationship of the

remaining true butterfly families to this group has been a source of some controversy. Scott (1985) has demonstrated a sister relationship between the Nymphalidae and Lycaenidae on one hand, and a sister relationship between the Pieridae and Papilionidae on another hand, thus supporting the relationship: (Nymphalidae + Lycaenidae) + (Pieridae + Papilionidae). The sister relationship of the Papilionidae and Pieridae was largely supported on the basis of the adult musculature and eyes (Scott and Wright, 1990) as well as pupal structures such as the peculiar arrangement of proboscis extensor muscles (Schmitt, 1938). On the other hand, Kristensen (1976) has demonstrated a close relationship between the Nymphalidae and Lycaenidae groups with the Pieridae on the basis of morphological characteristics, including the absence of a fore-tibial epiphysis, loss of the prospinasterno-procoxal muscle, unsegmented or absent maxillary pulp, and a secondary sclerite behind the metascutellum. However, these morphological characteristics were controversial because they are either found in other families, confirmed for only limited number of taxa, or more distinct in other families (Ackery et al., 1999). Nevertheless, the hypothesis of Kristensen (1976) was supported by a combination of morphological and molecular characteristics (Wahlberg et al., 2005a; Weller et al., 1996) and--in this study only--consistently by molecular characteristics.

Wahlberg et al. (2005a) has reported previously that the solely molecular-data based phylogeny yielded several unconventional relationships, and most nodes were supported very weakly by parsimony analysis. Further, they detected several clades that had never been previously suggested in the literature by Bayesian methodology. This includes the placement of the Papilionidae as a sister to the rest of the butterflies, skippers, and hedylids with the maximum BPP, and also the placement of the nymphalid satyrine clade as a sister to the Riodinidae + Lycaenidae clade with the maximum BPP. These results were attributed to large quantities of homoplasious characteristics in the molecular data, and signified the crucial role of the morphological data in the phylogenetic reconstruction of butterflies (Wahlberg et al., 2005a). In fact, the inclusion of morphological data combined with the molecular data strongly supports Kristensen's hypothesis (((Nymphalidae + Lycaenidae) + Pieridae) + Papilionidae) (Wahlberg et al., 2005a). On the other hand, it has been suggested previously that the morphological phylogeny data also play a limited role (Scotland et al., 2003). Wortley and Scotland (2003) have shown that the addition of a morphological dataset to an existing molecular dataset had a significantly positive effect on the resolution, thereby increasing the number of supported clades, but not upon support. This was attributed to the conflict between the two data types, resulting in a reduction in support at some nodes (Aagesen and Sanso, 2003; Scotland et al., 2003). Nazari et al. (2007) previously evaluated the utility of the morphological data (236 characters) and seven mitochondrial and nuclear genes (a total of 5,775 bp) in a phylogenetic reconstruction of swallowtail butterflies belonging to the Parnassiinae subfamily. They determined that the combined morphological and molecular data yielded the same topology to the tree obtained from only the combined molecular data under the parsimony criteria, and also showed no increase in the overall node support provided by the combined morphological and molecular data (Nazari et al., 2007). Taking these results into consideration, the molecular data alone still appear to constitute an important source of evolutionary signals for the phylogenetic reconstruction of butterflies, although the extent to which the molecular data solely influence tree construction may differ according to taxonomic level. In summary, phylogenetic analysis of the true butterfly families (Papilionidae, Pieridae, Lycaenidae, and Nymphalidae) solely using molecular data consistently supported a topology consistent with Kristensen's hypothesis (1976): (((Lycaenidae + Nymphalidae) + Pieridae) + Papilionidae).

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

ACKNOWLEDGMENTS

This work was supported by a research grant from National Institute of Biological Resources, "Origin of Biological Diversity of Korea: Molecular Phylogenetic Analyses of Major Korean Taxa" awarded to Iksoo Kim.

REFERENCES

Aagesen, L., and Sanso, A.M. (2003). The phylogeny of the Alstroemeriaceae, based on morphology, *rps* 16 intron, and *rbc*L sequence data. Syst. Bot. *28*, 47-69.

Ackery, P.R. (1984). Systematic and faunistic studies on butterflies. In The Biology of Butterflies, R.I. Vane-Wright, and P.R. Ackery, eds. (London, United Kingdom: Academic Press), pp. 9-21.

Ackery, P.R., de Jong, R., and Vane-Wright, R.I. (1999). The butterflies: Hedyloidea, Hesperioidea, and Papilionoidea. In Lepidoptera: Moths and Butterflies. 1. Evolution, systematics, and biogeography, Handbook of Zoology Vol. IV, Part 35, N.P. Kristensen, ed. (Berlin and New York: De Gruyter), pp. 264-300.

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

Brower, A.V.Z. (2000). Phylogenetic relationships among the Nymphalidae (Lepidoptera), inferred from partial sequences of the wingless gene. Proc. Biol. Sci. *267*, 1201-1211.

Brock, J.P. (1971). A contribution towards an understanding of the morphology and phylogeny of the ditrysian Lepidoptera. J. Nat. Hist. *5*, 29-102.

Brock, J.P. (1990). Origins and phylogeny of butterflies. In Butterflies of Europe 2, O. Kudrna, ed. (Wiesbaden, Germany: Aula Verlag), pp. 209-233.

Campbell, D.L., Brower, A.V.Z., and Pierce, N.E. (2000). Molecular evolution of the wingless gene and its implications for the phylogenetic placement of the butterfly family Riodinidae (Lepidoptera: Papilionoidea). Mol. Biol. Evol. *17*, 684-696.

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

Caterino, M.S., Cho, S., and Sperling, F.A.H. (2000). The current state of insect molecular systematics: a thriving Tower of Babel. Annu. Rev. Entomol. *45*, 1-54.

Caterino, M.S., Reed, R.D., Kuo, M.M., and Sperling, F.A.H. (2001). A partitioned likelihood analysis of swallowtail butterfly phylogeny (Lepidoptera: Papilionidae). Syst. Biol. *50*, 106-127.

Corbet, A.S., Pendlebury, H.M., and Eliot, J.N. (1992). The butterflies of the Malay Peninsula (Kuala Lumpur: Malayan Nature Society).

de Jong, R., Vane-Wright, R.I., and Ackery, P.R. (1996). The higher classification of butterflies (Lepidoptera): problems and prospects. Entomol. Scand. *27*, 65-101.

Ehrlich, P.R. (1958). The comparative morphology, phylogeny and higher classification of the butterflies (Lepidoptera: Papilionoidea). Kanas Univ. Sci. Bull. *39*, 305-370.

Ehrlich, P.R., and Ehrlich, A.H. (1967). The phenetic relationships of the butterflies. I. Adult taxonomy and the non-specificity hypothesis. Syst. Zool. *16*, 301-317.

Eliot, J.N. (1973). The higher classification of the Lycaenidae (Lepidoptera): a tentative arrangement. Bull. Br. Mus. Nat. Hist. Entomol. *28*, 371-505.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. Evolution *39*, 783-791.

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.

Freitas, A.V.L., and Brown, K.S. (2004). Phylogeny of the Nymphalidae (Lepidoptera). Syst. Biol. *53*, 363-383.

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.
Hajibabaei, M., Singer, G.A.C., and Hickey, D.A. (2006). Benchmarking DNA barcodes: an assessment using available primate sequences. Genome *49*, 851-854.
Harvey, D.J. (1987). The higher classification of the Riodinidae (Lepidoptera). Ph.D. thesis, University of Texas, Austin.
Häuser, C.L. (1993). The internal female genital organs in butterflies (Rhopalocera): comparative morphology and phylogenetic interpretation (Insecta, Lepidoptera). Zoologische Jahrbücher, Abteilung für Systematik *120*, 389-439.
Huelsenbeck, J.P., and Ronquist, F. (2001). MrBayes: Bayesian inference of phylogeny. Bioinformatics *17*, 754-755.
Joo, H.Z., Kim, S.S., and Sohn, J.D. (2005). Butterflies of Korea. (Seoul, Korea: Kyo-Hak Publishing Co., Ltd), pp. 402-411.
Jordan, K. (1898). Contributions to the morphology of Lepidoptera. Novit. Zool. *5*, 374-415.
Kandul, N.P., Lukhtanov, V.A., Dantchenko, A.V., Coleman, J.W.S., Sekercioglu, C.H., Haig, D., and Pierce, N.E. (2004). Phylogeny of *Agrodiaetus* Hübner 1822 (Lepidoptera: Lycaenidae) inferred from mtDNA sequences of COI and COII and nuclear sequences of EF-1α: Karyotype diversification and species radiation. Syst. Biol. *53*, 278-298.
Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059-3066.
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, M.I., Baek, J.Y., Kim M.J., Jeong, H.C., Kim, K.G., Bae, C.H. Han, Y.S., Jin, B.R., and 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.
Kristensen, N.P. (1976). Remarks on the family-level phylogeny of butterflies (Insecta, Lepidoptera, Rhopalocera). Zool. Syst. Evol. Res. *14*, 25-33.
Lanave, C., Preparata, G., Saccone, C., and Serio, G. (1984). A new method for calculating evolutionary substitution rates. J. Mol. Evol. *20*, 86-93.
Lin, C.P., and Danforth, B.N. (2004). How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. Mol. Phylogenet. Evol. *30*, 686-702.
Martin, J.A., and Pashley, D.P. (1992). Molecular systematic analysis of butterfly family and some subfamily relationships (Lepidoptera: Papilionidea). Ann. Ent. Soc. Am. *85*, 127-135.
Massana, R., Castresana, J., Balagué, V., Guillou, L., Romari, K., Groisillier, A., Valentin, K., and Pedrós-Alió, C. (2004). Phylogenetic and ecological analysis of novel marine stramenopiles. Appl. Environ. Microbiol. *70,* 3528-3534.
Monteiro, A., and Pierce, N.E. (2001). Molecular phylogeny of *Bicyclus* butterflies (Satyridae) using COI, COII and EF1α. Mol. Phylogenet. Evol. *18*, 264-281.
Nazari, V., Zakharov, E.V., and Sperling, F.A. (2007). Phylogeny, historical biogeography, and taxonomic ranking of Parnassiinae (Lepidoptera, Papilionidae) based on morphology and seven genes. Mol. Phylogenet. Evol. *42*, 131-156.
Niehuis, O., Yen, S.H., Naumann, C.M., and Misof, B. (2006). Higher phylogeny of zygaenid moths (Insecta: Lepidoptera) inferred from nuclear and mitochondrial sequence data and the evolution of larval cuticular cavities for chemical defence. Mol. Phylogenet. Evol. *39*, 812-829.
Posada, D., and Crandal, K.A. (1998). Modeltest: testing the model of DNA substitution. Bioinformatics *14*, 817-818.
Robbins, R.K. (1982). How many butterfly species? News Lepid. Soc. *1982*, 40-41.
Robbins, R.K. (1988). Comparative morphology of the butterfly foreleg coxa and trochanter (Lepidoptera) and its systematic implications. P. Entomol. Soc. Wash. *90*, 133-154.
Savard, J., Tautz, D., Richards, S., Weinstock, G.M., Gibbs, R.A., Werren, J.H., Tettelin, H., and Lercher, M.J. (2006). Phylogenomic analysis reveals bees and wasps (Hymenoptera) at the base of the radiation of Holometabolous insects. Genome Res. *16*, 1334-1338.
Schmitt, J.B. (1938). The feeding mechanism of adult Lepidoptera. Smithson Misc. Coll. *97*, 1-28.
Scotland, R.W., Olmstead, R., and Bennett, J.R. (2003). Phylogeny reconstruction: the role of morphology. Syst. Biol. *52*, 539-548.
Scott, J.A. (1985). The phylogeny of butterflies (Papilionoidea and Hesperioidea). J. Res. Lepid. *23*, 241-281.
Scott, J.A., Wright, D.M., and Kudrna, O. (1990). Butterfly phylogeny and fossils. Butterflies of Europe: Introduction to lepidopterology *2*, 152-208.
Smart, P. (1989). The illustrated encyclopedia of the butterfly world (New York, USA: Chartwell Books).
Strimmer, K., and Rambaut, A. (2002). Inferring confidence sets of possibly misspecified gene trees. Proc. Biol. Sci. *269*, 137-142.
Swofford, D.L. (2002). PAUP*. Phylogenetic analysis using parsimony (*and Other Methods) ver 4.0b10 (Sunderland, Massachusetts, USA: Sinauer Associates).
Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) software ver 4.0. Mol. Biol. Evol. *24*, 1596-1599.
Wahlberg, N., Weingartner, E., and Nylin, S. (2003). Towards a better understanding of the higher systematic of Nymphalidae (Lepidoptera: Papilionidae). Mol. Phylogenet. Evol. *28*, 473-484.
Wahlberg, N., Braby, M.F., Brower, A.V.Z., de Jong, R., Lee, M.-M., Nylin, S., Pierce, N.E., Sperling, F.A.H., Vila, R., Warren, A.D., et al. (2005a). Synergistic effects of combining morphological and molecular data in resolving the phylogeny of butterflies and skippers. Proc. Biol. Sci. *272*, 1577-1586.
Wahlberg, N., Brower, A.V.Z., and Nylin, S. (2005b). Phylogenetic relationships and historical biogeography of tribes and genera in the subfamily Nymphalinae (Lepidoptera: Nymphalidae). Biol. J. Linn. Soc. *86*, 227-251.
Weller, S.J., Pashley, D.P., and Martin, J.A. (1996). Reassessment of butterfly family relationships using independent genes and morphology. Ann. Entomol. Soc. Am. *89*, 184-192.
Zakharov, E.V., Caterino, M.S., and Sperling, F.A.H. (2004). Molecular phylogeny, historical biogeography, and divergence time estimates for swallowtail butterflies of the genus *Papilio* (Lepidoptera: Papilionidae). Syst. Biol. *53*, 193-215.