

# The mitochondrial genome of the butterfly *Papilio xuthus* (Lepidoptera: Papilionidae) and related phylogenetic analyses

Xia Feng · Dian-Feng Liu · Nai-Xin Wang ·  
Chao-Dong Zhu · Guo-Fang Jiang

Received: 27 December 2009 / Accepted: 24 February 2010 / Published online: 8 March 2010  
© Springer Science+Business Media B.V. 2010

**Abstract** The nearly complete mitochondrial genome of the butterfly *Papilio xuthus* (Lepidoptera: Papilionidae) was sequenced for its nucleotide sequence of 13,964 bp. The genome has a typical gene order identical to other lepidopteran species. All tRNAs showed same stable canonical clover-leaf structure as those of other insects, except for tRNA<sup>Ser</sup> (AGN), in which the dihydrouracil arm (DHU arm) could not form stable stem-loop structure. Anomalous initiation codons have been observed for the *cox1* gene, where the ATTACG hexa-nucleotide was believed to be involved in the initiation signaling. Twelve mitochondrial protein-coding gene sequence data were used to infer the phylogenetic relationships among the insect orders. Even though the number of insect orders represented by complete mitochondrial genomes is still limited, several well-established relationships are evident in the phylogenetic analysis of the complete sequences. Monophyly of the Homometabola was not supported in this paper. Phylogenetic analyses of the available species of Bombycoidea, Pyraloidea, Papilionoidea and Tortricidea bolstered the current morphology-based hypothesis that Bombycoidea, Pyraloidea and Papilionoidea are monophyletic (Obtectomera). Bombycoidea (*Bombyx mandarina*

and *Antheraea pernyi*) and Papilionoidea (*P. xuthus* and *Coreana raphaelis*) formed a sister group.

**Keywords** *Papilio xuthus* · Mitochondrial genome · Molecular phylogeny · Papilionidae · Lepidoptera

## Abbreviations

<i>atp6</i> and <i>atp8</i>	ATPase subunits 6 and 8
bp	Base pair(s)
<i>cox1–3</i>	Cytochrome oxidase subunits I–III
<i>cob</i>	Cytochrome <i>b</i> gene
<i>nad1–6</i> and <i>4L</i>	NADH dehydrogenase subunits 1–6 and 4L genes
rRNA	Ribosomal RNA
tRNA	Transfer RNA
PCG	Protein coding gene

## Introduction

Mitochondrial genomes (mitogenomes) form units of genetic information, evolving independently from nuclear genomes. There are many unique features for mitogenomes, which include their small sizes, fast evolutionary rates, relatively conserved gene contents and organization, maternal inheritance and limited recombination events [1–4]. Thus, mitogenomes offer a broad range of characters to study phylogenetic relationships of animal taxa. Besides nucleotide and amino acid sequences, several features of mitogenomes have been successfully used as various sources for phylogenetic inferences, which include tRNA secondary structures [5], deviations from the universal genetic code [6, 7], as well as changes in the mitochondrial gene orders [8, 9].

X. Feng · D.-F. Liu · N.-X. Wang · G.-F. Jiang (✉)  
Jiangsu Key Laboratory for Biodiversity and Biotechnology,  
College of Life Sciences, Nanjing Normal University, No. 1,  
Wenyuan Road, 210046 Nanjing, Jiangsu,  
People's Republic of China  
e-mail: cnjgf1208@163.com

C.-D. Zhu  
Key Laboratory of Animal Evolution and Systematics, Institute  
of Zoology, Chinese Academy of Sciences, Beijing 100101,  
People's Republic of China

Several genes encoded in the mitochondrion, particularly *cox1*, *cox2*, and *rrnS*, have been widely used in molecular phylogenetics [10, 11]. More recently, there has been growing attentions paid to the gene orders, in which the genes appear along the molecule, regarding it as a much informative genetic marker to resolve phylogenetic relationships between distantly related taxa [9, 12–15].

The size of insect mtDNA varies from 14 to 19 kb [16–19], although some sequenced mitogenomes possess an exceptionally large genome size (e.g., 36 kb) [20]. The representative metazoan mtDNA contains genes for a complete set of 22 tRNAs, 2 rRNAs (small- and large-subunit rRNAs) and 13 proteins components of the oxidative phosphorylation system. For these latter 13 components, there are subunits I, II, and III of the cytochrome oxidase (*cox1*, *cox2* and *cox3*), the cytochrome *b* (*cob*), the subunits 6 and 8 of the ATPase complex (*atp6* and *atp8*), and the subunits 1–6 and 4L of the NADH dehydrogenase (*nad1–6* and *nad4L*) [21]. Additionally, it has a control region known as the A(adenine) + T(thymine)-rich region in insect mtDNA, which includes the sequences responsible for the origin of heavy-strand mtDNA replication in vertebrates [22] and those for the replication origin for both mtDNA strands in *Drosophila* species [21]. The length of this region is highly variable in insects due to the indels and the presence of variable copy numbers of tandemly repeated elements [23–25].

The butterfly *Papilio xuthus* has been a famous pest of agriculture in Asia [26]. It is a species of the family Papilionidae (Lepidoptera: Papilionoidea) recorded in Burma, Philippines, Korea, Mongolia, Japan and China, etc. Several protein sequences of *P. xuthus* have been sequenced [27] and some partial mtDNA sequences are listed in GenBank. However, the mitogenome sequence of this species was not yet available before this paper. In this study, we report the nearly complete mitogenome sequence of the butterfly *P. xuthus*. Its sequence was compared with

other insect mitogenomes and used for the reconstruction of insect phylogeny.

## Materials and methods

### Biological material

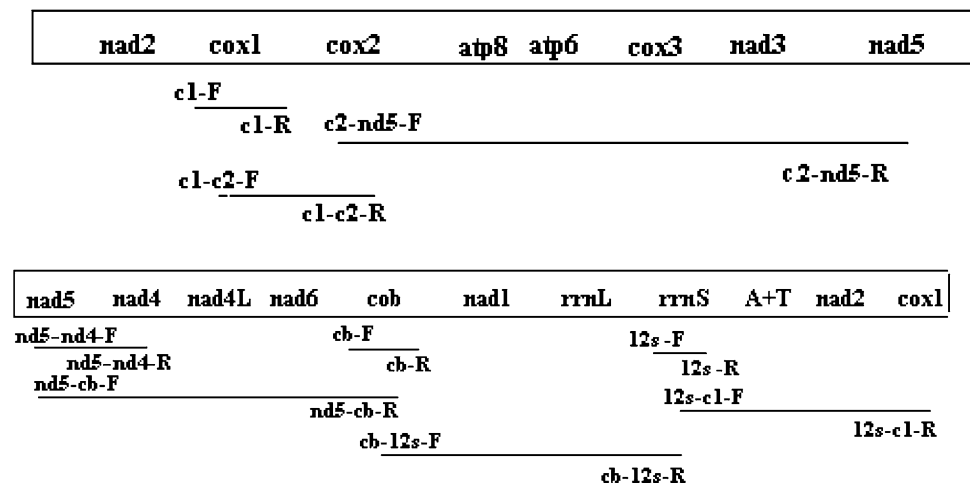
Specimens of *P. xuthus* were collected from the park of Hongshan Zoo in Nanjing, Jiangsu Province, People's Republic of China.

### DNA extraction, primer design, PCR amplification, cloning and sequencing

After the brief examination of the external morphology for the identification of the species, *P. xuthus*, the midgut and wings were removed. Total genomic DNA was extracted from the thorax of adult specimens using a simple proteinase K/SDS method. Scissored tissues were re-suspended in 400  $\mu$ l 0.01 mol/l Tris (pH 8.0), 0.1 mol/l EDTA (pH 8.0), 0.05 mol/l NaCl, 1% SDS, 5  $\mu$ l Proteinase K and incubated at 50°C for 8–10 h. The digested samples were phenol-extracted, ethanol-precipitated again before they were diluted in 30  $\mu$ l ddH<sub>2</sub>O, pH 8.0. DNA quality was checked on a 1% agarose/Tris–borate–EDTA gel. All DNA samples were stored at –20°C. to be used as a template for subsequent PCR reactions.

There are some partial sequences of *cox1*, *cox2*, *nad1*, *rrnS*, *rrnL*, *nad5* and *cob* genes published for *P. xuthus* [28–30]. Basing on known sequences, we designed three pairs of primers (c1-F & c1-R, cb-F & cb-R, 12s-F & 12s-R) (Fig. 1) and amplified three short fragments of *cox1*, *cob* and *rrnS*. Based on above sequenced fragments, five new pairs of primers (c1-c2-F & c1-c2-R, c2-nd5-F & c2-nd5-R, nd5-cb-F & nd5-cb-R, cb-12s-F & cb-12s-R, 12s-c1-F & 12s-c1-R) (Table 1) were designed to amplify

**Fig. 1** Summary of the sequencing strategy. Horizontal lines indicate the fragments which were amplified and cloned. Primers above each line were designed on the sequence of the strand for which the majority of genes were transcribed; primers below each line were designed on the opposite strand



**Table 1** List of PCR primers used in this study

Name	Direction	Sequence
12s-	F	5'-gaa agc gac ggg caa tat gt-3'
12s-	R	5'-gtg cca gca gtt gcg gtt ata c-3'
cb-	F	5'-ctt aat cca aca tcg agg tcg c-3'
cb-	R	5'-gta cat att gcc cgt cgc ttt c-3'
c1-	F	5'-cgt cca gga act tta gct gtt cg-3'
c1-	R	5'-gca tga tga gct caa gtt act g-3'
c1-c2-	F	5'-gca gga gct atc act ata c-3'
c1-c2-	R	5'-ctt gag aat agg ttt gtt tac-3'
c2-nd5-	F	5'-gat gct aat cct ggt cgt ct-3'
c2-nd5-	R	5'-cac tct tct acg tta gtg act gc-3'
nd5-cb-	F	5'-gca gtc act aat gTA gaa gag tga ac-3'
nd5-cb-	R	5'-ctg gat ctc cta ata gat aag g-3'
cb-12s-	F	5'-caa gga att caa ttt tac cc-3'
cb-12s-	R	5'-cag atc aag atg cag att ata a-3'
12s-c1-	F	5'-gaa agc gac ggg caa tat gta cat-3'
12s-c1-	R	5'-cta tca gaa taa tcc ttt ttg tca ggc-3'
nd5-nd4-	F	5'-gta tta gaa tca gca gta tct c-3'
nd5-nd4-	R	5'-ggg gat gat ttt tct cga tct g-3'

F forward and R reverse primers, respectively

the *P. xuthus* mitogenome using the standard Takara LA TATM protocol (Takara) and the following Long-PCR conditions: an initial denaturation for 4 min at 94°C, followed by 15 cycles of 40 s at 94°C and 2 min 30 s at 58°C

then 15 cycles of 40 s at 94°C and 2 min 30 s with 5 s added per cycle at 58°C, at last, a subsequent 10 min final extension at 72°C. Longer PCR products were sequenced by the primer walking strategy with internal primers using ABIPRISM 310 sequencer.

Because of the special structure PolyA/T located in *nad4*, we designed one more pair of primers (n5-n4-F & n5-n4-R). The fragment was sequenced after cloning using the TA cloning method, using the plasmid vector pUC19, by cloning kit (Takara) following the manufacturer's protocol.

#### Sequence analysis

Genes were identified from nucleic acid or the derived protein sequences by BLAST [31], the National Centre for Biotechnology Information (NCBI). The 22 tRNA genes were identified using the software tRNA Scan-SE 1.21 (<http://lowelab.ucsc.edu/tRNAscan-SE>) and their cloverleaf secondary structure and anticodon sequences were identified using DNASIS (Ver.2.5, Hitachi Software Engineering). The nearly complete mitogenome sequence was submitted to GenBank.

#### Phylogenetic analyses

Among available insect mitogenomes including *P. xuthus*, 20 ones were included for phylogenetic analysis in this

**Table 2** Sequences used in this study

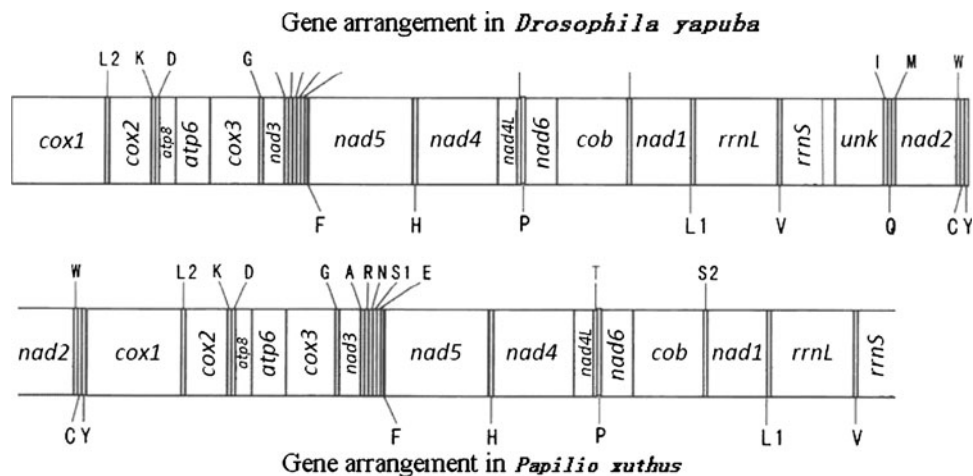
Order	Species	GenBank accession no.	Reference
Lepidoptera	<i>Adoxophyes honmai</i>	NC_008141	Lee et al. [67]
	<i>Bombyx mandarina</i>	NC_003395	Yukuhiro and Sezutsu [69]
	<i>Antheraea pernyi</i>	NC_004622	Liu et al. (Unpublished)
	<i>Coreana raphaelis</i>	NC_007976	Kim et al. [57]
	<i>Ostrinia furnacalis</i>	NC_003368	Coates and Sumerford [70]
	<i>Ostrinia nubilalis</i>	NC_003367	Coates and Sumerford [70]
	<i>Papilio xuthus</i>	EF_621724	This study
Coleoptera	<i>Tribolium castaneum</i>	NC_003081	Friedrich and Muqim [71]
	<i>Crioceris duodecimpunctata</i>	NC_003372	Stewart and Beckenbach [72]
	<i>Pyrocoelia rufa</i>	NC_003970	Bae et al. [73]
Blattaria	<i>Periplaneta fuliginosa</i>	NC_006076	Yamauchi and Miya [74]
Mantodea	<i>Tamolanica tamolana</i>	NC_007702	Cameron and Barker [75]
Hemiptera	<i>Philaenus spumarius</i>	NC_005944	Stewart and Beckenbach [76]
Diptera	<i>Simosyrphus grandicornis</i>	NC_008754	Cameron et al. [77]
	<i>Trichophthalma punctata</i>	NC_008755	Cameron et al. [77]
	<i>Haematobia irritans irritans</i>	NC_007102	Lessinger et al. (Unpublished)
	<i>Bactrocera oleae</i>	NC_005333	Nardi et al. [40]
	<i>Chrysomya putoria</i>	NC_002697	Junqueira and Lessinger [78]
Zygentoma	<i>Thermobia domestica</i>	NC_006080	Cook et al. [79]
Archaeognatha	<i>Petrobius brevistylis</i>	NC_007688	Podsiadlowski [80]
Decapoda	<i>Penaeus monodon</i>	AF217843	Wilson et al. [46]
	<i>Pagurus longicarpus</i>	AF150756	Hickerson and Cunningham [81]

**Table 3** Organization of the *P. xuthus* mitochondrial genome

Gene	Direction	Nucleotide no.	Size	Intergenic nucleotides	Start codon	Stop codon
Partial <i>nad2</i>	F	1–568	568	0		T–
<i>tRNA<sup>Trp</sup></i>	F	569–632	65	–8		
<i>tRNA<sup>Cys</sup></i>	R	625–691	66	–1		
<i>tRNA<sup>Tyr</sup></i>	R	691–757	67	–4		
<i>cox1</i>	F	755–2295	1541	–5	ATTACG	TAA
<i>tRNA<sup>Lue(UUR)</sup></i>	F	2291–2358	68	0		
<i>cox2</i>	F	2359–3075	717	–35	ATG	TAA
<i>tRNA<sup>Lys</sup></i>	F	3041–3111	71	–1		
<i>tRNA<sup>Asp</sup></i>	F	3111–3176	66	12		
<i>atp8</i>	F	3189–3344	156	–7	ATA	TAA
<i>atp6</i>	F	3338–4018	681	8	ATG	TAA
<i>cox3</i>	F	4027–4815	789	3	ATG	TAA
<i>tRNA<sup>Gly</sup></i>	F	4819–4883	65	1		
<i>nad3</i>	F	4885–5236	352	0	ATT	T–
<i>tRNA<sup>Ala</sup></i>	F	5237–5300	64	–1		
<i>tRNA<sup>Arg</sup></i>	F	5300–5364	65	–1		
<i>tRNA<sup>Asn</sup></i>	F	5364–5427	64	–1		
<i>tRNA<sup>Ser(AGN)</sup></i>	F	5427–5487	61	5		
<i>tRNA<sup>Glu</sup></i>	F	5493–5559	67	–1		
<i>tRNA<sup>Phe</sup></i>	R	5559–5620	62	1		
<i>nad5</i>	R	5622–7355	1734	0	ATA	TAA
<i>tRNA<sup>His</sup></i>	R	7356–7423	68	0		
<i>nad4</i>	R	7424–8760	1337	0	ATG	TA–
<i>nad4L</i>	R	8761–9050	291	2	ATG	TA–
<i>tRNA<sup>Thr</sup></i>	F	9053–9117	65	0		
<i>tRNA<sup>Pro</sup></i>	R	9118–9183	66	0		
<i>nad6</i>	F	9184–9717	534	15	ATG	TAA
<i>cob</i>	F	9733–10881	1149	3	ATG	TAA
<i>tRNA<sup>Ser</sup></i>	F	10885–10950	66	13		
<i>nad1</i>	R	10964–1905	942	2	ATG	T–
<i>tRNA<sup>Leu</sup></i>	R	11908–11977	70	–2		
<i>rrnL</i>	R	11976–13307	1332	–5		
<i>tRNA<sup>Val</sup></i>	R	13303–13365	63	1		
Partial <i>rrnS</i>	R	13367–13964	598			

Negative numbers indicate that adjacent genes overlap

**Fig. 2** Compared with the gene arrangement of *Drosophila yakuba* and *P. xuthus*, protein and rRNA genes are transcribed from left-to-right except genes indicated by *underbars*, tRNA genes are designated by *single-letter amino acid codes*, those encoded by the J and N strands are show above and below the gene map. UNK, (A + T) rich region



paper. Others were excluded due to unusual sequence evolution reported previously [32] (e.g., phthirapteran *Heterodoxus macropus* belonging to hemipteroid assemblage [33] and the hymenopteran *Apis mellifera* [34], or because of high gene rearrangements (e.g., species of hemipteroid assemblage, the thysanopteran *Thrips imaginis* [35] and the psocopteran *Lepidopsocis* sp. [33]).

The *nad2* gene was not entirely sequenced in this study. Except for *nad2*, all protein-coding genes (PCGs) of insect mitogenomes chosen were used for the phylogenetic analyses. All nucleotide sequences of each PCG were retro-aligned using the RevTrans 1.4 server available through the DTUCBS website [36]. Positions encoding proteins were translated to amino acids using MEGA 3.0 [37] for confirmation of alignment. Unalignable regions were excluded manually from the phylogenetic analyses. The 5' and 3' unalignable ends of the PCGs were trimmed from the alignments to have the same length.

Phylogenetic analyses were performed on the complete amino acid sequences of 12 PCGs using bayesian approach as implemented in MrBayes, ver. 3.1 [38, 39]. Two species, *Penaeus monodon* and *Pagurus longicarpus* (Crustacea), were set as outgroups in all analyses in this paper [40, 41]. Selection of suitable nucleotide substitution models for Bayesian analyses was guided by results of hierarchical likelihood ratio tests calculated using Modeltest 3.06 [42] in PAUP\* version 4.0b10 [43]. Bayesian inference was done with MrBayes 3.0b4 software [38] using the GTR + I + G model suggested by Modeltest. Bayesian analyses were launched with random starting trees and run for  $4 \times 10^6$  generations, sampling the Markov chains at intervals of 100 generations. Four chains were run simultaneously, three hot and one cold, with the initial 200 cycles discarded as burn-in. To determine whether the Bayesian analyses had reached stationarity, likelihoods of sample points were plotted against generation time. Sample points generated before reaching stationarity were discarded as “burnin” samples. After the removal of burnin, a majority-rule consensus topology of all trees was constructed, with the

percentage of trees where each node was found expressed on the tree as posterior probabilities (Table 2).

## Results and discussion

### Gene content and genome organization

The nearly complete mitogenome of *P. xuthus* is 13,964 bp (GenBank accession Number: EF621724), including the standard 12 protein-coding genes, 19 tRNA genes, whole *rrnL* gene and partial *rrnS* gene sequence (Table 3). The gene order is identical to that of *Drosophila yakuba* [44] (Fig. 2), which is conserved in divergent insect orders and even some crustaceans [40, 45, 46], and reflects the presumed ancestral condition for Pancrustacea. Some of the genes overlap as those in other animal mtDNAs. In *P. xuthus*, overlaps between genes occur 13 times and involve a total 72 bp. In the sequence, nucleotide composition of each major-coding strand was shown in Table 4.

Compared with other mitogenome of Lepidoptera [47, 48], A, T, or A + T compositions of this species are slightly lower than other insects. The reason might be that we did not get the sequence of the control region which known as A + T-rich region. Actually the primers (12s-c1-F & 12s-c1-R) were used to amplify the fragment from *rrnS* to *cox1* gene (Fig. 1), which in most insect mtDNA is inclusive of the A + T-rich region. Unfortunately, we were unable to sequence the amplified fragment. The method of TA cloning has also been tried to sequence this fragment, but failed. So finally, we only successfully sequenced the partial *nad2* sequence. We guess that the failure in sequencing the A + T-rich region was possibly because of its tandem-repetitive nature and relatively large size.

### Gene initiation and termination

The 11 PCGs are observed to have a putative, inframe ATR methionine or ATT isoleucine codons as start signals,

**Table 4** Nucleotide composition (%) of selected Lepidoptera mitochondrial genomes

Species	T (%)	C (%)	A (%)	G (%)	A + T (%)	G + C (%)	Total nucleotide
<i>Papilio xuthus</i>	40.45	12.17	39.53	7.85	79.98	20.02	13964
<i>Bombyx mori</i>	38.3	11.4	43.0	7.3	81.3	18.7	15643
<i>Bombyx mandarina</i>	38.6	11.1	43.1	7.2	81.7	18.3	15928
<i>Adoxophyes honmai</i>	40.2	11.7	40.2	7.9	80.4	19.6	15679
<i>Antheraea pernyi</i>	40.9	12.1	39.2	7.8	80.1	19.9	15575
<i>Coreana raphaelis</i>	43.3	10.0	39.4	7.3	82.7	17.3	15314
<i>Ostrinia furnacalis</i>	38.9	11.7	41.5	7.9	80.4	19.6	14535
<i>Ostrinia nubilalis</i>	38.8	11.8	41.4	8.0	80.2	19.8	14535

Nucleotide composition of major-coding strands are compared

which are the triplets that usually initiate metazoan mitochondrial genes. Canonical initiation codons (ATA or ATG), encoding the amino acid methionine, are used in 10 PCGs (*atp8*, *nad5*, *cox2*, *atp6*, *cox3*, *nad4*, *nad4L*, *nad6*, *cob* and *nad1*) except *nad3* (Table 3), which appears to use the nonstandard start codon as it often happens in animal mtDNAs [21]. The use of ATG as a start codon is limited to the protein genes encoded immediately downstream of either another protein gene (*cox1-cox2*, *atp8-atp6*, *atp6-cox3*, *nad6-cob*, *nad4L-nad4*)—and in all such cases, the downstream protein gene has an ATG start codon—or a

tRNA gene encoded on the opposite strand (*trnT-nad4L*) [49].

In *cox1*, however, a typical ATN initiator for PCGs is not found in the start site for *cox1* or the neighbouring tRNA<sup>Tyr</sup>. None of the triplets known to act as initiation codons could be found in the vicinity of the supposed *cox1* initiation: The amino acid sequence at the beginning of *cox1* is well conserved in all arthropods, and the possible initiation signal is likely to be found in an area of four to five triplets [50]. In this location, no ATAA signal, proposed to initiate *cox1* in *Drosophila* and *Locusta*, was

**Fig. 3** Predicated secondary clover-leaf structure for the 19 tRNA genes of *P. xuthus*. The tRNAs are labelled with the abbreviations of their corresponding amino acids. Nucleotide sequences from 5' to 3' as indicated for tRNA<sup>Ala</sup>. Dashes (–) indicate Watson–Crick basepairing, and plus sign (+) 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 dihydrouridine (DHU or D) arm

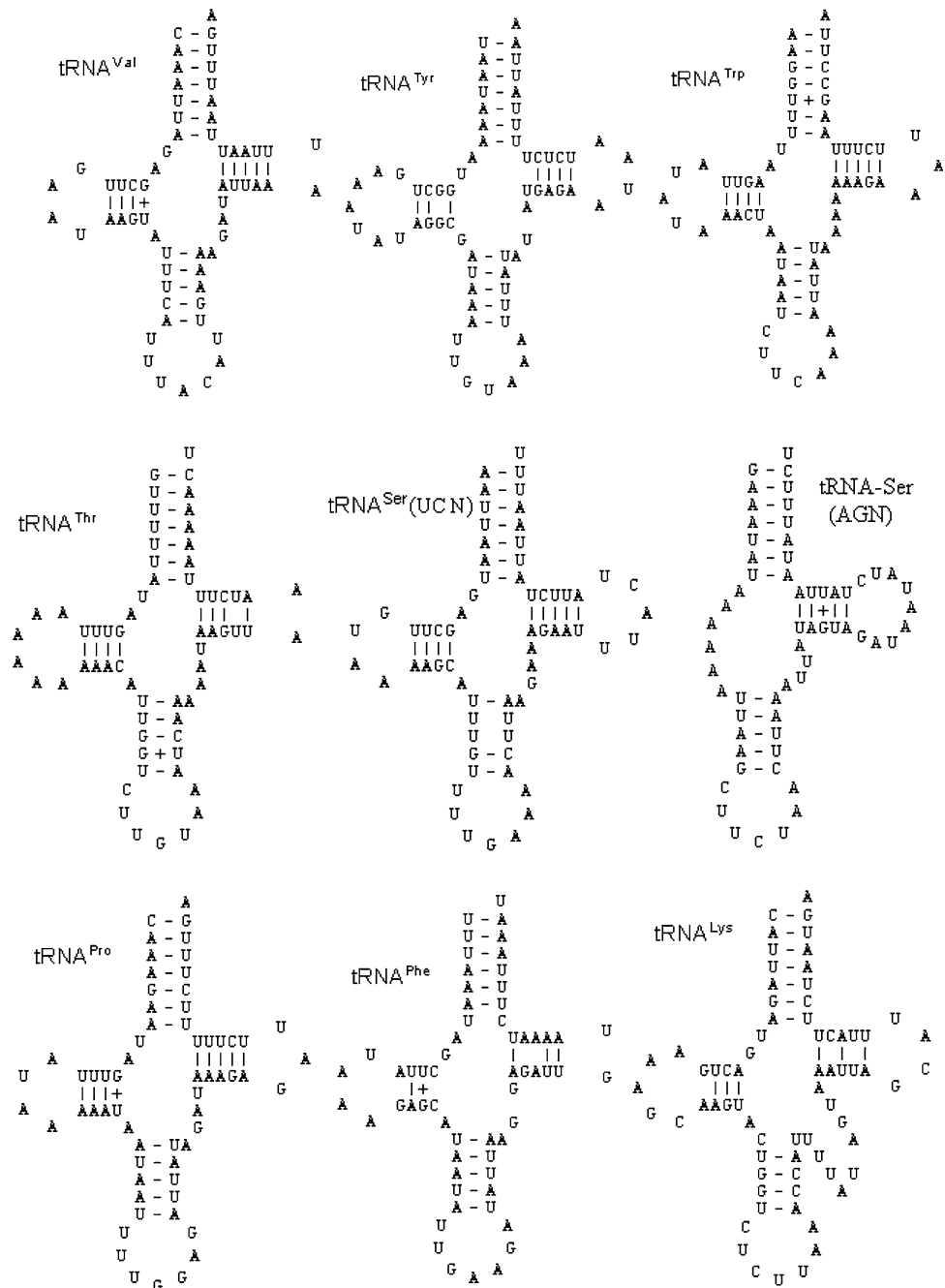
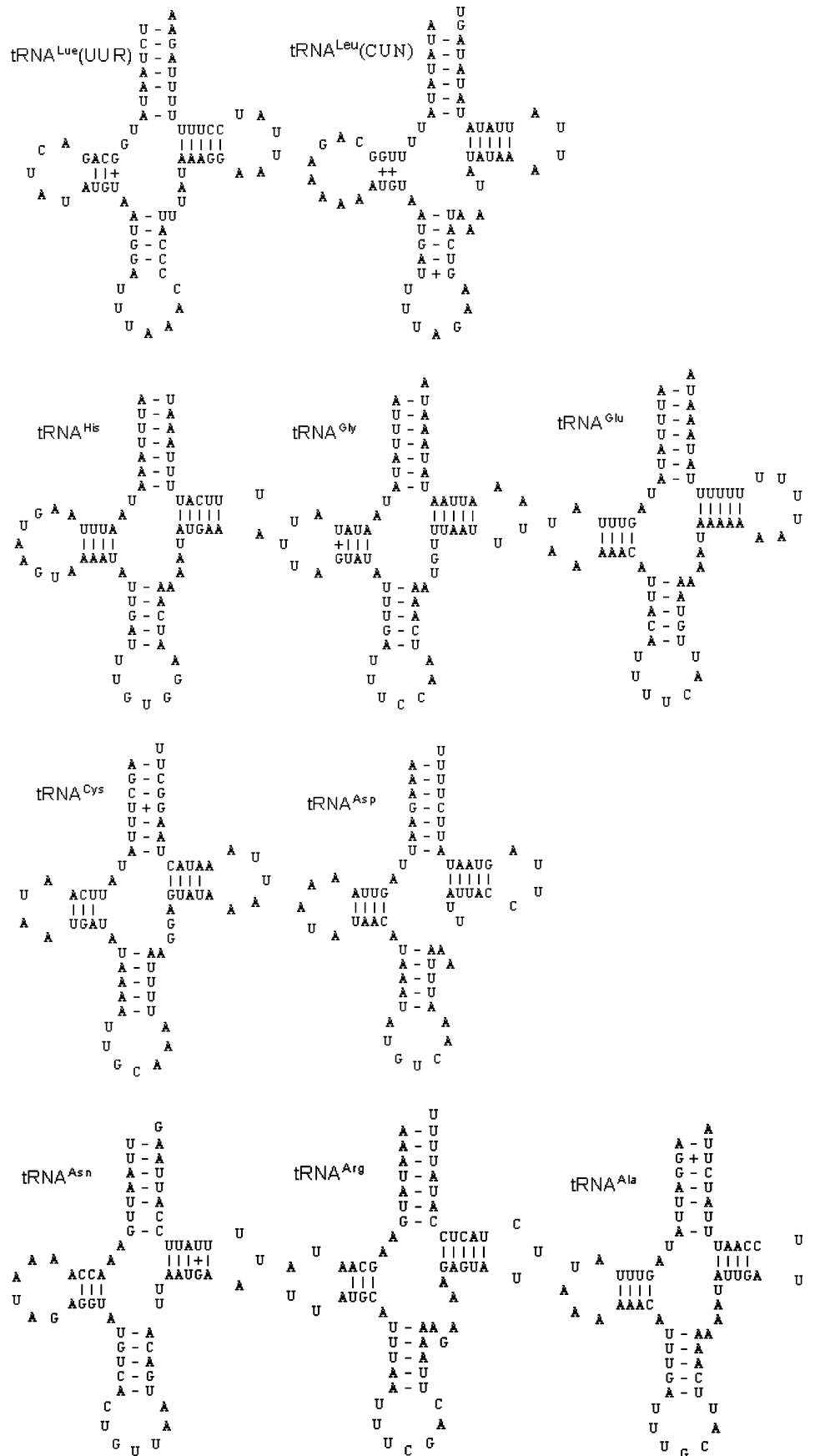


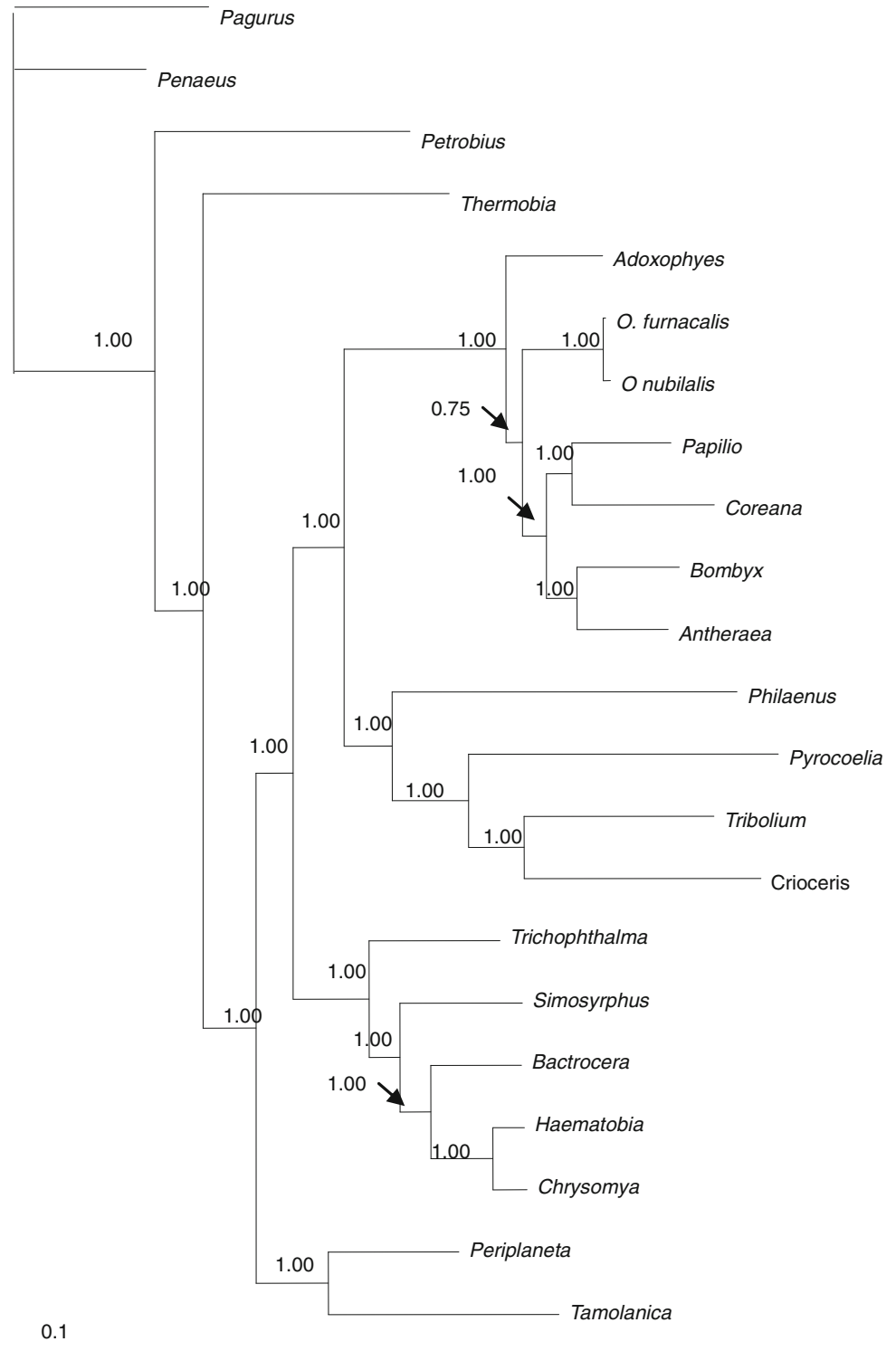
Fig. 3 continued



present. A hexa-nucleotide ATTACG flanks the beginning of *cox1*, and was followed by a CGA triplet in *P. xuthus*. Similarly, the hexa-nucleotide ATTTAA has been proposed in a collembolan *Tetrodontophora bielanensis* [50]. All lepidopteran species examined to date use R (coded by CGA) as the initial amino acid for *cox1* and the use of

non-canonical start codon for this gene is common across insects [51, 52]. In the current situation where no mRNA expression data for *P. xuthus* are available, we tentatively designated the hexanucleotide ATTACG as an initiation codon for *P. xuthus cox1*. It is unclear why the sequence of *cox1* is usually the most conserved among metazoan

**Fig. 4** Bayesian tree based on amino acid sequences of the 12 protein-coding genes. Number at nodes indicate posterior probabilities





mitochondrial genes. A mechanism, which would permit translation to start at this sequence, was suggested, without any experimental documents, that the anticodon of the initiating *N*-formylmethionine tRNA might permit the ATTACG sequence to be recognized as a single codon [53].

Eight PCGs terminate with the complete termination codon TAA (*cox1*, *cox2*, *atp8*, *atp6*, *cox3*, *nad5*, *nad6* and *cob*). In all other cases, stop codons are truncated (T or TA) and their functionality probably recovered after a post-transcriptional polyadenylation [54]. These abbreviated stop codons are found in PCGs (*nad2*, *nad3* and *nad4*) that are followed by a downstream tRNA gene, suggesting that the secondary structure information of the tRNA genes could be responsible for the correct cleavage of the polycistronic transcript [55]. Thus, it is highly probable that, during mRNA processing the U is exposed at the end of a mRNA molecule and polyadenylated, forming the complete UAA termination signal. tRNA genes are usually interspersed among PCGs, which secondary structure acting as a signal for the cleavage of the polycistronic primary transcript [54, 56]. However, there is also a direct junction between two PCGs (*nad4L/nad4*) where other cleavage signals, different from tRNA gene secondary structures, may be involved in the processing of the polycistronic primary transcript [57]. For the last three genes (*cox1*, *cox2* and *atp8*), the complete termination codons are all within the next gene or tRNA, for the overlaps between them are more than three nucleotides.

#### Ribosomal RNAs and transfer RNAs

As all other metazoan mtDNAs sequenced, *P. xuthus* mtDNA contains genes for both small and large ribosomal subunit RNAs (*rnmS* and *rnlL*). Both genes are encoded by the heavy (H) strand and are separated by tRNA<sup>Val</sup>, which is identical to the arrangement in many other metazoans. The size of the inferred *rnlL* is 1,332 bp, and the partial *rnmS* is about 598 bp.

The series of 19 tRNAs typical of metazoan mitogenomes were found, and secondary structures were drawn for each one (Fig. 3). The 18 tRNAs showed typical cloverleaf secondary structures except for the tRNA<sup>Ser</sup>(AGN). The tRNA<sup>Ser</sup>(AGN) gene observed in *P. xuthus* could not form a stable stem loop structure in the DHU arm as shown in many other insect tRNA<sup>Ser</sup>(AGN)s. Despite of the unusual secondary structure (Fig. 3), the tRNA is still predicted to adopt an appropriate tertiary structure, based on the folding rules proposed by Steinberg and Cedergren [58].

In 10 cases, tRNA coding did show overlaps with the sideward tRNA or gene length ranging from 1 to 35 bp. The 8 bp overlapping was observed for tRNA<sup>Trp</sup>/tRNA<sup>Cys</sup>,

as reported by Lessinger et al. [59], producing separate transcripts with their opposite directions like other insect species [34, 44, 59, 60].

#### Phylogenetic analysis

A 22-taxon data set, 3,830 characters after removal of gap-experiencing sites, of which 2,539 were variable and 1,989 were parsimony-informative, was analyzed using Bayesian method. The topology of the Bayesian tree is shown in Fig. 4.

The most striking result of this analysis is that Insecta (Microcoryphia + Zyentoma + Pterygota), Microcoryphia, Zyentoma, Pterygota, Diptera, Lepidoptera and Coleoptera are monophyletic. Within Pterygota, the Dictyoptera and the holometabolans orders Diptera, Lepidoptera and Coleoptera are monophyletic, Homometabola per se are not, because the Coleoptera (*Pyrocoelia*, *Tribolium* and *Crioceris*) is sister to the Hemiptera (*Philaenus*) of the Hemimetabola. These results confirm most recent molecular analyses [41, 61] but which the Holometabola do not form a monophyletic clade is in open disagreement with most morphological [62, 63] and molecular [64] analyses.

The seven lepidopteran mitogenome sequences represent four lepidopteran superfamilies within the lepidopteran suborder, Ditrysia: *B. mandarina* and *A. pernyi* for the Bombycoidea, *P. xuthus* and *C. raphaelis* for the Papilionoidea, *O. furnacalis* and *O. nubilalis* for the Pyraloidea, and *A. honmai* for the Tortricidae. This phylogenetic analysis led to well supported monophyletic groups, Bombycoidea, Papilionoidea, Pyraloidea, and Obtectomera. This result illuminated traditional classification system very well [65, 66]. Although further studies are needed for more diverse species, the result supports the relationship of (Apoditrysia (Obtectomera (Macro-lepidoptera))) [67, 68].

**Acknowledgements** We are grateful to Dr. Jie Yan and Mr. Hao Shi (College of Life Sciences, Nanjing Normal University) for technical assistance, and to David Lees (Department of Entomology, The Natural History Museum, London) for improving this manuscript. Finally we thank Dr. Welch D. M. and one anonymous reviewer for their helpful comments on previous of this manuscript. This work was supported by a grant from the National Natural Sciences Foundation of China (No. 30160015).

#### References

1. Brown WM (1983) Evolution of animal mitochondrial DNA. In: Nei M, Koehn RK (eds) Evolution of genes and proteins. Sinauer, Sunderland, pp 62–88
2. Avis JC (1994) Molecular markers, natural history and evolution. Chapman & Hall, New York, p 511
3. Jiang JP, Zhou KY (2001) Evolutionary relationships among Chinese ranid frogs inferred from mitochondrial DNA sequences of 12S rRNA gene. Acta Zool Sin 47:38–44 (in Chinese)

4. Wu XB, Wang YQ, Zhou KY, Zhu WQ, Nie JS, Wang CL (2003) Complete mitochondrial DNA sequence of Chinese alligator, *Alligator sinensis* and phylogeny of crocodiles. *Chin Sci Bull* 48:2050–2054
5. Macey JR, Schulte JA, Larson A (2000) Evolution and phylogenetic information content of mitochondrial genomic structural features illustrated with acrodont lizards. *Syst Biol* 49:257–277
6. Castresana J, Feldmaier-Fuchs G, Paabo S (1998) Codon reassignment and amino acid composition in hemichordate mitochondria. *Proc Natl Acad Sci USA* 95:3703–3707
7. Telford MJ, Herniou EA, Russell RB, Littlewood DT (2000) Changes in mitochondrial genetic codes as phylogenetic characters: two examples from the flatworms. *Proc Natl Acad Sci USA* 97:11359–11364
8. Boore JL, Collins TM, Stanton D, Daehler LL, Brown WM (1995) Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. *Nature* 376:163–165
9. Boore JL, Lavrov DV, Brown WM (1998) Gene translocation links insects and crustaceans. *Nature* 392:667–668
10. Simon C, Frati F, Bekenbach A, Crespi B, Liu H, Flook PK (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain-reaction primers. *Ann Entomol Soc Am* 87:651–701
11. Caterino MS, Cho S, Sperling FAH (2000) The current state of insect molecular systematics: a thriving Tower of Babel. *Annu Rev Entomol* 45:1–54
12. Blanchette M, Kunisawa T, Sankoff D (1999) Gene order breakpoint evidence in animal mitochondrial phylogeny. *J Mol Evol* 49:193–203
13. Boore JL, Brown WM (2000) Mitochondrial genomes of Galathealinea, Helobdella, and Platynereis: sequence and gene arrangement comparisons indicate that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. *Mol Biol Evol* 17:87–106
14. Kurabayashi A, Ueshima R (2000) Complete sequence of the mitochondrial DNA of the primitive opisthobranch gastropod *Pupa strigosa*: systematic implication of the genome organization. *Mol Biol Evol* 17:266–277
15. Scouras A, Smith MJ (2001) A novel mitochondrial gene order in the crinoid echinoderm *Florometra serratissima*. *Mol Biol Evol* 18:61–73
16. Boore JL (1999) Animal mitochondrial genomes. *Nucleic Acids Res* 27:1767–1780
17. Hua J, Li M, Dong P, Xie Q, Bu W (2009) The mitochondrial genome of *Protohermes concolorus* Yang et Yang 1988 (Insecta: Megaloptera: Corydalidae). *Mol Biol Rep* 36(7):1757–1765
18. Zhou Z, Huang Y, Shi F, Ye H (2009) The complete mitochondrial genome of *Deracantha onos* (Orthoptera: Bradyporidae). *Mol Biol Rep* 36(1):7–12
19. Wolstenholme DR (1992) Animal mitochondrial DNA: structure and evolution. *Int Rev Cytol* 141:173–216
20. Boyce TM, Zwick ME, Aquadro CF (1989) Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. *Genetics* 123:825–836
21. Wolstenholme DR (1992) Genetic novelties in mitochondrial genomes of multicellular animals. *Curr Opin Genet Dev* 2:918–925
22. Brown WM (1985) The mitochondrial genome of animals. In: MacIntyre RJ (ed) *Molecular evolution genetics*. Plenum, New York, pp 95–130
23. Fauron CMR, Wolstenholme 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
24. Inohira K, Hara T, Matsuura ET (1997) Nucleotide sequence divergence in the A + T-rich region of mitochondrial DNA in *Drosophila simulans* and *Drosophila mauritiana*. *Mol Biol Evol* 14:814–822
25. Wei SJ, Tang P, Zheng LH, Shi M, Chen XX (2009) The complete mitochondrial genome of *Evania appendigaster* (Hymenoptera: Evaniidae) has low A + T content and a long intergenic spacer between atp8 and atp6. *Mol Biol Rep*. [Epub ahead of print] PubMed PMID: 19655273
26. Kong HR, Liang XC, Luo YZ (2006) Anatomy of the reproductive system of *Papilio xuthus* Linnaeus. *J Yunnan Agr Univ* 21:459–462
27. Ozaki K, Utoguchi A, Yamada A, Yoshikawa H (2008) Identification and genomic structure of chemosensory proteins (CSP) and odorant binding proteins (OBP) genes expressed in foreleg tarsi of the swallowtail butterfly *Papilio xuthus*. *Insect Biochem Mol Biol* 38:969–976
28. Caterino MS, Sperling FA (1999) Papilio phylogeny based on mitochondrial cytochrome oxidase I and II genes. *Mol Phylogenet Evol* 11:122–137
29. Aubert J, Legal L, Descimon H, Michel F (1999) Molecular phylogeny of swallowtail butterflies of the tribe Papilionini (Papilionidae, Lepidoptera). *Mol Phylogenet Evol* 12:156–167
30. Takashi Y, Go S, Hiraku T (1999) Phylogeny of Japanese Papilionid butterflies inferred from nucleotide sequences of the mitochondrial ND5 gene. *J Mol Evol* 48:42–48
31. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
32. Foster PG, Hickey DA (1999) Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. *J Mol Evol* 41:284–290
33. Shao R, Campbell NJ, Barker SC (2001) Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthiraptera). *Mol Biol Evol* 18:858–865
34. Crozier RH, Crozier YC (1993) The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics* 133:97–117
35. Shao R, Barker SC (2003) The highly rearranged mitochondrial genome of the plague thrips, *Thrips imaginis* (Insecta: Thysanoptera): convergence of two novel gene boundaries and an extraordinary arrangement of rRNA genes. *Mol Biol Evol* 20:362–370
36. Wernersson R, Pedersen AG (2003) RevTrans—constructing alignments of coding DNA from aligned amino acid sequences. *Nucl Acids Res* 31:3537–3539
37. Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 5:150–163
38. Huelsenbeck JP, Ronquist F (2001) MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755
39. Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19:1572–1574
40. Nardi F, Spinsanti G, Boore JL, Carapelli A, Dallai R, Frati F (2003) Hexapod origins: monophyletic or paraphyletic? *Science* 299:1887–1889
41. Carapelli A, Pietro L, Nardi F, der Wath E, van Frati F (2007) Phylogenetic analysis of mitochondrial protein coding genes confirms the reciprocal paraphyly of Hexapoda and Crustacea. *BMC Evol Biol* 7(Suppl 2):S8
42. Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics* 14(9):817–818
43. Swofford DL (2002) PAUP\*: Phylogenetic analysis using parsimony (\*and other methods), version 4.0. Sinauer Associates, Sunderland

44. Clary DO, Wolstenholme DR (1985) The mitochondrial DNA molecular of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *J Mol Evol* 22:252–271
45. Crease TJ (1999) The complete sequence of the mitochondrial genome of *Daphnia pulex* (Cladocera: Crustacea). *Gene* 233:89–99
46. Wilson K, Cahill V, Ballment E, Benzie J (2000) The complete sequence of the mitochondrial genome of the crustacean *Penaeus monodon*: are malacostracan crustaceans more closely related to insects than to branchiopods? *Mol Biol Evol* 17(6):863–874
47. 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(7):1871–1880
48. 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(6):1441–1449
49. Lavrov DV, Boore JL, Brown WM (2000) The complete mitochondrial DNA sequence of the horseshoe crab *Limulus polyphemus*. *Mol Biol Evol* 17:813–824
50. Nardi F, Carapelli A, Fanciulli PP, Dallai R, Frati F (2001) The complete mitochondrial DNA sequence of the basal hexapod *Tetrodontophora bielanensis*: evidence for heteroplasmy and tRNA translocations. *Mol Biol Evol* 18:1293–1304
51. Hu J, Zhang D, Hao J, Huang D, Cameron S, Zhu C (2009) The complete mitochondrial genome of the yellow coaster, *Acraea issoria* (Lepidoptera: Nymphalidae:Heliconiinae: Acraeini): sequence, gene organization and a unique tRNA translocation event. *Mol Biol Rep* [Epub ahead of print] PubMed PMID: 20091125
52. Fenn JD, Cameron SL, Whiting MF (2007) The complete mitochondrial genome sequence of the Mormon cricket (*Anabrus simplex*: Tettigoniidae: Orthoptera) and an analysis of control region variability. *Insect Mol Biol* 16(2):239–252
53. Krzywinski J, Grushko OG, Besansky NJ (2006) Analysis of the complete mitochondrial DNA from *Anopheles funestus*: an improved dipteran mitochondrial genome annotation and a temporal dimension of mosquito evolution. *Mol Phylogenet Evol* 39(2):417–423
54. Ojala D, Merkel C, Gelfand R, Attardi G (1980) The tRNA genes punctuate the reading of genetic information in human mitochondrial DNA. *Cell* 2:393–403
55. Carapelli A, Comandi S, Convey P, Nardi F, Frati F (2008) The complete mitochondrial genome of the Antarctic springtail *Cryptopygus antarcticus* (Hexapoda: Collembola). *BMC Genom* 9:315
56. Montoya J, Gaines GL, Attardi G (1983) The pattern of transcription of the human mitochondrial rRNA genes reveals two overlapping transcription units. *Cell* 34:151–159
57. Kim I, Lee EM, Seol KY, Yun EY, Lee YB, Hwang JS, Jin BR (2006) The mitochondrial genome of Korean hairstreak *Coreana raphaelis* (Lepidoptera: Lycaenidae). *Insect Mol Biol* 15:217–225
58. Steinberg S, Cedergren R (1994) Structural compensation in atypical mitochondrial tRNAs. *Nat Struct Biol* 1:507–510
59. Lessinger AC, Martins Junqueira AC, Lemos TA, Kemper EL, Silva FR, Vettore AL, Arruda P, Azeredo-Espin AM (2000) The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae). *Insect Mol Biol* 9:521–529
60. Campebl NJH, Barker SC (1999) The novel mitochondrial gene arrangement of the cattle tick, *Boophilus microplus*: fivefold tandem repetition of a coding region. *Mol Biol Evol* 16:732–740
61. Giribet G, Ribera C (2000) A review of arthropod phylogeny: new data based on ribosomal DNA sequences and direct character optimization. *Cladistics* 16:204–231
62. Hennig W (1981) In: Pont A (ed) *Insect phylogeny*. John Wiley and Sons, New York
63. Kristensen NP (1999) Phylogeny of endopterygote insects, the most successful lineage of living organisms. *Eur J Entomol* 96:237–253
64. Whiting MF (2002) Phylogeny of holometabolous insect orders: molecular evidence. *Zool Sci* 31:3–15
65. Kristensen NP (2003) Skeleton and muscles: adults. In: Kristensen NP (ed) *Lepidoptera: Moths and Butterflies 2*. *Handbuch der Zoologie/Handbook of Zoology IV/36*, xii + 564 pp. Walter de Gruyter, Berlin & New York, pp 39–131
66. Kristensen NP, Scoble MJ, Karsholt O (2007) Lepidoptera phylogeny and systematics: the state of inventorying moth and butterfly diversity. *Zootaxa* 1668:699–747
67. Lee ES, Shin KS, Kim MS, Park H, Cho S, Kim CB (2006) The mitochondrial genome of the smaller tea tortrix *Adoxophyes honmai* (Lepidoptera: Tortricidae). *Gene* 373:52–57
68. 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
69. Yukuhiro K, Sezutsu H, Itoh M, Shimizu K, Banno Y (2002) Significant levels of sequence divergence and gene rearrangements have occurred between the mitochondrial genomes of the wild mulberry silkmoth, *Bombyx mandarina*, and its close relative, the domesticated silkmoth, *Bombyx mori*. *Mol Biol Evol* 19:1385–1389
70. 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
71. Friedrich M, Muqim N (2003) Sequence and phylogenetic analysis of the complete mitochondrial genome of the flour beetle *Tribolium castaneum*. *Mol Phylogenet Evol* 26:502–512
72. Stewart JB, Beckenbach AT (2003) Phylogenetic and genomic analysis of the complete mitochondrial DNA sequence of the spotted asparagus beetle *Crioceris duodecimpunctata*. *Mol Phylogenet Evol* 26:513–526
73. Bae JS, Kim I, Sohn HD, Jin BR (2004) The mitochondrial genome of the firefly, *Pyrocoelia rufa*: complete DNA sequence, genome organization, and phylogenetic analysis with other insects. *Mol Phylogenet Evol* 32:978–985
74. Yamauchi MM, Miya MU, Nishida M (2004) Use of a PCR-based approach for sequencing whole mitochondrial genomes of insects: two examples (cockroach and dragonfly) based on the method developed for decapod crustaceans. *Insect Mol Biol* 13:435–442
75. Cameron SL, Barker SC, Whiting MF (2006) Mitochondrial genomics and the new insect order Mantophasmatodea. *Mol Phylogenet Evol* 38:274–279
76. Stewart JB, Beckenbach AT (2005) Insect mitochondrial genomics: the complete mitochondrial genome sequence of the meadow spittlebug *Philaenus spumarius* (Hemiptera: Auchenorrhyncha: Cercopoidae). *Genome* 48:46–54
77. Cameron SL, Lambkin CL, Barker SC, Whiting MF (2007) A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad time-scales with high precision. *Syst Entomol* 32:40–59
78. Junqueira AC, Lessinger AC, Torres TT, da Silva FR, Vettore AL, Arruda P, Azeredo-Espin AM (2004) The mitochondrial genome of the blowfly *Chrysomya chloropyga* (Diptera: Calliphoridae). *Gene* 339:7–15

79. Cook CE, Yue Q, Akam M (2005) Mitochondrial genomes suggest that hexapods and crustaceans are mutually paraphyletic. *Proc Biol Sci* 272:1295–1304
80. Podsiadlowski L (2006) The mitochondrial genome of the bristletail *Petrobius brevistylis* (Archaeognatha: Machilidae). *Insect Mol Biol* 15:253–258
81. Hickerson MJ, Cunningham CW (2000) Dramatic mitochondrial gene rearrangements in the hermit crab *Pagurus longicarpus* (Crustacea, Anomura). *Mol Biol Evol* 17(4):639–644
82. Zhang YY, Xuan WJ, Zhao JL, Zhu CD, Jiang GF (2009) The complete mitochondrial genome of the cockroach *Eupolyphaga sinensis* (Blattaria: Polyphagidae) and the phylogenetic relationships within the Dictyoptera. *Mol Biol Rep* [Epub ahead of print] PubMed PMID: 20012368