Available online at www.sciencedirect.com
SCIENCE DIRECT
Organisms, Diversity & Evolution 6 (2006) 17–32
ORGANISMS DIVERSITY & EVOLUTION
www.elsevier.de/ode

Identification of evolutionary conserved structural elements in the mt SSU rRNA of Zygaenoidea (Lepidoptera): A comparative sequence analysis

Oliver Niehuis*[,1], Clas M. Naumann[†], Bernhard Misof

Alexander Koenig Research Institute and Museum of Zoology, Adenauerallee 160, 53113 Bonn, Germany

Received 10 November 2004; accepted 23 March 2005

Abstract

Knowledge of the secondary structure of ribosomal RNA (rRNA) molecules has become increasingly important in phylogenetic analyses. Advances in RNA substitution models have underlined the need for reliable secondary-structure models for individual taxonomic groups. The present investigation aims to infer a secondary-structure model of the mt SSU (12S) rRNA of Zygaenoidea using a comparative approach. Structural variation of the 12S rRNA molecule proves to be minor among the investigated species, although at least two helices exhibit taxon-specific deviations. The consensus structure of the zygaenoid mt SSU rRNA clearly differs from the structure published for *Bombyx mori* and challenges some helices proposed in the silk moth model. Our analyses demonstrate the need for taxon-specific rRNA models, which can capture evolutionary patterns in these molecules far better than general eukaryotic consensus structures and thus provide an improved basis for phylogenetic analyses incorporating secondary-structure information.

Keywords: Lepidoptera; 12S rRNA; Multiple alignment; Comparative analysis; Secondary-structure models

Introduction

Ribosomal RNA (rRNA) molecules play a fundamental role in almost all stages of cellular protein synthesis (Dahlberg 1989; Hill et al. 1990). Their complex three-dimensional structure is considerably conserved across distantly related taxa, yet individual taxonomic groups regularly display unique features (Van de Peer et al. 1999, 2000; Cannone et al. 2002; Wuyts et al. 2004). Secondary-structure information has been applied successfully to tackle systematic problems (e.g. Billoud et al. 2000; Lydeard et al. 2000; Ouvrard et al. 2000; Misof and Fleck 2003) and significantly improves the fit of sequence evolution models in analyses of rDNA sequences (Savill et al. 2001; Jow et al. 2002; Hudelot et al. 2003). Nevertheless, most investigations still make exclusive use of the primary structure as a source of information.

Incorporating secondary-structure information in RNA sequence substitution models depends on inferred rRNA consensus structures, which are an expression of the evolutionary plasticity and stability of the molecule. In phylogenetic sequence analyses, when considering rRNA structure information, an investigation of the structural stability is required. Catalogues with rRNA structure information are available for large sets of eukaryotic sequences (Cannone et al. 2002; Wuyts et al.

*Corresponding author. Fax: +1 480 965 6899.
E-mail address: o.niehuis@web.de (O. Niehuis).
[1]Current address: Arizona State University, School of Life Sciences, PO Box 874501, Tempe, AZ, 85287-4501, USA.
[†]Deceased.

1439-6092/$ - see front matter
doi:10.1016/j.ode.2005.03.001

2004), but taxon-specific analyses facilitating the incorporation of these data into phylogenetic analyses of narrower scope are frequently missing. We think that this is a reason why characteristics of rDNA sequence evolution (e.g. co-variation of paired nucleotides) are frequently ignored in analyses of ribosomal sequences.

In molecular phylogenetic analyses, secondary-structure models of rRNA have received more attention by providing a framework for the alignment of rDNA sequences by means of adding structural information as additional criteria of homology (Hickson et al. 1996). The advantages of structurally based sequence alignments have been demonstrated repeatedly (Kjer 1995; Titus and Frost 1996; Hickson et al. 2000), and several alignment algorithms and programs are available that take secondary structures into account (Corpet and Michot 1994; Notredame et al. 1997; Thompson et al. 1997; Lenhof et al. 1998).

Adoption of structural information is not a trivial task. Manually adjusting new sequences to an already existing model of a more or less closely related taxon involves the danger that erroneous base pairings are proposed if the applied model deviates from the structure of the investigated group or if the model is too general (Kjer 1995; Page 2000; Page et al. 2002). Such adjustments are most critical with advanced sequence evolution models (see above), which take correlation in paired sites of an rRNA molecule into account and which therefore require explicit statements on all base pairings in a given data set (Jow et al. 2002; Hudelot et al. 2003). The availability of reliable secondary-structure models in individual taxonomic groups thus is essential for an appropriate phylogenetic analysis using rDNA sequences.

The most successful approach in deriving the RNA secondary structure for a particular taxonomic group is a comparative sequence analysis (Gutell et al. 1992, 1994, 2002; Woese and Pace 1993). The method is based on the assumption that RNA molecules with the same function in related taxa should have the same structure. Individual nucleotide interactions are derived by searching for co-varying sites in the alignment of the primary sequences. Consistent (e.g. AU→GU) and compensatory (e.g. CG→AU) substitutions, which maintain the base pairing ability at the corresponding site, are an indication for a particular base pairing in the secondary structure (Higgs 2000). Recent X-ray crystallographic studies (Ban et al. 2000; Schluenzen et al. 2000; Wimberly et al. 2000; Yusupov et al. 2001) have provided a direct test of the accuracy of rRNA comparative-structure models and confirmed almost all predicted secondary-structure base pairings in model organisms (Gutell et al. 2002), thus demonstrating the reliability of the comparative approach.

The mitochondrial small subunit (mt SSU or 12S) rRNA is a regularly applied marker in insect molecular systematics (Caterino et al. 2000), but its secondary structure has received little attention so far (Page 2000). Currently, there is a single 12S rRNA secondary-structure model (for *Drosophila virilis*) available on The Comparative RNA Web Site (Cannone et al. 2002), and only a few more (including one for the silk moth, *Bombyx mori*) can be obtained from The European Ribosomal RNA Database (Wuyts et al. 2004). In an exemplary publication, Page (2000) inferred a core set of base pairing interactions of the 12S rRNA among insects to automatically generate structures of other hexapod sequences. However, his investigation was restricted to domain III, the most frequently sequenced section of the 12S rRNA gene (Simon et al. 1994). Hence, comparative analyses dealing with domains I, II, and IV are not available.

In order to elucidate the evolutionary and biogeographic history of burnet moths (Zygaeninae), we compiled a large set of new, nearly complete or complete 12S rRNA sequences of the Zygaenoidea, a superfamily of ditrysian Lepidoptera, which comprises more than 2300 described species (Epstein et al. 1999). The sequences cover the domains I, II, III, and IV, although the 5′-end of domain I is missing in all but one sequence. When we applied the 12S rRNA secondary-structure model for the silk moth *B. mori* (Wuyts et al. 2004) to sequences of Zygaenoidea, we found evidence for major differences between the model and the potential actual structure in our investigated group. These differences concern the existence and extension of stems, the size of loops, as well as the presence of internal bulges.

In the present study, we propose a secondary-structure model of the 12S rRNA of Zygaenoidea derived from comparative sequence analysis. We discuss the differences between structural elements found in the Zygaenoidea and those assumed in the models for *B. mori* and *D. virilis*, and assess the reliability of the silk moth model in light of the new sequence data. The secondary-structure model provided will help improve the fit of parameters incorporated in doublet sequence evolution models. This, in turn, will hopefully lead to more accurate inferences of genealogical relationships (cf. Kjer 2004) and to realistic values of tree robustness (cf. Galtier 2004) in phylogenetic analyses of Zygaenoidea and other Lepidoptera.

Material and methods

Taxon sampling

12S rRNA sequences included in the present study are listed in Appendices A and B. Since the primary goal of our sequence compilation was a systematic study of burnet moths, the taxon sampling represents major families of Zygaenoidea with a specific focus on the subfamily Zygaeninae. For outgroup comparison, we also sequenced one species each of the superfamilies Sesioidea and Tortricoidea. In addition to our own data (Appendix A), we obtained 12S rRNA sequences for the bombycoid moths *Antheraea pernyi*, *Bombyx mandarina*, and *B. mori* (the only currently available complete sequences of this gene in Lepidoptera), as well as the homologous sequence of the fruit fly *D. virilis*, from GenBank (Appendix B). As the 5′-end of domain I is missing in most of our sequences, we retrieved d-loop entries for Lepidoptera from GenBank that cover this part of the 12S rRNA gene. These entries represent the superfamilies Hesperioidea, Noctuoidea, and Papilionoidea (Appendix B).

Molecular procedures

Total genomic DNA was extracted from muscle tissue by applying either the Qiagen DNeasy® Tissue kit or an equivalent system (Macherey-Nagel NucleoSpin® Tissue kit). In cases where only a single leg was available, we used a CHELEX extraction method (Gerken et al. 1998). If possible, voucher specimens were stored in absolute ethanol at −20 °C and deposited in the Alexander Koenig Research Institute and Museum of Zoology (ZFMK) in Bonn, Germany.

Nearly complete (or complete) 12S rDNA sequences were accomplished in two (or three) steps. In the first polymerase chain reaction (PCR), we amplified a stretch comprising the 5′-end of the 16S rRNA, tRNA-Val, and the 3′-end of the 12S rRNA by applying the oligonucleotide primers 16Sf5a and 16Sr5a (Table 1). In a few cases we used the primers 16Sf5b and/or 16Sr5b (Table 1) instead of those mentioned above. With a second PCR, the main section of the 12S rRNA sequence was amplified utilizing the primers 12Sf1a (rarely 12Sf1b) and 12Sr2 (Table 1). To achieve a complete sequence for our model species, *Zygaena sarpedon lusitanica*, the 5′-end of the 12S rRNA and the flanking d-loop (AT-rich region) were finally amplified with the (otherwise unsuccessful) primer combination 12Sf4 and 12Sr3 (Table 1).

PCR amplifications were performed in 50 µl volumes (0.75 U Taq polymerase (Sigma), 5 µl 10 × PCR buffer without $MgCl_2$ (Sigma), 7 µl $MgCl_2$ (25 mM), 4 µl dNTPs (2 mM), 0.8 µl of each primer (10 µM), 1 µl template DNA, filled up to 50 µl with sterile water), and carried out on a GeneAmp® PCR System 2700 or 9600 (Applied Biosystems) or a TGradient (Biometra®). The PCR temperature profile started with an initial 3 min denaturation step at 94 °C, followed by 15 cycles of 35 s at 94 °C, 30 s at 55–40 °C, and 1 min 30 s at 72 °C. Within the first 15 cycling steps, the annealing temperature was decreased by 1 °C each cycle, starting at 55 °C and ending at 40 °C. An additional 25 cycles followed with a constant annealing temperature of 50 °C. The profile ended with a 10 min extension step at 72 °C. PCR products were subsequently purified using Macherey-Nagel NucleoSpin® Extract kits.

All fragments were sequenced in both directions by using the specific PCR primers. In certain cases, however, we additionally employed the internal primers 12Sr1, 12Sf2, and 12Sf3 (Table 1) to ensure a high sequence quality of the 12S rRNA main section. Sequencing reactions were carried out using BigDye ReadyMix (Applied Biosystems) following the manufacturer's recommendations. After cleaning the sequencing products utilizing a standard ethanol-precipitation protocol, we separated and recorded them on an ABI PRISM® 377 sequencer (Applied Biosystems). Complement strands and overlapping fragments were finally assembled into contiguous arrays and trimmed to just

Table 1. Primers used to amplify and sequence mt SSU (12S) rRNA in moths

Name	Direction	Sequence (5'→3')	Source
12Sf1a	Forward	*TATAAAATGAAAGCGACGGGC*	Niehuis, present study
12Sf1b[a]	Forward	*AAGAGCGACGGGCGATGTGT*	Simon et al. (1994)
12Sf2	Forward	*TTAAGTAAATTTAATCGTGG*	Niehuis, present study
12Sf3	Forward	*CAATTATTARACAGATTCCTCT*	Niehuis, present study
12Sf4	Forward	*ACGGTATCTAATCCTAGTCT*	Niehuis, present study
12Sr1	Reverse	*TAGTTCATTTAGAGGAATCTG*	Niehuis, present study
12Sr2[b]	Reverse	*GACAAAATTCGTGCCAGCAGT*	Simon et al. (1994)
12Sr3	Reverse	*AAATAATCCTTWWTCAGGCA*	Niehuis, present study
16Sf5a	Forward	*ATTAATAAACTCTGATACAC*	Niehuis, present study
16Sf5b	Forward	*AAACTCTGATACACAAGATAC*	Niehuis, present study
16Sr5a	Reverse	*AAAATTAAATCAGATCAAGATG*	Niehuis, present study
16Sr5b	Reverse	*AAATTAAATCAGATCAAGATGC*	Niehuis, present study

[a]Alias SR-J-14233.
[b]Alias SR-N-14756.

the 12S rRNA gene. All sequences have been submitted to the EMBL Data Library (for accession numbers see Appendix A).

Sequence and structure analysis

Sequences were initially aligned using Clustal X 1.8 (Thompson et al. 1997). The alignment was subsequently checked visually and corrected for obviously misaligned positions to maximize primary sequence homology utilizing BioEdit 7.0.0 (Hall 1999). In a third step, we made use of the *Escherichia coli* secondary-structure model of the SSU rRNA (Gutell et al. 1994; Cannone et al. 2002) to search for conserved motives, which are associated with specific structural elements. The resulting skeleton served as a starting point for the structural investigation of the more variable parts of the molecule. We studied patterns of co-variation by calculating the frequencies of nucleotide pairs and mutual information indices $M(x,y)$ (Gutell et al. 1992) in BioEdit. By searching for consistent and compensatory substitutions (CCS), we inferred core base pairing interactions among the studied species. We recognized standard Watson–Crick base pairs and non-canonical G:U interactions, but considered other types of non-canonical base pairings when they were proposed in the models for *E. coli* and *D. virilis* at corresponding sites. Currently assumed phylogenetic relationships of higher taxa considered in the present data set (Minet 1991; Fänger 1999; Fänger et al. 1999; Fänger and Naumann 2001), as well as morphologically circumscribed species groups within the genus *Zygaena* (Hofmann and Tremewan 1996, 2003) were used to assess the frequency of independent CCS. To explore possible folds in highly variable and difficult-to-align sections of the rRNA molecule, we used the MFOLD web server (Zuker 2003) to predict structures of minimum free energy (Mathews et al. 1999). In sections where the present sequence compilation did not show enough variation to derive and/or support structural features, we adopted the secondary structure assumed in the *E. coli* model. All secondary structures were drawn using the program RnaViz 2.0 (De Rijk et al. 2003). Homologous helices are annotated according to Wuyts et al. (2002) and base pairs named by the combination of the individual figures of involved nucleotides (i.e. 5′:3′). Structure logos (Schneider and Stephens 1990; Gorodkin et al. 1997) were used to summarize sequence variation, the relative frequency of nucleotides, and the information content of selected helices. The complete sequence alignment with co-notated structure information is available upon request.

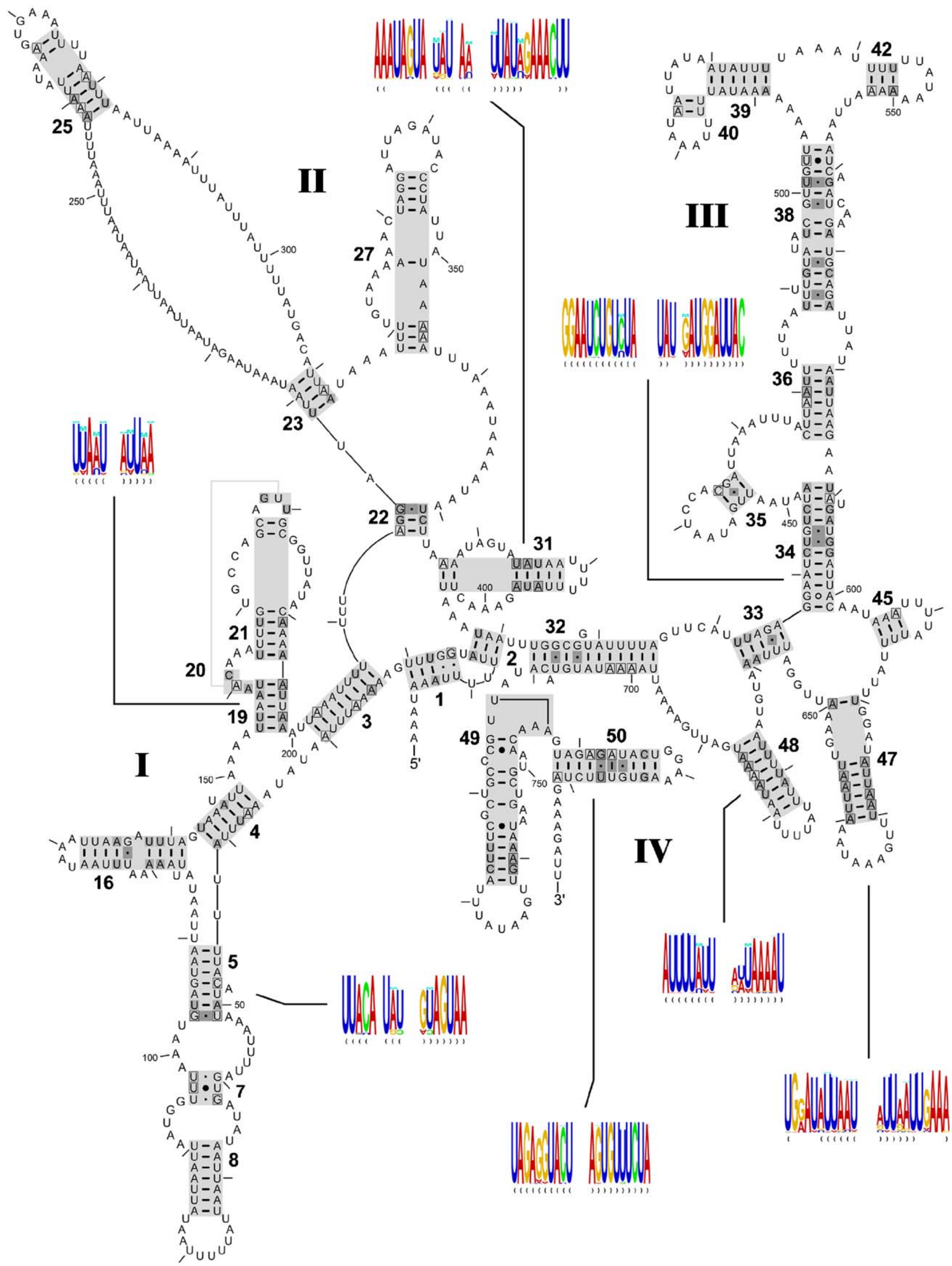
Results

With sequence distances ranging from 0.00 to 0.21 (mean divergence 0.07; SD 0.04) among the Zygaenoidea, it was generally possible to apply the comparative approach in both variable and conserved sections of the mt SSU rRNA molecule. Our proposed secondary-structure model of the 12S rRNA is shown in Fig. 1. The rRNA of the model species selected, *Z. sarpedon lusitanica*, consists of 789 nucleotides and probably folds into 30 helices. Most of these received justification from the present sequence compilation by CCS. The only exception is helix 8, currently based exclusively on thermodynamic considerations, since homologous sequences from other zygaenoid moths were missing for this highly variable section of the molecule, and sequences from other ditrysian Lepidoptera unfortunately contributed little to a conclusive result.

All helices in the model for *Zygaena sarpedon* proved to be homologous with regions paired also in the 12S rRNA models for *B. mori* (Wuyts et al. 2004) and *D. virilis* (Cannone et al. 2002). This allowed a convenient comparative evaluation of different structural hypotheses using consistent terminology.

The 12S rRNA model proposed here differs from that for the silk moth, *B. mori*, at first glance by three additionally assumed stems in domain I (i.e. helices 5, 7, and 8) and by three missing helices (24, 26, and 41) in domains II and III, respectively. The presumed absence of helices 24 and 26 is uncertain, however. High sequence variation and an elevated AT content in the corresponding sections of the molecule confounded a comparative sequence analysis (see below). Further and significant discrepancies between the two models concern individual base pair interactions in helices 22, 23, 25, 31, 39, 40, and 49.

Fig. 1. Proposed secondary structure of mt SSU (12S) rRNA of *Zygaena* (*Mesembrynus*) *sarpedon lusitanica* Reiss, 1936 (Lepidoptera: Zygaenidae; AJ785727). Nucleotides are continuously numbered beginning at 5′-end of the molecule; tick marks identify every tenth base. Pale shading indicates helical structures numbered according to Wuyts et al. (2002). Canonical Watson–Crick interactions represented by dashes, non-canonical guanine–uracil interactions by dots, guanine–adenine interaction by an open circle, all other non-canonical interactions by solid circles. Dominant G:U pairings are darkly shaded. Boxed nucleotides indicate positions displaying consistent substitutions; compensatory substitutions are specified by dark shading. The consensus sequence, relative frequency of nucleotides, and information content of selected helices is displayed by structure logos (height of a nucleotide symbol is proportional to its frequency; letter M indicates amount of mutual information). Roman numerals specify domains I–IV.


I
II
III
IV
5'
3'

In the following section, we will set forth the individual evidence for proposed base pair interactions. Note that the comparative analysis of domain I is predominantly based on sequences of Bombycoidea, Hesperioidea, Noctuoidea and Papilionoidea, as only one complete sequence of a zygaenoid moth was available.

Domain I

Helix 1: This helix is probably five base pairs long in Zygaenoidea. Confidence in the structure emerged from a compensatory change at site 7:18 between the fruit fly and the burnet moth sequence. The other ditrysian Lepidoptera sequences also support the existence of this helix. However, the alignment in this section is ambiguous and does not allow assessing the exact number of consistent substitutions. Whether an additional base pairing between nucleotides 4 and 21 occurs in Zygaenoidea is unclear. Further ingroup sequences are necessary to answer this question.

Helix 2: A three base-pair-long stem is assumed to occur in Zygaenoidea. A compensatory change at site 14:408 between Diptera and ditrysian Lepidoptera gave some support. But lack of complete 12S rRNA sequences precluded a more detailed analysis. Sequence deviations on the 3′-end side of the helix in *H. penella* and *S. bembeciformis* make these two species attractive candidates for future investigations. The silk moth model suggests a fourth base pair at site 15:407. However, due to missing variation at the corresponding sites in the sequence compilation, no conclusions have been drawn. We adopted the structure of the *E. coli* model here, which assumes only three base pairs. Confirmation of a compensatory change at position 15:407 in moths would favour the hypothesis of an additional hydrogen bond here.

Helix 3: We hypothesize that eight base pairs are interacting in ditrysian Lepidoptera. Whether two additional base pairings occur in zygaenoid moths at the distal end of the helix is unclear. The assumption of these two hydrogen bonds would require to suppose internal U:U interactions in many species especially at site 31:201. Helix 3 is supported by consistent substitutions at nucleotide positions 24, 26, 30, 203 and by compensatory substitutions at site 24:208. But there are also taxa which have mismatches or U:U interactions at sites 26:206 and 24:208: *A. pernyi* (U:U, 26:206), *J. evagoras* (A:A, 26:206), *E. oeme* (A:C, 24:208), and a further eight moth or butterfly species with U:U pairings at site 24:208. The fruit fly model assumes identical base pair interactions but hypothesizes two additional hydrogen bonds at the distal end of the helix (see above). The silk moth model, on the other hand, supposes two more hydrogen bonds at the base of the helix and a single nucleotide bulge at position 207. The present sequence compilation of ditrysian Lepidoptera is mostly incompatible with this model.

Helix 4: This helix is probably six base pairs long in Zygaenoidea. Helix 4 received support from two compensatory (sites 38:147 and 41:143) and three consistent substitutions (position 37). In the silk moth model, an additional interaction between nucleotides 35 and 150 has been proposed, but the substitution pattern in the present data set was contradictory, as none of the observed substitutions were consistent or compensatory.

Helix 5: Helix 5 is not recognized in the *B. mori* model. However, three independent compensatory substitutions at site 50:104 and consistent substitutions at positions 47, 51, and 103 suggested nucleotide interactions. We propose a pairing of nucleotides 44–51 (but not 48) with 103–109. Identical base pair interactions have been assumed in the models for *E. coli* and *D. virilis*.

Helices 7 and 8: The section of the 12S rRNA molecule enclosed by helix 5 is highly variable among ditrysian Lepidoptera and difficult to align (Fig. 2). It forms ten prominent stems in *E. coli* (Cannone et al. 2002), but in insects most of these are reduced. In the *D. virilis* model, only two stems (helices 7 and 8) are hypothesized, whereas the *D. melanogaster* model provided by the European Ribosomal RNA Database suggests four. Comparative sequence analysis of the ditrysian Lepidoptera sequences revealed no convincing evidence for any helix in this section of the molecule, but thermodynamic considerations implied a seven base-pair-long stem at the position of helix 8 in the *D. virilis* model. Consideration of helix 7 in Lepidoptera requires assuming non-canonical nucleotide interactions (U:U and A:G) in the majority of species. We considered both helices in our present burnet moth secondary-structure model, but emphasize that their occurrence in Zygaenoidea and other ditrysian Lepidoptera is poorly supported and needs confirmation.

Helix 16: This stem is highly conserved and well supported in ditrysian Lepidoptera. Consistent substitutions between Diptera and Ditrysia confirmed base pair interactions at sites 119:139, 120:138, and 124:135. The homologous helix in the silk moth model is almost identical but assumes no lateral single nucleotide bulge at the base of the helix. The present data set is congruent with both hypotheses.

Helix 19: Compensatory substitutions at sites 154:199, 155:198, 157:196, and 158:195 and two single consistent substitutions at positions 195 and 198 gave high confidence in the presence of this helix in ditrysian Lepidoptera.

Helix 20: This small stem received support from two compensatory substitutions at sites 159:180 and 161:178, as well as from a single consistent substitution at position 160.

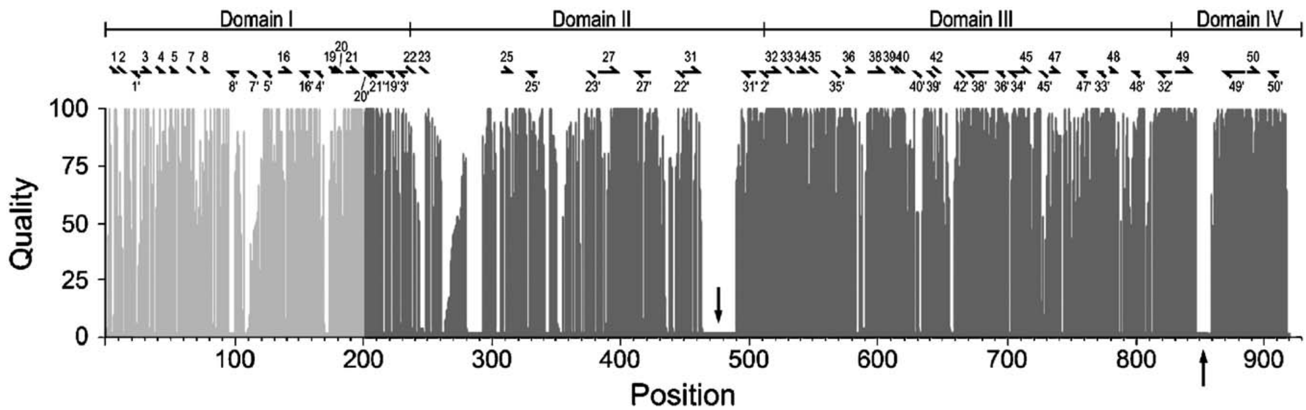


Fig. 2. Quality plot of the 12S rRNA alignment. Quality scores correspond with values of alignment-quality analysis in Clustal X. Dark shading indicates quality scores based on the complete taxon sampling: pale shading indicates 5′-end of domain I for which only a limited number of sequences were available. Horizontal arrows specify locations of stems in secondary-structure model of mt SSU rRNA in *Zygaena sarpedon lusitanica*. Downward-pointing arrow marks terminal loop of helix 31 extended in *Neurosymploca caffra* and *N. concinna*; upward-pointing arrow denotes enlarged terminal loop of helix 49 in *Himantopterus dohertyi* and *Somabrachys aegrota*.

Helix 21: Helix 21 consists of a five base-pair-long basal stem, an internal bulge, and two additional distal nucleotide pairings. An identical structure has been assumed in the models for *E. coli* and *D. virilis* on The Comparative RNA Web Site. The *B. mori* model from the European Ribosomal RNA Database, on the other hand, does not assume the two distal base pairs. Compensatory (165:194, 168:191) and consistent substitutions (position 191) supported the basal stem, but did not allow us to discern whether or not the distal nucleotides interact.

Domain II

Helix 22: We assume a three base-pair-long stem, although in some taxa (e.g. *B. mori, H. penella, S. bembeciformis*) the distal nucleotide interaction (215:371) seems to be absent. Confidence in the existence of this helix emerged from compensatory substitutions at sites 214:372 and 215:371 as well as from consistent substitutions at nucleotide positions 213 and 215. In the genus *Neurosymploca,* the insertion of a single base between nucleotides 212 and 213 implied an additional fourth pairing with nucleotide 374. The silk moth model suggests a three base-pair-long helix 22, but individual nucleotide interactions deviate from our model; CCS indicated that these interactions are most likely incorrect.

Helix 23: The primary structure of the 12S rRNA molecule between helices 22 and 27 is extremely variable among the studied taxa and only aligned ambiguously (cf. Fig. 2). An exception is the enclosed section of helix 25 discussed below. The nucleotide sequence variation and the high AT bias, which reduced information content to almost only two character states, effectively precluded comparative analysis in the present case. We currently propose only a four base-pair-long helix 23, which received some support from compensatory substitutions at site 218:313 and two consistent substitutions at position 312. The models for *B. mori* and *D. virilis* additionally suggest the presence of helices 24 and 26, but we have not been able to unequivocally infer any consensus structure besides the one mentioned above. We emphasize, however, that the present structure is a preliminary hypothesis.

Helix 25: This helix is well conserved in moths. Compensatory substitutions indicated base pair interactions at sites 258:282 and 260:280; consistent substitutions at nucleotide positions 258, 259, 260, 267, and 279 added further confidence. However, in eight species internal U:U base pairings must be hypothesized to ensure a continuous helical structure (e.g. *O. nebulosa* at site 266:275, *A. infausta* at site 260:280, *S. bembeciformis* at site 159:281). In *H. dohertyi*, the first nucleotide pair of the stem (i.e. 258:282) is absent. Whether helix 25 is extended proximally by two additional base pairs remains unclear. In most species, thermodynamic considerations supported two additional base pairings, but the substitution pattern at the corresponding sites was not unequivocally supportive (possible compensatory and non-compensatory substitutions co-occurred). Compared with the fruit fly model, we assume a more prominent internal bulge in Zygaenoidea. We realized that a pairing of nucleotides 263 and 277 is possible in most taxa, but none of the observed substitutions at these sites were consistent or compensatory. The *B. mori* model assumes entirely different base pair interactions in helix 25 (Fig. $3A_{I}$). The present sequence compilation clearly contradicted this structure and supported the alternative folding shown in Fig. $3A_{II}$.

Helix 27: The distal section of this helix is extremely conserved. In contrast to the *E. coli* model, but in

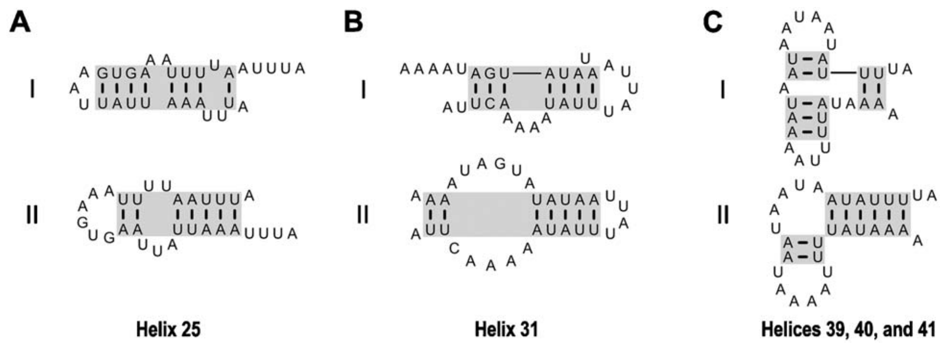


Fig. 3. Major differences between base pair interactions assumed in mt SSU rRNA secondary-structure models of the moths *Bombyx mori* (top; Wuyts et al. 2004) and *Zygaena sarpedon lusitanica* (bottom; present study); helices indicated by shading.

congruence with the structure assumed for *D. virilis*, we propose a single central base pair in moths (i.e. 327:351) that is flanked by loops. The proximal extension of helix 27 is uncertain. Although at least four base pairings could be hypothesized in the majority of Zygaenoidea, the substitution pattern supported only three. The silk moth model, on the other hand, assumes two internal nucleotide interactions and a five base-pair-long proximal stem.

Helix 31: Helix 31 can be divided into a proximal and a distal section separated by an internal bulge. The proximal section consists of two base pairs. A third pairing, as assumed in the *E. coli* and *D. virilis* models, would have required postulating a non-canonical A:C interaction. Only one (less likely two independent) consistent substitution(s) at position 376 supported a base pairing at site 376:404. The distal section of helix 31 seems to be length-variable among taxa. The substitution pattern suggests seven hydrogen bonds in the genus *Neurosymploca*. However, only six base pairs are supported by the comparative analysis in most of the remaining zygaenid moth taxa. In some groups (Chalcosiinae, *C. splendens, Z. loti*, *Z. johannae-felix* species group), thermodynamic considerations imply a further reduction to only four, respectively three hydrogen bonds. In a few taxa (*C. pronubana*, *Neurosymploca caffra*, *Neurosymploca concinna*, and *S. aegrota*), U:U pairs at site 384:397 lead us to assume U:U interactions. The terminal loop of helix 31 is considerably extended in *N. caffra* and *N. concinna*, which is a synapomorphy of these two species (Fig. 2). Proposed base pairings in the *B. mori* model are entirely different (Fig. $3B_{I}$). The present data set contradicted the silk moth model and supported the structure shown in Fig. $3B_{II}$. Similar nucleotide interactions have been hypothesized in the models of *E. coli* and *D. virilis*.

Domain III

Helix 32: This stem is highly conserved among moths. Sequence comparisons revealed few variable sites (pos. 700, 701, 702) with all substitutions being consistent. The distal extension of the helix remains unclear. In Sesioidea, Tortricoidea, and Zygaenoidea, three additional base pairings were conceivable (426:697, 427:696, 428:695), but the nucleotides of the first two pairs were invariant and the substitution pattern in the third potential pair was ambiguous, leaving the question open. The sequence data in *Bombyx* spp. contradicted an expansion, which has not been assumed in the silk moth model either.

Helix 33: This helix is well supported in insects by CCS (Page 2000). Compensatory substitutions in the two proximal base pairs (432:660, 433:659) also supported its occurrence in moths.

Helix 34: Compensatory substitutions confirmed most of the inner (442:596, 443:595) and distal nucleotide interactions (445:593, 446:592, 448:589); a single substitution at nucleotide position 590 can be interpreted as consistent. In *N. caffra* and *N. concinna*, the usually single base bulge at position 591 is expanded by an additional nucleotide; a further synapomorphy of these two species. In congruence with the *E. coli* model, a non-canonical G:A pairing is assumed at site 438:600. In the secondary-structure model for the silk moth as well as in the model for *D. yakuba* published by Page (2000), this pairing has been avoided by shortening the stem by two base pairs. However, the lack of variation at these sites prevented a definitive decision, therefore we adopted the *E. coli* structure.

Helix 35: Page (2000) found high values of mutual information for the nucleotide pairs 454:465 and 455:464. However, two additional proximal pairings have been assumed in *D. virilis*. In accordance with the *B. mori* model, three nucleotide pairs are conceivable in Zygaenoidea, but this was not supported by the comparative analysis. Assuming a fourth pairing at the base of the helix would have been possible in many taxa, but substitutions observed at the corresponding sites were neither compensatory nor consistent.

Helices 36 and 38: Both helices are highly preserved among insects (Page 2000). The structures assumed for Zygaenoidea are in accordance with the silk moth model. In *D. virilis*, helix 38 is one base pair shorter at its distal end. In one species (*R. simonyi*), an internal U:U binding had to be supposed to ensure a continuous helical structure.

Helix 39: Page (2000) found evidence for this six base-pair-long stem in insects. This is in congruence with the nucleotide interactions proposed for *E. coli* and *D. virilis*. Helix 39 can be drawn in all investigated moths, and comparative sequence analysis yielded at least some support by compensatory substitutions at site 509:535. However, in one species (*Z. carniolica*) an A:C base pair had to be assumed at site 511:533. The silk moth model proposes only a two base-pair-long helix 39. These base pairs correspond to nucleotide binding at sites 509:535 and 510:534 in the *Z. sarpedon* model.

Helix 40: We assume a two base-pair-long helix 40 in Zygaenoidea. A single consistent substitution at site 523 gave some confidence for a pairing of nucleotides 516 and 523. A third base pairing, as hypothesized in *D. virilis*, *D. yakuba*, and *E. coli*, seemed unlikely as it would have required supposing U:U base pair interactions in more than 90% of the taxa. Strongly deviating base pair interactions have been proposed in the *B. mori* model, which also suggests an additional helix 41 (Fig. $3C_{I}$). Due to the evidence Page (2000) found for helices 39 and 40, we favour his predicted structures here (Fig. $3C_{II}$) and do not adopt those proposed in the silk moth model.

Helix 42: As recently stated (Page 2000), there is little evidence for helix 42 in insects as a whole. In moths, CCS suggested a binding of nucleotides 543 and 550. Reliance on an interaction of nucleotides 541 and 553 came from two (less likely three) independent consistent substitutions at position 553. In *Z. sarpedon*, we draw helix 42 as shown in the *D. virilis* model. However, we assume only three base pairs, not four. Nonetheless, in eight taxa of moths (e.g. *A. infausta*, *C. splendens*, *P. sinica*, *S. bembeciformis*, *Z. storaiae*) even the third pairing is absent, resulting in a stem only two base pairs long. Our inferred structure resembles that proposed by Page (2000), but in the proposed secondary-structure model of *Z. sarpedon*, helix 42 is shifted one nucleotide towards the terminal loop. The structure assumed in our burnet moth model is identical with that of the silk moth model.

Helix 45: This prominent helix in the *E. coli* model is difficult to establish in insects (Page 2000). A three base-pair-long stem can be assumed in all studied moths, but a fourth base pairing, as suggested in the *B. mori* model, seems unlikely in Zygaenoidea, as in more than half of the investigated species a U:U binding would have to be assumed. In the *D. virilis* model, a stem of only two base pairs is assumed, but confidence in a third base pair in moths is based on consistent substitutions at site 607.

Helix 47: Helix 47 consists of a lone nucleotide pair at the base of the helix (620:650), separated from a distal stem by an internal bulge. The single base pair was not well supported, but some confidence emerged from consistent substitutions and a compensatory change correlated with the split Diptera/Zygaenoidea. In one species (*N. remota*), however, the single binding is most probably lost by a substitution leading to a non-canonical A:A pair. No nucleotide interaction at site 620:650 has been presumed in the silk moth model. The distal stem of helix 47 consists of at least six base pairs in almost all taxa and is well supported by CCS. However, in some species U:U base pairs are likely to occur, especially at the ends of the stem. In *H. dohertyi*, a lateral bulge consisting of two nucleotides had to be hypothesized. In *R. brandti* and *R. pruni*, a non-canonical A:A base pair at site 627:643 interrupts the distal section of the helix. Whether the stem is longer than six base pairs in some taxa (e.g. *Z. sarpedon*) is unclear. Thermodynamic considerations supported up to eight pairings, but almost all substitutions observed were non-compensatory. In the silk moth model, only five base pair interactions have been proposed for the distal stem section.

Helix 48: As in the models for *D. virilis* and *E. coli*, eight base pairings can be assumed in almost all groups of Zygaenoidea. Most of them received support from CCS. In a few taxa (e.g. *G. flavivitella*, *P. atratus*, *R. brandti*) internal U:U base pairs had to be proposed, but in one species (*T. ampellophaga*) an internal A:A nucleotide pair probably interrupts the helical structure. Less than eight base pairs seem to be present in the genus *Neurosymploca* and in some subordinate taxa of *Zygaena* (e.g. *Z. johannae-felix* group). The silk moth model suggests only six base pairs, omitting the two distal interactions.

Domain IV

Helix 49: This helix is 11 base pairs long in moths, and CCS supported the distal part of the stem (sites 725:741, 726:740, and 727:739). In the fruit fly and the silk moth models, two additional base pair interactions are proposed at the distal end of the stem, but we found no evidence for this in the comparative analysis. Two additional nucleotide interactions would also have required assuming U:U base pairings at site 729:737 in all species and further non-canonical A:A pairs at site 730:736 in about 20%of the taxa. The terminal loop of helix 49 comprises nine nucleotides in all species except *H. dohertyi* and *S. aegrota*, in which the loop is considerably enlarged; a potential synapomorphy of these two taxa. The silk moth model shows only a six base-pair-long helix 49 corresponding to nucleotides 725–730 and 776–741 in the burnet moth model. Base pair interactions in the proximal part of helix 49 as hypothesized here are entirely adopted from the *E. coli* model, since lack of sequence variation in this section of the molecule prevented a clear decision between both hypotheses.

Helix 50: Helix 50 was well supported by CCS and is identical with the homologous helices in the models for *B. mori* and *D. virilis*.

Discussion

The present investigation aimed at deriving a secondary-structure model of the mt SSU rRNA of Zygaenoidea using a comparative approach. The structure inferred closely resembled that of other arthropods, in particular that of *D. virilis*. It thus confirmed most of the previously proposed base pair interactions in the 12S rRNA molecule, but also revealed some unique features. A comparison with the secondary-structure model for the silk moth, for instance, indicated strikingly different base pair interactions in some of the helices (e.g. 22, 23, 25, 31, 39, 40, and 49). The present sequence compilation strongly suggested modifying the silk moth model with respect to helices 22, 23, 25, and 31, and implied the existence of helix 5 in both, Zygaenoidea and Bombycoidea. Evidence from previous investigations on the secondary structure of domain III in insects (Page 2000) also pointed indirectly to alternative base pair interactions in the *B. mori* model (i.e. helices 39 and 40; Fig. $3C_{I,II}$).

Not all discrepancies between the silk moth and burnet moth models (e.g. helices 2, 21, 34, 49) are the result of new evidence. Comparative analysis critically depends on nucleotide variation (i.e. CCSs); therefore lack of variation poses a problem. As a consequence of a more local taxonomic scope, for example, comparative analysis may not help in the more conserved sections of a molecule. In the present study, we treated this problem by adopting secondary-structure elements of the *E. coli* model in cases where variation was absent; because this model was derived by considering much more divergent sequences, thus providing information for the most conserved molecule sections. By doing so, we rely on the accuracy of the *E. coli* model and on the assumption that base pairings in the *E. coli* structure are correct in the burnet moth model as well. Crystallographic studies of the ribosomal subunits (Ban et al. 2000; Schluenzen et al. 2000; Wimberly et al. 2000; Yusupov et al. 2001), however, almost entirely confirmed the secondary-structure models of the reference organism, *E. coli* (Gutell et al. 2002).

Structural variation within Zygaenoidea appeared to be minor and was restricted to quantitative characters. In this sense, the 12S rRNA molecule proved to be highly conserved. Nonetheless, noteworthy structural variation was observed in helices 31, 34, 47, and 49. Two species groups received support from structural characters derived independently of strict phylogenetic preassumptions: a taxon consisting of *N. caffra* and *N. concinna*, and a group comprising *Himantopterus dohertyi* (Himantopteridae) and *Somabrachys aegrota* (Somabrachyidae). The former group is characterized by a significantly enlarged terminal loop in helix 31 and an extra nucleotide in the lateral bulge of helix 34, whereas the latter group has an extended terminal loop in helix 49. In helix 47, the comparative analysis also suggested stem lengths deviating among subordinate groups (i.e. the substitution pattern clearly supported a nucleotide interaction in one taxon, but the pattern in another species group contradicted such an interaction). However, this estimation is based on the hypothesis that the investigated species groups are monophyletic. As this need not be the case, the derived molecular–morphological characters may not be regarded as independent evidence for a monophyly of these groups.

Inferring a secondary-structure model for the 12S rRNA in Zygaenoidea had been motivated by the intention to apply doublet substitution models in phylogenetic analyses of burnet moths; these models take correlation in paired sites of a molecule into account. The secondary-structure model provided here may help to apply these sequence evolution models in Zygaenoidea and other ditrysian Lepidoptera. The limitations of the burnet moth model discussed above are largely insignificant in this context. Erroneously assumed base pair interactions in invariant parts of the molecule may have only minor effects on substitution model parameters, and the influence of a wrongly proposed lack of helices 24 and 26 in Zygaenoidea may also be negligible: the paired nucleotide sites obviously are less strongly correlated, probably due to slipped-stand mispairing, and the ambiguous sequence alignment suggests rejecting evidence from this section of the data set. However, a certain degree of inaccuracy in the specification of paired sites in the sequence alignment is inevitably due to the observed structural plasticity, as current doublet sequence evolution models assume parameters stationary along lineages.

Acknowledgments

For providing specimens or tissue samples, we are grateful to E. Aistleitner, U. Aistleitner, V. Anikin, Y. Arita, S. Assad, W. Billen, W. Bischoff, G. Broquet, G. Buntebarth, M. Danilevsky, E. Drouet, F. Fernández-Rubio, H. Fänger, X. Fangsen, H. Geertsema, J. Grosser, W. ten Hagen, A. Hofmann, M. Kalashian, T. Keil, T. Khan, C. Klütsch, H. Krenn, A. Kreuzberg, E.O. Krüger, P. Langhoff, A. Mochizuku, B. Mollet, M. Niehuis, M. Owada, G. Pape, K. Schurian, E. Sepp, D. Tarkhnishvili, M.R. Tarrier, W.G. Tremewan, F.

Vegliante, R. Vila, M. Wiemers, F. Wilhelm, S.-H. Yen, and A. Zilli. Field collecting by O.N. in North Africa was kindly supported by the Alexander Koenig Foundation; O.N. further acknowledges a Doctoral Fellowship from the German National Academic Foundation. We thank M. Bleidißel, B. von Reumont, and B. Ullrich for valuable aid in the laboratory, and C. Etzbauer for technical assistance. Special credit goes to P. Langhoff for her help sequencing the d-loop in *Z. sarpedon*. We finally appreciate helpful comments from E. Bettag, J.R. Codd, A. Hofmann, K.M. Kjer, M. Niehuis, B. Sinclair, B. Ullrich, and an anonymous reviewer. This research has been supported by the "Evolution and Biodiversity in Time and Space" Research Training Group of the German National Research Foundation at the Rheinische Friedrich-Wilhelms University in Bonn. The present study is a contribution from the Molecular Systematics Unit at the ZFMK.

Appendix A. Species names and EMBL accession numbers of taxa sequenced in the present study

Systematic category[a]	Species[b]	Accession no.
Sesioidea		
Sesiidae	*Sesia bembeciformis* (Hübner, [1806])	AJ785615
Tortricoidea		
Tortricidae	*Cacoecimorpha pronubana* (Hübner, 1799)	AJ785616
Zygaenoidea		
Heterogynidae	*Heterogynis penella* (Hübner, 1819)	AJ785617
Himantopteridae	*Himantopterus dohertyi* (Elwes, 1890)	AJ785618
Lacturidae	*Gymnogramma flavivitella* (Walsingham, 1881)	AJ785619
Limacodidae	*Apoda limacodes* Hufnagel, 1766	AJ785620
Phaudidae	*Phauda mimica* Strand, 1915	AJ785627
Somabrachyidae	*Somabrachys aegrota* (Klug, 1830)	AJ785621
Zygaenidae		
Callizygaeninae	*Callizygaena splendens* Candeze, 1927	AJ785622
Chalcosiinae	*Aglaope infausta* (Linnaeus, 1767)	AJ785623
	Aglaope labasi Oberthür, 1922	AJ785624
	Neochalcosia remota (Walker, 1854)	AJ785625
	Pidorus atratus Butler, 1877	AJ785626
Procridinae	*Adscita geryon* (Hübner, [1813])	AJ785628
	Adscita mannii (Lederer, 1853)	AJ785629
	Adscita mauretanica (Naufock, 1932)	AJ785630
	Jordanita hector (Jordan, 1907)	AJ785631
	Rhagades brandti (Alberti, 1938)	AJ785632
	Rhagades pruni ([Denis and Schiffermüller], 1775)	AJ785633
	Theresimima ampellophaga (Bayle-Barelle, 1808)	AJ785634
	Thyrassia penangae (Moore, 1859)	AJ785635
	Zygaenoprocris persepolis (Alberti, 1938)	AJ785636
Zygaeninae	*Pryeria sinica* Moore, 1877	AJ785637
	Epizygaenella caschmirensis caschmirensis (Kollar, 1844)	AJ785638
	Neurosymploca caffra (Linnaeus, 1764)	AJ785639
	Neurosymploca concinna (Dalman, 1823)	AJ785640
	Neurosymploca sp. 1 (morphotype '*atomarina*')	AJ785641
	Neurosymploca sp. 2 (morphotype '*geertsemai*')	AJ785642
	Neurosymploca sp. 3 (morphotype '*magnifica*')	AJ785643
	Orna nebulosa (Guérin-Méneville, 1832)	AJ785644
	Praezygaena agria (Distant, 1892)	AJ785645
	Praezygaena ochroptera (Felder, 1874)	AJ785646
	Reissita simonyi yemenicola Tremewan, 1959	AJ785647
	Zygaena (*Agrumenia*) *afghana afghana* Moore, [1860]	AJ78564

Appendix A (*continued*)

Zygaeninae (cont.)	*Zygaena* (*Agrumenia*) *algira algira* Boisduval, 1834	AJ785649
	Zygaena (*Agrumenia*) *alluaudi alluaudi* Oberthür, 1922	AJ785650
	Zygaena (*Agrumenia*) *bakhtiyari* Hofmann and Tremewan, 2005	AJ78565
	Zygaena (*Agrumenia*) *beatrix metaxys* Dujardin, 1973	AJ785652
	Zygaena (*Agrumenia*) *carniolica virginea* Müller, 1766	AJ785653
	Zygaena (*Agrumenia*) *chirazica eckweileri* Naumann and Naumann, 1980	AJ785654
	Zygaena (*Agrumenia*) *cocandica minor* Erschoff, 1874	AJ785655
	Zygaena (*Agrumenia*) *escalerai escalerai* Poujade, 1900	AJ785656
	Zygaena (*Agrumenia*) *excelsa rosei* Hofmann, 1980	AJ785657
	Zygaena (*Agrumenia*) *fausta elodia* Powell, 1934	AJ785658
	Zygaena (*Agrumenia*) *fausta fassnidgei* Tremewan and Manley, 1965	AJ785659
	Zygaena (*Agrumenia*) *felix boursini* Dujardin, 1973	AJ785660
	Zygaena (*Agrumenia*) *felix hemerocallis* Dujardin, 1973	AJ785661
	Zygaena (*Agrumenia*) *formosa hesselbarthi* Junge, Naumann and Rose, 1977	AJ785662
	Zygaena (*Agrumenia*) *formosa molleti* Hofmann, in prep.	AJ785663
	Zygaena (*Agrumenia*) *fraxini fraxini* Ménétriés, 1832	AJ785664
	Zygaena (*Agrumenia*) *haberhaueri elbursica* Tremewan, 1975	AJ785665
	Zygaena (*Agrumenia*) *hilaris escorialensis* Oberthür, 1884	AJ785666
	Zygaena (*Agrumenia*) *johannae johannae* Le Cerf, 1923	AJ785667
	Zygaena (*Agrumenia*) *kavrigini* Grum-Grshimailo, 1887	AJ785668
	Zygaena (*Agrumenia*) *marcuna tingitana* Reiss, 1937	AJ785669
	Zygaena (*Agrumenia*) *maroccana tichkana* Wiegel, 1973	AJ785670
	Zygaena (*Agrumenia*) *occitanica huescacola* Tremewan and Manley, 1965	AJ785671
	Zygaena (*Agrumenia*) *olivieri dsidsilia* Freyer, 1851	AJ785672
	Zygaena (*Agrumenia*) *orana contristans* Oberthür, 1922	AJ785673
	Zygaena (*Agrumenia*) *pamira pamira* Sheljuzhko, 1919	AJ785674
	Zygaena (*Agrumenia*) *rosinae brandti* Reiss, 1937	AJ785675
	Zygaena (*Agrumenia*) *rosinae sengana* Holik and Sheljuzhko, 1956	AJ785676
	Zygaena (*Agrumenia*) *sedi sedi* Fabricius, 1787	AJ785677
	Zygaena (*Agrumenia*) *separata separata* Staudinger, 1887	AJ785678
	Zygaena (*Agrumenia*) *sogdiana sogdiana* Erschoff, 1874	AJ785679
	Zygaena (*Agrumenia*) *storaiae storaiae* Naumann, 1974	AJ785680
	Zygaena (*Agrumenia*) *transpamirina andarabensis* Koch, 1938	AJ785681
	Zygaena (*Agrumenia*) *transpamirina transpamirina* Koch, 1936	AJ785682
	Zygaena (*Agrumenia*) *truchmena esseni* Blom, 1973	AJ785683
	Zygaena (*Agrumenia*) *truchmena ferganica* Holik and Sheljuzhko, 1956	AJ785684
	Zygaena (*Agrumenia*) *youngi youngi* Rothschild, 1926	AJ785685
	Zygaena (*Agrumenia*) *youngi glaoua* Wiegel, 1973	AJ785686
	Zygaena (*Mesembrynus*) *aisha* Naumann and Naumann, 1980	AJ785687
	Zygaena (*Mesembrynus*) *alpherakyi alpherakyi* Sheljuzhko, 1936	AJ785688
	Zygaena (*Mesembrynus*) *aurata aurata* Blachier, 1905	AJ785689
	Zygaena (*Mesembrynus*) *brizae vesubiana* Le Charles, 1933	AJ785690
	Zygaena (*Mesembrynus*) *cacuminum* Christoph, 1877	AJ785691
	Zygaena (*Mesembrynus*) *cambysea cambysea* Lederer, 1870	AJ785692
	Zygaena (*Mesembrynus*) *centaureae* Fischer von Waldheim, 1832	AJ785693
	Zygaena (*Mesembrynus*) *contaminei contaminei* Boisduval, 1834	AJ785694
	Zygaena (*Mesembrynus*) *corsica* Boisduval, [1828]	AJ785695
	Zygaena (*Mesembrynus*) *cuvieri cuvieri* Boisduval, [1828]	AJ785696
	Zygaena (*Mesembrynus*) *cuvieri cuvieri* Boisduval, [1828]	AJ785697
	Zygaena (*Mesembrynus*) *cynarae samarensis* Holik, 1939	AJ785698
	Zygaena (*Mesembrynus*) *erythrus actae* Burgeff, 1926	AJ785699
	Zygaena (*Mesembrynus*) *favonia elissae* Hofmann, Reiss and Tremewan, 1994	AJ785700
	Zygaena (*Mesembrynus*) *graslini* Lederer, 1855	AJ785701
	Zygaena (*Mesembrynus*) *haematina aurora* Hofmann, 2000	AJ785702
	Zygaena (*Mesembrynus*) *haematina fusca* Hofmann, 2000	AJ785703

Appendix A (*continued*)

Zygaeninae (cont.)	*Zygaena* (*Mesembrynus*) *hindukuschi cishindukuschi* Naumann, 1974	AJ785704
	Zygaena (*Mesembrynus*) *huguenini* Staudinger, 1887	AJ785705
	Zygaena (*Mesembrynus*) *laeta laeta* (Hübner, 1790)	AJ785706
	Zygaena (*Mesembrynus*) *loyselis loyselis* Oberthür, 1876	AJ785707
	Zygaena (*Mesembrynus*) *loyselis ungemachi* Le Cerf, 1923	AJ785708
	Zygaena (*Mesembrynus*) *lydia lydia* Staudinger, 1887	AJ785709
	Zygaena (*Mesembrynus*) *manlia manlia* Lederer, 1870	AJ785710
	Zygaena (*Mesembrynus*) *manlia piti* Hofmann, 2000	AJ785711
	Zygaena (*Mesembrynus*) *manlia piti* Hofmann, 2000	AJ785712
	Zygaena (*Mesembrynus*) *manlia* cf. *pjotri* Hofmann, 1983	AJ785713
	Zygaena (*Mesembrynus*) *manlia turkmenica* Reiss, 1933	AJ785714
	Zygaena (*Mesembrynus*) cf. *manlia* Lederer, 1870	AJ785715
	Zygaena (*Mesembrynus*) *minos ingens* Burgeff, 1926	AJ785716
	Zygaena (*Mesembrynus*) *minos persica* Burgeff, 1926	AJ785717
	Zygaena (*Mesembrynus*) *nocturna meinekei* Hofmann and Tremewan, 2003	AJ785718
	Zygaena (*Mesembrynus*) *nocturna nocturna* Ebert, 1974	AJ785719
	Zygaena (*Mesembrynus*) *purpuralis austronubigena* Verity, 1946	AJ785720
	Zygaena (*Mesembrynus*) *purpuralis pseudorubicundus* Klír and Naumann, 2002	AJ785721
	Zygaena (*Mesembrynus*) *rubicundus* (Hübner, [1817])	AJ785722
	Zygaena (*Mesembrynus*) *rubricollis flavicola* Naumann, 1969	AJ785723
	Zygaena (*Mesembrynus*) *rubricollis ginnereissi* Hofmann, 2000	AJ785724
	Zygaena (*Mesembrynus*) *rubricollis kermanensis* Tremewan, 1975	AJ785725
	Zygaena (*Mesembrynus*) *rubricollis tenhageni* Hofmann and Tremewan, 2003	AJ785726
	Zygaena (*Mesembrynus*) *sarpedon lusitanica* Reiss, 1936	AJ785727
	Zygaena (*Mesembrynus*) *seitzi seitzi* Reiss, 1938	AJ785728
	Zygaena (*Mesembrynus*) *tamara fahima* Naumann and Naumann, 1980	AJ785729
	Zygaena (*Mesembrynus*) *zuleima harchaica* Dujardin, 1973	AJ785730
	Zygaena (*Zygaena*) *angelicae elegans* Burgeff, 1913	AJ785731
	Zygaena (*Zygaena*) *anthyllidis* Boisduval, [1828]	AJ785732
	Zygaena (*Zygaena*) *armena armena* Eversmann, 1851	AJ785733
	Zygaena (*Zygaena*) *dorycnii dorycnii* Ochsenheimer, 1808	AJ785734
	Zygaena (*Zygaena*) *ecki ecki* Christoph, 1882	AJ785735
	Zygaena (*Zygaena*) *ephialtes albaflavens* Verity, 1920	AJ785736
	Zygaena (*Zygaena*) *exulans exulans* (Hohenwarth, 1792)	AJ785737
	Zygaena (*Zygaena*) *filipendulae gemina* Burgeff, 1914	AJ785738
	Zygaena (*Zygaena*) *ignifera* Korb, 1897	AJ785739
	Zygaena (*Zygaena*) *lavandulae consobrina* Germar, [1836]	AJ785740
	Zygaena (*Zygaena*) *lonicerae kindermanni* Oberthür, 1910	AJ785741
	Zygaena (*Zygaena*) *lonicerae leonensis* Tremewan, 1961	AJ785742
	Zygaena (*Zygaena*) *loti macedonica* Burgeff, 1926	AJ785743
	Zygaena (*Zygaena*) *mana chaos* Burgeff, 1926	AJ785744
	Zygaena (*Zygaena*) *nevadensis interrupta* Boursin, 1923	AJ785745
	Zygaena (*Zygaena*) *niphona niphona* Butler, 1877	AJ785746
	Zygaena (*Zygaena*) *osterodensis validior* Burgeff, 1926	AJ785747
	Zygaena (*Zygaena*) *oxytropis oxytropis* Boisduval, [1828]	AJ785748
	Zygaena (*Zygaena*) *rhadamanthus grisea* Oberthür, 1909	AJ785749
	Zygaena (*Zygaena*) *romeo adumbrata* Burgeff, 1926	AJ785750
	Zygaena (*Zygaena*) *transalpina hippocrepidis* Hübner, [1799]	AJ785751
	Zygaena (*Zygaena*) *transalpina tilaventa* Holik, 1935	AJ785752
	Zygaena (*Zygaena*) *trifolii diffusemarginata* Rothschild, 1933	AJ785753
	Zygaena (*Zygaena*) *viciae confusa* Staudinger, 1881	AJ785754

[a]Systematics according to Epstein et al. (1999) and Hofmann and Tremewan (1996).
[b]Species taxonomy after Hofmann and Tremewan (1996, 2001, 2003) and Hofmann (2000a–d).

Appendix B. Species names, references, and accession numbers of mt SSU (12S) rRNA sequences included in the present study

Systematic category	Species	Domain	Reference	Accession no.
Diptera				
Ephydroidea				
Drosophilidae	*Drosophila virilis*	I–V	Clary and Wolstenholme (1987)	X05914
Lepidoptera				
Bombycoidea				
Bombycidae	*Bombyx mandarina*	I–IV	Yukuhiro et al. (2002)	AB070263
	Bombyx mori	I–IV	Lee et al. (unpublished)	AF149768
Saturniidae	*Antheraea pernyi*	I–IV	Liu et al. (unpublished)	AY242996
Hesperioidea				
Hesperiidae	*Daimio tethys*	I	Vila and Björklund (2004)	AY351421
	Erynnis montanus	I	Vila and Björklund (2004)	AY351420
Noctuoidea				
Noctuidae	*Helicoverpa armigera*	I	McKechnie et al. (1993)	U02678
	Helicoverpa punctigera	I	Taylor et al. (1993)	L17343
	Spodoptera frugiperda	I	Mans and Knebel-Mörsdorf (1999)	X97968
Papilionoidea				
Lycaenidae	*Aricia agestis*	I	Vila and Björklund (2004)	AY351427
	Jalmenus evagoras evagoras	I	Taylor et al. (1993)	L16849
Nymphalidae	*Arethusana arethusa*	I	Vila and Björklund (2004)	AY351412
	Coenonympha arcania	I	Vila and Björklund (2004)	AY346246
	Erebia epiphron	I	Vila and Björklund (2004)	AY346241
	Erebia euryale	I	Vila and Björklund (2004)	AY346237
	Erebia gorge	I	Vila and Björklund (2004)	AY346244
	Erebia ligea	I	Vila and Björklund (2004)	AY346240
	Erebia meolans	I	Vila and Björklund (2004)	AY346239
	Erebia oeme	I	Vila and Björklund (2004)	AY346243
	Erebia palarica	I	Vila and Björklund (2004)	AY346236
	Erebia pandrose	I	Vila and Björklund (2004)	AY346242
	Erebia triaria	I	Vila and Björklund (2004)	AY346235
	Inachis io	I	Vila and Björklund (2004)	AY351411
	Melitaea didymoides	I	Vila and Björklund (2004)	AY351419
	Melitaea latonigena	I	Vila and Björklund (2004)	AY346250
Papilionidae	*Parnassius apollo*	I	Vila and Björklund (2004)	AY351418
Pieridae	*Artogeia napi*	I	Vila and Björklund (2004)	AY351422
	Artogeia rapae	I	Vila and Björklund (2004)	AY351423

References

Ban, N., Nissen, P., Hansen, J., Moore, P.B., Steitz, T.A., 2000. The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. Science 289, 905–920.

Billoud, B., Guerrucci, M.-A., Masselot, M., Deutsch, J.S., 2000. Cirripede phylogeny using a novel approach: molecular morphometrics. Mol. Biol. Evol. 17, 1435–1445.

Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Muller, K.M., Pande, N., Shang, Z., Yu, N., Gutell, R.R., 2002. The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BMC Bioinformatics 3, 2.

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

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

Corpet, F., Michot, B., 1994. RNAlign program: alignment of RNA sequences using both primary and secondary structure. Comput. Appl. Biosci. 10, 389–399.

Dahlberg, A.E., 1989. The functional role of ribosomal RNA in protein synthesis. Cell 57, 525–529.

De Rijk, P., Wuyts, J., De Wachter, R., 2003. RnaViz2: an improved representation of RNA secondary structure. Bioinformatics 19, 299–300.

Epstein, M.E., Geertsema, H., Naumann, C.M., Tarmann, G.M., 1999. The Zygaenoidea. In: Kristensen, N.P. (Ed.), Handbook of Zoology, vol. 4 (35.1). De Gruyter, Berlin, New York, pp. 159–180.

Fänger, H., 1999. Comparative morphology of tergal phragmata occurring in the dorsal thoraco-abdominal junction of ditrysian Lepidoptera (Insecta). Zoomorphology 119, 163–183.

Fänger, H., Naumann, C.M., 2001. The morphology of the last instar larva of *Aglaope infausta* (Lepidoptera: Zygaenidae: Chalcosiinae). Eur. J. Entomol. 98, 201–218.

Fänger, H., Yen, S.-H., Naumann, C.M., 1999. External morphology of the last instar larva of *Phauda mimica* Strand, 1915 (Lepidoptera: Zygaenoidea). Entomol. Scand. 29, 429–450.

Galtier, N., 2004. Sampling properties of the bootstrap support in molecular phylogeny: influence of nonindependence among sites. Syst. Biol. 53, 38–46.

Gerken, T., Kurtz, J., Sauer, K.P., Lubjuhn, T., 1998. DNA preparation and efficient microsatellite analysis from insect hemolymph. Electrophoresis 19, 3069–3070.

Gorodkin, J., Heyer, L.J., Brunak, S., Stormo, G.D., 1997. Displaying the information contents of structural RNA alignments: the structure logos. CABIOS 13, 583–586.

Gutell, R.R., Power, A., Hertz, G.Z., Putz, E.J., Stormo, G.D., 1992. Identifying constraints on the higher-order structure of RNA: continued development and application of comparative sequence analysis methods. Nucleic Acids Res. 20, 5785–5795.

Gutell, R.R., Larsen, N., Woese, C.R., 1994. Lessons from an evolving rRNA: 16S and 23S rRNA structures from a comparative perspective. Microbiol. Rev. 58, 10–26.

Gutell, R.R., Lee, J.C., Cannone, J.J., 2002. The accuracy of ribosomal RNA comparative structure models. Curr. Opin. Struct. Biol. 12, 301–310.

Hall, T.A. (Ed.), 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, Nucleic Acids Symp. Ser. 41, 95–98.

Hickson, R.E., Simon, C., Cooper, A., Spicer, G.S., Sullivan, J., Penny, D., 1996. Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. Mol. Biol. Evol. 13, 150–169.

Hickson, R.E., Simon, C., Perrey, S.W., 2000. The performance of several multiple-sequence alignment programs in relation to secondary-structure features for an rRNA sequence. Mol. Biol. Evol. 17 (4), 530–539.

Higgs, P.G., 2000. RNA secondary structure: physical and computational aspects. Quart. Rev. Biophys. 33, 199–253.

Hill, W.E., Dahlberg, A., Garrett, R.A., Moore, P.B., Schlessinger, D., Warner, J.R., 1990. The Ribosome: Structure, Function, & Evolution. American Society for Microbiology, Washington, DC.

Hofmann, A., Tremewan, W.G., 1996. A Systematic Catalogue of the Zygaeninae (Lepidoptera: Zygaenidae). Harley Books, Colchester.

Hofmann, A., Tremewan, W.G., 2001. Contribution to the knowledge of the genus *Zygaena* Fabricius, 1775 in Iran (Lepidoptera, Zygaenidae). Part V: *Zygaena tamara*. Linneana Belg. 18, 125–136.

Hofmann, A., Tremewan, W.G., 2003. Contribution to the knowledge of the genus *Zygaena* Fabricius, 1775 in Iran (Lepidoptera, Zygaenidae). Part VI: *Zygaena nocturna*, *Z. aisha*, *Z. rubricollis* and *Z. fredi* (addenda based on results of fieldwork in 2001 and 2002). Linneana Belg. 19, 9–20.

Hofmann, A., 2000a. Contribution to the knowledge of the genus *Zygaena* Fabricius, 1775 in Iran (Lepidoptera, Zygaenidae). Part I: Introduction, systematic part: *Zygaena* (*Mesembrynus*) *seitzi*, *Z. seitzi nocturna*, *Z. manlia*. Linneana Belg. 17, 171–196.

Hofmann, A., 2000b. Contribution to the knowledge of the genus *Zygaena* Fabricius, 1775 in Iran (Lepidoptera, Zygaenidae). Part II: *Zygaena rubricollis*. Linneana Belg. 17, 227–232.

Hofmann, A., 2000c. Contribution to the knowledge of the genus *Zygaena* Fabricius, 1775 in Iran (Lepidoptera, Zygaenidae). Part III: *Zygaena haematina*. Linneana Belg. 17, 293–299.

Hofmann, A., 2000d. Contribution to the knowledge of the genus *Zygaena* Fabricius, 1775 in Iran (Lepidoptera, Zygaenidae). Part IV: *Zygaena haematina* (addendum), *Z. cacuminum*. Linneana Belg. 17, 339–347.

Hudelot, C., Gowri-Shankar, V., Jow, H., Rattray, M., Higgs, P.G., 2003. RNA-based phylogenetic methods: application to mammalian mitochondrial RNA sequences. Mol. Phylogenet. Evol. 28, 241–252.

Jow, H., Hudelot, C., Rattray, M., Higgs, P.G., 2002. Bayesian phylogenetics using an RNA substitution model applied to early mammalian evolution. Mol. Biol. Evol. 19, 1591–1601.

Kjer, K.M., 1995. Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: an example of alignment and data presentation from the frogs. Mol. Phylogenet. Evol. 4, 314–330.

Kjer, K.M., 2004. Aligned 18S and insect phylogeny. Syst. Biol. 53, 506–514.

Lenhof, H.-P., Reinert, K., Vingron, M., 1998. A polyhedral approach to RNA sequence structure alignment. J. Comput. Biol. 5, 517–530.

Lydeard, C., Holznagel, W.E., Schnare, M.N., Gutell, R.R., 2000. Phylogenetic analysis of molluscan mitochondrial LSU rDNA sequences and secondary structures. Mol. Phylogenet. Evol. 15, 83–102.

Mans, R.M.W., Knebel-Mörsdorf, D., 1999. Mitochondrial DNA acts as potential promoter of the baculovirus RNA polymerase. Biol. Chem. 380, 579–583.

Mathews, D.H., Sabina, J., Zuker, M., Turner, D.H., 1999. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. J. Mol. Biol. 288, 911–940.

McKechnie, S.W., Spackman, M.E., Naughton, N.E., Kovacs, I.V., Ghosn, M., Hoffmann, A.A., 1993. Assessing budworm population structure in Australia using the AT-rich region of mitochondrial DNA. In: Beltwide Cotton Insect Research Conference, vol. 2, New Orleans, pp. 838–839.

Minet, J., 1991. Tentative reconstruction of the ditrysian phylogeny (Lepidoptera: Glossata). Entomol. Scand. 22, 69–95.

Misof, B., Fleck, G., 2003. Comparative analysis of mt LSU rRNA secondary structures of Odonates: structural variability and phylogenetic signal. Insect Mol. Biol. 12, 535–547.

Notredame, C., O'Brien, E.A., Higgs, D.G., 1997. RAGA:RNA sequence alignment by genetic algorithm. Nucleic Acids Res. 25, 4570–4580.

Ouvrard, D., Campbell, B.C., Bourgoin, T., Chan, K.L., 2000. 18S rRNA secondary structure and phylogenetic position of Peloridiidae (Insecta, Hemiptera). Mol. Phylogenet. Evol. 16, 403–417.

Page, R.D.M., 2000. Comparative analysis of secondary structure of insect mitochondrial small subunit ribosomal RNA using maximum weighted matching. Nucleic Acids Res. 28, 3839–3845.

Page, R.D.M., Cruickshank, R., Johnson, K.P., 2002. Louse (Insecta: Phthiraptera) mitochondrial 12S rRNA secondary structure is highly variable. Insect Mol. Biol. 11, 361–369.

Savill, N.J., Hoyle, D.C., Higgs, P.G., 2001. RNA sequence evolution with secondary structure constraints: comparison of substitution rate models using maximum-likelihood methods. Genetics 157, 399–411.

Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F., Yonath, A., 2000. Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. Cell 102, 615–623.

Schneider, T.D., Stephens, R.M., 1990. Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 18, 6097–6100.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H., Flook, P., 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. 87, 651–701.

Taylor, M.F., McKechnie, S.W., Pierce, N., Kreitman, M., 1993. The lepidopteran mitochondrial control region: structure and evolution. Mol. Biol. Evol. 10, 1259–1272.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Titus, T.A., Frost, D.R., 1996. Molecular homology assessment and phylogeny in the lizard family Opluridae (Squamata: Iguania). Mol. Phylogenet. Evol. 6, 49–62.

Van de Peer, Y., De Rijk, P., Wuyts, J., Winkelmans, T., De Wachter, R., 2000. The European small subunit ribosomal RNA database. Nucleic Acids Res. 28, 175–176.

Van de Peer, Y., Robbrecht, E., de Hoog, S., Caers, A., De Rijk, P., De Wachter, R., 1999. Database on the structure of small subunit ribosomal RNA. Nucleic Acids Res. 27, 179–183.

Vila, M., Björklund, M., 2004. The utility of the neglected mitochondrial control region for evolutionary studies in Lepidoptera (Insecta). J. Mol. Evol. 58, 280–290.

Wimberly, B.T., Brodersen, D.E., Clemons Jr., W.M., Morgan-Warren, R.J., Carter, A.P., Vonrhein, C., Hartsch, T., Ramakrishnan, V., 2000. Structure of the 30S ribosomal subunit. Nature 407, 327–339.

Woese, C.R., Pace, N.R., 1993. Probing RNA structure, function and history by comparative analysis. In: Gesteland, R.F., Atkins, J.F. (Eds.), The RNA World. Cold Spring Harbor Laboratory Press, New York, pp. 91–117.

Wuyts, J., Parrière, G., Van de Peer, Y., 2004. The European ribosomal RNA database. Nucleic Acids Res. 32, D101–D103.

Wuyts, J., Van de Peer, Y., Winkelmans, T., De Wachter, R., 2002. The European database on small subunit ribosomal RNA. Nucleic Acids Res. 30, 183–185.

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.

Yusupov, M.M., Yusupova, G.Z., Baucom, A., Lieberman, K., Earnest, T.N., Cate, J.H.D., Noller, H.F., 2001. Crystal structure of the ribosome at 5.5 Å resolution. Science 292, 883–896.

Zuker, M., 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. 31, 3406–3415.