JOURNAL OF EXPERIMENTAL ZOOLOGY (MOL DEV EVOL) 306B:70–88 (2006)

A Hexapod Nuclear SSU rRNA Secondary-Structure Model and Catalog of Taxon-Specific Structural Variation

BERNHARD MISOF[1]*, OLIVER NIEHUIS[1], INGE BISCHOFF[1], ANDREAS RICKERT[1], DIRK ERPENBECK[2], AND ARNOLD STANICZEK[3]
[1]*Zoologisches Forschungsinstitut und Museum A. Koenig Adenauerallee 160, 53113 Bonn, Germany*
[2]*Department of Coelenterata and Porifera (Zoologisch Museum), Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, 1090 GT Amsterdam, The Netherlands*
[3]*Staatliches Museum für Naturkunde Stuttgart, Abt. Entomologie, Rosenstein 1, 70191 Stuttgart, Germany*

ABSTRACT RNA molecules and in particular the nuclear SSU RNA play an important role in molecular systematics. With the advent of increasingly parameterized substitution models in systematic research, the incorporation of secondary-structure information became a realistic option compensating interdependence of character variation. As a prerequisite, consensus structures of eukaryotic SSU RNA molecules have become available through extensive comparative analyses and crystallographic studies. Despite extensive research in hexapod phylogenetics, consensus SSU RNA secondary structures focusing on hexapods have not yet been explored. In this study, we compiled a representative hexapod SSU data set of 261 sequences and inferred a specific consensus SSU secondary-structure model. Our search for conserved structural motives relied on a combined approach of thermodynamic and covariation analyses. The hexapod consensus-structure model deviates from the canonical eukaryotic model in a number of helices. Additionally, in several helices the hexapod sequences did not support a single consensus structure. We provide consensus structures of these sections of single less-inclusive taxa, thus facilitating the adaptation of the consensus hexapod model to less-inclusive phylogenetic questions. The secondary-structure catalog will foster the application of RNA structure models in phylogenetic analyses using the SSU rRNA molecule, and it will improve the realism of substitution models and the reliability of reconstructions based on rRNA sequences. *J. Exp. Zool. (Mol. Dev. Evol.) 306B:70–88, 2006.*

Ribosomal RNAs are still an indispensable marker system in molecular systematics. Fragments of these genes or complete rDNA are readily amplified by means of PCR and are regularly used as primary sources of molecular character variation. Ribosomal RNAs are often hard to align due to frequent occurrences of indels. Secondary structural information can aid in the alignment procedure (Kjer, '95; Buckley et al., 2000; Hickson et al., 2000; Misof et al., 2002) and this, in turn, can improve the reliability of subsequent phylogenetic analyses (for example, Kjer, '95, 2004; Titus and Frost, '96). RNA sites can be grouped into paired and unpaired character classes, where base pairing causes interdependent character variation in contrast to nucleotide variation of unpaired sites. Thus, paired sites do not display independent phylogenetic signal, and in consequence ignored base pairing leads to inflated measurements of tree robustness, particularly in likelihood approaches (see, for example, Rzhetsky, '95; Tillier and Collins, '95, '98; Stephan, '96; Parsch et al., 2000). Recent theoretical and empirical results confirmed these considerations (Schoeniger and von Haeseler, '94; Savill et al., 2001; Jow et al., 2002; Hudelot et al., 2003; Galtier, 2004). It is therefore paramount to apply consensus structures in RNA-based phylogenetics via RNA substitution models.

*Correspondence to: Bernhard Misof, Abteilung für Entomologie, Zoologisches Forschungsinstitut und Museum A. Koenig, Adenauerallee 160, 53113 Bonn, Germany. E-mail: b.misof.zfmk@uni-bonn.de
Received 2 December 2004; Accepted 26 January 2005
Published online 13 September 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jez.b.21040.

TABLE 1. Included taxa of structural analyses

Taxon	Genus	Species	Subspecies/location	GenBank #
Diptera	*Tipula*	*sp.*		X89496
	Tipula	*paterifera*		AF136855
	Sergentomyia	*fallax*	*cypriotica*	AJ244427
	Sergentomyia	*minuta*		AJ244420
	Aedes	*albopictus*		X57172
Siphonaptera	*Archaeopsylla*	*erinacea*		X89486
	Ctenocephalides	*felis*		AF136859
	Hystrichopsylla	*schefferi*		L10185
Mecoptera	*Boreus*	*sp.*		X89487
	Apteropanorpa	*evansi*		AF286284
	Merope	*tuber*		AF286287
	Panorpa	*communis*		DQ008167
	Panorpa	*acuta*		DQ008168
	Panorpa	*maculosa*		DQ008169
	Panorpa	*alpine*		DQ008170
	Panorpa	*claripennis*		DQ008171
	Panorpa	*cognate*		DQ008172
	Panorpa	*fluvicaudaria*		DQ008173
	Panorpa	*helena*		DQ008174
	Panorpa	*nebulosa*		DQ008175
	Panorpa	*similes*		DQ008176
	Panorpa	*vulgaris*		DQ008177
	Panorpa	*multifasciata*		DQ008178
	Neopanorpa	*sp.*		DQ008179
	Brachypanorpa	*oregonensis*		DQ008180
	Apterobittacus	*apterus*		AF136858
	Bittacus	*chlorostigmus*		L10184
	Boreus	*brumalis*		AF136856
	Bittacus	*strigosus*		U65144
Lepidoptera	*Galleria*	*mellonella*		X89491
	Lymantria	*dispar*		AF136872
	Thyridopteryx	*ephemeraeformis*		AF136874
	Prodoxus	*quinquepunctellus*		AF136868
	Tegeticula	*yuccasella*		AF136869
	Sthenopis	*quadriguttatus*		AF136871
	Agathiphaga	*queenslandensis*		AF136864
	Micropterix	*calthella*		AF136863
Trichoptera	*Hydropsyche*	*sp.*		X89483
	Brachycentrus	*nigrosoma*		AF136860
	Wormaldia	*moesta*		AF136861
Hymenoptera	*Ephedrus*	*niger*		AJ009328
	Aphidius	*funebris*		AJ009322
	Xenostigmus	*bifasciatus*		AJ009353
	Protaphidius	*wissmannii*		AJ009348
	Pauesia	*pini*		AJ009344
	Lysiphlebus	*confuses*		AJ009331
	Polistes	*dominulus*		X77785
	Leptothorax	*acervorum*		X89492
	Bareogonalos	*canadensis*		L10176
	Mesopolobus	*sp.*		L10177
	Hartigia	*cressonii*		L10173
	Orussus	*thoracicus*		L10174
	Periclista	*linea*		L10172
	Epyris	*sepulchralis*		L10180
	Evania	*appendigaster*		L10175
Coleoptera	*Loricera*	*foveata*		AF012503
	Pamborus	*guerinii*		AF012508
	Calosoma	*scrutator*		AF002800

TABLE 1. Continued

Taxon	Genus	Species	Subspecies/location	GenBank #
	Ceroglossus	*chilensis*		AF012509
	Cychrus	*italicus*		AF012510
	Laccocenus	*ambiguous*		AF012486
	Blethisa	*multipunctata*	*aurata*	AF002803
	Oregus	*aereus*		AF012500
	Australphilus	*montanus*		AF199527
	Copelatus	*chevrolati*	*renovatus*	AF012524
	Hydaticus	*transversalis*		AF199545
	Hyderodes	*schuckardi*		AF199548
	Bidessus	*calabricus*		AF199581
	Bidessodes	*mjobergi*		AF199579
	Bidessus	*goudoti*		AF199582
	Hydroglyphus	*geminus*		AF199583
	Copelatus	*haemorrhoidalis*		AF199525
	Hydaticus	*leander*		AF199544
	Rhantaticus	*congestus*		AF199547
	Megadytes	*sp.*		AF199551
	Cybister	*lateralimarginalis*		AF199550
	Notaticus	*sp.*		AF199546
	Elaphrus	*clairvillei*		AF002802
	Notiophilus	*semiopacus*		AF002804
	Systolosoma	*lateritium*		AF012522
	Diplochaetus	*planatus*		AF002789
	Pericompsus	*laetulus*		AF002790
	Batesiana	*hilaris*		AF012489
	Trachypachus	*gibbsii*		AF002808
	Agonum	*albipes*		AF201403
	Catapiesis	*brasiliensis*		AF012476
	Meloe	*proscarabaeus*		X77786
	Tenebrio	*molitor*		X07801
	Aulonogyrus	*striatus*		AF199512
Neuropteroidea	*Anisochrysa*	*carnea*		X89482
	Chrysoperla	*plorabunda*		L10183
	Oliarces	*clara*		AF012527
	Corydalus	*cognatus*		U65132
	Hemerobius	*stigmata*		U65136
	Lolomyia	*texana*		U65134
	Mantisp.a	*pulchella*		U65135
	Sialis	*sp.*		X89497
	Phaeostigma	*notata*		X89494
Hemiptera	*Philaenus*	*sp. Umarius*		U06480
	Okanagana	*utahensis*		U06478
	sp.issistilus	*festinus*		U06477
	Prokelisia	*marginata*		U09207
	Rhaphigaster	*nebulosa*		X89495
Orthoptera	*Oedipoda*	*coerulescens*		Z97573
	Acheta	*domesticus*		X95741
	Trigonopteryx	*hopei*		Z97589
	Batrachideidae	*sp.*		Z97631
	Melanoplus	*sp.*		U65115
	Oxya	*chinensis*		AY037173
Phasmatodea	*Agathemera*	*crassa*		Z97561
	Carausius	*morosus*		X89488
	Anisomorpha	*buprestoides*		U65116
Dictyoptera	*Archimantis*	*latistylus*		AF220578
	Paraoxypilus	*tasmaniensis*		AF220577
	Creobroter	*pictipennis*		AF220576
	Kongobatha	*diademata*		AF246712

TABLE 1. Continued

Taxon	Genus	Species	Subspecies/location	GenBank #
	Tenodera	*angustipennis*		AF220579
	Coptotermes	*lacteus*		AF220564
	Serritermes	*serrifer*		AF220565
	Neotermes	*koshunensis*		AF220566
	Hodotermopsis	*japonica*		AF220567
	Microhodotermes	*viator*		AF220569
	Mastotermes	*darwiniensis*		AY121141
	Periplaneta	*americana*		AF220572
	Polyphaga	*aegyptiaca*		AF220575
	Cryptocercus	*relictus*		AF220570
	Cryptocercus	*punctulatus*		AF220571
	Blattella	*germanica*		AF220573
	Panesthia	*cribrata*		AB036194
	Blaberus	*sp.*		U65112
	Labidura	*riparia*		U65114
	Mantis	*religiosa*		U65113
Plecoptera	*Nemoura*	*meyeri*		Z97595
	Brachyptera	*seticornis*		AF311456
	Scopura	*montana*		AF311459
	Austrocerca	*tasmanica*		AF311462
	Diamphipnopsis	*samali*		AF311440
	Tasmanoperla	*thalia*		AF311441
	Acruroperla	*atra*		AF311439
	Cosmoperla	*sp.*		AF311444
	Stenoperla	*maclellani*		AF311445
	Zelandoperla	*agnetis*		AF311447
	Newmanoperla	*sp.*		AF311448
	Leptoperla	*sp.*		AF311446
	Paracapnia	*angulata*		AF311442
	Allocapnia	*vivipara*		AF311443
	Pteronarcella	*californica*		AF311465
	Pteronarcys	*badia*		AF311464
	Leuctra	*nigra*		AF311457
	Leuctra	*hippopus*		AF311458
	Notonemoura	*maculate*		AF311461
	Siphonoperla	*torrentium*		AF311463
	Claassenia	*sabulosa*		AF311469
	Beloneura	*georgiana*		AF311468
	Hesp.eraperla	*pacifica*		AF311470
	Megarcys	*signata*		AF311471
	Isoperla	*sp.*		AF311472
	Nemurella	*pictetii*		AF311451
	Megaleuctra	*stigmata*		AF311460
	Protonemoura	*praecox*		AF311449
	Ostrocerca	*albidipennis*		AF311450
	Amphinemura	*sulcicollis*		AF311452
	Malenka	*californica*		AF311453
	Sierraperla	*cora*		AF311466
	Taeniopteryx	*burski*		AF311454
	Taeniopteryx	*nebulosa*		AF311455
Ephemeroptera	*Ephemera*	*sp.*		X89489
	Caenis	*luctuosa*		AF461250
	Anthopotamus	*sp.*		AF461255
	Baetis	*buceratus*		AF461248
	Centroptilum	*luteolum*		AF461251
	Cloeon	*dipterum*		AF461249
	Stenonema	*sp.*	*RH2002*	AF461252
	Leucrocuta	*aphrodite*		AF461254

TABLE 1. Continued

Taxon	Genus	Species	Subspecies/location	GenBank #
	Hexagenia	*sp.*		AY121136
	Callibaetis	*ferrugineus*	*ferrugineus*	AF370791
	Hexagenia	*rigida*		AF461253
	Behningia	*sp.*		AY338703
	Ametropus	*neavei*		AY338700
	Lachlania	*saskatchewanensis*		AY338701
	Pseudiron	*centralis*		AY338699
	Siphlonurus	*croaticus*		DQ008181
	Ritrogena	*sp.*		DQ008182
	Ephemerella	*major*		DQ008183
	Polyplocia	*sp.*		AY338705
Odonata	*Aeschna*	*cyanea*		X89481
	Epiophlebia	*superstes*		AF461247
	Brachytron	*pretense*		AF461232
	Leucorrhinia	*pectoralis*		AF461240
	Sympetrum	*danae*		AF461243
	Celithemis	*eponina*		AF461233
	Lestes	*macrostigma*		AJ421950
	Lestes	*numidicus*		AJ421952
	Lestes	*virens*		AJ421951
	Sympecma	*fusca*		AJ421948
	Chalcolestes	*viridis*		AJ421949
	Lestes	*sponsa*		AF461244
	Erythromma	*najas*		AF461238
	Ischnura	*elegans*		AF461239
	Coenagrion	*sp.*		AF461235
	Pyrrhosoma	*nymphula*		AF461241
	Enallagma	*cyathigerum*		AJ420944
	Cordulia	*aenea*		AF461236
	Somatochlora	*flavomaculata*		AF461242
	Sympetrum	*sanguineum*		AF461245
	Sympetrum	*vulgatum*		AF461246
	Gomphus	*externus*		DQ008184
	Stylurus	*intricatus*		DQ008185
	Stylurus	*amnicola*		DQ008186
	Gomphus	*exilis*		DQ008186
	Arigomphus	*cornutus*		DQ008188
	Dromogomphus	*spinosus*		DQ008189
	Onychogomphus	*forcipatus*	*forcipatus/Nestos*	DQ008190
	Onychogomphus	*forcipatus*		DQ008191
	Ophiogomphus	*severus*		DQ008192
	Hagenius	*brevistylus*		DQ008193
	Oxygastra	*curtisi*		DQ008194
	Macromia	*splendens*		DQ008195
	Lindenia	*tetraphylla*		DQ008196
	Caliaeschna	*microstigma*		DQ008197
	Cordulegaster	*picta*		DQ008198
	Anaciaeschna	*isoceles*		DQ008199
	Crocothemis	*erythraea*		DQ008200
	Sympetrum	*vulgatum*		DQ008201
	Sympetrum	*flaveolum*		DQ008202
	Orthetrum	*albistylum*		DQ008203
	Libellula	*depressa*		DQ008204
	Libelula	*fulva*		DQ008205
	Tramea	*lacerata*		DQ008206
	Platycnemis	*pennipes*	*Nestos*	DQ008207
	Calopteryx	*splendens*	*Nestos*	DQ008208
	Aeshna	*juncea*		AF461231

TABLE 1. Continued

Taxon	Genus	Species	Subspecies/location	GenBank #
Zygentoma	*Lepisma*	*sp.*		AF005458
	Lepisma	*saccharina*		X89484
	Ctenolepisma	*longicaudata*		AY210811
	Thermobia	*sp.*	*ZG01*	AY338726
	Tricholepidion	*gertschi*		AF370789
	Tricholepidion	*sp.*	*ZG03*	AY338728
	Battigrassiella	*sp.*	*ZG02*	AY338727
	Ctenolepisma	*sp.*		DQ008209
	Thermobia	*domestica*		DQ008210
Archaeognatha	*Machilidae*	*sp.*		DQ008211
	Trigoniopthalmus	*alternatus*		U65106
	Allomachilis	*froggarti*		AF370788
	Petrobius	*brevistylis*		X89808
	Dilta	*littoralis*		AF005457
Diplura	*Campodea*	*tillyardi*		AF173234
	Lepidocampa	*weberi*		AY037167
	Parajapyx	*emeryanus*		AY037168
	Campodeidae	*sp.2*		AF005455
	Catajapyx	*sp.*		AF005456
Collembola	*Crossodonthina*	*koreana*		Z36893
	Podura	*aquatica*		AF005452
	Neanura	*latior*		AY037172
	Hypogastrura	*dolsana*		Z26765
	Onychiurus	*yodai*		AY037171
	Lepidocyrtus	*paradoxus*		U61301
Protura	*Neocondeellum*	*dolichotarsum*		AY037170
	Baculentulus	*tienmushanensis*		AY037169
	Acerentulus	*traegardhi*		AF173233
	Acerentulus	*traegardhi*		AF005453

Recently, several investigations report consensus structures of ribosomal RNA fragments within specific taxa (Billoud et al., 2000; Buckley et al., 2000; Lydeard et al., 2000; Ouvrard et al., 2000; Page, 2000; Wuyts et al., 2001; Misof et al., 2002; Page et al., 2002; Misof and Fleck, 2003). Gutell (’94), van de Peer et al. (’99, 2000), Cannone et al. (2002), and Wuyts et al. (2004) published consensus structures for the eukaryotic ribosomal SSU RNA based on comparative analyses of a multitude of sequences. However, less taxonomically inclusive models are presently not available. This can become a problem, since, as Page (2000) already showed, the application of general consensus-structure models in insect mt SSU molecules in some cases will force helical structures that are not supported by mutual information indices. Our preliminary comparative analyses of insect SSU sequences also suggested some deviations from the canonical eukaryotic SSU model that merit further investigations. In this respect, taxon-specific rRNA structures would advance phylogenetics based on rRNA sequence data.

A major drawback of former attempts to align sequences guided by secondary-structure information was the lack of automation having reasonable time requirements for the realization of representative data (also compare Kjer, ’95; Buckley et al., 2000; Misof et al., 2002). The utilization of structure masks in alignment packages provided a pragmatic solution to the problem (Thompson et al., ’97). New advances combine secondary-structure reconstruction and sequence alignment in one integrated process (Notredame et al., ’97) but still with prohibitively large calculation efforts for representative data.

The automation of alignment procedures using secondary-structure cues failed not only because of inadequate formalizations of the alignment process but also because of difficulties in generating secondary-structure models for rRNA sequences (Higgs, 2000). The formerly heavy reliance

on thermodynamic-folding models to reconstruct secondary structures of RNA sequences (reviewed in Higgs, 2000) is now routinely accompanied by comparative analyses (compare, for example, Gutell et al., '92, 2002; Gutell, '93; Higgs, 2000; Page, 2000; confirmed by results of Ban et al., 2000). The sole reliance on thermodynamic folding harbors the major drawback of yielding multiple solutions of obscure biological significance. In contrast, the analysis of covariation patterns has been successful in identifying biologically relevant structures in RNA molecules (compare Ban et al., 2000; Gutell et al., 2002) and has been implemented in several software packages (for example, Hofacker et al., '94, '98, 2002; Tabaska et al., '98). For the purpose of phylogenetics, the combination of thermodynamic folding with the analysis of covariation patterns in alignments is probably a superior way of choosing among multiple solutions and suboptimal foldings. This approach is implemented, for example, in the RNAalifold (Hofacker et al., 2002) and CIRCLE (Page et al., 2002) software. Most recently, the *European Ribosomal DataBase* (Wuyts et al., 2004) provides tools for automated alignment of rRNA sequences based on secondary-structure models. Taken together, the technical standard in secondary-structure reconstruction is advanced enough to facilitate routine inferences of structure models from large data sets.

A representative number of hexapod sequences can be retrieved from Gen Bank to infer reliable structural features of hexapod nuclear SSU RNAs (compare also Kjer, 2004). In the present study, we use the exhaustive sampling of major taxa within hexapods and characterize nearly complete hexapod SSU consensus structures. We also highlight fragments within the RNA molecule where a consensus model does not appear to be applicable. The analyses of taxonomically less inclusive groups resulted in taxon-specific consensus models in cases of sufficient sequence samples, which can serve as starting points in more restricted analyses. We expect that these SSU consensus structures should promote phylogenetic research within hexapods.

MATERIALS AND METHODS

Sequences and sequence alignment

Within Odonata, Ephemeroptera, ''apterygotes'', and Mecoptera we characterized sequences of additional taxa. These new sequences are indicated in Table 1.

Using DNeasy Tissue Kit (Qiagen), genomic DNA was extracted from thorax muscle tissue. PCRs were performed on a GeneAmp 2700 thermal cycler (Applied Biosystems) for 5 min at 94°C, followed by 35 cycles at 94°C for 45 sec, 50°C for 45 sec and 72°C for 1 min 45 sec, and finally at 72°C for 30 min. After cycle-sequencing reactions using BigDye ReadyMix (Applied Biosystems), amplification products were separated using an ABI PRISM® 377 sequencer (Applied Biosystems). Fragments were read from both sides and were assembled in BioEdit (Hall, '99). Primers used to amplify the almost entire SSU gene were 18SV and 18SR (Chalwatzis et al., '95) and internal primers 18Sai, 18Sbi (Maddison et al., '99). PCR primers were used for cycle sequencing as well.

Genbank accession numbers of new sequences will be provided after acceptance of the manuscript.

In total, we compiled a data set of 261 hexapod SSU rDNA sequences representing all major taxa within hexapods (see Table 1). Some insect orders have not been considered despite available SSU sequences, for example, the Embioptera or Strepsiptera. We omitted these sequences because sampling was insufficient within the order (Embioptera) or SSU sequences of the group exhibited highly aberrant characteristics (Strepsiptera). Genbank entries were rejected if they span less than 2/3 of the entire gene, except for the plecopteran sequences for which only one complete SSU entry was available: the remaining plecopteran sequences are drawn from Thomas et al. (2000) and span roughly 1,300 bp.

The alignment of sequences relies on the secondary-structure model available at the European Ribosomal DataBase (Wuyts et al., 2004). For each hexapod order, we selected a prealigned SSU sequence from the database and used this sequence as a profile to align all remaining sequences of this order in CLUSTAL X (Thompson et al., '97). Profile alignment to prealigned sequences maintains the alignment of the prealigned sequences drawn from the database in which structural information has been the guiding principle (van de Peer et al., 2000; Wuyts et al., 2001, 2004). This alignment procedure assumes that structural variation within hexapod orders is neglectable in comparison to variation between orders. This assumption will not be probed further in this investigation. There are just not enough sequences available within several

groups to assess structural variation within these hexapod orders. The CLUSTAL X alignment with annotated structural information is available upon request from the author.

Structural refinements

Helices and loops are annotated according to Wuyts et al. (2004). We studied patterns of

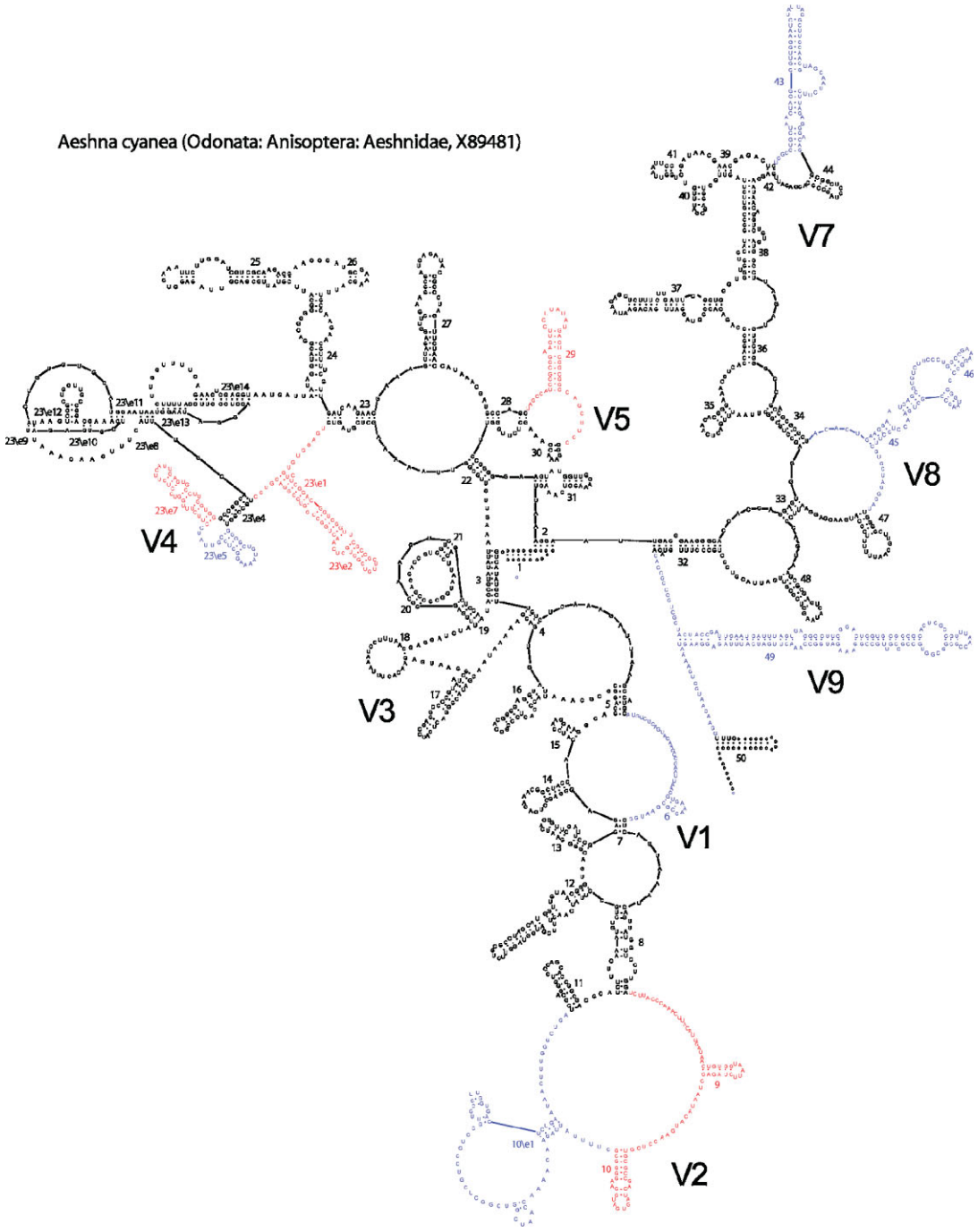


Fig. 1. Inferred hexapod consensus SSU rRNA structure with mayfly- and dragonfly-specific structural variation. The hexapod consensus structure is illustrated with co-notated helices using the odonate *Aeshna cyanea* SSU rRNA sequence. Helices and loops in black are identical to the general SSU-vdP model. Helices and loops in red are hexapod-specific consensus structures different from the SSU-vdP model, and helices and loops in blue are taxon-specific and, in the case of *Aeshna cyanea*, specific structures for Palaeoptera. Blue helices are not included in the general hexapod consensus model.

covariation in paired sites by calculating frequencies of nucleotide pairs and mutual-information indices $M(x.y)$ (Gutell et al., '92). The program BioEdit (Hall, '99) was used for these purposes. We started the analysis by using a data set of complete sequences. Only helices 1, 2, 3, and 50 were cut off since most Genbank entries were incomplete for these helices. Additionally, we used RNAalifold (Hofacker et al., 2002) to compare the general eukaryotic model of individual helices with predictions from a combined thermodynamic and covariation approach (compare, also Misof and Fleck, 2003).

The generated alignment contained blocks of high positional variation for which taxon-specific substructures appeared conceivable. We used the RNAalifold software to check for the occurrence of taxon-specific consensus structures. Taxa were selected as the most inclusive groups for which a consensus structure could be inferred. This led to different groups when compared between helices. For example, if all hexapod sequences did not support a consensus structure, the data were split into pterygote and ''apterygote'' sequences and reanalyzed. For each analyzed helix, the composition of groups is given in Table A1 of Appendix A. The taxa groups largely vary in their numbers of sequences, and consensus structures for taxa groups with only a couple of sequences are certainly less reliable compared with well-sampled groups. Notice that the taxon grouping was not congruent between helices due to heterogeneous distribution of sequence variation. The presence or absence of structures for taxa is given in Appendix A.

For each helix deviating from the canonical eukaryotic model (subsequently called the SSU-vdP model), we illustrate the RNAalifold reconstructions by one example, usually derived from the Odonata or Palaeoptera sequences. Exemplary structures are accompanied by base-pairing probability matrices, which provide an idea of the reliability of inferred structures. RNA secondary structures were drawn with RNAViz (de Rijk and de Wachter, '97) or directly provided in bracket notations by the RNAalifold software. In addition to the complete alignment, SSU sequences in DCSE format with co-notated structure information are available for every group. These sequence masters can be used to draw RNA secondary structures with the RNAViz software (de Rijkand de Wachter, '97).

Presentation

It is obvious that not every single structure can be pictured in this publication. We just give galleries and selected base-pairing probability matrices to illustrate specific results. Complete data on every inferred helix with additional information on base-pairing probability are collected and available at the journal's web site as supplementary material and upon request from the corresponding author.

RESULTS

Fig. 1 shows the inferred secondary structure of the SSU rRNA of *Aeshna cyanea* (Odonata: Anisoptera, Genbank accession: X89481). The illustrated secondary structure shows helices compatible with all hexapod sequences and some helices for which a general hexapod model appears inapplicable. The hexapod SSU model is largely compatible with the general eukaryotic SSU model. However, in some helices, our analysis implies structural differences (Fig. 2).

In the following, we refrain from discussing each helix separately, but instead present information on helices from which extensive sequence variation indicated possible taxon-specific structural variation.

Most of the helices receive some support from covariation analyses or can be folded due to the presence of Watson–Crick base pairs. It is obvious that for several of the highly conserved helices, hexapods do not display enough variation to yield informative covariation patterns. In these cases, we relied on the helical folding suggested by the general eukaryotic SSU model.

Section V1, helix 6

The eukaryotic SSU secondary-structure model assumes a short helical structure for helix 6 (Fig. 1). Within hexapods the pattern of covariation shows occasional non-Watson–Crick base pairs, which suggests deviations from a common consensus folding among taxa. Splitting the data set yielded several different consensus structures in RNAalifold analyses. Proturan sequences did not support any helical structure (see Appendix A). In Fig. A1 of Appendix A, taxon-specific structures inferred from RNAalifold analyses are illustrated in bracket notations. It is obvious that holometabolan and hemimetabolan insect sequences support different helical structures. The lack of clear support from covariation analyses and thermodynamic criteria leads us to assume that a helix 6 is not present in all hexapod SSU rRNAs. Phylogenetic analyses should rely on locally inferred base pairings instead.

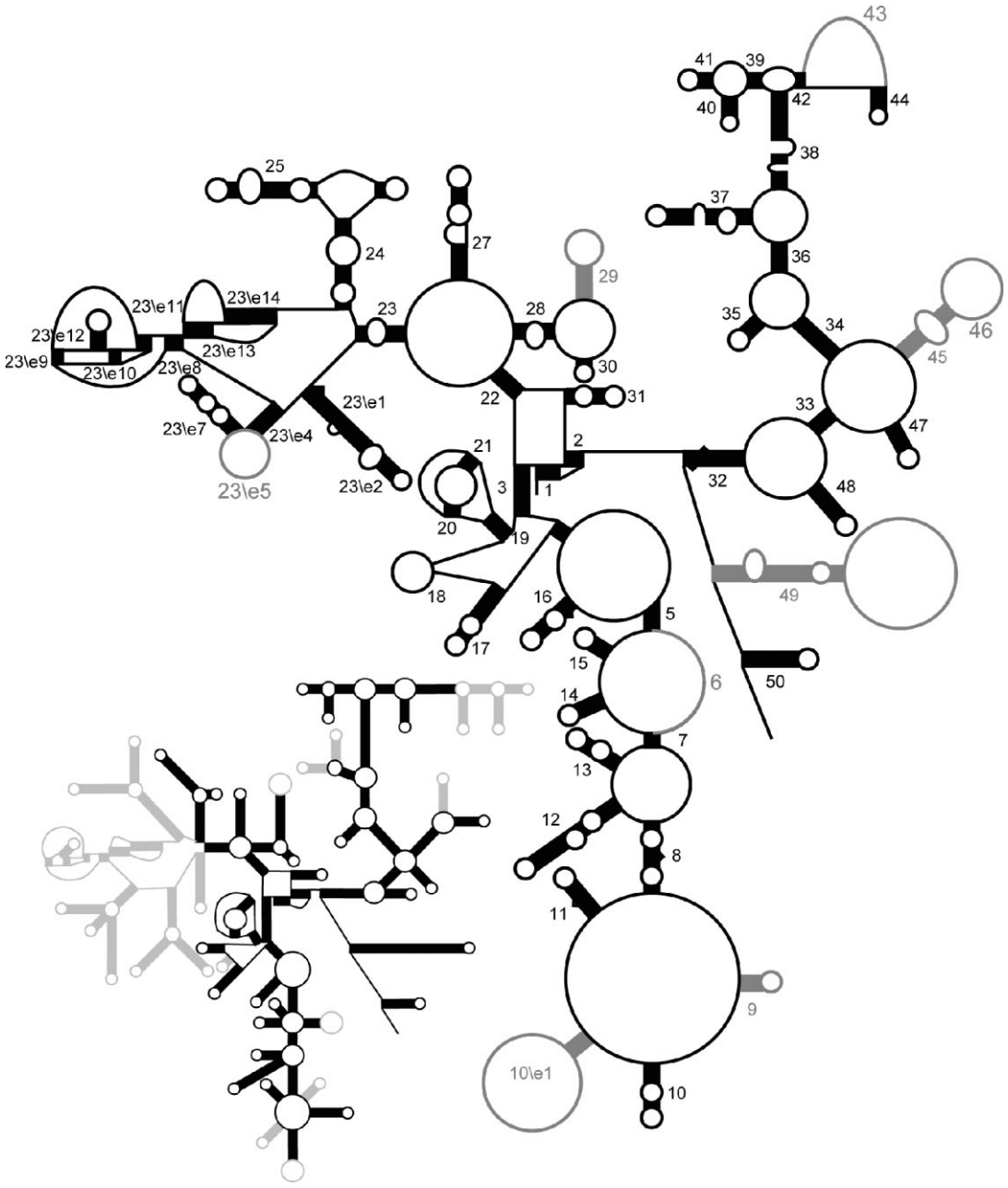


Fig. 2. Comparison of general eucaryotic and hexapod model. The small structure is drawn after the eukaryotic consensus model of Wuyts et al. (2001). In this drawing, structural sections that are known to be variable in eukaryotes are drawn in gray according to Wuyts et al. (2001). The drawing is a simplified version of the drawing of Wuyts et al. (2001). The large drawing represents the hexapod consensus structure as it was developed in this study. Red sections of the structure highlight helical regions in which the hexapod consensus model is largely different from the general eukaryotic structure.

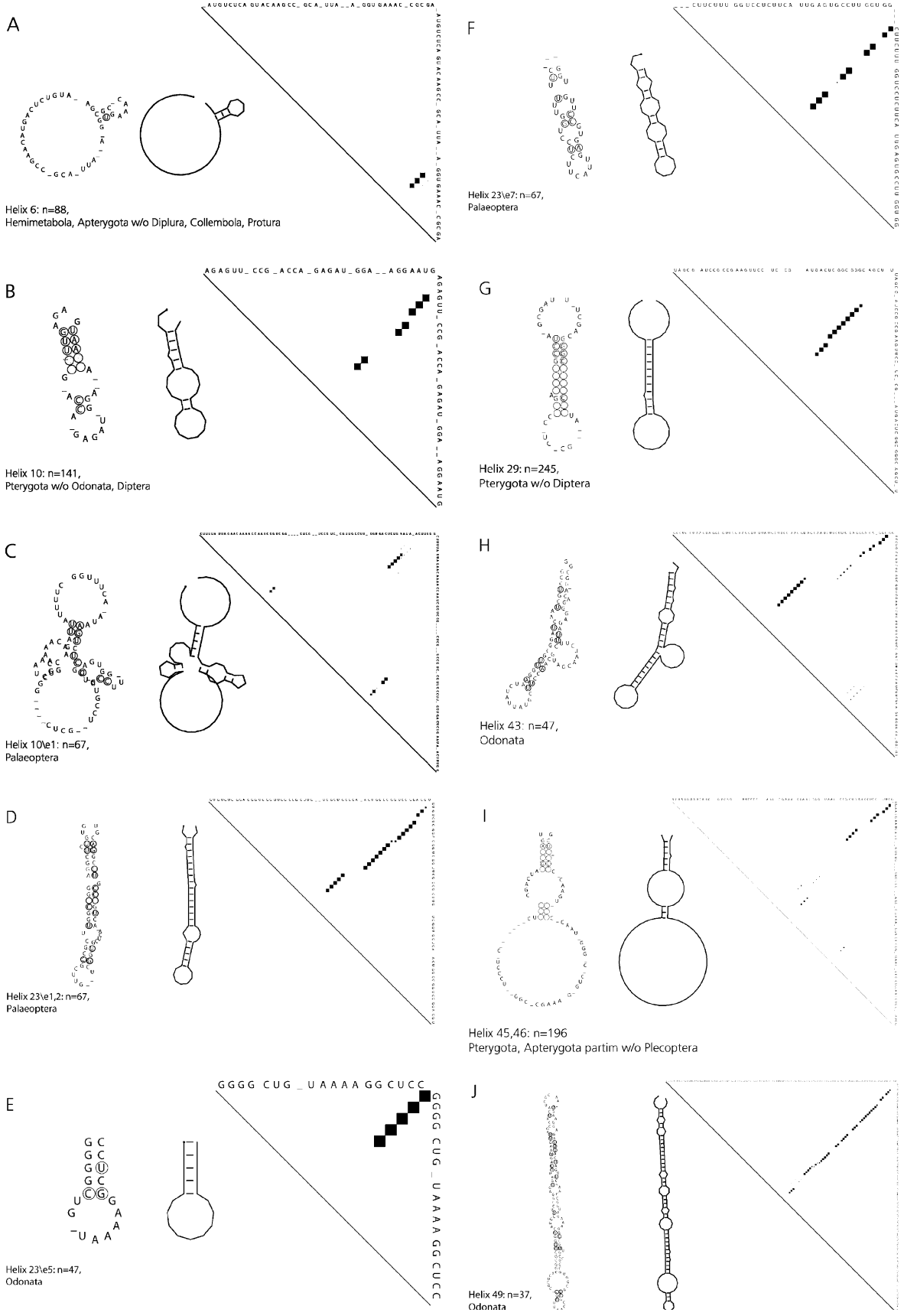
The hemimetabolan structure with base-pairing probabilities is illustrated in Fig. 3A.

Section V1, helix 9

The proposed structure of helix 9 in the SSU-vdP model is not confirmed by the complete hexapod data set. Instead, the hexapod sequences support a slightly different helix 9, which is adopted in the hexapod consensus SSU model (Fig. 1, also see Fig. A1 of Appendix A).

Section V2, helix 10

Wuyts et al. (2001) report a helix 10 homolog in all metazoan taxa with potentially additional helices in protists. The pattern of covariation showed considerable support for the proposed


A
Helix 6: n=88,
Hemimetabola, Apterygota w/o Diplura, Collembola, Protura
B
Helix 10: n=141,
Pterygota w/o Odonata, Diptera
C
Helix 10\e1: n=67,
Palaeoptera
D
Helix 23\e1,2: n=67,
Palaeoptera
E
Helix 23\e5: n=47,
Odonata
F
Helix 23\e7: n=67,
Palaeoptera
G
Helix 29: n=245,
Pterygota w/o Diptera
H
Helix 43: n=47,
Odonata
I
Helix 45,46: n=196
Pterygota, Apterygota partim w/o Plecoptera
J
Helix 49: n=37,
Odonata

eukaryotic helix within hexapods. However, in several taxa the pattern of positional variation indicated deviations from the consensus folding. Pterygote sequences excluding odonates support a structure comparable to the eukaryotic model (Fig. 3B). Odonates and ''apterygote'' taxa, on the other hand, show some deviations in base pairing, but there is a clear overlap in the helical structures between the eukaryotic model and the derived structure in hexapods (Fig. 1). Overall, a slight modification of the eukaryotic consensus structure of helix 10 appears appropriate for most hexapods, despite some taxon-specific deviations.

Section V2, helix 10\e1

The pattern of covariation supports the canonical eukaryotic model, but several mismatches indicate deviations from a single congruent model within hexapods. We found no well-supported consensus structure similar to the SSU-vdP for all hexapods (Fig. 1). This observation is in contrast to $M(x,y)$ values obtained for the base pairings of the canonical eukaryotic model. Helical structures between the general model and taxon-specific models partially overlap. In particular, helical structures in pterygote taxa, except Diptera and Odonata, largely overlap and receive support from covariation patterns. The Palaeoptera sequences yield a deviant structure supported by covariation patterns (Fig. 3C). In Zygentoma, Archaeognatha, and Collembola, we were unable to recover a consensus structure at all. The problems in identifying congruent secondary structures in helix 10\e1 might be due to sparse taxon sampling in individual subgroups. The $M(x,y)$ support for the proximal sections of the helix suggests using a consensus helical model for the proximal section different from the SSU-vdP folding until a more extensive sampling of subgroups can help to develop better supported, locally extended helices (Fig. 1).

Section V4, helix 23\e1–2

Wuyts et al. (2001) report eukaryotic consensus structures of helices 23\e1–2. Covariation analyses support this folding with an occasional occurrence of mismatches in both helices of hexapods. We were unable to reconstruct a consensus structure similar to the SSU-vdP model for all hexapod sequences using RNAalifold. But, except for the Diptera and the proturans, helical structures of other hexapods overlap and are well supported by thermodynamic and covariation criteria. Additionally, the number of sequences within individual taxa is high, lending confidence to the calculations (compare core structure for Palaeoptera in Fig. 3D). The deviant structures in Diptera and proturans suggest structural variation within hexapods, but the number of dipteran and proturan samples is too small to present well-supported alternative structures for these groups. The SSU-vdP model is an acceptable consensus model for hexapods, despite structural modifications in some taxa.

Section V4, helix 23\e5

Helix 23\e5 is absent in most hexapods. A possible helix 23\e5 homolog was present only in Odonata, Orthoptera, and Hymenoptera. These helices greatly differ. In Orthoptera, the first helical part appears well supported whereas the more distant base pairings receive little support in the probability matrix. Sequence variation within this region is extensive; therefore, the obvious sequence conservation of Odonata in this region is surprising (Odonate structure in Fig. 3E). A structure comparable to helix 23\e5 of the eukaryotic SSU model cannot be adopted in hexapods.

Section V4, helix 23\e7

A helix corresponding to the position of helix 23\e7 in the eukaryotic model was present in Palaeoptera, Hemimetabola, Hymenoptera, Coleoptera, Neuropterida, and ''apterygotes'' but not in diplurans and proturans. Dipluran sequences do not support any consensus folding and proturan lack this part completely. Structures were fairly well supported by consistent and compensatory substitutions (for example, in

Fig. 3. Examples of inferred helical structures of taxon specificity accompanied by base-pairing probability matrices. Examples of all newly inferred helices are illustrated and accompanied by the corresponding base-pairing probability matrices derived from the RNAalifold analyses. For all helices, the included taxa and the total number of sequences are given. In the graphical representation, sites that experience consistent or compensatory substitutions are indicated by black circles. Sites shaded in gray show inconsistent character states, but below a given threshold. The nucleotide sequence used to illustrate helices represents a consensus sequence from which structures were inferred. Base-pairing probability matrices depict the base4-pairing probability for a set of aligned sequences. The probability of pairing of two nucleotides is proportional to the size of the black square. Base-pairing matrices might show alternative pairings for a specific structure and their corresponding probabilities. (**A**) Helix 6, (**B**) helix 10, (**C**) helix 10\e1, (**D**) helices 23\e1,2, (**E**) helix 23\e5, (**F**) helix 23\e7, (**G**) helix 29, (**H**) helix 43, (**I**) helices 45, 46, and (**J**) helix 49.

Palaeoptera, Fig. 3F), and base-pairing probability matrices. The extent of helices overlaps between the canonical eukaryotic model and the inferred taxon-specific structures. The eukaryotic model is supported by covariation patterns and is adopted as an acceptable helical 23\e7 model within hexapods with certain limitations (see Proturans and Diplurans). A better supported inference of deviant structures in specific taxa requires additional data.

Section V5, helix 29

Covariation among hexapods supports a helix compatible with the eukaryotic SSU structure. However, the search for a common structure among all hexapod sequences using RNAalifold yielded no result. A helix compatible with the SSU model was found using all pterygote sequences. Base pairing is supported by covariation patterns and in base-pairing probability matrices (for example, in all pterygotes except for Diptera, Fig. 3G). The ''apterygote'' sequences fold into a slightly modified but compatible structure. Currently, a consensus model derived from RNAalifold analyses appears appropriate for the majority of taxa. Additional data are necessary to construct more detailed taxon-specific structures.

Section V7, helix 43

Covariation supports the general eukaryotic model for hexapod sequences. Using RNAalifold we were unable to derive unequivocal results. Structural variation between taxa is extensive (structures not shown). Analyses suggest that at least two different structural arrangements are stable within hexapods. These clearly different structures are partially well supported by covariation and base-pairing probabilities (Fig. 3H as an example). It appears that some comparable (homologous) helical motifs are present in several taxa. The extensive differences between some taxa prohibit the application of a general helical model in this region, and length variation in this part of the alignment suggests the occurrence of additional group-specific structural variation.

Section V8, helix 45, 46

We reconstructed a potentially homologous consensus structure to helices 45 and 46 of the eukaryotic SSU model for almost the complete data set (Fig. 3I), but the Plecoptera and Protura form vastly different structures and, if included in the alignment, destroy the consensus structure in hexapods (structures not shown). Our hexapod H45 consensus structure is compatible with the H45 eukaryotic model. A helix 45\e1 cannot be inferred from the present data. A helix 46 comparable to the general SSU model is not present in the majority of taxa and therefore not supported in the RNAalifold analyses. Currently, the data clearly support a helical folding similar, but not identical, to helix 45 in the eukaryotic model; consequently, this structure is adopted here. Base pairing in the region of a possible helix 46 is left unpaired for hexapods, given the available data.

Section V9, helix 49

The extent of helix 49 is extremely variable among taxa (compare collection of structures in the European Ribosomal RNA Database). The proximal stem region of the helix seems conserved between taxa, but covariation between paired sites does not unequivocally support the existence of helix 49. The complete alignment does not support a consensus structure in this helical region in RNAalifold either. Nevertheless, all taxa support individual helices with partly similar motifs (for example, helix 49 in Odonata, Fig. 3J). Base pairings between helices clearly overlap, but obvious variation between structures, particularly due to slippage in base pairings, does not support a consensus structure. Additional data are necessary to construct taxon-specific models of these highly length-variable regions. The stem section of the helix can be supported by covariation patterns and thermodynamic folding among the hexapod sequences. However, RNAalifold analyses and $M(x,y)$ indices are somewhat contradictory in the first section of the stem. The consensus structure in hexapods includes only this stem section of the helix.

In summary, the canonical eukaryotic SSU model fits well with hexapod sequences, but there are clearly taxon-specific deviations from the general SSU model. Structural variation ranges from length differences in helical extensions and differences in base pairings to complete loss of nucleotides. Structural variation is most prominent in helices 6, 10, 23, and 43–49. Exactly in these sections of the RNA molecules, the analyses of Wuyts et al. (2001) yielded high rates of site variation among eukaryotes. It could have been expected that deviations from the general model within hexapods were found in these sections.

At www.zfmk.de, extensive material is available: (a) a complete alignment of hexapod SSU sequences with RNA structure masks, both the eukaryotic SSU-vdPeer, the general hexapod RNA structure mask, and a helix code for rapid identification, (b) a complete list of RNA structure masks of

analyzed helices, (c) detailed information of all taxa-specific structures like taxon-consensus sequences and $M(x,y)$, (d) a complete list of $M(x,y)$ support for all hexapod SSU helices, and (e) complete data on the RNAalifold analyses of taxon-specific reconstructions. RNA structure masks and group-specific consensus sequences can serve for the analysis of additional sequence samples in more detailed, less taxonomically inclusive phylogenetic investigations.

DISCUSSION

Two lessons can be drawn from our analysis: (1) the general eukaryotic structure model is a more or less adequate model for hexapod sequences, but the approach is too coarse to present structural variation of less taxonomically inclusive data, and in turn (2) phylogenetic analyses incorporating secondary-structure information should rely on less-inclusive taxon-specific consensus models instead of general eukaryotic SSU models to profit most from the additional information. The generation of reliable secondary-structure models for specifically focused phylogenetic questions and corresponding data sets will promote even more realistic models of sequence evolution in molecular phylogenetics, thus further improving the reliability of fully parameterized approaches (compare Kjer, 2004). Our catalog of taxon-specific consensus structures (compare Fig. A1 of Appendix A) can serve as a starting point for phylogenetic analyses within hexapods.

A consensus model, as is developed here, is useful only if it proves (a) representative and (b) reliable. We think our analysis fulfills both requirements. First, with more than 250 sequences our results are most likely catching the essence of structural variation within hexapods. Second, the reliability of inferred structures depends predominantly on three points: (1) the quality of sequence alignments, (2) the presence of sequence variation, and (3) the biological significance of inference methods. The quality of sequence alignments is certainly critical in evaluating consensus structures. We decided to rely on structurally aligned SSU sequences from the European Ribosomal database, which incorporates structural information of more than 1,000 sequences. The quality of these alignments is most likely superior to any other automatically constructed alignment.

The comparison of the eukaryotic SSU model and hexapod SSU sequences was based on more than 250 sequences and most likely provides reliable information on deviant structural arrangements. However, sampling in certain taxa (orders) is admittedly insufficient to infer reliably taxon-specific structures because site variation will be insufficiently represented in these small samples. For example, in caddisflies only three sequences were available, which clearly limits the potential of the comparative approach. In these instances, sequences were combined with sequences of the presumptive sister clade. By doing this, we will obviously not discover caddisfly-specific structural variation. A better sampling will alleviate this problem in the future. In other cases, sampling was sufficient, but sequence variation was low, again limiting the power of a comparative approach.

Taxon-specific structural variation might contain phylogenetic signal (see, for example, Billoud et al., 2000; Collins et al., 2000; Lydeard et al., 2000; Ouvrard et al., 2000; Caetano-Anollés, 2002a, b). We searched structural variation within hexapods for phylogenetic patterns. As briefly mentioned in the Results section, there is indeed a great deal of autapomorphic group-specific variation. However, we were unable to identify clear phylogenetic signals within this variation (data not shown). It is quite striking that several helices display extensive within-group-specific variation and are completely conserved in other taxa. Obviously, selection regimes shift within the SSU rRNA molecule in a taxon-specific way. A similar pattern was observed in mitochondrial SSU rRNA sequences of dragonflies (Misof and Fleck, 2003). We guess that further comparative structural analyses could elucidate the rules governing the shift from rapid structural change and evolutionary conservation.

We refrained from using entire SSU rRNA molecules in RNAalifold analyses because it is unlikely to infer correct secondary-structure elements from entire molecules. The partitioning of the sequences into homologous sections, for which roughly appropriate helical models are available, seemed a more promising approach. It tremendously reduces the number of equally possible solutions and increases the chance of detecting biologically meaningful secondary structures. Our approach illustrates the need of further improved automation of structural analyses in phylogenetics. The intellectual background is present, but the realization of fully automated analyses is still lacking.

In the Introduction, we have argued that the inference of RNA secondary-structure models will

be decisive for the proper treatment of character dependence in phylogenetic analyses. Several RNA secondary-structure models have been published based on covariation analyses and thermodynamic criteria. So far, their taxon-specific applicability within hexapods has never been systematically explored. Our presentation of general hexapod consensus models and of subordinated taxa pursues the goal of developing aids for phylogenetic analyses at various taxonomic levels. Published phylogenetic investigations made use of general RNA secondary-structure models derived from comparison of eukaryotic sequences with sometimes minor adjustments for taxon sampling (for a most recent example, see Kjer, 2004). However, these general secondary-structure models might be inadequate for specific groups or at least will not capture the group-specific features of the rRNA molecules. Eukaryotic consensus models might be simply too general to serve as RNA structure masks within these parameterized approaches. A consequence could be the application of RNA substitution models on sequence sections that are not paired at all or the application of DNA substitution models on actually paired sites with the above-mentioned detrimental effects. It is obvious that consensus models specifically inferred from the taxon under consideration will improve the reality of model parameters. The secondary-structure models of hexapod orders inferred in the present study should help to reach this goal.

ACKNOWLEDGMENTS

We acknowledge the assistance of Maria Orland, Christiane Wallnisch, and Sabrina Bleidißel in our sequencing efforts. This work was supported by the DFG (MI 649/1–1).

APPENDIX A

For details see Table A1 and Fig. A1.

```
H6, alignment positions: 152 - 202
SSUvdPeer   ...((........(((.((.............))))....))...))...
Sequence    -AUGUCUCAGUACAAG-CC---nnA-UUA--A--GGU-GAAAC--CGCG-A
H6Ib        ..............(.((.............))).............
H6Ic        ....................................(((.....)))..
H6II        ....((((.........(((....((((.....))))).))).)))).....
H6III       ....................................(((....)))..

H9, alignment positions: 267 - 305
SSUvdPeer   .((........(...(((.....)))..)......))..
Sequence    CUUGG--A-UAACUGUGGUAAUUCUAGAGCUAAUACAUG
Model       ..........(.(.(((.....))).)).........

H10, alignment positions: 315 - 390
SSUvdPeer   ....((..((...(((.............................................).))......))))..
Sequence    nGAGCU--CCG-ACCn--------------GnGA------------------------n-GGA----AGGAGCG
Model       ....((..((...((.................................................))......))))..
H10I        ....((..((...((.................................................))......))))..
H10II       ...(((..((...((.................................................)).....)))))..
H10III      ..(((...((...((.................................................))......)).)))
H10IV       ....((..((...(((................................................)))....))))..
H10V        ...(((..(....((................................................))........))))..
H10VI       ...(((..((....(((..............................................)))....)))))..
H10VII      ...(((..((...((.................................................))......)))))..
H10IIX      ...(((...(...((..............((((.........................)))..)))....).)))..

H23\e7, alignment positions: 1771 - 1876
SSUvdPeer   .((((.........................(.......(((.(((.((.....................).).)))..))...................))))))
Sequence    -CUGU------------------------CU------CGG-U-C-UC-----UUA-AUU---------G-A-GUG--UCU-----------------UGGUGG
Model       .((((.........................(.......(((.(((.((.....................).).)))..))...................))))))
H23\e7Ic    .((((.........................(........((...(.((.....................).).)....))...................))))).
H23\e7Id    .((((.........................(...........(((.((.....................).).)))....................)))))
H23\e7Ie    .((.(.........................(........((....(((.....................).).)....))...................)).))
H23\e7If    ....................................((....(((...........).........).).)...)...)....................

H29, alignment positions: 2284 - 2343
SSUvdPeer   .(((...((((((..........(((.............)))..)))))))..))))..
Sequence    UGGCG-AUCCGnC--GnA-----GUUnCn--UU-nn---A-UGnCUCGnCGGGCAGCC-U
Model       .......((((((..((.........................))))))))........
H29I        .(((..(((((((...(((.....(((((.............)))))))))).)))..
H29II       .......((((((..((......(((.............)))))))))).........
H29III      .(((...((((((..(.(((.....................)))))))))..))))..
H29IV       .(((...((((((..(.......((((.............)))).)))))))..))))..
H29V        .......(((((((.((.......(((.............)))))))).))).......
```

Fig. A1. Examples of group-specific structures in bracket notation in comparison to the SSU-vdP model and the hexapod consensus.

TABLE A1. Compilation of analyzed helices and corresponding taxon groupings

Helix	Subset	*n*	Structure	Taxa
H6				
	H6Ib	78	y	Holometabola
	H6Ic	88	y	Hemimetabola, Palaeoptera, Apterygota w/o Diplura, Collembola, Protura
	H6II	5	y	Diplura
	H6III	7	y	Collembola
	H6IV	5	n	Protura
H10				
	H10I	6	y	Diptera
	H10II	141	y	Pterygota w/o Odonata, Diptera
	H10III	47	y	Odonata
	H10IV	9	y	Zygentoma
	H10V	4	y	Archaeognatha
	H10VI	6	y	Diplura
	H10VII	7	y	Collembola
	H10VIII	5	y	Protura
H10\e1				
	I	6	y	Diptera
	II	50	y	Mecoptera, Lepidoptera, Hymenoptera
	III	34	y	Coleoptera
	IVb	30	y	Hemimetabola w/o Palaeoptera
	IVc	67	y	Palaeoptera
	V	26	n	Apterygota w/o Protura
	VI	5	y	Protura
H23\e1,2				
	Ib	6	y	Diptera
	Ic	96	y	Holometabola
	Id	33	y	Hemimetabola w/o Palaeoptera, Plecoptera
	Ie	67	y	Palaeoptera
	If	35	y	Mecopterida w/o Diptera
	Ig	61	y	Holometabola w/o Mecopterida
	II	34	y	Plecoptera
	III	14	y	Zygentoma, Archaeognatha
	IV	6	y	Diplura
	V	7	y	Collembola
	VI	4	y	Protura
H23\e5				Only Pterygota, missing in Apterygota
	I	6	n	Diplura
	II	25	n	Mecoptera, Siphonaptera
	III	11	n	Lepidoptera (four sequences do not have a helix here)
	IV	15	y	Hymenoptera
	V	33	n	Coleoptera
	VI	12	n	Neuropteroidea
	VII	6	n	Hemiptera
	VIII	5	y	Orthoptera
	IX	3	n	Phasmatodea
	X	19	n	Dictyoptera
	XI	34	n	Plecoptera
	XII	20	n	Ephemeroptera
	XIII	47	y	Odonata
H23\e7				Missing in Protura
	I	257	n	Pterygota, Collembola, Zygentoma
	Ib	42	n	Mecopterida
	Ic	61	y	Hymenoptera, Coleoptera, Neuropteroidea
	Id	64	y	Hemimetabola w/o Palaeoptera
	Ie	67	y	Palaeoptera
	If	21	y	Apterygota w/o Diplura, Protura
	II	6	n	Diplura

TABLE A1. Continued

Helix	Subset	*n*	Structure	Taxa
H29				
	I	6	y	Diptera
	II	245	y	Pterygota w/o Diptera
	IIb	97	y	Holometabola w/o Diptera
	IIc	68	y	Hemimetabola w/o Palaeoptera
	IId	80	y	Palaeoptera, Apterygota w/o Diplura, Collembola, Protura
	III	6	y	Diplura
	IV	7	y	Collembola
	V	5	y	Protura
H43				
	I	5	y	Diptera
	II	19	y	Mecoptera, Siphonaptera
	III	2	y	Lepidoptera
	IV	8	y	Hymenoptera
	V	34	n	Coleoptera
	VI	7	n	Neuropteroidea
	VII	6	y	Hemiptera
	VIII	4	n	Orthoptera
	IX	19	y	Phasmatodea, Dictyoptera
	X	34	y	Plecoptera
	XI	20	y	Ephemeroptera
	XII	47	y	Odonata
	XIII	12	y	Zygentoma, Archaeognatha
	XIV	4	y	Diplura
	XV	8	y	Collembola
	XVI	4	y	Protura
H45,46				
	I	34	y	Plecoptera
	II	196	y	Pterygota, Apterygota partim w/o Plecoptera
	IIb	75	n	Holometabola
	IIc	96	n	Hemimetabola
	IId	25	y	Apterygota w/o Protura
	IIe	34	y	Mecopterida, Hymenoptera
	IIf	41	y	Coleoptera, Neuropteroidea
	IIg	29	y	Hemimetabola w/o Plecoptera, Palaeoptera
	IIh	67	y	Palaeoptera
	III	5	y	Protura
H49				
	I	5	y	Diptera
	II	19	y	Mecoptera, Siphonaptera
	III	2	y	Lepidoptera
	IV	8	y	Hymenoptera
	V	33	y	Coleoptera
	VI	5	y	Neuropteroidea
	VII	6	y	Hemiptera
	VIII	4	y	Orthoptera
	IX	2	y	Phasmatodea
	X	13	y	Dictyoptera
	XI	34	y	Plecoptera
	XII	17	y	Ephemeroptera
	XIII	37	y	Odonata
	XIV	12	y	Zygentoma, Archaeognatha
	XV	6	y	Diplura
	XVI	7	y	Collembola
	XVII	5	y	Protura

LITERATURE CITED

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

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

Buckley TR, Simon C, Flook PK, Misof B. 2000. Secondary structure and conserved motifs of the frequently sequenced domains IV and V of the insect mitochondrial large subunit rRNA gene. Insect Mol Biol 9:565–580.

Caetano-Anollés G. 2002a. Evolved RNA secondary structure and the rooting of the universal tree of life. J Mol Evol 54: 333–345.

Caetano-Anollés G. 2002b. Tracing the evolution of RNA structure in ribosomes. Nucleic Acids Res 30:2575–2587.

Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, Feng B, Lin N, Madabusi LV, Muller KM, Pande N, Shang Z, Yu N, Gutell RR. 2002. The comparative RNA web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BioMed Central Bioinform 3:2. [Correction: BioMed Central Bioinformatics. 3:15.], http://www.rna.icmb.utexas.edu/.

Chalwatzis N, Baur A, Stetzer E, Kinzelbach R, Zimmermann FK. 1995. Strongly expanded 18S rRNA genes correlated with a particular morphology in the insect order Strepsiptera. Zoology 98:115–126.

Collins LJ, Moulton V, Penny D. 2000. Use of RNA secondary structure for studying the evolution of RNase P and RNase MRP. J Mol Evol 51:194–204.

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

Gutell RR. 1993. Comparative studies of RNA: inferring higher-order structure from patterns of sequence variation. Curr Opin Struct Biol 3:313–322.

Gutell RR. 1994. Collection of small subunit (16S- and 16S-like) ribosomal RNA structures. Nucleic Acids Res 22: 3502–3507.

Gutell RR, Power A, Hertz GZ, Putz EJ, Stormo GD. 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 RR, Lee JC, Cannone JJ. 2002. The accuracy of ribosomal RNA comparative structure models. Curr Opin Struct Biol 12:301–310.

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

Hickson RE, Simon C, Perrey SW. 2000. The performance of several multiple-sequence alignment programs in relation to secondary-structure features for an rRNA sequence. Mol Biol Evol 17:530–539.

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

Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P. 1994. Fast folding and comparison of RNA secondary structures. Monatsh Chem 125: 167–188.

Hofacker IL, Fekete M, Flamm Ch, Huynen MA, Rauscher S, Stolorz PE, Stadler PF. 1998. Automatic dedection of conserved RNA structure elements in complete RNA virus genomes. Nucleic Acids Res 26:3825–3836.

Hofacker IL, Fekete M, Stadler P. 2002. Secondary structure prediction for aligned RNA sequences. J Mol Biol 319: 1059–1066.

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

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

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

Kjer KM. 2004. Aligned 18S and insect phylogeny. Syst Biol 53:506–514.

Lydeard C, Holznagel WE, Schnare MN, Gutell RR. 2000. Phylogenetic analysis of molluscan mitochondrial LSU rDNA sequences and secondary structures. Mol Phylogenet Evol 15:83–102.

Maddison DR, Baker MD, Ober KA. 1999. Phylogeny of the carabid bettles as inferred from 18S ribosomal DNA (Coleoptera: Carabidae). Syst Entomol 24:103–138.

Misof B, Fleck G. 2003. Comparative analysis of mt 16S rRNA secondary structures of Odonates and its relevance to phylogenetic problems in insect systematics. Insect Mol Biol 12:535–547.

Misof B, Anderson CL, Buckley TR, Erpenbeck D, Rickert A, Misof K. 2002. An empirical analysis of mt 16S rRNA covarion-like evolution in insects: site-specific rate variation is clustered and frequently detected. J Mol Evol 55:460–469.

Notredame C, O'Brien EA, Higgins DG. 1997. RAGA: RNA sequence alignment by genetic algorithm. Nucleic Acids Res 25:4570–4580.

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

Page RDM. 2000. Comparative analysis of secondary structure of insect mitochomdrial small subunit ribosomal RNA using maximum weighted matching. Nucleic Acids Res 28:3839–3845.

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

Parsch J, Braverman JM, Stephan W. 2000. Comparative sequence analysis and patterns of covariation in RNA secondary structures. Genetics 154:909–921.

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.

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.

de Rijk P, de Wachter R. 1997. RnaViz, a program for the visualisation of RNA secondary structure. Nucleic Acids Res 25:4679–4684.

Rzhetsky A. 1995. Estimating substitution rates in ribosomal RNA. Genetics 141:771–783.

Savill NJ, Hoyle DC, Higgs PG. 2001. RNA sequence evolution with secondary structure constraints: comparison of substitution rate models using maximum likelihood methods. Genetics 157:399–411.

Schoeniger M, von Haeseler A. 1994. A stochastic model for the evolution of autocorrelated DNA sequences. Mol Phylogenet Evol 3:240–247.

Stephan W. 1996. The rate of compensatory evolution. Genetics 144:419–426.

Tabaska JE, Cary RB, Gabow HN, Stormo GD. 1998. An RNA folding method capable of identifying pseudoknots and base triples. Bioinformatics 14:691–699.

Thomas MA, Walsh KA, Wolf MR, McPheron BA, Marden JH. 2000. Molecular phylogenetic analysis of evolutionary trends in stonefly wing structure and locomotor behavior. Proc Natl Acad Sci USA 97:13178–13183.

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882.

Tillier ERM, Collins RA. 1995. Neighbor joining and the maximum likelihood with RNA sequences: addressing the interdependence of sites. Mol Biol Evol 12:7–15.

Tillier ERM, Collins RA. 1998. High apparent rate of simultaneous compensatory base pair substitutions in ribosomal RNA. Genetics 148:1993–2002.

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

Wuyts J, Van de Peer Y, de Wachter R. 2001. Distribution of substitution rates and location of insertion sites in the tertiary structure of ribosomal RNA. Nucleic Acids Res 29: 5017–5028.

Wuyts J, Perrière G, Van de Peer Y. 2004. The European ribosomal RNA data base. Nucleic Acid Res 32: D101–D103.