

18S rRNA Secondary Structure and Phylogenetic Position of Peloridiidae (Insecta, Hemiptera)

David Ouvrard,* Bruce C. Campbell,† Thierry Bourgoïn,* and Kathleen L. Chant

* Laboratoire d'Entomologie and ESA 8043 du CNRS, Muséum National d'Histoire Naturelle, 45 Rue Buffon, F-75005 Paris, France; and

† Western Regional Research Center, USDA-ARS, 800 Buchanan Street, Albany, California 94710-1100

Received September 22, 1999; revised March 10, 2000

A secondary structure model for 18S rRNA of peloridiids, relict insects with a present-day circumantarctic distribution, is constructed using comparative sequence analysis, thermodynamic folding, a consensus method using 18S rRNA models of other taxa, and support of helices based on compensatory substitutions. Results show that probable *in vivo* configuration of 18S rRNA is not predictable using current free-energy models to fold the entire molecule concurrently. This suggests that refinements in free-energy minimization algorithms are needed. Molecular phylogenetic datasets were created using 18S rRNA nucleotide alignments produced by CLUSTAL and rigorous interpretation of homologous position based on certain secondary substructures. Phylogenetic analysis of a hemipteran data matrix of 18S rDNA sequences placed peloridiids sister to Heteroptera. Resolution of affiliations between the three main euhemipteran lineages was unresolved. The peloridiid 18S rRNA model presented here provides the most accurate template to date for aligning homologous nucleotides of hemipteran taxa. Using folded 18S rRNA to infer homology of character as morpho-molecular structures or nucleotides and scoring particular sites or substructures is discussed. © 2000 Academic Press

INTRODUCTION

The 18S rRNA molecule of the ribosomal small subunit is frequently used to infer phylogenetic affiliations of ancient (>100 Mya) eukaryotic relationships. Its suitability as a phylogenetic tool is two-fold. First, it is a good source of phylogenetic information based on conservation of function, variable mutation rates depending on substructure position, and its ubiquity in all taxa. Second, the molecule provides readily obtainable nucleotide sequences because of high rRNA transcript copy number in eukaryotes and ease of PCR primer design (Woese, 1987; Kjer, 1995). The 18S rRNA molecule contains certain stable sections having low substitution rates. As such, 18S rRNAs can provide

informative characters for assessing affiliations of evolutionarily distant taxa. However, 18S rRNAs also possess variable regions having high substitution rates. In some organisms these rapidly evolving expansion regions (or helices) provide phylogenetic signals for discerning relationships between evolutionarily closer clades.

Genetic changes (characters) used to predict phylogenetic affiliations are of greatest utility if homology of characters can be inferred as accurately as possible. Determining homology between molecular characters is a significant issue. In some instances, definitions for homology used in morphology (Remane, 1956; Patterson, 1982) are not exactly applicable to nucleotide data. For example, in comparing rRNA nucleotide sequences between taxa, character to character homology (primary homology) is frequently proposed by aligning respective sequences according to an evaluation of similarity. Individual base positions are aligned to maximize overall position likeness along the entire sequence. Variations of this approach are found in the computer programs MALIGN (Wheeler and Gladstein, 1994) and POY (Gladstein and Wheeler, 1997). These programs perform iterations of progressively optimized data matrices based on gap costs and tree length or character insertion and deletion, respectively. Cladistic analysis of such restructured character matrices may or may not support initially proposed homologies according to congruency of apomorphies (i.e., secondary homology: De Pinna, 1991; Bourgoïn, 1996). An altogether different approach to determining homology of character is secondary reconstruction of rRNA molecules to improve recognition of primary homology (De Rijk and De Wachter, 1993; Kjer, 1995), applied in this study.

Peloridiids (common name moss bugs) are rare insects believed to be relictual members of an ancient preheteropteran lineage of Hemiptera (*sensu lato*) (Evans, 1981). The family includes 21 extant species having a classically Gondwanan distribution currently confined to small isolated populations in South America, Australia, New Zealand, New Caledonia, and certain remote islands in the southern Pacific (Evans,

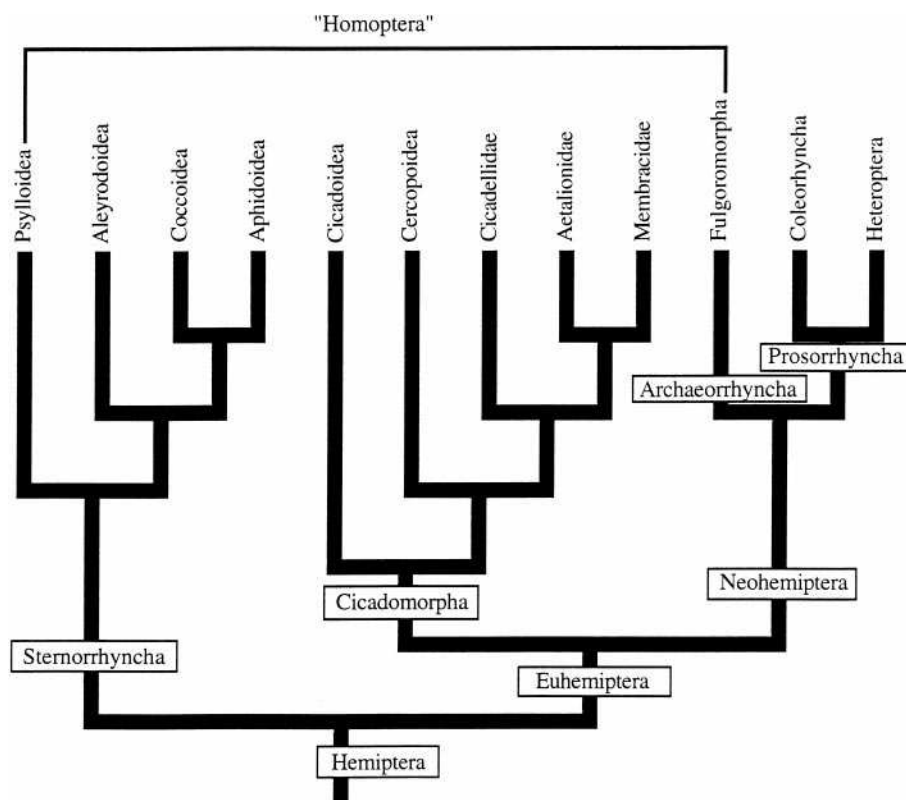


FIG. 1. Inferred phylogeny of Hemiptera (*sensu lato*) showing taxonomic scheme based on cladification (after Sorensen *et al.*, 1995).

1981; Burckhardt and Agosti, 1991). The phylogenetic position of the Peloridiidae among Hemiptera has been continually debated. Breddin (1897) first placed Peloridiidae in Heteroptera. Myers and China (1929) transferred them to Homoptera, as members of a newly created series, the Coleorrhyncha. China (1962) later suggested transferring them to Auchenorrhyncha. Schlee (1969) placed Peloridiidae and Heteroptera as sister groups (adelphotaxa) in the superorder Heteropteroidea. Cobben (1978) considered Schlee's synapomorphies as superficial and argued against monophyly of Heteroptera + Coleorrhyncha. Evans (1981) regarded Peloridiidae as an independent suborder of Hemiptera with Homoptera and Heteroptera. Finally, based on genital morphology, Bourgoin (1993) argued that Homoptera and Auchenorrhyncha were paraphyletic and Peloridiidae + Fulgoromorpha might form a clade sister to Heteroptera. However, recent results suggest a closer relationship of Prosorrhyncha to Clypeorrhyncha than to Fulgoromorpha (Bourgoin *et al.*, 1999). Studies using molecular data also indicate that Homoptera and Auchenorrhyncha are paraphyletic (Sorensen *et al.*, 1995; von Dohlen and Moran, 1995; Campbell *et al.*, 1995).

Clade Heteropteroidea (*sensu* Schlee, 1969) was recently supported with morphological (see Sorensen *et al.*, 1995) and 18S rDNA data, the latter consisting of two studies showing four (Wheeler *et al.*, 1993; Schuh

and Slater, 1995) and two (Campbell *et al.*, 1995) synapomorphic sites. The new suborder name Prosorrhyncha was proposed to replace Heteropteroidea, in addition to other new names to conform all hemipteran suborders with a “-rrhyncha” suffix (Sorensen *et al.*, 1995). However, in these molecular studies, only partial 18S rDNA sequences were used. Campbell *et al.* (1995) and Bourgoin *et al.* (1997) found only one synapomorphic site to infer Prosorrhyncha as sister to Archaeorrhyncha (=Fulgoromorpha), forming clade Neohemiptera *sensu* Sorensen *et al.* (1995) (Fig. 1). To date, morphological and anatomical studies have not agreed on a definitive phylogenetic framework of these hemipteran subgroups. Because peloridiids possess a combination of ancestral and derived “homopteran” and heteropteran morphological features, knowing their evolutionary position in Hemiptera (*sensu lato*) might help resolve affiliations of the major hemipteran lineages. In this study, we present full 18S rRNA sequences of two peloridiids, *Hackeriella veitchi*, first described by Hacker in the genus *Hemiodocus* in 1932, and *Hemiowoodwardia wilsoni* (Evans, 1936). We incorporate these sequences into a phylogenetic dataset to deduce evolutionary affiliations of hemipteran lineages. We also use these sequences to perform a rigorous inference of the secondary structure of peloridiid 18S RNA. Preliminary examination of this structure shows that it will serve as a reliable template

TABLE 1

Glossary of Terms for Secondary Structures of RNAs

Term	Definition
Stem	A right-handed double helix composed of a succession of complementary hydrogen-bonded nucleotides between paired strands.
Single strand	Unpaired nucleotides separating stems.
Hairpin loop	Stem closed distally by a loop of unpaired nucleotides.
Terminal bulge	Succession of unpaired nucleotides at the end of a hairpin loop.
Lateral bulge	Succession of unpaired nucleotides on one strand of a stem.
Internal bulge	Group of nucleotides from the two opposite strands unable to form canonical base pairs.
Junction or multibranched loop	Succession of groups of unpaired nucleotides joining the last proximal pairing of several stems.
Compensatory substitution	Mutation on one strand of a stem following initial mutation of a complementary base to maintain structure.
Semiconsensatory substitution	Mutation of a base on one strand not affecting structure (A•U or G•C replaced by G•U), except in terms of thermodynamics.

to improve understanding of homology of characters. These characters can be in the form of nucleotides or “morpho-molecular” substructures to be incorporated into a larger dataset, containing other hemipteran exemplars.

MATERIAL AND METHODS

Insects

Peloriid nymphs and adults were collected from moss associated with *Nothofagus*. Specimens of *H. veitchi* were collected in the Springbrook Plateau, S.E. Queensland, 14 March 97, and those of *H. wilsoni* in Melba Gully, Otways, Victoria, 5 November 97. Insects were collected from moss into ethylene glycol using a Berlese funnel. Peloriids were transferred to 95% ethanol, shipped to BCC, and stored at -80°C prior to extraction of genomic DNA.

Isolation of Genomic DNA, PCR Amplification, Cloning, and Sequencing 18S rDNA

Insects were removed from alcohol and dried for 10 min using a Speed Vac Concentrator (Savant). Total genomic DNA was purified by homogenizing two insects in 200 μl buffer [10 mM Tris (pH 8.0), 2.5 mM MgCl_2 , 50 mM KCl], 200 μl phenol, and 20 μl 10% SDS. Liquid phases were separated by centrifugation. DNA was precipitated from the top phase using 95% ethanol (3:1). The pellet was dried, washed in 70% ethanol, dried again, and resuspended in 20 μl TE buffer [10 μM Tris (pH 8.0), 1 mM EDTA]. The polymerase chain reactions (PCR: Saiki *et al.*, 1988) were performed with components of the Platinum *Taq* DNA Polymerase High Fidelity kit (Life Technologies, Inc.) in 25- μl reactions: 1 μl DNA template, 2.5 μl PCR buffer, 0.5 μl each dNTP, 2 μl each forward and reverse primer, 1 μl MgSO_4 , 0.125 μl *Taq* DNA polymerase, and 14.375 μl water. Full 18S rDNA were amplified using the forward primer: 5'-CTG GTT GAT CCT GCC AGA GT-3' and reverse primer: 5'-TCC TTC CGC AGG TTC ACC-

3'. The PCR cycling program was 94°C for 1 min, followed by 30 cycles 94°C for 30 s, 55°C for 30 s, 68°C for 1 min, with 68°C for 7 min after last cycle. PCR products were cloned using plasmids and competent cells supplied in the TA Cloning System (Invitrogen). Putative clones containing 18S rDNA were selected using supplies and protocols supplied with the RPM Rapid Pure Miniprep Kit (BIO 101). Both top and bottom strands of 18S rDNA were fully sequenced by the dideoxynucleotide-termination method (Sanger *et al.*, 1977) on a LI-COR automated DNA sequencer using components and protocols from SequiTherm EXCEL II DNA Sequencing Kit-LC (Epicentre Technologies).

Predicting Secondary Structure

For a complete review of techniques used for predicting secondary structures see Chastain and Tinoco (1991). Descriptions of seven types of substructures of a folded rRNA molecule are provided for reference (Table 1). The three approaches used to estimate secondary structure for peloriid 18S rRNA are outlined as follows.

Comparative sequence analysis. The comparative sequence method is initially used to find regions in 18S rRNA shared by a large number of taxa and to preliminarily infer secondary structure of relatively conserved portions of the molecule. According to Chastain and Tinoco (1991), this method provides the best results when sequences have approximately 70% base similarity. With this method, secondary structure is inferred based on the principle that substitutional changes not affecting function may be relatively conserved. The basic postulate of this comparative method is that structural conservation between different species, despite base substitutions in the sequence, is reliable evidence of an important functional role for that structure. During the first step, rRNA sequences from different organisms are aligned according to primary structure. The second step is to confirm paired regions as structural elements by positional covariance be-

tween available sequences. For example, A · U base pairs might replace G · C base pairs (or vice versa) to maintain stem substructures of helical regions. Semi-compensatory substitutions can also conserve structural features, but indicate a stem which is not as energetically stable as one having full compensatory substitutions. As number and diversity of full 18S rDNA sequences available from databases such as GenBank have increased, complementary base pairings in conserved regions have become reasonably established according to positional covariance (Gutell, 1993).

Secondary structure determination of full 18S rRNAs has also been proposed by comparing closely related sequences using the DCSE program (De Rijk and De Wachter, 1993). This program maps a secondary structural prediction established for one organism onto the sequence of another organism. It then checks whether this second sequence can realistically fold according to the initially predicted model. Another program, the RNAlign program (Corpet and Michot, 1994), uses a similar approach but for small relatively conserved regions. Refinement of aligned sequences based on secondary structure can also be attempted manually following the procedures outlined by Kjer (1995).

Published models of 18S rRNA secondary structures of arthropods used as references in the comparative sequence method in our study included Arachnida—*Aphonopelma* sp. (Hendriks *et al.*, 1988a); Crustacea—*Artemia salina* (Nelles *et al.*, 1984); Insecta: Coleoptera—*Tenebrio molitor* (Hendriks *et al.*, 1988b); Hemiptera—*Acyrtosiphon pisum* (Kwon *et al.*, 1991); Diptera—*Drosophila melanogaster* (Hancock *et al.*, 1988; Gutell, 1994; van de Peer *et al.*, 2000).

For constructing the model of peloridiid 18S rRNA using comparative sequence analysis we also examined full sequences of 18S rRNAs of other hemipterans (Campbell *et al.*, 1995), including *Okanagana utahensis* Davis, 1919 (Cicadidae, U06478); *Philaenus spumarius* L., 1758 (Cercopidae, U06480); *Pealius kelloggii* (Bemis, 1904) (Aleyrodidae, U06479); *Lygus hesperus* Knight, 1917 (Miridae, U06476). The comparative sequence analysis is impractical for determining secondary structure of strictly conserved sequences (no compensatory substitutions) and highly variable regions (difficulty of alignment).

Thermodynamic folding. This second method is based on predicting thermodynamically stable rRNA molecules by calculating free energies of all possible secondary structures and retaining one of lowest energy (i.e., the most stable). In some instances, structures with free energies slightly higher than one with lowest energy are also of interest. Currently used algorithms for thermodynamic folding are predicted on incomplete experimental data so that the predicated probability of certain structures can be irresolute

(Chastain and Tinoco, 1991). Furthermore, there may be stabilizing tertiary interactions between helices *in vivo* that are not incorporated into currently available algorithms used to predict structure based on lowest free energy (Fields and Gutell, 1996).

The Zuker–Steigler algorithm (Zuker and Steigler, 1981) was used to infer secondary structures of variable regions <50 bases in length for which no model was published. This algorithm is a subprogram of MacDNASIS (Hitachi Software). For folding longer variable regions, the MUFOLD program v. 3 (Mathews *et al.*, 1999; Zuker *et al.*, 1999) was used from the mfold server, <http://mfold2.wustl.edu/~mfold/rna/form1.cgi>. In some cases, a consensus structure was chosen instead of a structure predicated on lowest energy (optimal thermodynamic structure). In such cases, this consensus structure was defined by comparing several structures of low energy (suboptimal structures) for the same molecule (automated procedure in Konings and Hogeweg, 1989). Retained consensus helices were folded for several insects so that the final consensus structure was one interpreted from both phylogenetic comparison and thermodynamic folding methods.

Reliability of this method depends greatly on thermodynamic parameters. Currently, experimental thermodynamic data are missing for some substructures such as bulges, hairpins, and internal loops (Table 1). Also, there are no experimental thermodynamic values for junctions or multibranched loops (Jaeger *et al.*, 1989). Prediction of secondary structures having numerous multibranched loops should be accepted with caution. Furthermore, tertiary and quaternary relations, not inferred in lowest energy models, may stabilize true helices *in vivo* (Noller, 1984). There are more than 20 proteins present in the small subunit of the ribosome that could play a role in constraining structural conformations which are not predicted using available algorithms.

The thermodynamic folding technique appears to give best results (i.e., agree using broad phylogenetic comparisons) when pairing base positions relatively near to each other in the primary sequence, particularly those bases forming a hairpin loop (Konings and Gutell, 1995). It is for this reason that an entire 18S rRNA cannot be inferred from a single folding using any of the available thermodynamic algorithms. Attempts to fold entire primary sequences of 18S rRNAs between even relatively closely related taxa (e.g., among insects in the same family) renders secondary structural models too different for reliable comparison, and even basic postulates on structure cannot be tested.

Consensus method. Chastain and Tinoco (1991) suggested successive use of thermodynamic and phylogenetic comparison techniques. In our study, structures proposed by other authors for other insects were used as templates to overlay a structure for the pelo-

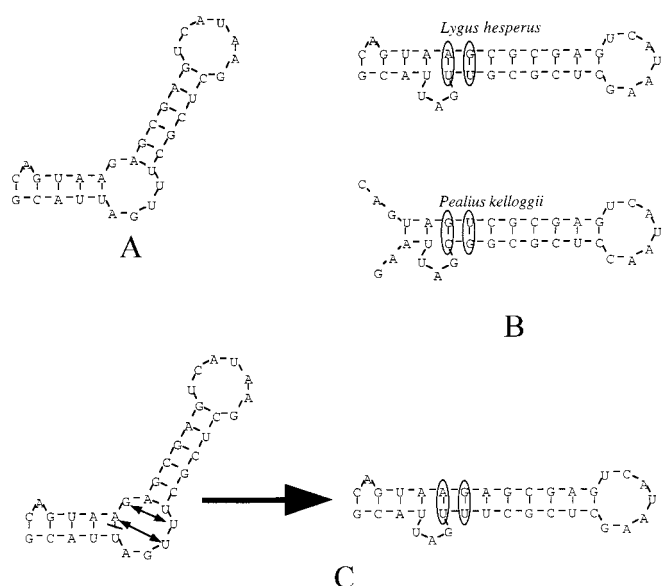


FIG. 2. Steps in folding of a helix combining techniques of thermodynamic folding and phylogenetic comparison. (A) First step, helix 48 of *Hackeriella veitchi* (Hemiptera, Prosorrhyncha, Peloridiidae) is folded using Zuker–Steigler MacDNASIS subprogram. (B) Second step, two compensatory substitutions are found in helices 48 of *Lygus hesperus* (Hemiptera, Prosorrhyncha, Heteroptera) and *Pealius kelloggii* (Hemiptera, Sternorrhyncha, Aleyrodoidea). (C) Third step, the four bases in the internal bulge of helix 48 in *H. veitchi* are forced to pair and render a final inference for the structural configuration of this helix.

ridiid sequences to determine degree of congruence. Figure 2 shows how a comparison between different taxa can be used to refine the structure of an unknown helix for a new taxon. If more than one structure could be inferred, the interpretation based on least free-energy was retained when calculation of free-energy values was possible with MacDNASIS v 3.6. Helical structures were inferred using thermodynamics only in regions where rRNA sequences were highly variable among taxa and no consensus model was available. Drawings of secondary structures were created using simple drawing software (Deskdraw, MK Software) or were imported from MacDNASIS files.

Steps and decisions followed to infer secondary structure for 18S rRNAs in our study were:

(1) Comparison between sequences. Can the selected sequence of peloridiids be folded into structures proposed in the literature? Yes → 2; No → 3.

(2) Comparison between structures proposed in the literature and thermodynamic folding. Does thermodynamic folding result in more stable and compatible helices, including examination of homologous sequences of other insects (e.g., *T. molitor* and *D. melanogaster*)? If so: → Refinement; If not: published helix retained. → 4.

(3) Peloridiid substructure sequence folded using Zuker–Steigler algorithm. Then → 4.

(4) In each case, supporting evidence of each helix is sought by searching for presence or absence of compensatory or semicompensatory substitutions.

Phylogenetic Analysis

Phylogenetic positions of Peloridiidae and other major lineages within Hemiptera were assessed using 18S rDNA nucleotide sequences of other hemipteran taxa, as presented in Campbell *et al.* (1995). The sequences of *H. veitchi* and *H. wilsoni* and other insect taxa were aligned initially using CLUSTAL W (Thompson *et al.*, 1994). Three data matrices were analyzed: full (all informative sites), attenuated (excluding ambiguously aligned sites), and polarized (excluding sites homoplasious in outgroup taxa), as defined in Campbell *et al.* (1995). Phylogenetic analyses were performed with PAUP* (v. 4.0b2a) (Swofford, 1998). Indels were treated as missing data.

RESULTS AND DISCUSSION

Primary Structure

Full 18S rRNA sequences of *H. veitchi* and *H. wilsoni*, 1909 bp each, are deposited in GenBank, under Accession Nos. AF004766 and AF131198, respectively.

Secondary Structure Model

Numbering of helices is based on van de Peer *et al.* (2000). Highly variable expansion regions of primary sequences and secondary structures among eukaryotes are often noted as E6, E9, E10, E23, or E41. The “E” refers to Eukaryota. Such expansion regions are absent from bacterial 16S and 16S-like rRNAs.

Helices 1 to 21 (Fig. 3, 18S rRNA model Part I). (1 and 2) No variation between peloridiid sequences and published 18S rRNA sequences of other insects. (3) Agrees with *T. molitor* model, but does not agree with *A. pisum*. The former model was chosen in accordance with Gutell (1994). (4) Same as the predicted structure for *T. molitor*, the strongest thermodynamically. (5) Each base in this helix is conserved in more than 95% of eukaryotes (Gutell, 1994). The retained arrangement is strongly supported. (6) In peloridiids this helix does not show great variation in length, unlike other eukaryotic expansion regions. Despite not expanding in length there is great variation between taxa in base composition of this helix. As a guiding aid, Gutell (1994) shows that the 1st base pairing of helix 6 should be a G · C pair. In the peloridiid model this G is at base position 54 and the C is at position 86. If mfold is constrained to predict the structure with this G · C pair at the base of this helix, the obtained structure for the peloridiid differs from the structure of helix 6 predicted for *T. molitor* (Hendriks *et al.*, 1988b), but our configuration for this helix shows greater stability (≈3.7 kcal/mol; less free-energy than in the *T. molitor* model) because of a greater number of G · C pairs. (7) A short

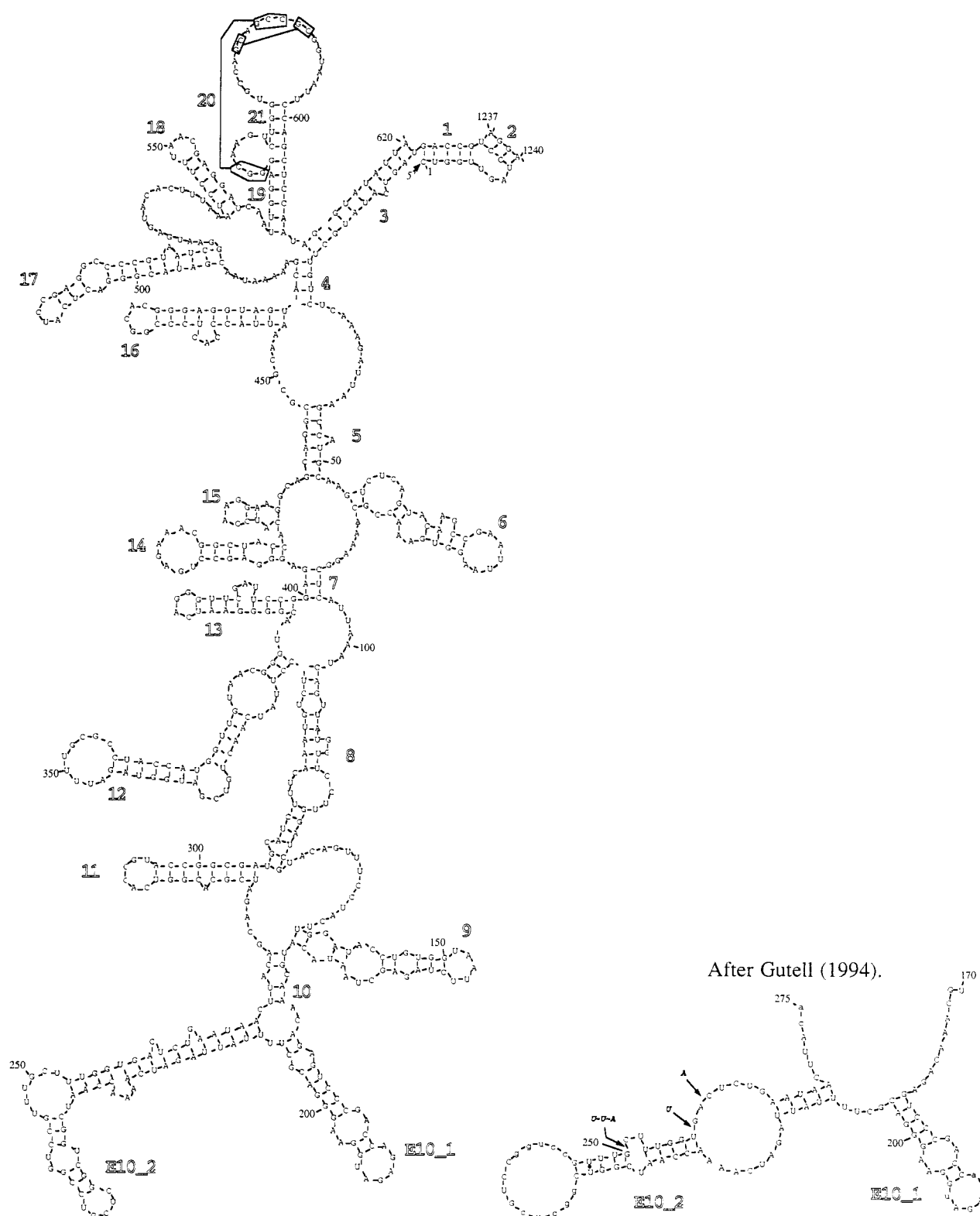


FIG. 3. Proposed secondary structure model of 18S rRNA of *Hackeriella veitchi*, Part I—Helices 1 to 21.

helix with only 3 base pairings. These pairings are conserved throughout eukaryotes. (8) The base pairing distal to the second internal bulge of this helix shows a

semicompensatory mutation replacing U · A in *T. molitor* by U · G in the peloridiids. The last base pairing is also a semicompensatory mutation replacing C · G in *T.*

molitor by U · G in the peloridiids. (9) Only one base substitutional difference between *T. molitor* and the peloridiids occurs in this helix. An A (position 147, *T. molitor*) is replaced by C (144). This change has no effect on the structural configuration because this site is in an internal bulge. (10) This helix comprises one of the more variable regions of eukaryotic 18S rRNA. The structure proposed for *T. molitor* by Hendriks *et al.* (1988b) cannot be reconstructed using the peloridiid sequences. Only the thermodynamic approach is available to infer secondary structure until there are more insect sequences to take into consideration. Calculations using the Zuker–Steigler algorithm yields helix 10 and two subhelicies enumerated as E10_1 and E10_2. Free-energy minimization also predicts similar two-subhelicie models in other hemipteran 18S rRNAs, *O. utahensis*, *P. spumarius*, and *L. hesperus*, examined by us. A recently available model of this region (van de Peer *et al.*, 2000) shows only 2 base pair differences compared with the model for *D. melanogaster* (Gutell, 1994). However, folding *H. veitchi* 18S RNA following the Gutell model ruptures a number of pairings and fails the “comparative sequence analysis” test. Thus, we opted to retain our more stably folded configuration. For comparison, helices E10_1 and E10_2 in Gutell’s model are provided (Fig. 3). (11) The terminal bulge of this hairpin loop is composed of five bases in *T. molitor* and six bases in the peloridiids. The character states of the bases in the loop are different in the two insects but the same overall configuration for this helix is the same in these insects. This indicates that constraints are less important in bulges than in stems. (12) The basal base pairings of this structure in peloridiids differ from the homologous helix of *T. molitor*, but, the predicted structures in the 12 terminal pairings of the hairpin loop, the lateral bulge, and the terminal loop are the same for both taxa. The sequences of these taxa in this portion of the molecule differ by only two base changes in the terminal loop and do not affect secondary structure. (13–15) These helices are conserved regions among all eukaryotes (Gutell, 1994) and retain the same configuration in the peloridiids. (16) The absence of a G at position 478 in the peloridiids, as opposed to its presence in *T. molitor*, transforms an internal bulge to a lateral bulge in the peloridiids. (17) No ambiguities exist in this helix since both thermodynamic results and comparative sequence analysis predict the same structures. (18) A congruent structure using the two folding techniques was not found. The structure with the least energy was retained. Only the final hairpin tetraloop UAAC and 4 pairings in the stem could be constructed in common. (19) The retained structure is the same as that predicted for the basal portion of helix 19 of *T. molitor* (Hendriks *et al.*, 1988b). (20) The retention of this pseudoknot has been experimentally confirmed (Gutell *et al.*, 1994). (21) The 4 pairs comprising this stem are presented by Hendriks *et al.* (1988b) in the apical part of what is defined as helix 19.

(Another pseudoknot has been described as being located here, but has not been numbered; Gutell *et al.*, 1994).

Helices 22 to 29 (Fig. 4, 18S rRNA model Part II).

(22) This helix consists of only 3 paired bases as predicted by Gutell (1994). The bases proposed to form helix 20 in the *T. molitor* model (Hendriks *et al.*, 1988b) are now placed in helix 26 of the peloridiid model. (23) The structural configuration of this helix was inferred from thermodynamics alone. No consensus secondary structure was available from known sequences of other eukaryotes. (E23) This helix and subhelicies generally incorporate the longest expansion region of eukaryotic 18S rRNA. Our model of the helix was constructed using mfold. Refinements of the mfold algorithm have improved its efficacy in rRNA folding, particularly for sequences longer than 50 bases. This particular structure is composed of eight subhelicies designated in our model as E23_1 to E23_8. (24) The predicted secondary structures of this helix for the peloridiids and *T. molitor* are identical. This helix possesses a number of compensatory substitutions that signify structural integrity of the model. There is one compensatory substitution at the base of the large internal bulge, three semicomplementary substitutions, and one reinforcement of a noncanonical A · C base pair by a canonical A · U base pair (Fig. 6). Three substitutional differences occur in the large internal bulge that do not affect folding configuration. (25 and 26) Thermodynamic folding of peloridiid sequences and comparison with *T. molitor* results in the same structural configuration for these two helices. Only two substitutional differences occur in unpaired bases of these helices in these taxa. (27) This retained structure is the most stable, with a total secondary structure energy of –16 kcal/mol. (28–30) These three helices were folded using MacDNASIS and are more stable than those proposed for *T. molitor* by Hendriks *et al.* (1988b). (31) This helix was folded following the instructions of Leontis and Westhof (1998), who isolated a recursive motif in rRNAs using supportive information from chemical probes. According to these authors, helix 31 contains 2 noncanonical A · A base pairs.

Helices 32 to 50 (Fig. 5, 18S rRNA model Part III).

(32) The configuration of this helix is energetically robust and concurs with secondary structure models proposed by Gutell *et al.* (1994) and Hendriks *et al.* (1988b). (33) In peloridiids this helix is 2 base pairs shorter than the *T. molitor* model. The reduced number of base pairs lowers the degree of thermodynamic support of the longer helix. A shorter helix is also inferred in the Gutell model (1994) for *D. melanogaster*. (34) This helix conforms with the proposed model for *T. molitor* by Hendriks *et al.* (1988b). (35) The nucleotides in this helix show no variation between the taxa used in our study and a comparative sequence analysis was not possible. As such, the most thermodynamically sta-

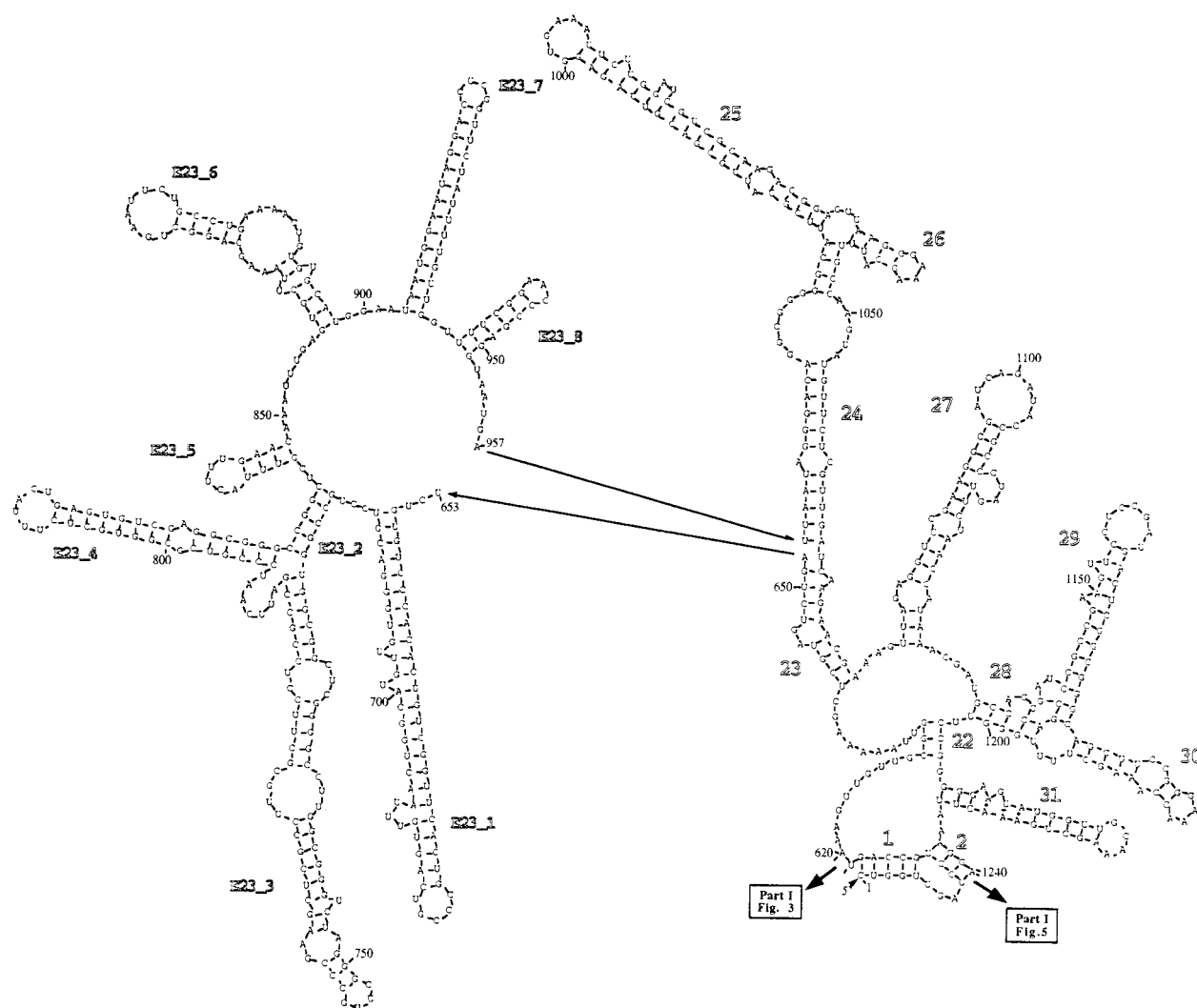


FIG. 4. Proposed secondary structure model of 18S rRNA of *Hackeriella veitchi*, Part II—Helices 22 to 31.

ble structure is retained. (36) This structural configuration is as in Hendriks *et al.* (1988b), with the exception of the distal base pair in the stem. There is a semicompensatory substitution; a G · U (1273 · 1566) in *T. molitor* is replaced by a G · C (1311 · 1563) in the peloridiids. (37) This helix is constructed according to the *T. molitor* model. A compensatory substitution in this helix, a 4th pair G · C in the main stem of the peloridiids, is replaced by A · U in *T. molitor*. The one semicompensatory substitution is represented by the 10th pair G · U in the main stem replaced by G · C in *T. molitor* (Fig. 5). (38) The retained helix is similar to the structure proposed for *T. molitor*, but, a C · G (1372 · 1548) is constrained to pair in our model, making it more thermodynamically stable than the model for *T. molitor* (−6.6 kcal/mol instead of −3.1 kcal/mol, respectively). (39–41) Folding of this region using the Zuker–Steigler algorithm created one hairpin loop. Since this result did not agree with the three-helice configuration proposed for *Drosophila* (Gutell, 1994)

and *T. molitor* (Hendriks *et al.*, 1988b) models, the more thermodynamically stable configuration proposed for *T. molitor* was retained (−2.8 kcal/mol compared to −0.3 kcal/mol). (42) This short helix consists of only 3 paired bases as in Hendriks *et al.* (1988b). (43) This helix is a component of an expansion region and is highly variable among insects. As such, we relied on thermodynamic folding to predict its configuration. (44) This helix was folded in concert with helix 41 using thermodynamic folding. (45 and 46) There are four substitutional differences distinguishing the peloridiids and *T. molitor* in this region, but, these changes are found in bulges and do not alter configuration. (47) Our calculations for this helix (−11.5 kcal/mol) support the same optimal structure as in Hendriks *et al.* (1988b). (48) This helix possesses one compensatory substitution, a C · G (1708 · 1726) in *T. molitor* replaced by A · U (1701 · 1719) in the peloridiids. The helix is 4 base pairs longer in the proximal portion of the stem of the peloridiids than that proposed for *T. molitor* by Hen-

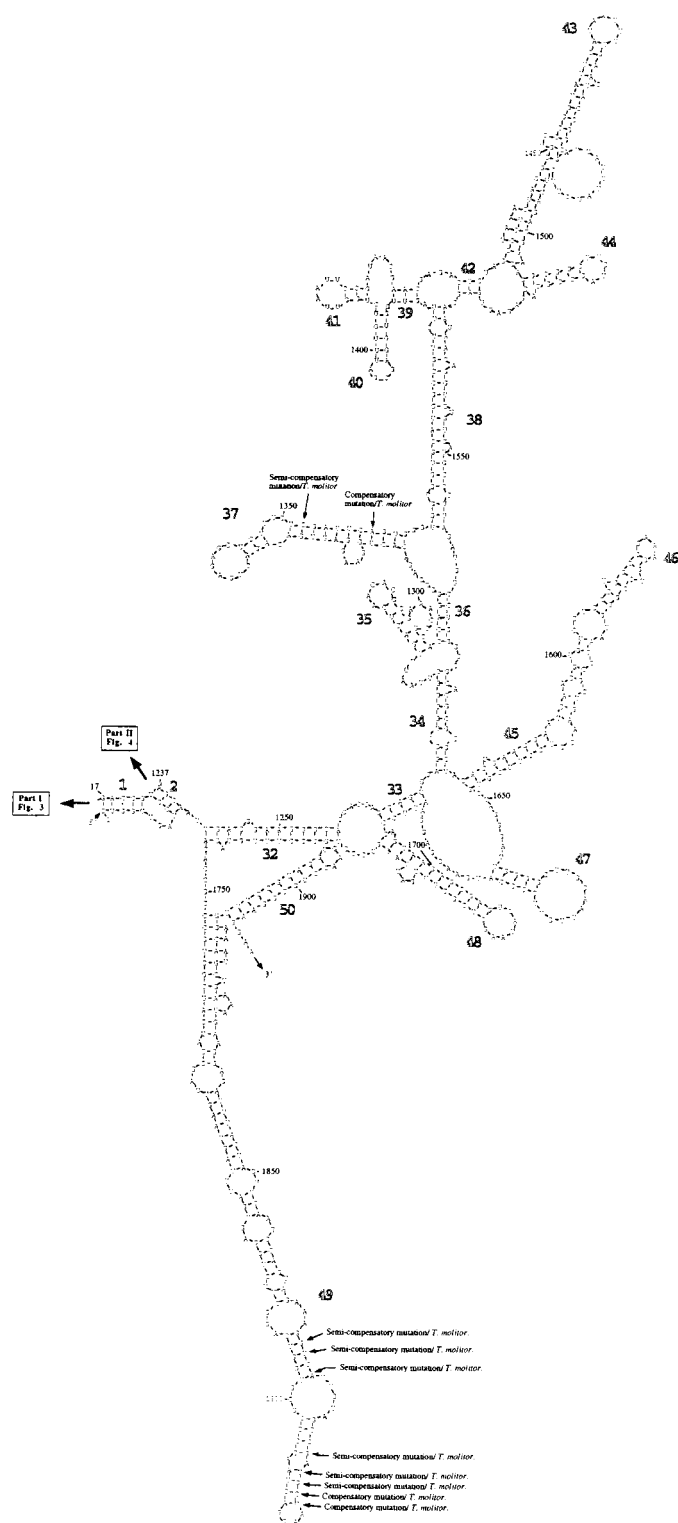


FIG. 5. Proposed secondary structure model of 18S rRNA of *Hackeriella veitchi*, Part III—Helices 32 to 50.

driks *et al.* (1988b). (49) Thermodynamic folding in this region infers one long hairpin loop with 131 base pairs, the longest helix of the entire 18S rRNA model. Variation between the sequences of *T. molitor* and the

peloridiids is most evident in the distal portion of this helix. As such, the proposed models differ. The configuration proposed here is supported, however, by two compensatory and six semicompensatory substitutions (Fig. 5). Seven other substitutional changes between these taxa do not affect our proposed structure since all occur in bulges and terminal loops. If helix 49 proposed by Hendriks *et al.* (1988b) for *T. molitor* is constrained to fit the model based on the peloridiid, then the number of compensatory and semicompensatory substitutions are the same. Six of these substitutions do not alter configuration, but two noncompensatory substitutions nullify the structure for *T. molitor*. (50) This extremity of the 18S rRNA gene is conserved among eukaryotes (Gutell, 1994) and our thermodynamic-based results are the same as those published for the *T. molitor* model.

Phylogenetic Implications

Procurement of full 18S rRNA sequences of peloridiids provided a means to assess phylogenetic affiliations of Coleorrhyncha and robustness of support for Neohemiptera. A previous study, using partial 18S rRNA sequences (Campbell *et al.*, 1995), inferred two most-parsimonious trees placing Prosorrhyncha either sister to Archaeorrhyncha (fulgoromorphs), forming clade Neohemiptera, or sister to Clypeorrhyncha (cicadomorphs). Similar results were reached by Bourgoin *et al.* (1997) but with clade Neohemiptera and Auchenorrhyncha equally supported, according to the outgroup. These studies did not include an 18S rDNA sequence of a peloridiid long enough to reveal the character state at the singular neohemipteran synapomorphic site found. In the full peloridiid sequences, base position 1439 is the homologous site of this neohemipteran synapomorphy. The character state of the peloridiids at this site is an A, as in other neohemipteran insects, as opposed to a G in nonneohemipteran insects. This synapomorphy

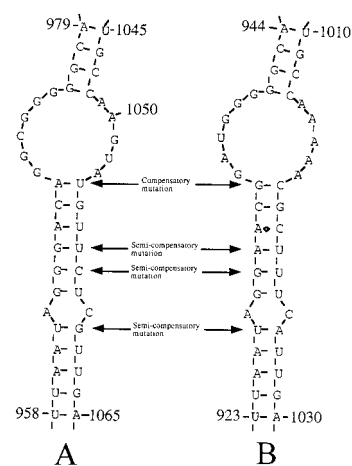


FIG. 6. Helix 24 showing positions of compensatory substitutions of bases. (A) *Hackeriella veitchi*. (B) *Tenebrio molitor*.

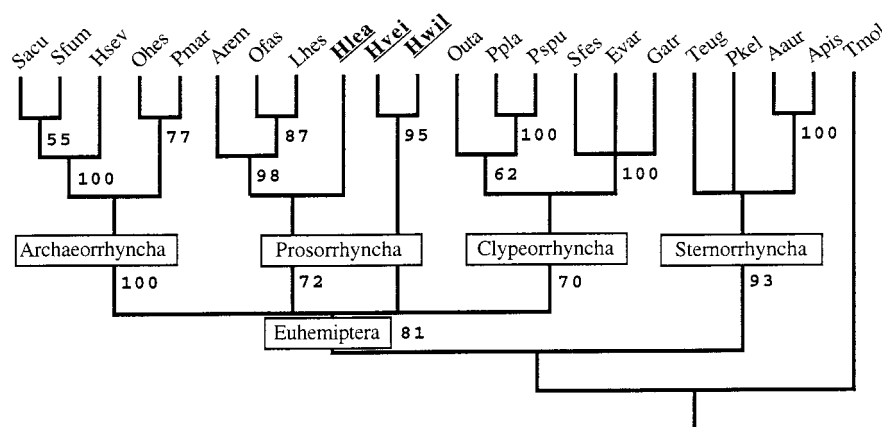


FIG. 7. Majority-rule consensus tree of full dataset based on an heuristic analysis (1000 bootstrap replicates) of informative sites. (Length = 927; CI = 0.540; HI = 0.460; RI = 0.530; RC = 0.287). Taxa abbreviations: Archaeorrhyncha: Sacu, *Siphanta acuta* (Flatidae, Accession No. U06476); Hsev, *Hysteropterum severini* (Issidae, U15214); Sfum, *Scolops fumida* (Dictyopharidae, U15216); Ohes, *Oliarus hesperinus* (Cixiidae, U15215); Pmar, *Prokelisia marginata* (Delphacidae, U09207); Prosorrhyncha: Aren, *Aquarius remigis* (Gerridae, U15691); Ofas, *Oncopeltus fasciatus* (Lygaeidae, U15188); Coleorrhyncha (boldface, underlined) Hlea, *Hemiodocus leai* (Peloridiidae) (partial sequence from Wheeler *et al.*, 1993); Hvei, *Hackeriella veitchi* (Peloridiidae, AF004766); Hwil, *Hemiwoodwardia wilsoni* (Peloridiidae, AF131198); Clypeorrhyncha: Outa, *Okanagana utahensis* (Cicadidae, U06478); Ppla, *Prosapia plagiata* (Cercopidae, U16264); Pspu, *Philaenus spumarius* (Cercopidae, U06480); Sfes, *Spissistilus festinus* (Membracidae, U06477); Evar, *Euscelidius variegatus* (Cicadellidae, U15148); Gatr, *Graphocephala atropunctata* (Cicadellidae, U15213); Sternorrhyncha: Teug, *Trioza eugeniae* (Psyllidae, U06482); Pkel, *Pealius kelloggi* (Aleyrodidae, U06479); Aaur, *Aonidiella aurantii* (Diaspididae, U06475); Apis, *Acyrtosiphon pisum* (Aphididae, X62623); Outgroup: Tmol, *Tenebrio molitor*, X07801.

suggests, based on this limited molecular evidence, that peloridiids are members of Neohemiptera.

However, a cladistic analysis taking into account full 18S rRNA sequences is necessary to confirm congruency of homologous and independent characters for inferring evolutionary affiliations among hemipteran taxa. Use of rRNA secondary structure enables assessment of such character independence. For example, presence of compensatory mutations shows that character state of a base may be dependent on the character state of a site elsewhere in the sequence. This issue of character independence is particularly striking in helix 11. In this helix unpaired bases in the terminal bulge are quite variable between *T. molitor* and peloridiids, whereas the primary sequences of these taxa are conserved in its stem. Some authors suggest treating bases differently in datasets. Bases in paired (stems) and unpaired sites could be differentiated by excluding one type (Wheeler and Honeycutt, 1988) or distinguishing them in a matrix to weigh one set of characters (Smith, 1989). Intuitively, it could be argued that paired nucleotides forming stems could be considered as candidates for weighting by one-half. Such nucleotides are interdependant characters, as shown through compensatory mutations, and should not be assessed as possessing as strong a phylogenetic signal as independent characters. However, to score such characters with half a weight *a priori* is not satisfactory because the interdependence does not necessarily render such paired nucleotides with equal weights. Evolutionary studies of transitional states in compensatory mutated sites might indicate that one of these sites should be

given a greater weight. The first mutation signals an important phylogenetic event, but a second compensatory mutation is necessary for maintaining the functional integrity of the ribosome. Hence, weighting the "responding" compensatory mutation more than the first mutation has justification.

Using secondary structure as a guide to initially align homologous regions of rRNAs between taxa, followed by use of parsimony and direct optimization (ALIGN and POY), has been suggested as one approach to minimize evolutionary events between respective nucleotide sequences or parts of sequences (Gladstein and Wheeler, 1996). However, employing such iterative programs as POY and ALIGN may contradict homology based on secondary structure, the homology hypotheses being totally different in the two approaches. Homology proposed by aligning sequences using secondary structure could be jeopardized by a new homology hypothesis using the principle of parsimony to explain variations between these sequences.

Cladistic Analysis

Phylogenetic analysis of the full dataset portrays a basal Sternorrhyncha sister to Euhemiptera (*sensu* Zrzavy, 1992), the latter represented as a quadratomous node (Fig. 7). Whereas clades Archaeorrhyncha and Clypeorrhyncha are monophyletic, Coleorrhyncha appears paraphyletic. There is little doubt that unique morphological features of peloridiids (Evans, 1963) establishes their monophyly. This anomalous paraphyly of Peloridiidae using the full dataset possibly originated from use of a published partial sequence of

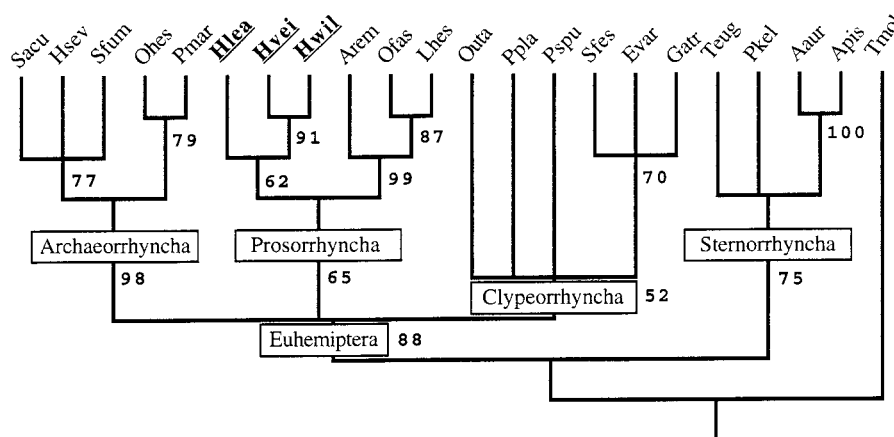


FIG. 8. Majority-rule consensus tree of attenuated dataset based on a branch-and-bound analysis (1000 bootstrap replicates) of informative sites. (Length = 182; CI = 0.599; HI = 0.401; RI = 0.678; RC = 0.406). Numbers at nodes represent level of boot strap support. Taxa abbreviations as in Fig. 7.

Hemiodoecus leai, a peloridiid sequence consisting of approx 450 bp (Wheeler *et al.*, 1993). The other peloridiid taxa were represented by full sequences. This anomaly and the fact that sections of the full data matrix could only be aligned ambiguously (based on a criterion of sequence similarity) indicated that an analysis using a dataset of unambiguously aligned and conclusively homologous sites was important.

Next, sites from the data matrix that were ambiguously aligned, because of either extensive homoplasy between all taxa (saturation) or presence of indels, were excluded. This treatment resulted in an attenuated dataset. The analysis of this dataset rendered a consensus tree with four distinct hemipteran lineages (Fig. 8). Sternorrhyncha was inferred sister to Euhemiptera that was represented as a trichotomy of archaeorrhynchan, prosorrhynchan, and clypeorrhynchan lineages. The peloridiids were grouped in a clade (Coleorrhyncha) sister to Heteroptera, thus forming clade Prosorrhyncha. The trichotomous euhemipteran node was still unresolved.

To resolve the trichotomous node, homoplasious sites among outgroup taxa (relative to euhemipterans) were removed. This was achieved by first excluding sites having different character states in the psyllid (the basal sternorrhynchan lineage) and *T. molitor*. Then, sites homoplasious within Sternorrhyncha (the most basal hemipteran lineage) were excluded from the data matrix. These filtrations of characters resulted in retention of only symplesiomorphic sites of a theoretical basal lineage represented by the psyllid as outgroup to euhemipterans. This dataset was termed the polarized dataset. Analysis of this dataset resulted in six most-parsimonious trees (Fig. 9A). Half of these trees supported Neohemiptera and half supported an arrangement in which Prosorrhyncha and Clypeorrhyncha were sisters; the consensus being an unresolved phylogeny for Euhemiptera (Fig. 9B). To assess nodal de-

cay (Bremer, 1994), trees one and two steps longer than the most-parsimonious trees were evaluated. A consensus of these 3528 longer trees did not support a monophyletic Neohemiptera, but recognized a Prosorrhyncha + Clypeorrhyncha clade (Fig. 9C). Interestingly, such a prosorrhynchan + clypeorrhynchan arrangement agrees with Bourgoin *et al.* (1999) and with paleontological interpretations proposed by Shcherbakov (1988).

A monophyletic Prosorrhyncha was supported in all trees and was represented in the data matrix by two synapomorphic sites. These synapomorphies are actually compensatory substitutions in the stem of helix 37 as an A · U (1321 · 1358) in Prosorrhyncha compared to a G · C in other insects. Whereas Neohemiptera has one supporting synapomorphic site in a short conserved portion of the 18S rRNA molecule (and retained in the polarized dataset), the potential affiliation of Prosorrhyncha + Clypeorrhyncha is supported by two synapomorphic sites retained in the full dataset. However, these sites were excluded from the other datasets because they were bracketed by sections of the sequence that were ambiguously aligned. The two sites are paired in the stem of helix E23_1 as an A · U (663 · 706) in the thermodynamically computed model of the variable E23 region. These homologous positions are replaced by a G · C in other insects (except in *A. pisum* where they are an A · C). The only synapomorphy supporting clade Neohemiptera is situated in a lateral bulge in helix 43 (site 1439, Fig. 5).

At this point, affiliations of the three clades of Euhemiptera are unresolved, despite increasing number of taxa and incorporating full peloridiid sequences. Construction of a new matrix taking into account secondary structures of all Hemiptera could resolve this trichotomy (T. Bourgoin *et al.*, unpublished). Such structurally based matrices should use models constructed with recent biochemical and molecular biolog-

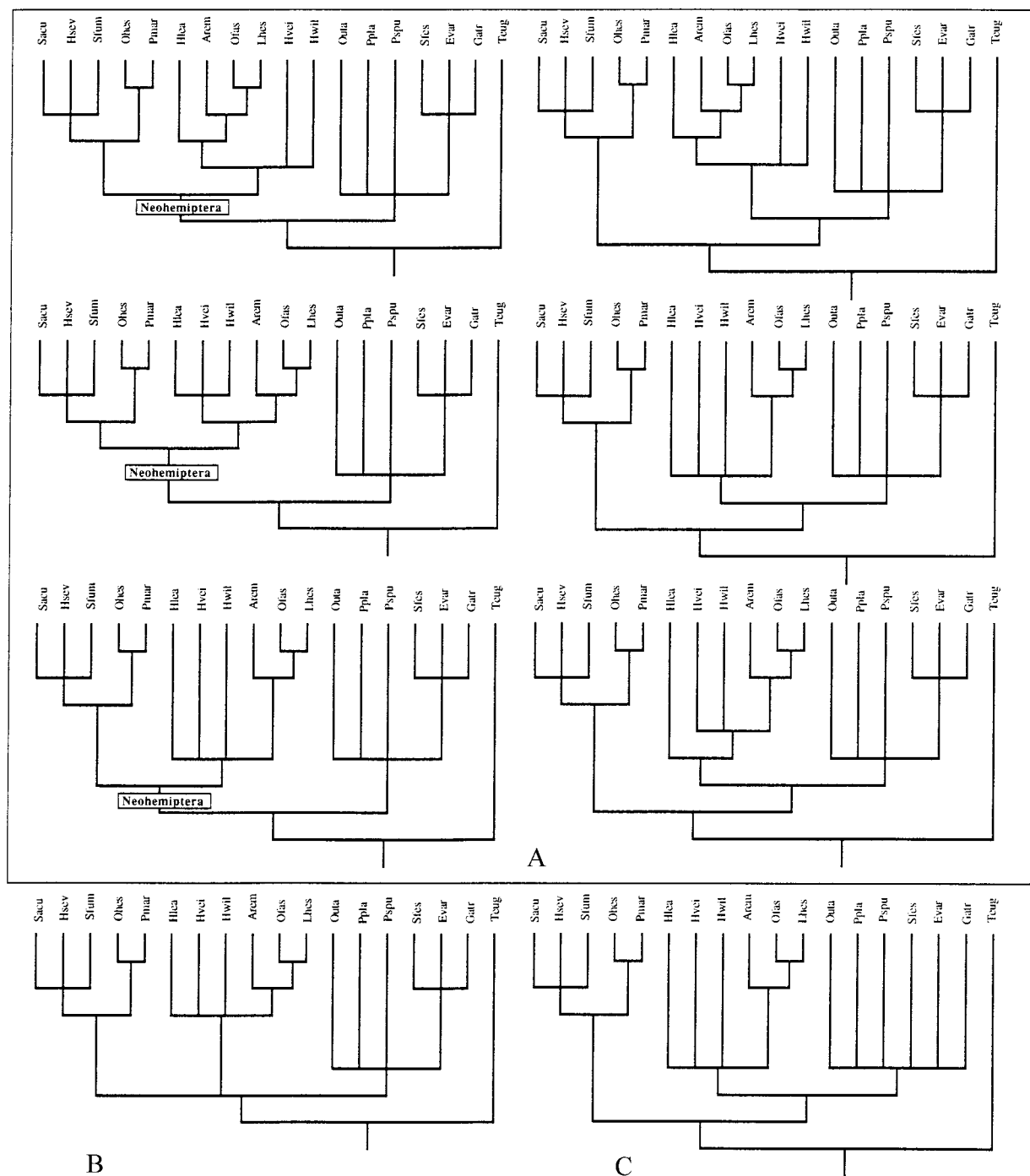


FIG. 9. Trees resulting from branch-and-bound analysis of the polarized hemipteran data matrix. (A) Six most-parsimonious trees showing equal inference for Neohemiptera (Prosothymus + Archaeothymus) and Prosothymus + Clypeothymus. (Length = 35; CI = 0.800; HI = 0.200; RI = 0.905; RC = 0.724). (B) Majority-rule consensus of the six trees. (C) Majority-rule consensus of 3528 trees one and two steps longer than the most-parsimonious trees.

ical data and not simply primary sequences aligned using rather limited algorithms (De Rijk and De Wachter, 1993). Secondary structure models should be broadly applied in a hemipteran data matrix to avoid arbitrary loss of phylogenetic (nonhomoplasious) information, as was possibly the case in our efforts to purge

disorder in the attenuated and polarized data matrices. The method employed by Sorensen *et al.* (1995) and applied in Campbell *et al.* (1995), Bourgoin *et al.* (1997), and in this study eliminated any biases introduced from ambiguously aligned regions in the full dataset rendering an attenuated dataset. However,

construction of the polarized dataset from the attenuated dataset using only symplesiomorphic sites in the outgroup led to a “generalized hypothetical Sternorrhyncha polarization” (or “sternorrhynchan archetype polarization”). Such a polarization cannot define a strict outgroup comparison according to Hennig (1966).

Proposed Model

Use of thermodynamic algorithms alone to infer secondary structure can result in erroneous interpretations. However, this method has the advantage of being rapid, with duration of analysis limited only by CPU speed to run the algorithms. To achieve improvements, phylogenetic, molecular biological, and physical chemistry data must be further assessed. An improvement in these algorithms is already recognizable in comparing the Zuker–Steigler algorithm (1981) with upgraded versions incorporating the latest refinements on the mfold server. We found the recent mfold algorithms to agree most frequently with those from comparative sequence analysis. As suggested by Gutell (1993), a “phylogenetic events” factor should be incorporated into folding algorithms so that base pairing can be manually constrained, or conversely remain unpaired in bulges or terminal loops, to enable phylogenetic comparisons.

However, examination of compensatory substitutions can furnish fundamental information on several helices. Compensatory substitutions are directed changes that reflect constraints applied to the two- or three-dimensional configuration of the 18S rRNA molecule. Phylogenetic studies overlooking secondary structure potentially lose valuable phylogenetic information inherent to the structural constraints of the molecule. Comparison of primary sequences might reveal portions undergoing a high rate of evolution, but preservation of a particular helix in such regions would show that functional constraints exist. For example, when helix 48 is compared among five arthropods there are five compensatory mutations in the 11 pairs forming its stem (Fig. 10). Of course, number of compensatory mutations would increase if a larger number of phylogenetically distant taxa were added, but, comparison of more closely related taxa furnishes better evidence for confirming helix structure because risk of comparing nonhomologous events is reduced.

The evolutionary dynamics of accumulating compensatory mutations is believed to first involve a noncompensatory mutation that may become fixed in a population through molecular drive (depending on number of deleterious changes in copies of the molecule). This is followed, at some later time, by a selective compensatory substitution that restores robustness of a structure and, therefore, fitness of descendants. Hancock *et al.* (1988) consider that hydrogen bonding between G · U pairs, while suboptimal, maintains a structure at an intermediate level prior to an eventual compensatory mutation. Rousset *et al.* (1991) show that G · U pairs

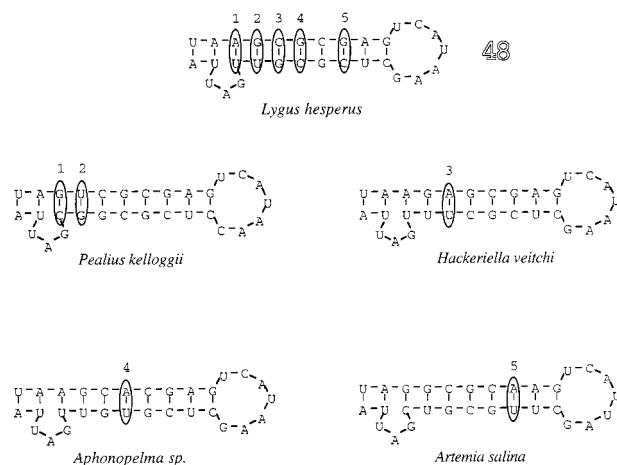


FIG. 10. Helix 48 showing positions of five different compensatory mutations among hemipterans and a crustacean.

are deleterious in some positions, but also may be advantageous in other positions where they become fixed and no compensatory mutation is required to transform a G · U to an A · U pairing. If G · U pairings did not exist in several different intermediate lineages, then it is likely that lack of a compensatory mutation was not a selective disadvantage. At the species level certain helices may be constant in length, number, and position of bulges along a stem and length of terminal loop.

Our analysis of 18S rRNA secondary structure using peloridiid sequences reflects the fact that stabilizing factors for the molecule are not completely known. The molecular biological environment in the ribosome probably contributes to configuration of ribosomal RNAs in some yet to be learned fashion and there are likely subtle differences of the *in vivo* structure and one predicted by current least free-energy algorithms. For example, intertaxon variability of helices 6, E_10, E_23, and 43 suggests that there is no widely shared structural conformation of these helices. Closely related taxa might have sequences in these helices that are incompatible to folding in the same configuration. This absence of overall similarity could indicate presence of specific features that are recognition sites for associated proteins or RNA–RNA interactions. This was recently demonstrated in helix 23 of prokaryotes for which *in vivo* structural variation reflected the flexibility of the helix to interact with ribosomal proteins S8 and S17 to maintain functional integrity (Clemons *et al.*, 1999). Thus, such variable helices might result in structural configurations that are not necessarily congruent with phylogeny. Studies on translation processing could identify key zones maintaining normal molecular biological functions of the ribosome that reflect evolutionary patterns. A recent report showed how functionality of secondary structure in variable helices of small subunit ribosomal RNA in prokaryotes is af-

affected by tertiary forces (Garrett, 1999; Clemons *et al.*, 1999). Such associations between functionality and molecular forces could further impact our philosophy of character treatment in rRNA datasets.

Finally, a potential field of interest concerns use of some structures to characterize and distinguish groups of taxa: these structures are "signatures" as defined in Winker and Woese (1991). The 18S rRNAs of several insects show a morpho-molecular feature in helix 24 which could be interpreted as a synapomorphy for the Fulgoromorpha (Archaeorrhyncha). A mispairing after the third pair is present in the five studied species of fulgones belonging to five different families, but is replaced by a conventional base pair in other Hemiptera studied. Such homologous structures are numerous enough in the long rRNA molecule to show variation and may be used as independent characters in a cladistic analysis.

CONCLUSION

When inferring evolutionary affiliations using molecular phylogenetics, it is important to recognize particular structural characteristics of molecules if these molecules are associated with significant functional processes. The results presented in this paper emphasize the importance of taking into account molecular changes that affect folding of 18S rRNA and possibly its function. Appreciation of such factors that affect folding, and formation of various helices and their substructures, improves ability to infer homology of character, whether in a broad sense as a particular structure or acutely as individual nucleotides. In addition, by taking into account particular aspects of secondary and tertiary structures, one is in a better position to visualize where distorted weighting of a phylogenetic signal might have occurred within a dataset. Moreover, one could use structural homology of features within substructures of the molecule as characters, similar to a morphological treatment, and analyze this dataset independently or in conjunction with a nucleotide dataset. The next step following such assessment of secondary structural elements for the current hemipteran dataset is translating these secondary structural elements into characters for a cladistic analysis to provide more robustness of phylogenetic signals (T. Bourgoin *et al.*, unpublished).

ACKNOWLEDGMENTS

We greatly appreciate the efforts of G. B. and S. R. Monteith, Queensland Museum, Brisbane, Queensland, Australia in collecting and providing specimens of the peloridiids used in this study. The authors thank C. D'Haese for providing documentation on MALIGN and POY software and for useful discussion and K. Kjer and an anonymous reviewer for their thorough and helpful reviews of the manuscript.

REFERENCES

- Bourgoin, T. (1993). Female genitalia in Hemiptera Fulgoromorpha, morphological and phylogenetic data. *Ann. Soc. Entomol. Fr. (N.S.)* **29**: 225–244.
- Bourgoin, T. (1996). Phylogénie des hexapodes. La recherche des synapomorphies ne fait pas toujours le cladisme! *Bull. Soc. Zool. Fr.* **121**: 5–20.
- Bourgoin, T., Steffen-Campbell, J. D., and Campbell, B. C. (1997). Molecular phylogeny of Fulgoromorpha (Insecta, Hemiptera, Archaeorrhyncha). The enigmatic Tettigometridae: Evolutionary affiliations and historical biogeography. *Cladistics* **13**: 207–224.
- Bourgoin, T., Chan, K., Ouvrard, D., and Campbell, B. C. (1999). Character Analysis Using Ribosomal RNA Secondary Structure and Its Application to Hemiptera and Fulgoromorpha Phylogeny. 10th International Auchenorrhyncha Congress, Cardiff.
- Bredden, G. (1897). Hemipteren. In "Ergebnisse der Hamburger Magalhaensischen Sammelreise 1982/93. II. Band. Arthropoden" (Naturhistorischen Museum zu Hamburg, Ed.), pp. 10–13. Friedrichsen, Hamburg.
- Bremer, K. (1994). Branch support and tree stability. *Cladistics* **10**: 295–304.
- Burckhardt, D., and Agosti, D. (1991). New records of South American Peloridiidae (Homoptera: Coleorrhyncha). *Rev. Chilena Entomol.* **19**: 71–75.
- Campbell, B. C., Steffen-Campbell, J. D., Sorensen, J. T., and Gill, R. J. (1995). Paraphyly of Homoptera and Auchenorrhyncha inferred from 18S rDNA nucleotide sequences. *Syst. Entomol.* **20**: 175–194.
- Chastain, M., and Tinoco, I., Jr. (1991). Structural elements in RNA. *Prog. Nucleic Acid Res. Mol. Biol.* **41**: 131–177.
- China, W. E. (1962). South American Peloridiidae (Hemiptera–Homoptera: Coleorrhyncha). *Trans. R. Entomol. Soc. Lond.* **114**: 131–161.
- Clemons, W. M., Jr., May, J. L. C., Wimberly, B. T., McCutcheon, J. P., Capel, M. S., and Ramakrishnan, V. (1999). Structure of a bacterial 30S ribosomal subunit at 5.5 Å resolution. *Nature* **400**: 833–840.
- Cobben, R. H. (1978). Evolutionary trends in Heteroptera. Part II. Mouthpart-structures and feeding strategies. *Mededelingen Landbouwhogeschool Wageningen Nederland* 78–5, 407 pp.
- Corpet, F., and Michot, B. (1994). RNAlign program: Alignment of RNA sequences using both primary and secondary structures. *Comput. Appl. Biosci.* **10**: 389–399.
- De Pinna, M. C. (1991). Concepts and tests of homology in the cladistic paradigm. *Cladistics* **7**: 367–394.
- De Rijk, P., and De Wachter, R. (1993). DCSE, an interactive tool for sequence alignment and secondary structure research. *Comput. Appl. Biosci.* **9**: 735–740.
- Evans, J. W. (1963). The phylogeny of Homoptera. *Annu. Rev. Entomol.* **8**: 77–94.
- Evans, J. W. (1981). A review of present knowledge of the family Peloridiidae and new genera and new species from New Zealand and New Caledonia (Hemiptera: Insecta). *Rec. Austral. Mus.* **34**: 381–406.
- Fields, D. S., and Gutell, R. R. (1996). An analysis of large RNA sequences folded by a thermodynamic method. *Folding Design* **1**: 419–430.
- Garrett, R. (1999). Mechanics of the ribosome. *Nature* **400**: 811–812.
- Gladstein, D. S., and Wheeler, W. C. (1997). "POY: Phylogeny Reconstruction via Direct Optimization," anon ftp.amnh.org/pub/wheeler/programs/poy. American Museum of Natural History, New York.
- Gutell, R. R. (1993). Comparative studies of RNA: Inferring higher-

- order structure from patterns of sequence variation. *Curr. Opin. Struct. Biol.* **3**: 313–322.
- Gutell, R. R. (1994). Collection of small subunit (16S- and 16S-like) ribosomal RNA structures. *Nucleic Acids Res.* **22**: 3502–3507.
- Gutell, R. R., Larsen, N., and Woese, C. R. (1994). Lessons from an evolving RNA: 16S and 23S RNA structures from a comparative perspective. *Microbiol. Rev.* **58**: 10–26.
- Hancock, J. M., Tautz, D., and Dover, G. A. (1988). Evolution of the secondary structures and compensatory mutations of the ribosomal RNAs of *Drosophila melanogaster*. *Mol. Biol. Evol.* **5**: 393–414.
- Hendriks, L., Van Broeckhoven, C., Vandenberghe, A., Van De Peer, Y., and De Wachter, R. (1988a). Primary and secondary structure of the 18S ribosomal RNA of the bird spider *Eurypelma californica* and evolutionary relationships among eukaryotic phyla. *Eur. J. Biochem.* **177**: 15–20.
- Hendriks, L., De Baere, R., Van Broeckhoven, C., and De Wachter, R. (1988b). Primary and secondary structure of the 18S ribosomal RNA of the insect species *Tenebrio molitor*. *FEBS Lett.* **232**: 115–120.
- Hennig, W. (1966). "Phylogenetic Systematics," Univ. of Illinois Press, Urbana.
- Jaeger, J. A., Turner, D. H., and Zuker, M. (1989). Improved predictions of secondary structures for RNA. *Proc. Natl. Acad. Sci. USA* **86**: 7706–7710.
- 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.
- Konings, D. A. M., and Hogeweg, P. (1989). Pattern analysis of RNA secondary structure: Similarity and consensus of minimal-energy folding. *J. Mol. Biol.* **207**: 597–614.
- Konings, D. A. M., and Gutell, R. R. (1995). A comparison of thermodynamic foldings with comparatively derived structures of 16S and 16S-like rRNAs. *RNA* **1**: 559–574.
- Kwon, O.-Y., Ogino, K., and Ishikawa, H. (1991). The longest 18S ribosomal RNA ever known: Nucleotide sequence and presumed secondary structure of the 18S RNA of the pea aphid, *Acyrtosiphon pisum*. *Eur. J. Biochem.* **202**: 827–833.
- Leontis, N. B., and Westhof, E. (1998). A common motif organizes the structure of multi-helix loops in 16S and 23S ribosomal RNAs. *J. Mol. Biol.* **283**: 571–583.
- Mathews, D. H., Sabina, J., Zuker, M., and Turner, D. H. (1999). Expanded sequence dependence of thermodynamic parameters provides robust prediction of RNA secondary structure. *J. Mol. Evol.* **288**: 911–940.
- Myers, J. G., and China, W. E. (1929). The systematic position of the Peloridiidae as elucidated by a further study of the external anatomy of *Hemiodocus leai* China. *Ann. Mag. Nat. Hist.* **105**: 282–294.
- Nelles, L., Fang, B.-L., Volckaert, G., Vandenberghe, A., and De Wachter, R. (1984). Nucleotide sequence of a crustacean 18S ribosomal RNA gene and secondary structure of eukaryotic small subunit ribosomal RNAs. *Nucleic Acids Res.* **12**: 8749–8768.
- Noller, H. F. (1984). Structure of the ribosomal RNA. *Annu. Rev. Biochem.* **53**: 119–162.
- Patterson, C. (1982). Morphological characters and homology. In "Problems of Phylogenetic Reconstruction" (K. A. Joysey and A. F. Friday, Eds.), pp. 21–74. Academic Press, London.
- Remane, A. (1956). "Die Grundlagen des Natürlichen Systems des Vergleichenden Anatomie und der Phylogenetik," Akademische Verlagsgesellschaft Geest & Portig, Leipzig.
- Rousset, F., Pélandakis, M., and Solignac, M. (1991). Evolution of compensatory substitutions through G · U intermediate state in *Drosophila* RNA. *Proc. Natl. Acad. Sci. USA* **88**: 10032–10036.
- Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487–491.
- Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**: 5436–5467.
- Schlee, D. (1969). Morphologie und Symbiose: Ihre Beweiskraft für die Verwandtschaftsbeziehungen der Coleorrhyncha. *Stuttg. Beitr. Naturk.* **210**: 1–27.
- Schuh, R. T., and Slater, J. A. (1995). "True Bugs of the World (Hemiptera: Heteroptera): Classification and Natural History," Cornell Univ. Press, Ithaca, NY.
- Shcherbakov, D. E. (1988). Origin and evolution of Auchenorrhyncha (Homoptera) based upon fossil evidence. *Proc. 28th Int. Congr. Entomol.*, p. 8.
- Smith, A. B. (1989). RNA sequence data in phylogenetic reconstruction: Testing the limits of its resolution. *Cladistics* **5**: 321–344.
- Sorensen, J. T., Campbell, B. C., Gill, R. J., and Steffen-Campbell, J. D. (1995). Non-monophyly of Auchenorrhyncha ("Homoptera"), based upon 18S rDNA phylogeny: Eco-evolutionary and cladistic implications within pre-Heteropteroidea Hemiptera (S. L.) and a proposal for new monophyletic suborders. *Pan-Pacific Entomol.* **71**: 31–60.
- Swofford, D. L. (1998). PAUP*. "Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4," Sinauer, Sunderland, MA.
- Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.
- van de Peer, Y., De Rijk, P., Wuyts, J., Winkelmans, T., and De Wachter, R. (2000). The European Small Subunit Ribosomal RNA database. *Nucleic Acids Res.* **28**: 175–176.
- von Dohlen, C. D., and Moran, N. A. (1995). Molecular phylogeny of the Homoptera: A paraphyletic Aascon. *J. Mol. Evol.*, **41**: 211–223.
- Wheeler, W. C., and Gladstein, D. S. (1994). MALIGN: A multiple sequence alignment program. *J. Hered.* **85**: 417.
- Wheeler, W. C., and Honeycutt, R. L. (1988). Paired sequence difference in ribosomal RNAs: Evolutionary and phylogenetic implications. *Mol. Biol. Evol.* **5**: 90–96.
- Wheeler, W. C., Schuh, R. T., and Bang, R. (1993). Cladistic relationships among higher groups of Heteroptera: Congruence between morphological and molecular data sets. *Entomol. Scand.* **24**: 121–137.
- Winker, S., and Woese, C. R. (1991). A Definition of the domains *Archaea*, *Bacteria* and *Eucarya* in terms of small subunit ribosomal RNA characteristics. *Syst. Appl. Microbiol.* **14**: 305–310.
- Woese, C. R. (1987). Bacterial evolution. *Microbiol. Rev.* **51**: 221–271.
- Zrzavy, J. (1992). Evolution of antennae and historical ecology of the hemipteran insects (Paraneoptera). *Acta. Entomol. Bohem.* **89**: 77–86.
- Zuker, M., Mathews, D. H., and Turner, D. H. (1999). Algorithms and thermodynamics for RNA secondary structure prediction: A practical guide. In "RNA Biochemistry and Biotechnology" (J. Barciszewski and B. F. C. Clark, Eds.), NATO ASI Series, Kluwer Academic, Dordrecht.
- Zuker, M., and Steigler, P. (1981). Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* **9**: 133–148.