Syst. Biol. 53(3):506–514, 2004

ISSN: 1063-5157 print / 1076-836X online
DOI: 10.1080/10635150490445922

Aligned 18S and Insect Phylogeny

KARL M. KJER

Department of Ecology Evolution and Natural Resources, 14 College Farm Road, Cook College, Rutgers University, New Brunswick, NJ 08901, USA; E-mail: kjer@aesop.rutgers.edu

Abstract.—The nuclear small subunit rRNA (18S) has played a dominant role in the estimation of relationships among insect orders from molecular data. In previous studies, 18S sequences have been aligned by unadjusted automated approaches (computer alignments that are not manually readjusted), most recently with direct optimization (simultaneous alignment and tree building using a program called "POY"). Parsimony has been the principal optimality criterion. Given the problems associated with the alignment of rRNA, and the recent availability of the doublet model for the analysis of covarying sites using Bayesian MCMC analysis, a different approach is called for in the analysis of these data. In this paper, nucleotide sequence data from the 18S small subunit rRNA gene of insects are aligned manually with reference to secondary structure, and analyzed under Bayesian phylogenetic methods with both GTR+I+G and doublet models in MrBayes. A credible phylogeny of Insecta is recovered that is independent of the morphological data and (unlike many other analyses of 18S in insects) not contradictory to traditional ideas of insect ordinal relationships based on morphology. Hexapoda, including Collembola, are monophyletic. Paraneoptera are the sister taxon to a monophyletic Holometabola but weakly supported. Ephemeroptera are supported as the sister taxon of Neoptera, and this result is interpreted with respect to the evolution of direct sperm transfer and the evolution of flight. Many other relationships are well-supported but several taxa remain problematic, e.g., there is virtually no support for relationships among orthopteroid orders. A website is made available that provides aligned 18S data in formats that include structural symbols and Nexus formats. [18S; alignment; doublet model; Insecta; Paleoptera.]

Conclusions from molecular data about the relationships among insect orders have been dominated by the nuclear small subunit rRNA (18S: Wheeler, 1989; Carmean et al., 1992; Pashley et al., 1993; Chalwatzis et al., 1996; Whiting et al., 1997; Wheeler et al., 2001). A fragment of the large subunit (28S; D3 region) has been included in several studies (Whiting et al., 1997; Wheeler et al., 2001; Hovmöller et al., 2002), but this fragment is small and ambiguous to align across Insecta, and the published relationships generated from analysis of the D3 alone are not reasonable. Mitochondrial genes, such as 16S and COI, have been examined in many insects, as has the nuclear EF-1α, but these markers have been shown to be homoplasious, even within orders (Flook and Rowell, 1997; Carapelli et al., 2000; Kjer et al., 2001; Misof et al., 2001; Johnson and Whiting, 2002). Analyses that include nuclear rRNA fragments combined with morphological data (Whiting et al., 1997; Wheeler et al., 2001) offer a reasonable picture of insect phylogeny, if one defines "reasonable" as recovering the most basic groups (Dicondylia, Pterygota, Neoptera, Holometabola; defined in Fig. 1), and the monophyly of most of the orders. However, the contribution of the 18S data to these studies has not been clearly established for several reasons. Wheeler et al. (2001) used direct optimization (POY; Wheeler, 1996; Gladstein and Wheeler, 1997) and analytical parameters that were based on minimizing character incongruence with the morphological data. However, POY does not produce an alignment and thus does not permit an assessment of homology nor allow the simultaneous visualization of character support of multiple nodes. Thus for nodes where the 18S had little to contribute, conclusions from the molecular partition would be highly dependent on the morphological data, and/or noise from the homoplasious data sets, and/or arbitrarily optimized homology of unalignable data. An analysis that optimized 18S alignment on character congruence with the D3 (Wheeler et al., 2001; their Fig. 12a) resulted in a phylogeny that no insect systematist would find credible (i.e., polyphyletic stoneflies, termites, neuropteroids; paraphyletic stick insects; *Agnetina* [a stonefly] + Diptera (flies); Zoraptera + Amphiesmenoptera [caddisflies and moths]). What does the 18S really have to say about the relationships among the non-holometabolous insects?

To address this question from an alternative perspective would require the presentation of a hypothesis of homology for each site in the 18S; in other words, an alignment. Direct optimization (POY) avoids fixed hypotheses of homology. It is true that there are cases in "unalignable regions" where homology statements across lineages are inappropriate because independent gains and losses may have occurred with such frequency that their history could only potentially be recovered with a dense taxon sample, reconstructing ancestral nodes on a tree. POY may be able to reconstruct homology pathways for these unalignable regions under ideal conditions (i.e., homogeneity of both gap-cost/change ratios at all sites and homogeneous nucleotide composition), but although such conditions may exist in some intron sequences, they are not typically found in the variable regions of rRNA, which is characterized by both among site rate variation, and nucleotide compositional heterogeneity. A different approach is taken here, with the specification of homology for each site in a structurally aligned data set that is publically available on the *Systematic Biology* website for adjustment and reassessment as additional data are collected. This alignment should clarify the contributions of individual characters of the 18S to insect phylogeny and facilitate the alignment of other insect 18S sequences. Structural alignment, although conceptually simple, is labor intensive. This template of structurally aligned data, with representatives of most insect orders, will be periodically

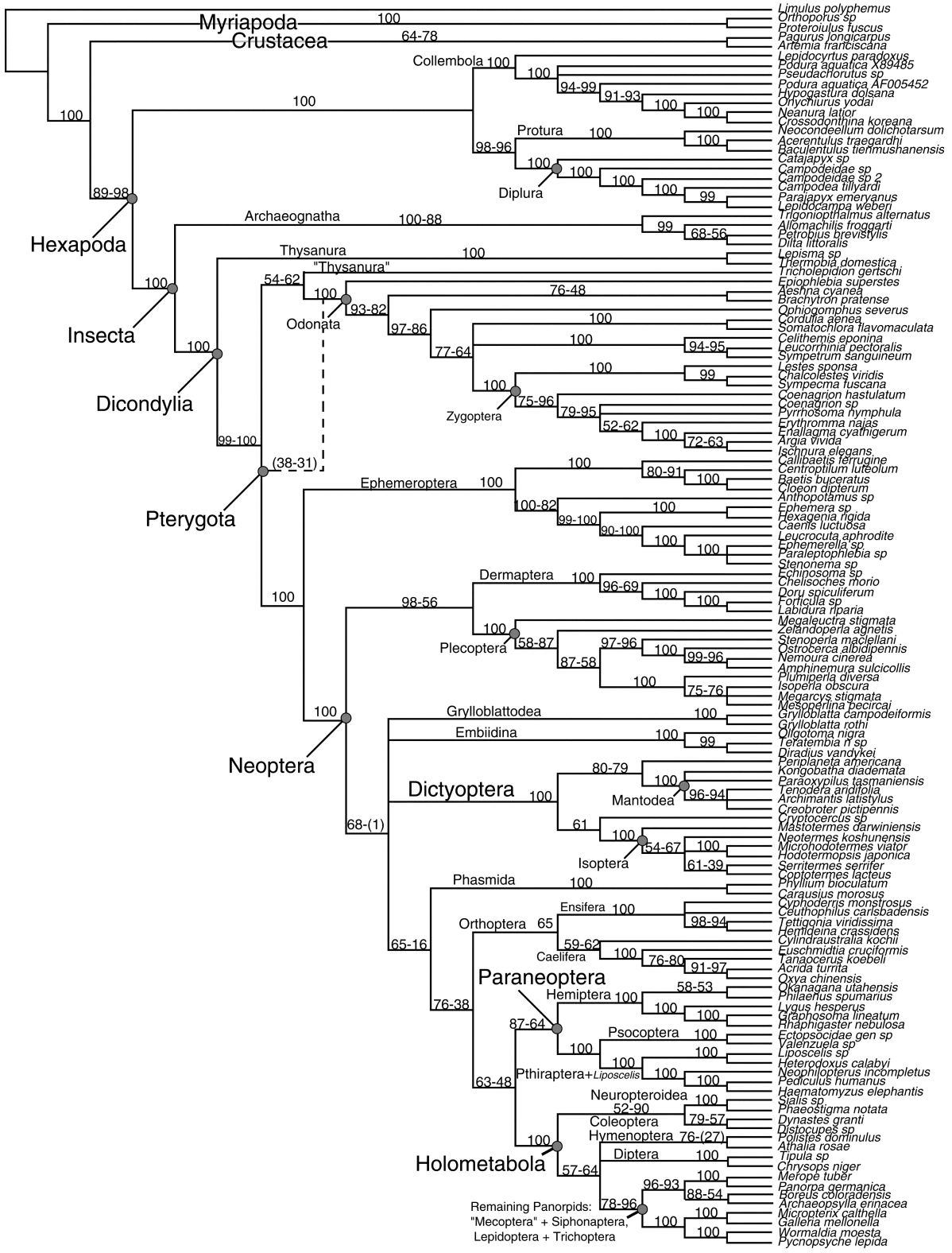

FIGURE 1. Phylogeny of Insecta from 18S rRNA data. This is 50% majority rule consensus tree of Bayesian MCMC samples taken from near the likelihood peak. Bayesian posterior probabilities are given above the node. The first numeral refers to the support value from the MCMC GTR+I+G analysis. The second numeral refers to the posterior probability from the MCMC analysis with the doublet model+I+G. Single numerals indicate that the values were the same for both analyses. Values in parentheses indicate that these nodes were not recovered as the highest percentage among trees. The dotted line refers to the position of Pterygota, which was not recovered in the majority of trees in either analysis.

updated and make the alignment of additional insect data easier.

Different alignment methods result in different hypotheses. Here the alignment is manual, or "by-eye," with reference to the conservative secondary structure of the rRNA. Many studies have shown that manual alignments that utilize secondary structure result in phylogenies that are favored or more congruent with other sources of data than other alignments (Kjer, 1995; Titus and Frost, 1996; Hickson et al., 1996; Morrison and Ellis, 1997; Notredame et al., 1997; Whitfield, 1997; Hwang et al., 1998; Uchida et al., 1998; Whitfield and Cameron, 1998; Cunningham et al., 2000; Gonzalez, 2000; Hickson et al., 2000; Lydeard et al., 2000; Morin, 2000; Mugridge et al., 2000; Xia, 2000; Goertzen et al., 2003; Xia et al., 2003). Ribosomal RNA functions on the basis of its structure, and this structure is conserved to a higher degree than are the nucleotides (Kjer, 1995). In addition, manual alignments permit the investigator to vary permissiveness for the insertion of gaps in hypervariable regions, and locate the boundaries of unalignable regions according to repeatable criteria (Kjer, 1997). Here, this approach to the analysis of the 18S is compared to the results from unadjusted computer alignments; a comparison that has not previously been made.

Chalwatzis et al. (1996) used a variety of parsimony and model-based distance analyses to explore the relationships among Insects using rRNA data. Huelsenbeck (1998) and Hwang et al. (1998) have explored the properties of insect rRNA with likelihood analyses of a small number of taxa. Bayesian inferences have resulted in improved phylogenetic estimates among Ecdysozoans (animals, such as Arthropods and others that moult their exoskeletons) from nuclear rRNA data (Mallatt et al., 2004). However, likelihood and Bayesian analyses have been left out of the comprehensive works of Whiting et al. (1997) and Wheeler et al. (2001) on insect phylogeny. I do not wish to argue the merits of likelihood over parsimony here, but would like to present an alternative view, particularly in the light of the recently available MrBayes 3.0 (Huelsenbeck and Ronquist, 2001; Huelsenbeck and Ronquist, 2002). MrBayes 3.0 includes the option to consider separate data partitions under separate models, and also includes the doublet model (Schöniger and von Haeseler, 1994) that incorporates information on possible character correlations.

METHODS

Taxon Sampling

Complete or nearly complete 18S sequences were selected from GenBank. Most of the previously published work has involved 18S fragments, and thus the majority of these sequences have never been analyzed together. The goal of the current paper was to obtain as complete a sampling of nonholometabolous insects as possible, while also limiting the size of the taxon sample. My laboratory's interest in Paleoptera resulted in an extended sampling of Odonata and Ephemeroptera. Other orders have been subject to extensive analysis with 18S data, and for these orders, randomly selected divergent taxa were used. No complete 18S was available for Zoraptera (all presently existing sequences are chimeric), Thysanoptera, or Mantophasmatodea. Holometabola was considered a single taxon, with 16 divergent species selected to represent this group without attempting to solve relationships among all of its orders. Strepsiptera was not included because the placement of this problematic, autapomorphic taxon and the properties of its 18S have already generated a thorough debate to which there is little to add. The 18S is an entirely inappropriate marker for addressing the placement of Strepsiptera because of its extreme substitution rate acceleration in this taxon, resulting in highly AT-rich hypervariable regions (Huelsenbeck, 1998; Hwang et al., 1998); conditions shared with Diptera. It is quite possible that these shared biases represent true synapomorphies, but matching non-homologous bases in these AT-rich insertions always artificially inflate support for grouping these taxa together.

Alignment

Sequences were aligned manually with reference to secondary structure. Alignments followed the secondary structure models of Gutell et al. (1994), downloaded from the website http://www.rna.icmb.utexas.edu, and modified where compensatory substitutions confirmed a custom arthropod rRNA secondary structural model. Regions that could not be aligned were excluded from the analysis. The criterion used for data exclusion follows Kjer (1997), delimiting unalignable regions flanked by hydrogen-bonded stems.

Analysis

Bayesian likelihood analyses were completed with MrBayes 3.0 (Huelsenbeck and Ronquist, 2001, 2002). Modeltest (Posada and Crandall, 1998) called for a GTR model, and showed that among-site rate variation was best modelled with a gamma correction (Yang, 1993, 1994a, 1994b, 1996) with invariant sites (Gu et al., 1995). However, Modeltest does not evaluate the doublet model (which takes into account correlations among designated characters). Data were analyzed with both the GTR+I+G model, and the doublet model (Schöniger and von Haeseler, 1994). The doublet model was included for rRNA stem sites. For each model, two independent runs of 1.5 million iterations were performed, each with four chains, three hot, one cold, sampling one tree in 500. Plots from the MrBayes "sump" command were used to determine the appropriate "burnin," and after discarding the burnin, trees were pooled. Other parameters recorded in the ".p" files were plotted in Microsoft Excel to insure that all parameters had reached a plateau. Sites were treated with a general time-reversible six-rate model plus gamma plus invariant sites (GTR+I+G) (Tavare, 1986; Gu et al., 1995). Unaligned regions were excluded from likelihood analyses. Equally and differentially weighted parsimony analyses were also performed. In the parsimony analyses, nucleotide motifs

in unaligned (deleted) regions were recoded as single multistate characters with INAASE (Lutzoni et al., 2000). The number of character states was limited to 10, so when there were more than 10, various "outgroup" states where replaced with "?." Successively designated "ingroups" included Pterygota, Neoptera, and Hololmetabola. Odonata, Ephemeroptera, and Dictyoptera were also considered separate "ingroups," with other taxa coded as missing data in order to reduce the number of character states to 10 when necessary. Site-secific differential weighting of the nucleotideswas performed as in Kjer et al. (2001, 2002).

RESULTS

The data file contains 2319 positions, of which 468 were excluded as unaligned. Of the remaining 1851 characters, 770 were constant, and 808 were parsimony informative. Primer sites for most of these sequences were designed at the 3′ and 5′ ends of the 18S, eliminating approximately 55 nucleotides for many taxa, so the estimated length of the 18S ranges from 1808 to 2231 nucleotides. Collembolans are among the shortest sequences, whereas *Parajapyx*, stoneflies, neuropteroids, and lice are among the longest.

For the GTR+I+G analysis, one run leveled off after 200,000 iterations (400 trees discarded), the other after 325,000 iterations (650 trees discarded). The doublet model analyses took much longer to stabilize, with one run leveling off after 650,000 iterations (1300 trees discarded), the other after 585,000 iterations (1170 trees discarded). Parameters are shown in Table 1.

TABLE 1. Parameters from the likelihood analysis. Symbols are taken directly from the MrBayes output; "r" refers to rates between the listed nucleotides (separated by arrows), with "r(G ↔ T)" set to 1, "pi" refers to the proportion of each pair or nucleotide, "alpha" is the shape parameter of the gamma distribution, and "pinvar" refers to the proportion of sites assumed to be invariable. "{all}" refers to parameters that apply to all sites, while "{1}" refers to paired sites, and "{2}" refers to unpaired sites.

	Doublet+I+G	GTR+I+G
lnL	−25796.70 ± 14.26	−27241.06 ± 14.20
TL{all}	7.44 ± 0.32	6.51 ± 0.24
r(A ↔ C){all}	1.03 ± 0.10	1.52 ± 0.16
r(A ↔ G){all}	2.30 ± 0.18	3.45 ± 0.27
r(A ↔ T){all}	0.70 ± 0.07	1.23 ± 0.11
r(C ↔ G){all}	0.65 ± 0.08	0.75 ± 0.08
r(C ↔ T){all}	2.84 ± 0.24	5.17 ± 0.40
r(G ↔ T){all}	1.00 ± 0.00	1.00 ± 0.00
pi(AA){1}	0.02 ± 0.004	
pi(AC){1}	0.01 ± 0.003	
pi(AG){1}	0.01 ± 0.002	
pi(AT){1}	0.18 ± 0.006	
pi(CA){1}	0.01 ± 0.001	
pi(CC){1}	0.01 ± 0.003	
pi(CG){1}	0.21 ± 0.011	
pi(CT){1}	0.01 ± 0.001	
pi(GA){1}	0.01 ± 0.003	
pi(GC){1}	0.24 ± 0.013	
pi(GG){1}	0.01 ± 0.003	
pi(GT){1}	0.03 ± 0.002	
pi(TA){1}	0.18 ± 0.014	
pi(TC){1}	0.01 ± 0.002	
pi(TG){1}	0.03 ± 0.004	
pi(TT){1}	0.02 ± 0.005	
pi(A){2}	0.33 ± 0.010	0.26 ± 0.01
pi(C){2}	0.20 ± 0.008	0.21 ± 0.01
pi(G){2}	0.20 ± 0.008	0.27 ± 0.01
pi(T){2}	0.27 ± 0.009	0.26 ± 0.01
alpha{all}	0.55 ± 0.027	0.58 ± 0.04
pinvar{1}	0.11 ± 0.024	
pinvar{2}	0.15 ± 0.023	0.17 ± 0.02

Results from the Bayesian analyses, rooted with the horseshoe crab (Chelicerata: *Limulus*), are shown in Figure 1. Significantly, Paraneoptera (true bugs, plant-sucking bugs, lice, thrips, and others) are shown as the sister taxon to a monophyletic Holometabola (Insects with complete metamorphisis including a pupal stage) but with weak support. Among the ancestrally wingless orders, Diplura was sister to Protura, and these taxa were sister to Collembola, as part of a monophyletic Hexapoda (as in Delsuc et al., 2003; Mallatt et al., 2004) with strong support, in contrast to recent results from mitochondrial COI, COII, and Cyt *b* data (Nardi et al., 2003), which placed Crustacea closer to Insecta than Collembola. Dictyoptera (mantids, roaches, and termites) are monophyletic, confirming the molecular results of others (Liu and Beckenbach, 1992; Whiting et al., 1997; Wheeler et al., 2001). Roaches are paraphyletic in this analysis, with the communal, wood feeding roach, *Cryptocercus*, as the sister taxon to the termites (as in Lo et al., 2000) *Mastotermes*, the termite that has hindwings that are shaped like those of roaches is at the base of the termites (reviewed in Eggleton, 2001). The wingless lipscelid psocopteran is more closely related to parasitic lice than it is to other winged psocopterans (book lice) (as in Lyal, 1985), and the wingless mecopteran (Boreidae) is more closely related to fleas than it is to other mecopterans (scorpion flies), in agreement with Whiting (2002). Ephemeroptera (mayflies) are recovered as the sister taxon to Neoptera (insects with a unique wing folding mechanism characteristic of most insect orders) with high support (Fig. 1: 100%). No trees were recovered, either with Odonata as sister to Neoptera, or with a monophyletic Paleoptera (Ephemeroptera and Odonata together). This is in contrast to the conclusions of Hovmöller et al. (2002), who used an unadjusted Clustal alignment analyzed with parsimony, and found strong support for Paleoptera. Ogden and Whiting (2003) found that the molecular results were inconclusive, in that results were sensitive to alignment parameters. Hovmöller et al. (2002) also included the D3 in their parsimony analysis, as did Ogden and Whiting (2003), who also included a much larger 28S fragment and data from the Histone 3 gene. A parsimony analysis of the D3 (not shown) recovers only the monophyly of insect orders in an otherwise polytomous tree.

The parsimony analysis, both equally and differentially weighted (not shown: data set available at www.rci.rutgers.edu/∼insects/indexpersonnel.htm and systbiol.org) recover the "long-branch" dipteran taxa as the sister taxon to Crustacea. Parsimony analyses also recover Anisoptera (dragonflies) as follows, (Epiophlebia((Aesnidae + Gomphidae) (Libellulidae))), and a

monophyletic Pterygota (winged insects), with the rest of the tree similar to the likelihood tree.

The websites include the aligned data in three formats. First, a Microsoft Word file is included that uses structural symbols as in Kjer et al. (1994) for the designation of hydrogen bonds at every position used in this paper. Second, a periodically updated version of this file is included, along with a third Nexus file with the data for this paper formated for MrBayes with the doublet model. The Word file includes instructions for converting it to a Nexus format. Once in a Nexus format, importing the data into MacClade 4 (Maddison and Maddison, 2000) permits a systematic visualization of character support for alternative hypotheses. In the tree window of MacClade 4, using the "trace changes" option with graphic display, "show bar for each change," and bar display option "color bars as a function of CI," one can evaluate the characters that support alternative arrangements of the orders, and the homoplasy in these characters. For example, Phasmida (walking sticks) can be successively moved to a sister taxon with each of the other insect orders, and it can be deduced that all character support for the placement of this taxon is highly homoplasious in any phylogeny. Very little support is shown for any relationship among orders of nonholometabolous neopterans (Fig. 1). By making the alignment publicly available, with instructions on how to convert the file into the Nexus format, continued refinement of the alignment, and the exploration of other models will be encouraged.

Discussion

An independent examination of this important gene permits some insight into several vexing questions. Ephemeroptera (mayflies) as the sister taxon to Neoptera offers an interesting scenario in the role of direct sperm transfer in the evolution of winged insects. Ephemeroptera and Neoptera both possess direct sperm transfer. Male Odonata (dragonflies and damselflies), on the other hand, transfer sperm from their primary genitalia on the terminal part of their abdomens to their secondary genitalia, on the basal part of their abdomens. However, the secondary genitalia among the three suborders of Odonata are derived from different structures, and are thus not homologous (Schmidt, 1915), leading to the intriguing possibility of independent evolution of this complex and unusual morphologically dependent behavior. The extinct Meganisoptera (the giant, familiar Carboniferous dragonfly-like fossils) has been shown to lack secondary genitalia (Brauckmann and Zessen, 1989), and Carle (1982a, 1982b) had previously postulated that Meganisoptera probably guided the female to spermatophores, as do the nonwinged insects. Carle (1982a, 1982b) considered Odonata to be the sister taxon of the rest of the winged insects (including the extinct Paleopdictyoptera), based on the absence of both direct sperm transfer and male abdominal copulatory forceps in Odonata. Under this scenario, the unique tandem hold, in which male Odonata grasp the females near the neck, is not homologous with the genital forceps of other pterygotes. Grasping the female in Odonata could have originally been a male adaptation to avoid female predation while directing females to the spermatophore (Carle, 1982a, 1982b). But spermatophores would be harder to reach as insects moved from two dimensions to three as they developed flight. This might explain the selection for direct sperm transfer in winged insects that resulted in a homologous form shared by Ephemeroptera and Neoptera, and at least one independent origin of it in Odonata. Although this hypothesis is speculative, coming from a single gene, it is worthy of further investigation. Morphological interpretations of this question are conflicting and somewhat circular, in that they are dependent on the importance one places on direct sperm transfer and its homology, and whether or not the unique form of sperm transfer in Odonata can be linked to the primitive state.

The relationships among Odonata are unexpected. Many studies (Carle, 1982c; Bechly, 1994; Carle, 1995; Rehn, 2003) have shown both Anisoptera (dragonflies) and Zygoptera (damselflies) to be monophyletic, as do unpublished data from the 28S and EF-1α from my laboratory. There are very few variable characters among Odonata in the entire 18S. This may explain why the Bayesian analyses fail to recover Anisoptera, while parsimony (see web resources) succeeds; because a group with virtually no change on the terminal branches is both "unlikely," and parsimonious.

One of the advantages of a Bayesian analysis is that alternative hypotheses can be evaluated, and support for these alternatives can be compared to support for the majority-rule consensus. The placement of the thysanuran *Tricholepidion* as the sister taxon of Odonata, implying multiple origins of flight in insects (or loss of wings in *Tricholepidion*), should not be taken seriously, given its weak posterior probability (54%), and a more reasonable alternative as the sister taxon to Pterygota, supported at 38%, and also supported in parsimony trees (see Web resources). Most of the remaining trees place *Tricholepidion* as sister to Ephemeroptera + Neoptera (8%). *Tricholepidion* (Lepidotrichidae) separated from the other thysanurans, represented by *Thermobia* and *Lepisma* (Lepsmatidae), is not unreasonable, and has been supported by many as the sister taxon of Dicondylia (Stys et al., 1993; Stys and Zrzavy, 1994; Kristensen, 1997; Bitsch and Bitsch, 2000).

Whereas several papers on insect phylogeny have been published with evidence from multiple genes plus morphology (Whiting et al., 1997; Wheeler et al., 2001), the molecular data from these studies has been dominated by a fragment of the larger and/or more conservative 18S. Support from the 18S for any relationship among the nonholometabolous neopteran orders is almost nonexistent, and thus, the conclusions from these studies rely heavily on either noise or on the morphological data, largely contributed by Kristensen (1975, 1981, 1991, 1995, 1997). Although I favor combined analyses, including morphological data and multiple genes, the impression that there are large molecular data sets supporting the

conclusions of Wheeler et al. (2001) is false, particularly with the use of a strategy that optimizes the alignment of rRNA based on the character congruence with the morphological data. Conclusions from automated alignments (Whiting et al.,1997; Wheeler et al., 2001; Ogden and Whiting, 2003) vary with the selection of alignment parameters such as fixed gapcost to change ratios. I have argued (Kjer, 1995) that there are no appropriate fixed gapcost to change ratios, even when selected by a thorough "sensitivity analysis" (Wheeler, 1995). This is because the probability of insertions and deletions in large rRNA molecules vary considerably from one region of the molecule to another (i.e., different regions have different gapcosts).

It is difficult to compare these results with previous results. A decision over which of the hypotheses is "best" is dependent on one's beliefs about how the data should be analyzed. Figure 2 shows the results of a direct optimization analysis (POY) of the 18S, using equal costs for gaps, and all substitution types (Wheeler et al., 2001; their Fig. 13). This is not a fair comparison, because the 18S fragment used in these analysis is considerably

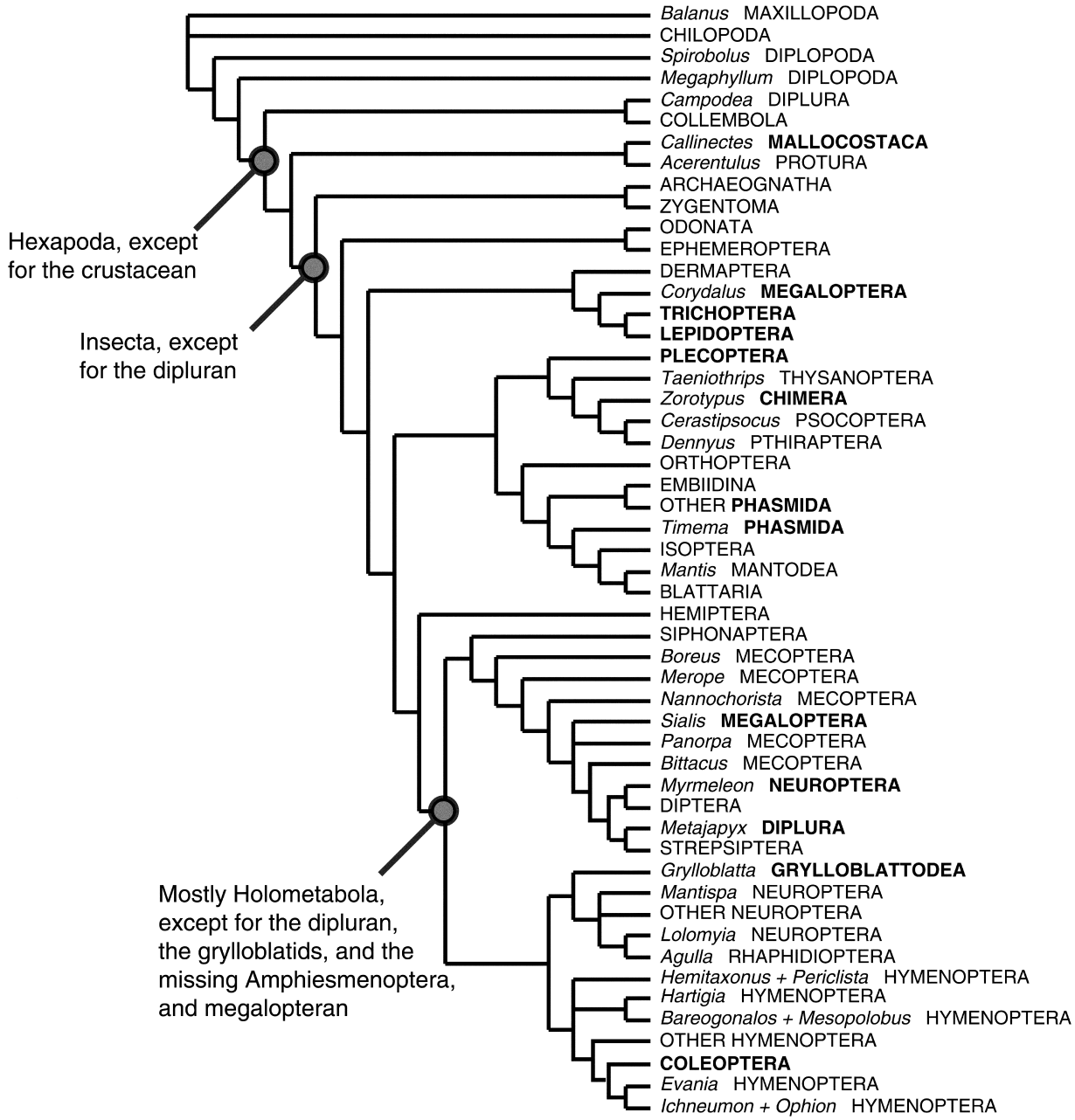

FIGURE 2. Phylogenetic relationships resulting from a direct optimization approach, using uniform gapcost to change ratios, and equal weighting for transitions and transversions, modified from Figure 13 of Wheeler et al. (2001). Some of the taxa that most systematists would consider misplaced are in bold.

smaller than the nearly complete 18S data used to generate Figure 1. Alternatively, adding additional data in a combined molecular analysis (the 18S fragment, plus the D3 fragment from 28S), Wheeler et al. (2001: their Fig. 12a) presented an equally improbable tree (discussed in the introduction). However, when the morphology was used to discover appropriate analytical parameters for the 18S data (Wheeler et al., 2001; their Fig. 12b), and when combined analyses were presented, an entirely reasonable phylogeny was recovered. Unfortunately, we cannot assess these different trees without some means to assess the different parameters. If sensitivity analysis as it is most frequently applied is arbitrary, then so are the trees derived from it. Ogden and Whiting (2003) argue that molecular data cannot yet resolve the relative positions of Odonata and Ephemeroptera + Neoptera because different input parameters result in different trees. An alternative interpretation is that because no single set of automated alignment parameters can be objectively favored, both sensitivity analysis and direct optimization have failed in this case (until additional data can be added). In other words, the methods failed, not the data. What is needed to assess support for alternative hypotheses from the data we have is some reference to homology, and homology statements are found in alignments. The question systematists must decide is whether a phylogenetic debate should center on alternative hypotheses of homology, or focus on alternative gapcosts and transversion weights. Although gaps may shift positions across lineages (favoring direct optimization in regions of ambiguous homology), there are not multiple histories for those positions according to gapcosts. Homologous nucleotides have never been subject to shift positions with alternative gapcosts. In other words, 300 million years ago, when a deletion happened in some specific lineage, it didn't happen one way under a gapcost of 2, and another way under a gapcost of 3.

Concern for the possibility of contaminant sequences entering an analysis is particularly relevant when considering analyses that favor automated approaches to alignment of large combined datasets. Several nonhomoplastic synapomorphies exist between a region of the Zoraptera sequence (AF372432), used in Wheeler et al. (2001), and acarines (mites). This region of several hundred nucleotides, near helix 36 (labeled as in van de Peer et al., 1994), including the sequence AAAACTTACCCGGCC, when subjected to a BLAST search returns acarines as its closest matches (Altschul et al., 1997). Most insects have "YCA" at the three underlined positions. A phylogenetic analysis of the suspect region recovers the "zorapteran" outside the insects, with the ticks. Two unpublished "zorapteran" sequences sent to us were also contaminants (one a Dragonfly, the other an Alga). It is still possible that the zorapteran sequence from Wheeler et al. (2001) is real (convergent with the acarine sequences), but whether it is or not, it is highly unusual and problematic. There may be something unusual about zorapteran 18S rRNA that makes it difficult to amplify. The combination of a chimeric contaminant/target sequence with a direct optimization approach in which ancestral sequences are reconstructed throughout the tree according to the combined analysis would seriously compromise the reconstruction of ancestral nodes. Similarly, the contaminant beetle/neuropteroid sequences used in Whiting et al. (1997), resulting in the reported polyphyly of Coleoptera, could have been easily recognized in a manual alignment. Visualization of individual sequences, required during manual alignments, permits a kind of proofreading (Kjer, 1997; Xia et al., 2003), flagging potentially problematic sequences for resequencing. As data sets become larger, proofreading will become both more important and more difficult. Just as an insect morphologist would never confuse a dragonfly with a spider (even if it were mislabeled), rRNA data have an unmistakable and recognizable structure that would be understood by morphologists and must be examined to be seen.

The assumption that each site is independent is clearly violated with rRNA, where a significant percentage of the sites in 18S are hydrogen-bonded and interacting with one another. This violation of assumptions has the potential to affect both parsimony and likelihood analyses, although perhaps in different ways. With parsimony, ignoring the nonindependence of sites results in a higher weight for stem sites, which may be inadvertently appropriate, given that stem sites tend to be more conservative (although this tendency is a very loose one; see van de Peer, 1994). In likelihood, ignoring the nonindependence of sites results in a model that overrepresents change at stem sites. Violation of the assumption of independence at sites has been examined with simulations by Tillier and Collins (1995) and Huelsenbeck and Nielsen (1999). Both studies concluded that likelihood is relatively robust to violation of the assumption of independence. However, the models used in their simulations did not evaluate the possibility that some correlated sites (e.g., long-range stems) may be more conservative than some single-stranded sites (e.g., unpaired regions not involved in function). In any case, the doublet model corrects for violations of assumptions of independence, and thus is worth investigating with empirical data. This model performed well here, mirroring the results from the GTR analysis, but showing a collapse in support for most relationships among nonholometabolous neopterans (Fig. 1). This collapse in support may indicate that some of the characters supporting relationships in the GTR analysis are not independent of one another. Another explanation for the decrease in support when the doublet model is used is simply that there has been an increase in model variance (Buckley and Cunningham, 2002).

It has been argued that manual alignments are not repeatable (Wheeler, 2001). This statement is here shown to be false. Anyone can repeat the analyses performed here by downloading the data and using the alignment. It may be that criticisms about repeatability stem from the fact that different manual alignments may not be identical, but it is hard to imagine why anyone would be interested in the exact reconstruction of an alignment as an ultimate product, whether by machine or otherwise, if the question of interest is the phylogeny, and the

alignment is made available. The fact that specific hypotheses of homology are made for each site makes it possible to challenge and upgrade these hypotheses. If it is thought that phylogenetic conclusions are biased by an alignment, such a possibility could be examined using the alignment. None of these challenges are possible with POY (direct optimization), because no fixed hypotheses of homology are offered. Implied alignments generated from POY analyses (Wheeler, 2003) are in my opinion of limited value because they do not represent the data from which phylogenetic hypotheses were based.

These results should not be interpreted as the phylogeny of Insecta. The inclusion of the morphological data (e.g., the excellent work done by Kristensen, 1981, 1991, 1995, 1997) and additional genes is the next step toward the resolution of this larger goal. This paper permits a comparison of partitions. The intent was to present a different view of the contribution of a large part of the molecular data to the question of the phylogeny of nonholometabolous insect orders.

ACKNOWLEDGEMENTS

I thank in advance anyone who informs me of adjustments to the web alignment. I thank Frank Carle for discussions on sperm transfer, and Joe Gillespie for discussions on the doublet model. I thank Frank Carle, Joe Gillespie, Elizabeth Jockusch, John LaPolla, Mike May, and Brian Wiegmann for comments on the manuscript. I thank Chris Simon, Thomas Buckley, Francesco Frati, and James Whitfield for their careful reviews and helpful comments. I thank NSF for grants to Duckett, C.N., and Kjer, Morse, J.C., and Kjer, and the New Jersey Agricultural Experiment Station for financial support.

REFERENCES

Altschul, S. F., T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res. 25:3389–3402.

Bechly, G. 1994. A cladistic study of the Odonata and a reconstruction of the wing venation in the ground plan of crown group Odonata. Petalura 1:1–24.

Bitsch, C., and J. Bitsch. 2000. The phylogenetic interrelationships of the higher taxa of apterygote hexapods. Zoologica Scripta 29:131–156.

Brauckmann, C., and W. Zessen. 1989. Neue Meganeuridae aus dem Namurium von Hagen-Vorhalle (BRD) und die Phylogenie der Meganisoptera (Insecta, Odonata) Deut. Entomol, Zeit. (N.F.) 36:177–215.

Buckley, T. R., and C. W. Cunningham. 2002. The effects of Nucleotide substitution model assumptions on estimates of nonparameteric bootstrap support. Mol. Biol. Evol. 19:394–405.

Carapelli, A., F. Frati, F. Nardi, R. Dallai, and C. Simon. 2000. Molecular phylogeny of the apterygotan insects based on nuclear and mitochondrial genes. Pedobiologia 44:361–373.

Carle, F. L. 1982a. Evolution of the odonate copulatory process. Odonatologica 11:271–286.

Carle, F. L. 1982b. Thoughts on the origin of insect flight. Ent. News 93:159–172.

Carle, F. L. 1982c. The wing vein homologies and phylogeny of the Odonata: A continuing debate. Soc. Int. Odonatol. Rapid Comm., 4:x + 66p.

Carle, F. L. 1995. Evolution, taxonomy and biogeography of ancient gondwanian Libelluloids, with comments on anisopteroid evolution and phylogenetics. Odonatologica 24:383–424.

Carmean D., L. S. Kimsey, and M. L. Berbee. 1992. 18S rDNA sequences and holometabolous insects. Mol. Phylogenet. Evol. 1:270–278.

Chalwatzis, N., J. Hauf, Y. van de Peer, R. Kinzelbach, and F. K. Zimmermann. 1996. 18S ribosomal RNA genes of insects: Primary structure of the genes and molecular phylogeny of Holometabola. Ann. Entomol. Soc. Amer. 89:775–787.

Cunningham, C. O., Aliesky, H., and C. M. Collins. 2000. Sequence and secondary structure variation in the Gyrodactylus (Platyhelminthes: Monogenea) ribosomal RNA gene array. J. Parasitol. 86:567–576.

Delsuc, F., M. J. Phillips, and D. Penny. Comment on "Hexapod Origins: Monophyletic or Paraphyletic." Science 301:1482d.

Eggleton, P. 2001. Termites and trees: A review of recent advances in termite phylogenetics. Insectes Soc. 48:187–193.

Flook, P. K., and C. H. F. Rowell. 1997. The effectiveness of mitochondrial rRNA sequences for the reconstruction of the phylogeny of an insect order (Orthoptera). Mol. Phylogenet. Evol. 8:177–192.

Gladstein, D. S., and W. C. Wheeler. 1997. "POY: The optimization of alignment characters." Program and documentation. New York. Available at ftp.amnh.org/pub/molecular.

Goertzen, L. R., Cannone, J. J., Gutell, R. R., and R. K. Jansen. 2003. ITS secondary structure derived from comparative analysis: Implications for sequence alignment and phylogeny of the Asteraceae. Mol. Phylogenet. Evol. 29:216–234.

Gonzalez, P., and J. Labarere. 2000. Phylogenetic relationships of *Pleurotus* species according to the sequence and secondary structure of the mitochondrial small-subunit rRNA V4, V6 and V9 domains. Microbiology 146:209–221.

Gu, X., Y.-X. Fu, and W.-H. Li. 1995. Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. Mol. Biol. Evol. 12:546–557.

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

Hickson, R. E., C. Simon, A. J. Cooper, G. Spicer, J. Sullivan, and D. Penny. 1996. Refinement of a secondary structure model for the third domain of animal 12S rRNA, with a comparison of alignment programs. Mol. Biol. Evol. 13:150–169.

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

Hovmöller, R., T. Pape, and M. Kallersjö. 2002. The Paleoptera problem: Basal Pterygote phylogeny inferred from 18S and 28S rDNA sequences. Cladistics 18:313–323.

Huelsenbeck, J. P. 1998. Systematic Bias in phylogenetic analysis: Is the Strepsiptera problem solved? Syst. Biol. 47:519–537.

Huelsenbeck, J. P, and R. Nielsen. 1999. Effect of nonindependent substitution on phylogenetic accuracy. Syst. Biol. 48:317–328.

Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755.

Huelsenbeck, J. P., and F. Ronquist. 2002. MrBayes 3: Bayesian analysis of phylogeny. Computer program distributed by the authors. Department of Ecology, Behavior and Evolution, University of California, San Diego.

Hwang, U. W., W. Kiim, D. Tautz, and M. Friedrich. 1998. Molecular phylogenetics at the Felsenstein zone: Approaching the Strepsipera problem using 5.8S and 28S rDNA sequences. Mol. Phylogenet. Evol. 9:470–480.

Johnson, K. P., and M. F. Whiting. 2002. Multiple genes and the monophyly of Ischnocera (Insecta: Phthiraptera). Mol. Phylogenet. Evol. 22:101–110.

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. Biol. Evol. 4:314–330.

Kjer, K. M. 1997. An alignment template for amphibian 12S rRNA, domain III: Conserved primary and secondary structural motifs. J. Herpetol. 31:599–604.

Kjer, K. M., G. D. Baldridge, and A. M. Fallon. 1994. Mosquito large subunit ribosomal RNA: Simultaneous alignment of primary and secondary structure. Biochim. Biophys. Acta 1217:147–155.

Kjer, K. M., R. J. Blahnik, and R. W. Holzenthal. 2001. Phylogeny of Trichoptera (Caddisflies): Characterization of signal and noise within multiple datasets. Syst. Biol. 50:781–816.

Kjer, K. M., R. J. Blahnik, and R. W. Holzenthal. 2002. Phylogeny of Caddisflies (Insecta, Trichoptera). Zoologica Scripta 31:83–91.

Kristensen, N. P. 1975. The phylogeny of hexapod "orders." A critical review of recent accounts. Z. Zool. Syst. Evollutionforsch. 13:1–44.

Kristensen, N. P. 1981. Phylogeny of insect orders. Annu. Rev. Entomol. 26:135–157.
Kristensen, N. P. 1991. Phylogeny of extant hexapods. Pages 125–140 *in* The insects of Australia, 2nd ed. (I. D. Nielsen, J. P. Spradberry, R. W. Taylor, M. J. Whitten, and M. J. Littlejohn, eds.). CSIRO, Melbourne University Press, Melbourne.
Kristensen, N. P. 1995. Forty years' insect phylogenetic systematics. Zool. Beitr. N. F. 36:83–124.
Kristensen, N. P. 1997. The ground plan and basal diversification of the hexapods. Pages 281–293 *in* Arthropod relationships (R. A. Fortey and R. H. Thomas, eds.). Chapman and Hall, London.
Liu, H., and A. T. Beckenbach. 1992. Evolution of the mitochondrial oxydase II gene among 10 orders of insects. Mol. Phylogenet. Evol. 1:41–52.
Lo, N., G. Tokuda, H. Watanabe, H. Rose, M. Slaytor, K. Maekawa, C. Bandi, and H. Noda. 2000. Evidence from multiple gene sequences indicates that termites evolved from wood-feeding cockroaches. Curr. Biol. 10:801–804.
Lutzoni F., P. Wagener, V. Reev, and S. Zoller. 2000. Integrating ambiguously aligned regions of DNA sequence in phylogenetic analyses without violating positional homology. Syst. Biol. 49:628–651.
Lyal, C. H. C. 1985. Phylogeny and classification of the Psocodea, with particular reference to the lice (Psocodea: Phthiraptera). Syst. Entomol. 10:145–165.
Lydeard, C., W. E. Holznagel, M. N. Schnare, and R. R. Gutell. 2000. Phylogenetic analysis of molluscan mitochondrial LSU rDNA sequences and secondary structures. Mol. Phylogenet. Evol. 15:83–102.
Maddison, D. R., and W. P. Maddison. 2000. MacClade 4. Computer program. Sinauer, Sunderland, MA.
Mallatt, J. M., J. R. Garey, and J. W. Schultz. 2004. Ecdysozoan phylogeny and Bayesian inference: First use of nearly complete 28S and 18S rRNA gene sequences to classify the arthropods and their kin. Mol. Phylogenet. Evol. 31:178–191.
Misof, B., A. M. Rickert, T. R. Buckley, G. Fleck, and K. P. Sauer. 2001. Phylogenetic signal and its decay in mitochondrial SSU and LSU rRNA gene fragments of Anisoptera. Mol. Biol. Evol. 18:27–37.
Morin, L. 2000. Long branch attraction effects and the status of "basal eukaryotes": Phylogeny and structural analysis of the ribosomal RNA gene cluster of the free-living diplomonad *Trepomonas agilis*. J. Eukaryot. Microbiol. 47:167–177.
Morrison, D. A., and J. T. Ellis. 1997. Effects of nucleotide sequence alignment on phylogeny estimation: A case study of 18S rDNAs of Apicomplexa. Mol. Biol. Evol. 14:428.
Mugridge, N. B., D. A. Morrison, T. Jakel, A. R. Heckeroth, A. M. Tenter, and A. M. Johnson. 2000. Effects of sequence alignment and structural domains of ribosomal DNA on phylogeny reconstruction for the protozoan family Scacocystidae. Mol. Biol Evol. 17:1842.
Nardi, F., G. Spinsanti, J. Boore, A. Carapelli, R. Dallai, and F. Frati. 2003. Hexapod Origins: Monophyletic or paraphyletic. Science 299:1887–1889.
Notredame, C., E. A. O'Brien, and D. G. Higgins. 1997. RAGA: RNA sequence alignment by genetic algorithm. Nucleic Acids Res. 25:4570–4580.
Ogden, T. H., and M. F. Whiting. 2003. The problem with "the Paleoptera Problem" sense and sensitivity. Cladistics 19:432–442.
Pashley, D. P., B. A. McPheron, and E. A. Zimmer. 1993. Systematics of holometabolous insect orders based on 18S ribosomal RNA. Mol. Phylogenet. Evol. 2:132–142.
Posada, D., and K. A. Crandall. 1998. ModelTest: Testing the model of DNA substitution. Bioinformatics 14:817–818.
Rehn, A. C. 2003. Phylogenetic analysis of higher-level relationships of Odonata Syst. Entomol. 28:181–241.
Schmidt, E. 1915. Vergleichende morphologie des 2. und 3. abdominalegmentes bei mannlichen Libellen. Zool. Jb. (Abt. Anat.) 39:87–196.
Schöniger, M., and A. von Haeseler. 1994. A stochastic model and the evolution of autocorrelated DNA sequences. Mol. Phylogenet. Evol. 3:240–247.
Stys, P., and J. Zrzavy. 1994. Phylogeny and classification of extant Arthropoda: Review of hypotheses and nomenclature. Eur. J. Entomol. 91:257–275.
Stys, P., J. Zrzavy, and F. Weyda. 1993. Phylogeny of the hexapoda and ovarian metamerism. Biol. Rev. 68:365–379.
Tavare, S. 1986. Some probabilistic and statistical problems on the analysis of DNA sequences. Lect. Math. Life Sci. 17:57–86, 368–376.
Titus, T., and D. R. Frost. 1996. Molecular homology assessment and phylogeny in the lizard family Opluridae (Squamata: lguania). Mol. Phylogenet. Evol. 6:49–62.
Uchida, H., K. Kitae, K. I. Tomizawa, and A. Yokota. 1998. Comparison of the nucleotide sequence and secondary structure of the 5.8S ribosomal RNA gene of *Chlamydomonas tetragama* with those of green algae. DNA Seq. 8:403–408.
Van de Peer, Y., J.-M. Neefs, P. De Rijk, and R. DeWacter. 1993. Reconstructing evolution om eukaryotic small-subunit RNA sequences: Calibration of the molecular clock. Mol. Evol. 37:221–232.
Wheeler, W. C. 1989. The systematics of insect ribosomal DNA. Pages 307–321 *in* The hierarchy of life. Molecules and morphology in phylogenetic analysis (B. Fernholm, K. Bremer, and H. Jörnvall, eds.). Elsevier, Amsterdam.
Wheeler, W. C. 1995. Sequence alignment, parameter sensitivity, and the phylogenetic analysis of molecular data. Syst. Biol. 44:321–331.
Wheeler, W. C. 1996. Optimization alignment: The end of multiple sequence alignment in phylogenetics. Cladistics 12:1–9.
Wheeler, W. C. 1999. Heuristic reconstruction of hypothetical-ancestral DNA sequences: Sequence alignment vs direct optimization. Pages 106–113 *in* Homology and Systematics: Coding characters for phylogenic analysis (R. W. Scotland, ed.). CRC Press.
Wheeler, W. C. 2001. Homology AND DNA sequence data. Pages 303–317 *in* The character concept in evolutionary biology (G. P. Wagner, ed.). Academic Press, San Diego, CA.
Wheeler, W. C. 2003. Implied alignment: A synapomorphy-based multiple-sequence alignment method and its use in cladogram search. Cladistics 19:261–268.
Wheeler, W. C., M. F. Whiting, Q. D. Wheeler, and J. M. Carpenter. 2001. The phylogeny of extant insect orders. Cladistics 17:113–169.
Whitfield, J. B. 1997. Molecular and morphological data suggest a single orignin of the polydnaviruses among brachonid wasps. Naturwissenschaften 84:502–507.
Whitfield, J. B. and S. A. Cameron. 1998. Hierarchical analysis of variation in the mitochondrial 16S rRNA gene among Hymenoptera. Mol. Biol. Evol. 15:1728–1743.
Whiting, M. F., J. C. Carpenter, Q. D. Wheeler, and W. C. Wheeler. 1997. The Strepsiptera problem: Phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. Syst. Biol. 46:1–68.
Whiting, M. F. 2002. Mecoptera is paraphyletic: Multiple genes and phylogeny of Mecoptera and Siphonaptera. Zoologica Scripta 31:93–104.
Xia, X. 2000. Phylogenetic relationship among horseshoe crab species: The effect of substitution models on phylogenetic analyses. Syst. Biol. 49:87–100.
Xia, X., Z. Xie, and K. M. Kjer. 2003. 18S ribosomal RNA and tetrapod phylogeny. Syst. Biol. 52:283–295.
Yang, Z. 1993. Maximum likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol. Biol. Evol. 10:1396–1401.
Yang, Z. 1994a. Estimating the pattern of nucleotide substitution. J. Mol. Evol. 39:105–111.
Yang, Z. 1994b. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: Approximate methods. Syst. Biol. 44:384–399.
Yang, Z. 1996. Among-site rate variation and its impact on phylogenetic analysis. TREE 11:367–372.

First submitted 3 October 2003; reviews returned 17 November 2003; final acceptance 4 February 2004
Associate Editor: Thomas Buckley