Pisciforma, Setisura, and Furcatergalia (Order: Ephemeroptera) Are Not Monophyletic Based on 18S rDNA Sequences: A Response to Sun et al. (2006)

T. HEATH OGDEN,[1] MICHEL SARTORI,[2] AND MICHAEL F. WHITING[3]

Sun et al. (2006) recently published an analysis of phylogenetic relationships of the major lineages of mayflies (Ephemeroptera). Their study used partial 18S rDNA sequences ($\approx$583 nucleotides), which were analyzed via parsimony to obtain a molecular phylogenetic hypothesis. Their study included 23 mayfly species, representing 20 families. They aligned the DNA sequences via default settings in Clustal and reconstructed a tree by using parsimony in PAUP*. However, this tree was not presented in the article, nor have they made the topology or alignment available despite multiple requests. This molecular tree was compared with previous hypotheses based on morphological data to "test" (but see below) which morphology-based relationships were not significantly different from the molecular topology. Although molecular data can help shed light on many of the fundamental questions in insect phylogenetics, it is important to perform analyses correctly and to accurately report results in a way that allows subsequent validation.

Sun et al. (2006) provided an adequate review of the major traditional hypotheses concerning higher level classification and relationships among mayflies. Some of these hypotheses are based on cladistic analysis of morphology, but others are intuitive phylogenies that were not derived from any formal repeatable and objective analysis. These authors failed to report previously published molecular hypotheses that shed light on mayfly phylogeny, many of which disagree with their results. For example, there have been previously published molecular studies investigating higher level relationships among mayflies and relatives (Hovmöller et al. 2002, Ogden and Whiting 2003, Ball et al. 2005, Danforth et al. 2005) and investigating relationships below the family level (Tojo and Matsukawa 2003, Monaghan et al. 2005, Rebora et al. 2005). All of these analyses include additional data and taxa that could have been included in the Sun et al. (2006) analysis to make a more robust estimate of mayfly phylogeny.

It is clear that the authors knew other ephemeropteran sequences were deposited on GenBank, because they downloaded and included sequence data for the genus *Rallidens* (Rallidentidae), which became available on GenBank October 2003. However, they chose not to include 34 other mayfly 18S rDNA sequences that were available 18 mo before submission of their manuscript (sequences available October 2003; their manuscript was submitted 1 March 2005). If the authors had included these additional taxa, they would have increased their generic and familial level sampling to include lineages such as Leptohyphidae, Potamanthidae, Behningiidae, Neoephemeridae, Ephemerellidae, and Euthyplociidae. Additionally, there were 194 sequences available (as of 1 March 2005) for other molecular markers, aside from 18S, that could have been used to investigate higher level relationships. These include Cytochrome oxidase I, Elongation factor I-α, Histone 3, 28S rDNA, and 16S mitochondrial rDNA.

The purpose of this short reply is to reexamine the Sun et al. (2006) analysis to investigate whether the data they present do in fact support 1) the monophyly of the suborders Setisura, Pisciforma, and Furcatergalia; 2) the placement of Baetiscidae as the sister lineage to all other represented clades; and 3) the elevation of *Pseudiron* and *Arthroplea* as separate monophyletic families. We will then present an analysis based on all the 18S mayfly data that were available to the authors to determine whether inclusion of these additional data result in conclusions contrary to those reported. We further examine their analytical methodology and focus on their use of the Shimodaira–Hasegawa test to conclude that the groups defined via morphology are not significantly different from the molecular topology. All of these results are then discussed in light of a more comprehensive phylogenetic analysis of mayflies as presented in Ogden and Whiting (2005).

Materials and Methods

Sun et al. (2006) have did not provide us their original sequence alignment, molecular topologies, or details related to parameters used in the reconstruction of molecular topologies. Consequently, we downloaded their sequences from GenBank (Table 1) in an attempt to reproduce their analysis. Because the authors were brief in their description of some analytical details, certain analytical parameters had to be assumed. For example, we treated gaps as missing data and all characters were equally weighted in our parsimony analysis. We constructed a second data set that consisted of these author's data and all mayfly 18S rDNA sequences that had been deposited on GenBank as of 1 October 2003, to gauge whether the

The authors of Sun et al. (2006) were notified of this critique of their paper, but did not submit a response.

[1] Corresponding author: Department of Biological Sciences, Idaho State University, Pocatello, ID 83209 (e-mail: ogdet@isu.edu).

[2] Museum of Zoology, P.O. Box, CH-1014, Lausanne, Switzerland.

[3] Department of Integrative Biology, Brigham Young University, 401 WIDB, Provo, UT 84602.

Table 1. Taxonomic sampling and GenBank accession numbers

Family	Genus species	GenBank
Lepismatidae	*Lepisma sp.*	AF005458
Acanthametropodidae	*Analetris eximia*	AY338697
	Analetris eximia	DQ648716
Ameletidae	*Ameletus sp.*	AY338712
	Ameletus celer	DQ648730
Ameletopsidae	*Chaquihua sp.*	AY338715
	Chiloporter sp.	DQ648721
Ametropodidae	*Ametropus neavei*	AY338700
	Ametropus neavei	DQ648734
Arthropleidae	*Arthroplea bipunctata*	DQ648727
Baetidae	*Baetis bicaudatus*	DQ648719
	Baetis buceratus	AF461248
	Baetis sp.	AY338695
	Callibaetis ferrugineus	DQ648714
	Callibaeits ferrugineus	AF370791
	Centroptilum luteolum	AF461251
	Cloeon dipterum	AF461249
Baetiscidae	*Baetisca sp.*	AY338707
	Baetisca lacustris	DQ648715
Behningiidae	*Behningia sp.*	AY338703
Caenidae	*Caenis sp.*	AY338710
	Caenis luctuosa	AF461250
	Caenis youngi	DQ648717
Coloburiscidae	*Coloburiscoides giganteus*	DQ648723
	Coloburiscus humeralis	AY338706
Ephemerellidae	*Drunella coloradensis*	AY338694
	Ephemerella sp.	U65107
Ephemeridae	*Ephemera sp.*	X89489
	Ephemera simulans	DQ648733
	Hexagenia sp.	AY121136
	Hexagenia rigida	AF461253
Euthyplociidae	*Polyplocia sp.*	AY338705
Heptageniidae	*Cinygmula sp.*	AY121137
	Epeorus grandis	DQ648729
	Heptagenia sp.	AY338709
	Heptagenia diabasia	DQ648731
	Leucrocuta aphrodite	AF461254
	Stenonema sp.	AF461252
Isonychiidae	*Isonychia sp.*	AY338708
	Isonychia rufa	DQ648728
Leptohyphidae	*Leptohyphes apache*	AY338714
Leptophlebiidae	*Paraleptophlebia memorialis*	DQ648722
	Choroterpes sp.	AY555525
Metretopodidae	*Metretopus borealis*	AY338698
	Metretopus borealis	DQ648718
Neoephemeridae	*Neoephemera youngi*	AY338702
Nesameletidae	*Ameletoides lacusalbinae*	DQ648732
Oligoneuriidae	*Lachlania saskatchewanensis*	AY338701
	Lachlania talea	DQ648725
Oniscigastridae	*Tasmanophlebia sp.*	DQ648720
Polymitarcyidae	*Tortopus primus*	DQ648724
Potamanthidae	*Anthopotamus sp.*	AY338711
	Anthopotamus sp.	AF461255
Pseudironidae	*Pseudiron centralis*	AY338699
	Pseudiron centralis	DQ648726
Rallidentidae	*Rallidens mcfarlanei*	AY338696
Siphlonuridae	*Paramaletus columbiae*	AY338713
	Siphlonurus columbianus	DQ648735

addition of these data might alter higher level relationships.

Phylogenetic Analysis. Sequences were aligned with ClustalW (Thompson et al. 1997) by using default parameters (as in Sun et al. 2006) for both data sets. Tree reconstruction consisted of 1,000 random additions and tree bisection-reconnection swapping (gaps treated as missing, nchuck = 20, and chuckscore = 1), a bootstrap analysis (nreps = 500 start = stepwise addseq = random nreps = 5 nchuck = 10 chuckscore = 1) in PAUP* version 4.0b10 (Swofford 2002). The Templeton (Templeton 1983, Larson 1994) and winning-sites tests (Prager and Wilson 1988) were used to compare Sun et al. (2006) accepted intuitive topologies to the most parsimonious tree (or consensus tree) computed in our reanalysis of their sequences.

Sun et al. (2006) chose to use the Shimodaira–Hasegawa (S/H) test (Shimodaira and Hasegawa 1999) to investigate whether their intuitive hypotheses of mayfly relationships were significantly different from the tree they reconstructed from the molecular data. Sun et al. (2006) compared their 18S molecular topology to each of their intuitive topologies one at a time. When the test showed that their intuitive phylogeny was not significantly different from the 18S topology, they accepted the intuitive topology's relationships as a credible estimate of mayfly phylogeny. However, the authors misused this test, because the S/H test is only appropriate for likelihood analyses when all a priori topologies are simultaneously compared (Goldman et al. 2000, Shimodaira 2002). More appropriate tests would have been the Templeton (Templeton 1983, Larson 1994) and winning-sites tests (Prager and Wilson 1988) for comparisons of their intuitive topologies to molecular topologies derived from parsimony analysis. We performed these latter tests on the three intuitive topologies deemed credible by Sun et al. (2006) by comparing these topologies to trees generated from the 18S data.

Results

Reanalysis of Sun et al. (2006) Sequences (Fig. 1). The reanalysis of the Sun et al. (2006) sequences resulted in an alignment consisting of 616 aligned characters: 422 characters were constant, 113 variable characters were parsimony-uninformative, and 81 characters were parsimony-informative. The heuristic search (6,846,620 rearrangements tried) resulted in 117 most parsimonious trees with a length of 323 steps. A 50% majority rule consensus topology is presented in Fig. 1. This topology does not support the monophyly of the suborders Furcatergalia, Pisciforma, or Setisura as suggested by Sun et al. (2006), nor does it support the position of Baetiscidae at the base of the tree, as sister group to the remaining mayfly taxa.

The intuitive trees 1, 9, and 13 of Sun et al. (2006) (depicted in their figs. 1, 5, and 7) correspond to three topologies they judged to be congruent with their 18S data based on these topologies passing the S/H test. These topologies were apparently based on a combination of morphological characters and intuitive notions, though they are vague on this point. We performed a Templeton test on these topologies by comparing them to the strict consensus of the 18S rDNA topology (Fig. 1), and we found that they were significantly different, having *P* values of 0.0243, 0.0166, and 0.0113, respectively. We performed the winning-sites tests on these topologies and found that although these trees were not significantly different
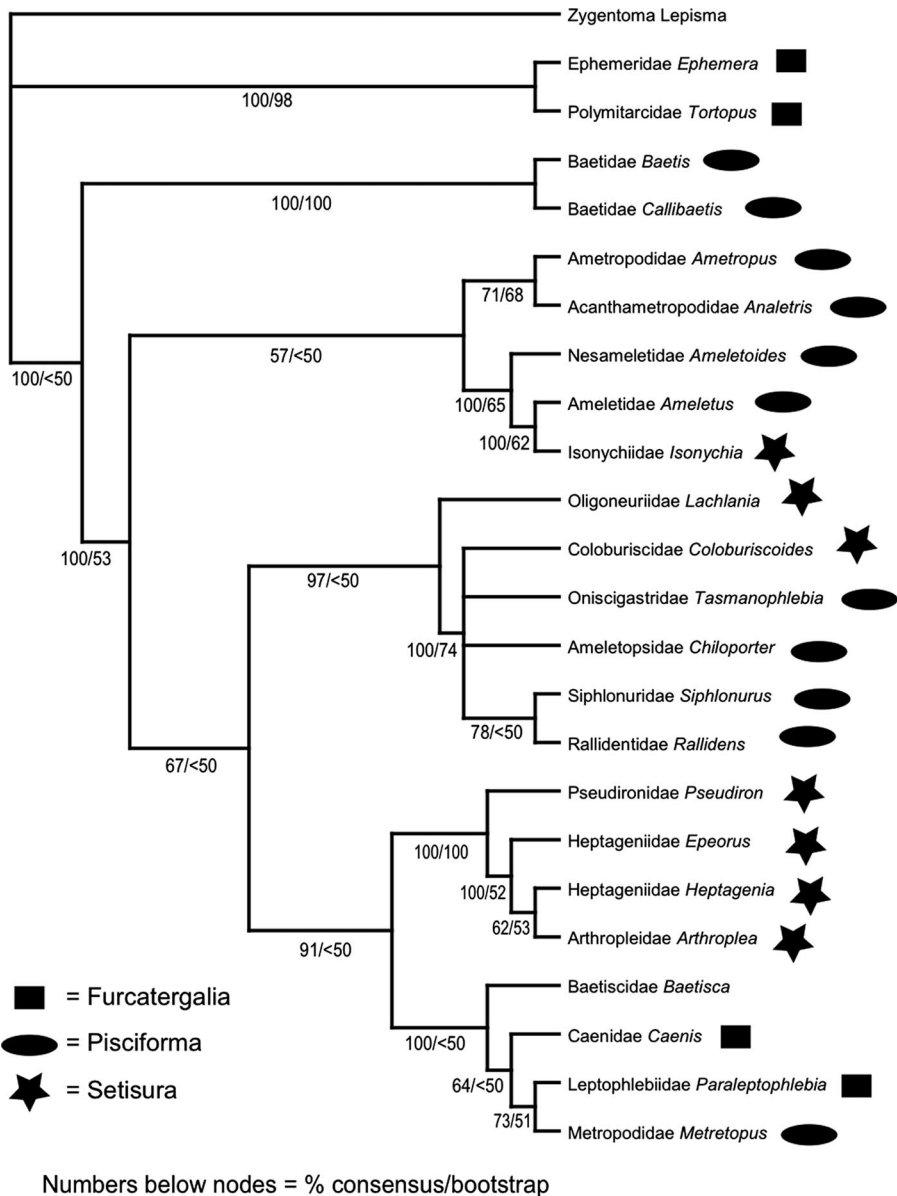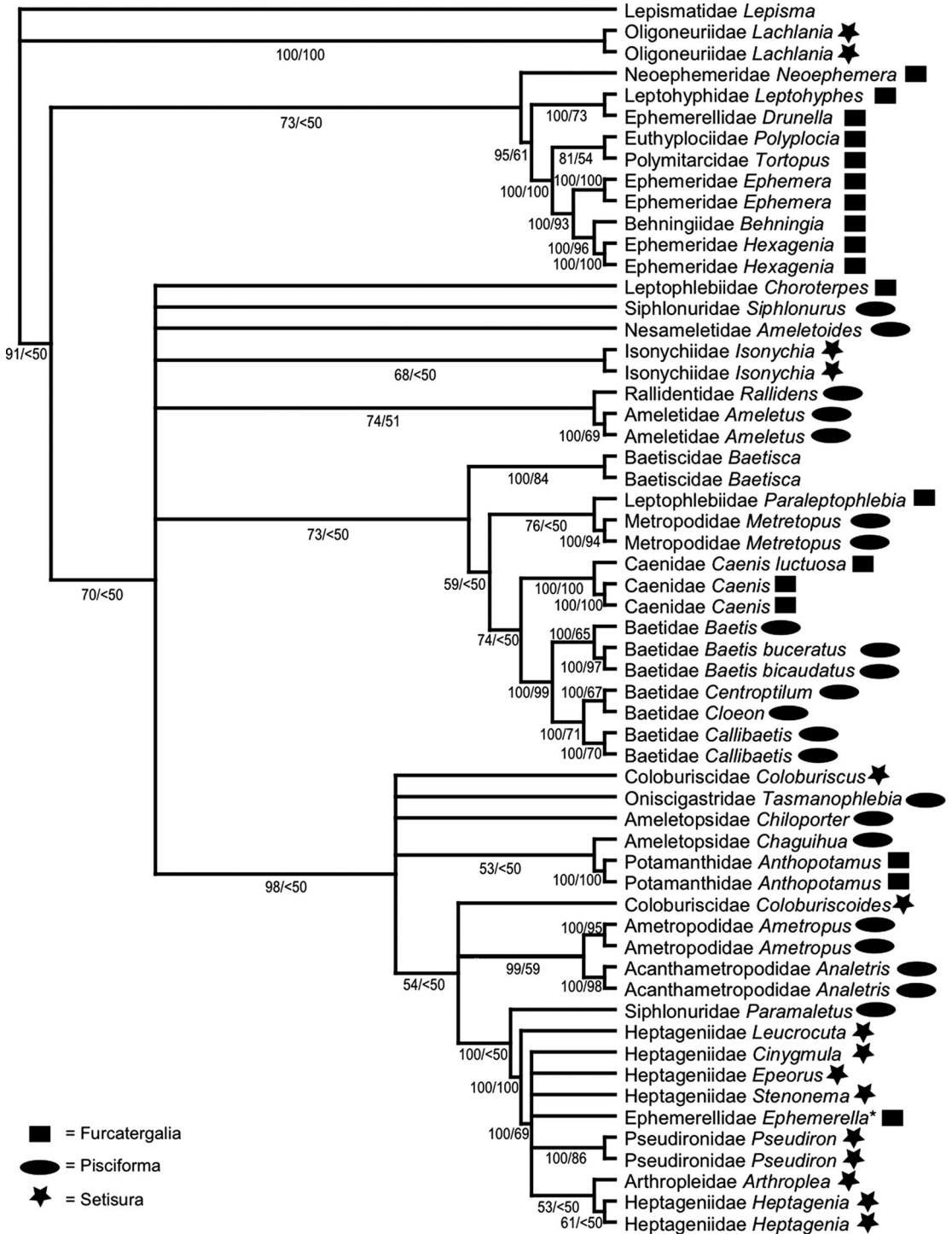
Fig. 1. Reanalysis of the Sun et al. (2006) original 18S rDNA data set. Majority rule consensus of the 117 most parsimonious topologies. The numbers below the nodes indicate the percent consensus (before the slash) and the bootstrap values (after the slash). Taxa belonging to the traditionally proposed suborders Furcatergalia, Pisciforma, and Setisura are indicated by the rectangle, oval, and star, respectively. Note that 18S data do not support the monophyly of the suborders.

($P = 0.1516$, 0.0931, and 0.0525, respectively), the P values were just barely above the threshold value of 0.05. When we take their 18S data matrix and map it on their intuitive topologies, the resultant trees are 368, 366, and 367 steps, respectively, compared with the length of 323 for the most parsimonious tree estimated from these data. Thus, their topologies require an additional 43 steps to explain these data, an increase of ≈6%. Contrary to Sun et al. (2006), we conclude that their 18S data do not support their intuitive hypotheses of higher level mayfly relationships.

Another relationship that is emphasized by the authors is of the monophyly of the families Heptageniidae, Arthropleidae, and Pseudironidae. Their trees 1 [fig. 1 in Sun et al. 2006)] and 9 [fig. 5 in Sun et al. (2006)] do not support a monophyletic Heptageniidae, because it is rendered polyphyletic by *Arthroplea*. However, their tree 13 [fig. 7 in Sun et al. (2006)] was manually adjusted to place *Arthroplea* outside the two heptageniid taxa, thus giving the appearance that Heptageniidae is supported as monophyletic by the molecular data. They conclude

Fig. 2. Analysis of all 18S rDNA sequences available to Sun et al. (2006) 18 mo before their manuscript submission. Majority rule consensus of the 8,160 most parsimonious topologies. The numbers below the nodes indicate the percent consensus (before the slash) and the bootstrap values (after the slash). Taxa belonging to the traditionally proposed suborders Furcatergalia, Pisciforma, and Setisura are indicated by the rectangle, oval, and star, respectively. Note that 18S data do not support the monophyly of the suborders. Asterisk (*) indicates the taxon *Ephemerella* is most likely misidentified, and based on Blast search hits and placement in this tree, is probably a genus within the family Heptageniidae.

that *Arthroplea* is deserving of familial status. However, our reanalysis of their sequences supports the nesting of *Arthroplea* within the Heptageniidae and thus contradicts their elevation of this group to familial status.

Analysis of All 18S Sequences Available to Sun et al. (2006) (Fig. 2). We performed a parsimony analyses of all 59 mayfly18S rDNA sequences, consisting of 37 sequences from GenBank that were available to Sun et al (2006) 18 mo before submission of their manuscript and the 22 sequences generated by Sun et al (2006). We aligned the sequences as described above, and this resulted in 1,945 aligned characters: 1,355 characters were constant, 236 variable characters were parsimony-uninformative, and 354 characters were parsimony-informative. Parsimony analysis with gaps treated as missing resulted in 8,160 most parsimonious topologies with a length of 1,485; Fig. 2 depicts a 50% majority rule consensus of these trees. Bootstrap values and consensus values are indicated for each node. Similarly, this topology does not support the monophyly of the suborders Furcatergalia, Pisciforma, or Setisura. It also does not support Baetiscidae as sister group to the remaining mayfly taxa. Finally, with an increased taxon sampling from Heptageniidae, the topology strongly supports *Pseudiron* and *Arthroplea* as nesting within Heptageniidae, and consequently does not support the elevation of these genera to familial status. Note that one taxon downloaded from GenBank (labeled *Ephemerella*) nests within the heptageniids. We think the identification of this specimen was in error as the top Blast hits for the 18S and 28S DNA sequence data from this specimen are all other heptageniids.

Discussion

Using the sequences generated by Sun et al (2006), we were not able to obtain a topology that supports the monophyly of the suborders Setisura, Pisciforma, and Furcatergalia, the placement of Baetiscidae as the sister lineage to all other represented clades, and *Pseudiron* and *Arthroplea* as separate monophyletic families. Likewise, the analyses of all 18S sequences available to the authors also did not support the monophyly of these groups. Because more data, particularly from multiple molecular markers, should offer a more accurate view of phylogeny (Pamilo and Nei 1988, Kluge 1989, Bull et al. 1993, Weller et al. 1994, Huelsenbeck et al. 1996, Baker and DeSalle 1997, Wiens 1998, Mitchell et al. 2000, Gadagkar et al. 2005), we want to emphasize a few points. Although our reanalyses of the 18S data as described above do not, in fact, support a monophyletic Furcatergalia, the results of Ogden and Whiting (2005), which consisted of a denser taxon sampling (94 species of mayflies) with five genetic markers (18S, 28S, 16S, 12S, and Histone3), support the monophyly of this group, but they contradict the notion of a monophyletic Setisura and Pisciforma. This conclusion, however, was reached because the additional data provided the signal to support this idea, not because of some a posteriori modification of a tree

that, in some way, does not agree with preconceived notions of relationships.

Further discussion of the most basal ephemeropteran clade (or the lineage that is sister to all other mayflies) also is merited. Sun et al. (2006) state that "Baetiscidae was just moved to the base." However, their data do not support Baetiscidae as sister to all remaining mayflies. In the reanalysis of their data, the burrowing mayflies (Ephemeridae + Polymitarcyidae) were supported as sister to all other lineages. The second larger 18S dataset supported Oligoneuriidae as the sister lineage to all other represented clades. The analysis of Ogden and Whiting (2005) support Baetidae as sister to other sampled families with Baetiscidae nesting more apically in the topology. This more comprehensive analysis, which included 10 additional heptageniid taxa, also demonstrated that both *Pseudiron* and *Arthroplea* nest well within the family Heptageniidae.

Many of the major conclusions of Sun et al. (2006) regarding relationships within Ephemeroptera are not supported by their own data, all 18S data available to them, or by more data for more taxa. The authors ignored data pertinent to their study and did not publish their own resultant molecular phylogeny. These authors did not use the appropriate test to compare their morphological "trees" with those generated by the molecular data, and when the appropriate test is used it is clear that the molecular data do not support their notions of phylogeny. Additionally, the practice of manually excluding taxa and moving branches around after the phylogenetic analysis is complete to get the desired relationships, is questionable at best.

References Cited

Baker, R. H., and R. DeSalle. 1997. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. Syst. Biol. 46: 654–673.

Ball, S. L., P.D.N. Hebert, S. K. Burian, and J. M. Webb. 2005. Biological identifications of mayflies (Ephemeroptera) using DNA barcodes. J. North Am. Benthol. Soc. 24: 508–524.

Bull, J. J., J. P. Huelsenbeck, C. W. Cunningham, D. L. Swofford, and P. J. Waddell. 1993. Partitioning and combining data in phylogenetic analysis. Syst. Biol. 42: 384–397.

Danforth, B. N., C.-P. Lin, and J. Fang. 2005. How do insect nuclear ribosomal genes compare to protein-coding genes in phylogenetic utility and nucleotide substitution patterns? Syst. Entomol. 30: 549–562.

Gadagkar, S. R., M. S. Rosenberg, and S. Kumar. 2005. Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. J. Exp. Zool. B Mol. Dev. Evol. 304B: 64–74.

Goldman, N., J. P. Anderson, and A. G. Rodrigo. 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. 49: 652–670.

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

Huelsenbeck, J. P., J. J. Bull, and W. Cunningham. 1996. Combining data in phylogenetic analysis. Trend. Ecol. Evol. 11: 152–158.

Kluge, A. J. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). Syst. Zool. 38: 7–25.

Larson, A. 1994. The comparison of morphological and molecular data in phylogenetic systematics, pp. 371–390. *In* B. Schierwater, B. Streit, G. P. Wagner, and R. DeSalle [lssb]eds.], Molecular ecology and evolution: approaches and applications. Birkhauser Verlag, Basel, Switzerland.

Mitchell, A., C. Mitter, and J. C. Regier. 2000. More taxa or more characters revisited: combining data from nuclear protein-encoding genes for phylogenetic analyses of Noctuoidea (Insecta: Lepidoptera). Syst. Biol. 49: 202–224.

Monaghan, M. T., J.-L. Gattolliat, M. Sartori, J.-M. Elouard, H. James, P. Derleth, O. Glaizot, F. de Moor, and A. P. Vogler. 2005. Trans-oceanic and endemic origins of the small minnow mayflies (Ephemeroptera, Baetidae) of Madagascar. Proc. R. Soc. B Biol. Sci. 272: 1829–1836.

Ogden, T. H., and M. Whiting. 2003. The problem with "the Paleoptera problem": sense and sensitivity. Cladistics 19: 432–442.

Ogden, T. H., and M. F. Whiting. 2005. Phylogeny of Ephemeroptera (mayflies) based on molecular evidence. Mol. Phylogenet. Evol. 37: 625–643.

Pamilo, P., and M. Nei. 1988. Relationships between gene trees and species trees. Mol. Biol. Evol. 5: 568–583.

Prager, E. M., and A. C. Wilson. 1988. Ancient origin of lactalbumin from lysozyme: analysis of DNA and amino acid sequences. J. Mol. Evol. 27: 326–335.

Rebora, M., L. Lucentini, A. Palomba, F. Panara, and E. Gaino. 2005. Genetic differentiation among populations of *Baetis rhodani* (Ephemeroptera, Baetidae) in three Italian streams. Ital. J. Zool. 72: 121–126.

Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51: 492–508.

Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications for phylogenetic inference. Mol. Biol. Evol. 16: 1114–1116.

Sun, L., A. Sabo, M. D. Meyer, R. P. Randolph, L. M. Jacobus, W. P. McCafferty, and V. R. Ferris. 2006. Tests of current hypotheses of mayfly (Ephemeroptera) phylogeny using molecular (18s rDNA) data. Ann. Entomol. Soc. Am. 99: 241–252.

Swofford, D. L. 2002. PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sinauer Associates, Sunderland, MA.

Templeton, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. Evolution 37: 221–244.

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

Tojo, K., and K. Matsukawa. 2003. A description of the second species of the family Dipteromimidae (Insecta, Ephemeroptera), and genetic relationships of two dipteromimid mayflies inferred from mitochondrial 16S rDNA gene sequences. Zool. Sci. 20: 1249–1259.

Weller, S. J., D. P. Pashley, J. A. Martin, and J. L. Constable. 1994. Phylogeny of noctuoid moths and the utility of combining independent nuclear and mitochondrial genes. Syst. Biol. 43: 194–211.

Wiens, J. J. 1998. Combining data sets with different phylogenetic histories. Syst. Biol. 47: 568–581.