Poor taxon sampling, poor character sampling, and non-repeatable analyses of a contrived dataset do not provide a more credible estimate of insect phylogeny: a reply to Kjer

T. Heath Ogden¹,²,*, Michael F. Whiting¹ and Ward C. Wheeler³

¹Department of Integrative Biology, 401 WIDB, Brigham Young University, Provo, Utah 84602 USA;
²School of Life Sciences, PO Box 874501, Arizona State University, Tempe, AZ 85287-4501, USA;
³Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024-5192, USA

Accepted 9 March 2005

Abstract

The wealth of data available for phylogenetic analysis of the insect orders, from both morphological and molecular sources, is steadily increasing. However, controversy exists among the methodologies one can use to reconstruct ordinal relationships. Recently, Kjer (2004) presented an analysis of insect ordinal relationships based exclusively on a single source of information: 18S rDNA sequence data. Kjer claims that his analysis resulted in a more “credible” phylogeny for the insect orders and strongly criticized our previous phylogenetic results. However, Kjer only used a subset of the data that are currently available for insect ordinal phylogeny, misrepresented our analyses, and omitted other analyses we have published on insect ordinal phylogeny. In our estimation, Kjer did a poor job of representing the current state of affairs in insect ordinal phylogenetics. Furthermore, we examine a number of analytical issues that are relevant not only for insect phylogeny, but systematics as a science, such as: repeatability and objectivity, locating alignment boundaries, secondary structure, goodness of fit measure, epistemological coherence, practicality and homology.

our analyses, and omitted other analyses we have published on insect ordinal phylogeny. In our estimation, Kjer did a poor job of representing the current state of affairs in insect ordinal phylogenetics.

**Empirical issues**

The first important point to recognize is that Kjer (2004) did not present any new evidence, but relied on a contrived matrix assembled from a subset of the 18S sequences available on GenBank. His “laboratory’s interest in Paleoptera” appears to be limited to downloading subsets of 18S sequences for Odonata and Ephemeroptera, as he brought no new information to the table, and actually neglected important taxa in his analysis for which 18S sequence data are freely available. Given that he presented no additional data, it seems odd to criticize us for not having “large molecular data sets” underlying our results, particularly since these data sets (Whiting et al., 1997; Wheeler et al., 2001), constituted some of the largest data to date addressing ordinal relationships. It is unclear to us why a treatment of published data but with fewer taxa, fewer loci, and absent morphological information represents an advance on the primary literature.

Kjer presented his paper as if only a single methodology (Direct Optimization) had been applied to questions in insect interordinal phylogeny by our group, and he is generally anachronistic and inaccurate in discussing our analyses and results. While Direct Optimization (DO) is currently our favored methodology (for reasons discussed below), in fairness it is important to recognize that multiple methods have been applied to different molecular and morphological datasets as these methods have been developed. For instance, Whiting et al. (1997) presented a multiple alignment for two genes (18S and 28S) that was derived from a multiple alignment program (MALIGN), not a DO program. These multiple alignments were published in the original paper as well as posted on the Systematic Biology website, and these “homology statements” have been widely reanalyzed in a variety of contexts (Huelsenbeck, 1997, 1998; Hwang et al., 1998; Whiting, 2002c). It is ironic that Kjer criticized us for not using Bayesian analysis in our 1997 paper, when Bayesian analysis was not made available until 2001 (Huelsenbeck and Ronquist, 2001). Likewise, it is inaccurate to insinuate that our data sets were not made generally available to the public allowing others to examine them in terms of a “hypothesis of homology”. In 2002, Whiting edited a volume on holometabolan relationships in which Kjer (Kjer et al., 2002) contributed results from his work on Trichoptera (referenced in Kjer, 2004), and Whiting (Whiting, 2002b) contributed results from a reanalysis of 182 18S sequences for insects (not referenced in Kjer, 2004).

In this analysis, POY was only used as a tool to generate multiple alignments, with the conserved portions of the alignment being used to resolve interordinal relationships. This paper discusses the pitfalls of using 18S as a single marker for interordinal relationships with some specific examples of where 18S fails, but Kjer did not discuss this paper nor place his results within the context of our larger analysis. In 2003, we published a paper placing stick insects among polyneopterous insects (Whiting et al., 2003) using three molecular markers (18S, 28S and H3) and multiple methods of tree reconstruction including DO, parsimony, Bayesian, and likelihood analysis. This paper provided additional insight into ordinal relationships among the basal Neoptera, but was also not referenced by Kjer. In 2003, we published an extensive analysis of Paleoptera using the three molecular markers above plus morphology, and comparing ClustalX (Thompson et al., 1997) versus DO methods of alignment, and multiple methods of tree reconstruction. Kjer (2004) cited this paper but ignored the underlying data (he failed to include these sequences in his analysis), and dismissed our finding that the current arsenal of molecular data does not provide a robust solution to the Paleoptera problem, by stating this is a case where “the methods failed, not the data.” This simplistic vision makes for a convenient argument—by dismissing DO, one can dismiss all of our prior work on insect phylogeny—but Kjer has failed to provide a complete appraisal of the data available for insect ordinal relationships and has ignored other analyses that would complicate his findings.

**A wealth of data**

What data are available for inferring insect ordinal phylogeny? Given Kjer’s description, a person unfamiliar with insect systematics would fail to recognize the wealth of data available from multiple sources. According to Kjer (2004), from a molecular standpoint there is only 18S data and a small fragment of 28S (d3 region, ~300 bp); the latter molecule he insisted is useless for interordinal phylogeny. As of November 2003, there were 1849 18S rDNA sequences of at least 637 bp (the shortest length included by Kjer) representing 1504 hexapod species. But there were also 1638 sequences of 28S of at least 637 bp for insects, and in many cases these are nearly complete sequences (2500 bp) for a wide diversity of insects. Moreover, previous studies have demonstrated that 28S carries much more signal than 18S for interordinal relationships (Whiting, 2002a; Ogden and Whiting, 2003; Svenson and Whiting, 2004; Whiting and Whiting, 2004), and we find the same pattern for our ongoing studies on Odonata, Ephemeroptera, Grylloblattodea, Plecoptera, Siphonaptera, Thysanoptera, and other insect groups. It is unfortunate
that Kjer did not also download 28S rDNA data to provide a molecular result that was more comparable to our own, and to truly demonstrate that his methods of data analysis produce more "credible" results.

Kjer also failed to discuss the morphological matrix of 275 characters that was coded for all insect orders, originally in Whiting et al. (1997), and more comprehensively by Wheeler et al. (2001). The value in this matrix is that we brought together, for the first time, character descriptions given by Kristensen, Boudreaux, Hennig, and other workers, provided additional character descriptions ourselves, and coded these across all of the insect orders. Coding matrices is vastly superior to simple narratives, which consist of character descriptions in a subset of which are never rigorously evaluated, since it allows the formal assessment of congruence and allows the direct combination of this evidence with molecular data. This matrix has been subsequently expanded and revised by other workers (Beutel and Gorb, 2001) who have understood the value of explicit character coding. It is thus ironic that Kjer should make claims about producing a topology more congruent with "traditional ideas", while providing no indication of what these traditional views might be nor how he assessed relative congruence, and all the while ignoring other assembled data matrices which speak directly to the issue.

### Taxon sampling

The taxon sampling strategy of Kjer was to obtain "as complete a sampling of non-holometabolous insects as possible", while also limiting the size of the taxon sample" with an "extended sampling of Odonata and Ephemeroptera". For the other orders "randomly selected divergent taxa were used". He is thus left with the challenge of including sufficient taxa to capture insect ordinal diversity, but not so many as to overwhelm his "labor intensive" process of manual alignment. This highlights a fundamental problem with his methodology: as one gathers more data, the ability to manually align the information becomes logistically more difficult, less objective, and less repeatable. While the proponents of manual alignment claim to handle effectively only around 100 sequences, how does one deal with hundreds or even thousands of sequences, when you can’t fit them all on a single computer screen? Recognizing the vast diversity represented by insect ordinal diversity, Kjer is always forced to deal with woefully inadequate sampling of taxa with no prospects of ever significantly increasing the size of the data set. Clearly it has been demonstrated that taxon sampling is crucial in phylogenetic studies (Pollock et al., 2002; Zwickl and Hillis, 2002).

How well does Kjer represent insect diversity with the GenBank data available to him? His analysis includes 132 sequences. This seems like a broad coverage, unless one views the data available on GenBank to him at the time of submission. Of the 1849 18S rDNA sequences available, he included roughly only 8% of the available species, but more critically omitted some important taxa that have been shown to play a pivotal role in insect phylogeny. For instance, he inexplicably omitted the order Thysanoptera, whose position in the Paraneoptera has been difficult to establish with 18S data alone. He similarly excluded the order Neuroptera, by far the most diverse clade within Neuropteroidea, even though 12 sequences were available. He purposely omitted Strepsiptera by arguing that other workers have demonstrated that 18S does not provide an adequate signal for the placement of Strepsiptera. However, he failed to recognize that the references he cited to argue this point (Huelsenbeck, 1998; Hwang et al., 1998) rely exclusively on the alignment published by Whiting et al. (1997), and thus he avoided an opportunity to demonstrate the superiority of his methods. He omitted *Timema*, a basal stick insect whose inclusion is critical in polarizing groups within Phasmida and linking it with Embiidiina (Whiting et al., 2003). He omitted *Nannochorista*, a basal mecopteran that may warrant ordinal status (Whiting, 2002a), and is important for establishing the monophyly of Antiaphora. If his method of analysis were truly superior, then why omit the orders which have been the greatest challenge to place, and why omit taxa whose inclusion have been shown to be critical towards understanding patterns of insect diversification?

The sampling of Kjer within orders is quite poor. For Coleoptera, he included two sequences (out of a possible 598), which appear as sister groups on his tree. He then launched into a diatribe and claims that the paraphyletic Coleoptera, as published by Whiting et al. (1997), was due to a contaminated 18S sequence. But he failed to cite the paper (Whiting, 2002c) which discusses these possible contaminants and also demonstrates that, regardless of analytical methodology, Coleoptera is always paraphyletic under 18S, and that their reported monophyly (based on molecular data) is an artifact of inadequate taxon sampling (Caterino et al., 2002; Whiting, 2002b; Whiting, 2002c). Within Phasmida, he included only two of the 14 major lineages available. Even within Odonata and Ephemeroptera, the two groups he claimed to have sampled the most densely, he only included 63% and 34% of the available sequences, respectively. We also find his claim of randomly selecting taxa within each order to be suspect, given the avoidance of sequences generated by our group (even if this means neglecting key taxa such as those discussed above). For instance, within Phasmida, there are 44 sequences, 41 of which we generated in a previous study (Whiting et al., 2003), and Kjer "randomly selected" two sequences generated in different studies. The same pattern is seen for Ephemeroptera, Odonata,
and other groups. A $\chi^2$ goodness of fit test indicates that it is highly improbable that this sample is random for these groups ($P < 0.05$). Moreover, his methodology eliminates a major portion of the data, and in this case he eliminated regions he designated as “unalignable”, which accounted for more than 50% of the possible parsimony informative sites given a ClustalX alignment. The Kjer (2004) data matrix is contrived and fails to take into account all of the complexity of the data available. While we agree that a limited taxon size may be required by the “labor intensive” manual alignment process, we do not think that 8% of available sequences, with half of the information removed, is adequate to the task.

**Basal pterygotes—group of emphasis**

Kjer’s emphasis concerning investigation of the relationships of the basal pterygotes prompts greater examination of the results from his methodology. The secondary structure manual alignment and Bayesian analysis of Kjer supports *Tricholepedion gertschi*, rendering Pterygota non-monophyletic in 62% of the trees. He suggested that this placement “should not be taken seriously” based on the results of other analyses. We similarly suggest that his result of Odonata as sister to (Ephemeroptera + Neoptera) should not be taken seriously, due to additional evidence and analyses that are available (Kristensen, 1991; Whiting et al., 1997; Fürst von Lieven, 2000; Stamiczek, 2000; Wheeler et al., 2001; Ogden and Whiting, 2003). It is not surprising that any one partition and any one methodology recovers a particular arrangement for the basal pterygotes, as was shown in Ogden and Whiting (2003). We demonstrated that, for these relationships, the current suite of molecular data, treated as partitions or simultaneously, did not come to a robust solution across various alignment and tree reconstruction methods such as DO, ClustalX, Parsimony, Likelihood, Bayesian, and MetaPIGA. Nevertheless, a combined analysis, which included morphological characters, strongly supported a robust topology recovering Ephemeroptera as sister to the remaining pterygotes under the various methods. Kjer stated that in this case “the methods failed, not the data”. This statement is incorrect because the analyses clearly showed that any unresolved or non-robust nodes resulted from a lack of signal from the data or conflict (homoplasy), not from a tree reconstruction methodology error. Ironically, his topology suggested that Anisoptera is not monophyletic. However, this is not because the method of Bayesian analysis failed, but because of a “lack of change on terminal branches” or in other words, insufficient molecular autapomorphies from the reduced Kjer 18S matrix. We agree that the relationships recovered in his analysis among odonate taxa were unexpected and are in contradiction to morphological analyses (Rehn, 2003). Likewise, the placement of the roach *Periplaneta* as sister group to Mantodea rather than to the Isoptera + Cryptocercus clade (Maekawa et al., 1999; Lo et al., 2000), the placement of Orthoptera as sister group to Holometabola + Paraneoptera (Hennig, 1969; Boudreaux, 1979; Kristensen, 1991; Kristensen, 1995; Kristensen, 1999), the placement of Phasmida in a clade separate from Embiidina (Matsuda, 1970; Rahle, 1970; Flook and Rowell, 1998; Whiting et al., 2003), and the placement of *Stenoperla* in a clade separate from *Zelanoperla* (Zwick, 2000), disagree with previous hypotheses and are in fact contradictory to “traditional ideas of insect ordinal relationships”. His monophyletic Mecopterida (which he calls “Remaining Panorpids”) and the Coleoptera + “Neuropteroidea” clade have never been supported in a molecular analysis that includes a broad sample of taxa (Whiting, 2002c,b) and are artifacts of his poor taxon sampling and data exclusion, and is a contrived result. Thus, his analysis of a subset of 18S sequences does not seem “credible” on all nodes, when compared to previous works, and he has provided no test of the accuracy of his method (simulation studies or comparisons to “known” phylogenies), and should be cautiously used to explore “vexing questions” in insect evolution.

**A specific example**

A clear example of the pitfalls of simple narratives emerges from Kjer’s discussion of the support for a basal placement of Odonata. Kjer discusses the “interesting scenario” of direct sperm transfer as a synapomorphy for (Ephemeroptera + Neoptera). We also coded this character in our morphological data matrix (Wheeler et al., 2001) and used it in our analysis of Paleoptera as discussed above (Ogden and Whiting, 2003). He omitted the discussion of morphological characters that contradict Odonata as sister to (Ephemeroptera + Neoptera), and neglected any discussion of characters in any explicit sense at all. For example, seven morphological characters, ignored by Kjer, support Ephemeroptera as sister to (Odonata + Neoptera): (1) the anterior articulation of the mandible is a non-permanent sliding groove and track system in Ephemeroptera, but in other pterygote lineages this articulation is more permanent; (2) subimago stage present in Ephemeroptera but absent in other pterygotes; (3) tracheation absent in arch of wing base and in posterior portion of the leg in Ephemeroptera, but present in other insects; (4) direct spiracular musculature absent in Ephemeroptera but present in odonates and neopterans; (5) never more than one tentorial-mandibular muscle in Odonata and Neoptera, but multiple muscles are present...
in Ephemeroptera; (6) annulated caudal filament presumably present in Archaeognatha, Monura, Zygentoma, and Ephemeroptera but absent in the remaining pterygotes; and (7) paired female genital openings retained in Ephemeroptera and nowhere else among Pterygota (Kristensen, 1991; Whiting et al., 1997; Fürst von Lieven, 2000; Staniczek, 2000; Wheeler et al., 2001). We agree that his “hypothesis is speculative, coming from a single gene” plus one morphological character, but emphasize that studies which take into account all available data—by coding the data in a formal matrix—contradict his conclusions. Kjer’s simple narratives are misleading and not fair to the body of data at hand.

Analytical issues

Systematics endeavors to achieve objective knowledge through hypothesis testing. How useful is Kjer’s methodology for furthering insect molecular systematics?

Repeatability and objectivity

Kjer’s method violates the criteria of repeatability and objectivity, in that it is not a transparent and explicit analytical procedure. Kjer’s method does not allow other investigators to repeat the experiment and test claimed results. Manual alignments generally lack any explicit discussion of how they are generated or the reasoning behind the chosen hypotheses of homology, and will be irreproducible and highly prone to bias, except in the most trivial of cases (Giribet et al., 2002). Kjer claimed that manual alignment is repeatable by stating “Anyone can repeat the analyses performed here by downloading the data and using the alignment.” But certainly this is not repeatability of alignment in any useful scientific sense, since it is the methodology that produces the alignment that must be replicated to make the alignment procedure repeatable, and not the subsequent analysis of a fixed alignment. Following his logic, any tree reconstruction method is repeatable—no matter how bizarre it may be—if one can download and examine the results from a website. Availability for download is not the hallmark of repeatability; independent investigators using a prescribed set of rules and arriving at the same end point is. Manual alignment could only be deemed repeatable if raw sequences, not the alignment, were downloaded (ideally, with the taxa names stripped, in order to blind the bias of the investigator) and identical alignments were reproduced time and time again.

Locating alignment boundaries

Kjer claimed that his method allows one to “locate the boundaries of unalignable regions according to repeatable criteria”, and cited his earlier paper (Kjer, 1997) on amphibians. The Kjer method locates these boundaries by “delimiting unalignable regions flanked by hydrogen-bonded stems” (Kjer, 2004), but provides no explanation of how this is actually accomplished, but he did reference his 1997 paper for a description of how this “repeatable” criterion is used. However, his 1997 paper provided no description of this criterion, but it does reference his 1995 paper (on frogs) which has an appendix with “Instructions on applying structural information to raw data” (Kjer, 1995). This appendix describes steps that allow one to take sequences and apply structural symbols to indicate hypothesized conserved stem and loop regions. However, this description lacks adequate information to explain how these regions are identified in the first place, how the boundaries are established, and relies more on intuition than algorithm (see his step 4). There is insufficient information in this appendix to code this methodology into any sort of automated algorithm, such that different workers would repeatedly find the same boundaries between alignable and unalignable data. Algorithmic approaches do already exist, in some form, for determining such boundaries (Castresana, 2000; Pei and Grishin, 2001).

Secondary structure

Kjer argued that ribosomal secondary structure provides an explicit, repeatable, and objectively defensible basis for performing manual alignments. However, several points should be considered further. First, secondary structure does not actually solve the problem of nucleotide homology. At best, it places constraints by establishing putative limits between loops and stems, but the nucleotides within each of those units must still be homologized (Giribet et al., 2002). Second, the determination of secondary structure is not nearly as simple and unambiguous as many studies suggest (Durbin et al., 1998). Indeed, in phylogenetic studies, secondary structure is typically inferred by aligning with a sequence of “known” secondary structure, although the basis of that knowledge remains uncertain in many cases. Kjer appropriately recognized several potential problems with secondary structure manual alignments such as “slippage”, “bulges”, “non-conserved stems”, and regions where the placements of nucleotides “remain arbitrary”, among others (Kjer, 1995, 1997). Third, although it might be reasonable to expect selective pressures to apply to secondary structure interactions (that is, requirements of compensatory changes), it is unclear just how relevant those interactions are, compared to selective pressures applied at other structural levels. Fourth, although functional constraint plays a role in preserving the pattern of shared ancestry, there is no necessary connection between functional considerations, including secondary structure, and the concept of
homology, which refers strictly to the historical identity of objects related through shared transformation events. Kjer (1995) claimed without evidence that, “structural features are more highly conserved than are nucleotides, and therefore structures are a better indication of homology than are nucleotides.” Recently, it has been shown that protein coding nucleotide sequences, while less conserved than the amino acids, were found to have a much greater phylogenetic signal (Kallersjo et al., 1999; Simmons et al., 2004). Conserved structural features are therefore not necessarily a better indication of homology (see Homology section below).

**Goodness of fit measure**

The most obvious thing that is lacking from the Kjer method is any sort of goodness of fit measure for a given alignment relative to a specific model of secondary structure. Aside from the issue of the applicability of his “custom arthropod rRNA secondary structural model” (which is never described) across all of insect diversity, if this methodology is to be useful, it must have some way of taking two different multiple alignments, comparing them head-to-head, and determining which one best fits a given secondary structure model. Kjer has never presented a metric that would allow an investigator to objectively “challenge and upgrade these hypotheses”, and it is unclear which specific criterion Kjer would use to demonstrate that one alignment is a better match to a model than another. Kjer stated that his hypothesized alignment “will be periodically updated”, but provided no way of determining if the new “updated” hypothesis was actually a superior alignment, based on any sort of measurable criterion. This simply underscores the fact that his methodology used neither an algorithmic criterion (such as ClustalX) or an optimality criterion (such as POY or MALIGN), but rather was dependent on some sort of intuition that lacks description and defies quantification.

**Epistemological coherence**

For scientific inferences to be valid, we believe that they must be methodologically, theoretically, and philosophically consistent. Empirical investigations must be firmly rooted in notions of evidence and inference, and they must describe and defend what is done, what is assumed, and why. These requirements, although crucial in science, are compromised by procedures such as the Kjer method of manual alignment of sequences, even in reference to secondary structure. Furthermore, how can the non-objective homology decision of manual alignment be carried over in a logically consistent framework to the tree reconstruction phase? For example, if the manual aligner really feels that a set of bases ought to be homologous, should a higher weight then be given to that character during tree reconstruction? Clearly there is no way to maintain an epistemological coherence throughout the entire process of manual alignment and tree reconstruction.

**Practicality**

While manual alignment falls short in these basic principles of scientific systematics, one can also question the practicality of manual alignment in the age of genomics. No automated approach to manual alignment is available and may never be, because the explicit rules of decision making have never been specifically articulated to allow for automation. Thus, manual methods used to align just one gene for a relatively small data set, would be futile for assembling the tree of life, particularly for insects, where data sets will undoubtedly reach to thousands of terminals for multiple genes in the near future. The issue of practicality most likely played a role in the paucity of 18S sequences selected by Kjer.

**Homology**

Cladograms imply statements of homology. Alternative cladograms might have alternative optimal homology statements and content. Features are homologous when their origins can be traced to a unique transformation on the branch of a cladogram leading to their most recent common ancestor. There can be no notion of homology without reference to a cladogram (albeit implicitly) and no choice among cladograms without statements of homology. So although Kjer suggests, “homology statements are found in alignments”, a cladogram is necessary to legitimize or test those generated statements. Although homology assessment often involves a two-stage procedure of first submitting each hypothesis of homology to a round of separate tests and then submitting the surviving, constrained set of hypotheses to the test of character congruence (that is, “static” homology assessment), this separation is neither a methodological nor epistemological necessity. POY embodies the concept of dynamic homology (Wheeler, 2001, 2003), in which the test of character congruence is applied to the entire, unconstrained set of hypotheses of homology, thereby allowing entire transformation series to be discovered on the basis of a single optimality criterion. That is, dynamic homology employs the same procedure to discover both the character (in the traditional sense) and the character-state transformations within the character. Since the same optimality criterion is employed in both cladogram assessment and homology assessment, the globally optimal explanation of the observed variation is achieved by the minimum-cost (or most likely, under a likelihood optimality criterion) cladogram-plus-homology-scheme combination. Kjer incorrectly stated that
POY does not produce an alignment, and therefore does not allow an assessment of homology. Recognizing that alignments may be useful as visual representations of nucleotide homology, POY can produce an implied alignment by taking the dynamic homologies established through direct optimization and tracing them back through the cladogram, linking the unaligned sequence positions through the respective transformation series. Thus, an implied alignment is really just a means of visualizing nucleotide transformation series, and the optimal set of nucleotide homologies for a given data set is topology and parameter specific. Dynamic homology is a powerful conceptual approach to the study of highly simplified data types, such as DNA and amino acid sequences or simple morphological structures like annelid segments, where structural or developmental evidence that could allow a defensible choice among competing hypotheses of homology is either nonexistent or unavailable.

In summary, manual alignment is not repeatable, not objective, not epistemologically coherent, and not a practical method in the genomics age. Even if one does not agree with DO as a method of tree reconstruction, surely it is clear that an automated, optimality criterion driven approach is more appropriate to test hypotheses of homology, than one which is deeply rooted in intuition and relies on a methodology that cannot be repeated by any other worker.

Conclusions

In our previous papers, we have demonstrated the importance of taxon sampling in higher level phylogeny, the use of multiple molecular markers, the formal combination of morphology with molecules, and we have argued for the need to make all stages of analysis as objective, transparent, and reproducible as possible (Whiting et al., 1997; Wheeler et al., 2001; Ogden and Whiting, 2003). With his paper, Kjer argued for the opposite in each case. He has done an inadequate job of sampling taxa, has myopically focused on a single molecular marker while ignoring other data sources, has eliminated a major portion of the data using a non-repeatable methodology, and champions an opaque, inefficient, and non-reproducible method of data analysis. His “credible” topology is suspect at many levels, and he has done a poor job of summarizing the current data available for inferring insect ordinal relationships.

There are certainly merits in exploring the influence of any particular partition such as 18S for deciphering insect ordinal relationships. Likewise, there are merits in trying to use secondary structure within the context of an alignment algorithm. The past two decades of systematic research have demonstrated the importance of optimality criteria and have focused on algorithms that allow discrimination among multiple hypotheses. While systematists are known to argue over what is the most appropriate criterion and how it should best be evaluated, there is a general consensus that a quantifiable approach is vastly superior to an intuitive or authoritarian one. Kjer failed to find any way to algorithmically describe his procedure such that it could be reproduced by other researchers, he has never performed any experiments to demonstrate repeatability, and never presented any metric to test alternative alignment hypotheses and compare relative accuracy. Moreover, since the Kjer method cannot be readily applied to large data sets, the prospects of it becoming a vital tool in these days of phylogenomics are increasingly dim. We would consider the widespread adoption of his methods as a major setback towards a full and robust understanding of insect ordinal phylogeny. Fortunately, such an adoption will not happen, since the method is wholly unrepeatable. We further argue that the most robust estimate of ordinal level phylogenetic relationships comes from using all the available data in a robust and repeatable phylogenetic analysis framework.

Acknowledgements

Our thanks to Keith Crandall, Alison Whiting, Matthew Terry, Gonzalo Giribet, Quentin Wheeler, James Carpenter, Cyrille D’Haese, Kurt Pickett, and Katherina Dittmar de la Cruz for valuable discussion and comments and thanks to David Whiting for assistance with the statistical analyses.

References


Hennig, W., 1969. Die Stammesgeschichte der Insekten. Krammer, Frankfurt am Main, Germany.


