Arthropod Systematics & Phylogeny 159
64 (2) 159–164

Structural Homology in Ribosomal RNA, and a Deliberation on POY

KARL M. KJER[1] *, JOSEPH J. GILLESPIE[2] & KAREN A. OBER[3]

[1] Department of Ecology, Evolution and Natural Resources, Rutgers University, New Brunswick, NJ, 08901 [kjer@AESOP.Rutgers.edu]
[2] Virginia Bioinformatics Institute, Bioinformatics Facility, Washington Street, Virginia Tech., Blacksburg, VA 24061 [pvittata@hotmail.com]
[3] Department of Biology, College of the Holy Cross, Worcester, MA 01610 [kober@holycross.edu]
* Corresponding author

Received 27.ii.2006, accepted 7.ix.2006.
Available online at www.arthropod-systematics.de

> Abstract

Computer alignments have been said to be more objective and repeatable than manual alignments. Here we suggest that computer alignment methods, particularly those using a program called POY, suffer from a series of practical problems, and philosophical inconsistencies. Subjective decisions are still a part of POY analyses, but are less transparent. We point out the problems POY has with ancestral state reconstruction under conditions of nucleotide compositional bias, its problems when gaps are not uniformly distributed, and its problems when gaps are not independent of one-another. In ribosomal RNA (rRNA), the individual nucleotides are less important than the structure within which they are associated. This higher level of conservation dictates that structures should be aligned before nucleotides. We show with an empirical example that manual alignments can be more repeatable, more objective, and more accurate than POY analyses, and call into question the conclusions drawn from POY analyses of rRNA data.

> Key words

Alignment, rRNA, secondary structure, POY, sensitivity analyses.

1. Introduction

For molecular data, alignments represent hypotheses of homology (topographic homology hypotheses sensu KLASS 2001: 230). While often trivial, alignment can be problematic if there is a high level of length heterogeneity across sequences. There are many ways to approach the alignment problem, with preferences based on philosophical, practical and operational considerations. For example, one might insist on strictly repeatable and algorithmic methods, while another may simply "eyeball" an alignment that "looks good." The argument between algorithmic methods (computer alignments) and manual alignments is often presented as one of objectivity versus intuition, repeatability versus guesswork, science versus authoritarianism. However, this dichotomy is not so clear when one considers the decisions that need to be made, even for the most ardent supporters of computer alignments. For example, applying the same principles of repeatability, we must question whether one should analyze morphological data in unison with molecules when it is highly unlikely that any two morphologists would come to identical morphological data matrices, even when considering the same set of taxa. The selection of which taxa to include is another decision that needs to be made, and will obviously influence phylogenetic results. Gene choice is still another decision. For example, sampling regions of the nuclear small subunit ribosomal RNA (18S rRNA) favors grouping Ephemeroptera with Neoptera (KJER 2004), while sampling mitochondrial gene regions suggests a monophyletic Palaeoptera (KJER et al. 2006). Many morphologists favor Odonata + Neoptera. Moreover, within the rRNA-encoding genes are regions that are difficult to align by any means, even among closely related taxa (recently reviewed in GILLESPIE 2004). Should one include regions of the data that cannot be confidently aligned across recent taxa (for example, 25 species of *Leptonema* [Trichoptera]), when considering the phylogeny of Hexapoda? If not, which nucleotides should be included, and which should be excluded? We all

agree that subjectivity should be minimized, but if decisions must be made, which ones should be considered justified?

Numerical taxonomy revolutionized systematics in the 1960s and 70s, just as cladistics did a decade later. These revolutions were a direct assault on the authoritarianism that characterized the old systematics, in which relationships were "proclaimed," by "the expert," sometimes without character support. The promise of cladistics, and then molecular phylogenetics, was that these systems would remove subjectivity, turning systematics into a "real science" (like physics). We have all benefited from the respect and increasing support that has come to our field through the rigors of prescribed methodology. However, it may have been naive to argue that the process of decision making has been effectively eliminated with molecular data. We argue that if subjectivity can't be completely eliminated, then we should draw attention to our decisions and be as transparent as possible about them. We should be skeptical of our own results, as well as open to the opinions of others. We should accept that experience and expertise still have the potential to influence our results. It is understood that we never "know the truth" in phylogenetics; we can never know phylogenetic relationships with certainty. They are hypotheses only. Therefore if a phylogeny is to be of any use at all, it becomes a matter of influencing beliefs, and these beliefs are supported by the strength of the presented evidence.

Phylogenetic hypotheses must be rigorously evaluated rather than just philosophically approved. Hypotheses can be qualitatively supported by corroboration of multiple independent datasets, and quantitatively supported through indices such as bootstraps or posterior probabilities on (combined) datasets. We "believe" in the monophyly of Pterygota, largely on the basis of a single character; wings, which is deemed to be unlikely to have evolved multiple times in insects. (Of course, "having wings" produced many structural modifications. A well-compiled matrix could include many characters just from this system.) It may be that Pterygota is not monophyletic, but again, in order to convincingly show this (for most of us) the data would have to influence our beliefs rather than appeal to philosophical arguments for epistemological consistency. Disagreements over alignment have been argued on largely philosophical grounds (KJER 1995; WHEELER 1996; KJER 2004; OGDEN et al. 2005). Here we explore these arguments, and present our views.

The first question to explore is whether the alignment argument is important enough to invest the time to understand it, and whether morphologists should enter the debate. We argue that it is, because all phylogenetic methods assume homology. Ribosomal RNA is typically difficult to align, and rRNA is now, and will likely continue to be one of our most important phylogenetic markers due to organismal universality, ease in PCR amplification, and the mass of data that has already accumulated. Alignments, whether static or dynamic, are the data from which phylogenies are drawn, and hypotheses may collapse based on the placement of a single nucleotide. Alignment is critical for phylogenetic inference as a statement of homology. Finally, structurally aligned data (KJER 2004) produce different trees than data that are analyzed with POY (WHEELER et al. 2001). This also applies to morphological character systems, as exemplified by the comparison of different hypotheses on cockroach phylogeny in KLASS (2001).

We consider two broad divisions in alignment approaches: computer alignments, in which parameters are input into a computer (with the resultant alignment unadjusted) and manual alignments, in which columns of nucleotides are aligned together by eye (reviewed in KJER et al. in press). Computer alignment methods include programs like Clustal (THOMPSON et al. 1994) and Malign (WHEELER & GLADSTEIN 1994), as well as POY (GLADSTEIN & WHEELER 1997). Unlike the other methods, POY is not strictly an alignment program, but rather an analysis program that simultaneously produces a dynamic alignment and a phylogeny. Manual alignments include those that are usually initially aligned with the assistance of a computer, and then manually adjusted. This includes POY analyses that manually eliminate portions of the data as "unalignable", or manually subdivide the data that is entered into the computer program in blocks (e.g., GILLESPIE et al. 2005a). Another form of manual alignment is based on using the secondary structure of the molecule to dictate decisions about homology, and these alignments are referred to as structural alignments (e.g., KJER 1995; GILLESPIE 2004). There are promising computer methods that use structural information to guide alignments as well (e.g., NOTREDAME et al. 1997; GORODKIN et al. 2001; MISOF et al. 2003; HOFACKER et al. 2004; HOLMES 2004; NIEHUIS et al. 2006). These subdivisions of "manual vs. computer" will change as our algorithms develop. Eventually, a fully automated structural alignment may be implemented, and when this happens, the most important contrast among methods will be whether primary sequences, or secondary structures dictate alignment decisions of rRNA.

The debate over alignment methods is framed by some as one of objective algorithm-based "science" against intuition-driven authoritarianism. Computer alignments are thought to be both more objective, and more repeatable, while manual alignments are thought to be more "accurate" as evidenced by the number of computer alignments that are subsequently manually adjusted (KJER et al. in press). Until recently, these assumptions have never been tested, and we wonder if they are true.

2. Background

Even sequences that are the same length may require decisions about alignment, but when two sequences differ in length, minimally, we must insert gaps into the shorter of the two sequences and make decisions about homology. Length differences are characteristic of rRNA sequences, yet are relatively rare in protein-coding genes because of their codon organization, wherein insertions or deletions in groups of 1 or 2 would result in a frame shift. Workers who study the evolution of genes across, e.g., Metazoa may find difficult-to-align protein-coding genes, just as population geneticists may find difficult-to-align intron or non-coding regions; however, these genes (or their corresponding divergence rates) are not as commonly used by systematists. For systematists, alignment problems are almost synonymous with rRNA, especially the less conserved regions of the molecule typically referred to as "expansion segments" or "variable regions."

For computer alignments, a variety of parameters must be set by the investigator. For example, the user must input how costly it is to insert a gap into one sequence if the nucleotides do not match another sequence to which it is being aligned. "Where the gaps belong" is dependent on a ratio between the gap cost and the substitution cost. Under Needleman-Wunsch algorithms (Needleman & Wunsch 1970), there are points given for lining up identical nucleotides, and points subtracted for inserting gaps or assuming substitutions. If the gap cost is trivially low, the computer will freely insert gaps until it matches nucleotides of an identical state. If the gap cost is prohibitively high, the computer will resist inserting gaps, even when, to the human eye, it is obvious that the data are offset (misaligned). There must be some optimal gap cost between these two extremes. Objectively finding this gap cost (and/or other input parameters) is what sensitivity analysis (Farris 1969; Wheeler 1995; Whiting et al. 1997) was meant to do. The idea behind sensitivity analysis is to analyze the data under a variety of input parameters (such as the gap cost, described above), and select among them by some criterion. One such criterion is to minimize character incongruence by subjecting partitioned data to an ILD test (Farris et al. 1994; c.f. PHT test of Swofford 1995). Briefly, an ILD test measures the sum of tree lengths of partitioned datasets, and then compares this value with the tree length of the combined dataset. Presumably, given the assumption that there is one phylogeny, the "best" set of parameters will be revealed when the partitioned datasets are least incongruent with the combined analysis. The fundamental flaw with this idea, however, is that, in the real world, the permissiveness for insertions and/or deletions (indels) among sites in rRNA is not randomly distributed, but rather, clustered. Some regions are extremely invariant in length, such that no insertions or deletions have ever been observed, even in comparisons across kingdoms. These regions should require a nearly infinite gap cost. Adjacent to these conserved sequences are regions in which indels are exceedingly common, even in comparisons among closely related species. These regions would best be aligned with a low gap cost. Between these extremes are regions best recovered with every gap cost between near-zero and infinity. There are no ideal average fixed gap costs for rRNA (Kjer 1995) because ideally, every position has its own currently undefined gap cost (Kjer 2004), and this gap cost is not necessarily an integer or half integer. During evolution, for example, a deletion occurred in a certain position, or it did not. Any fixed probability assigned to this process *a posteriori* (by the analysing phylogeneticist) and accordingly any fixed gap cost assigned to some sequence are gross oversimplifiactions of reality. It is illogical to seek an optimum from a variety of unjustifiable analyses. If gap costs in rRNA are not fixed among sites, then selecting an optimum from a large number of meaningless analyses is equally meaningless. Sensitivity analysis is inconsistent with a philosophy of avoiding subjectivity, because it leads to subjective decisions about gap costs that are less transparent than "eyeballing" the data. If one wished to be philosophically consistent, then gap costs and transversion weights should be set to one (Grant & Kluge 2003).

3. POY vs. Structure

The current debate is largely between direct optimization (as implemented in POY) and structural alignments, although the field of direct optimization is expanding (Redelings & Suchard 2005). Direct optimization is a broad field of analysis, of which POY is currently the most commonly used program. The use of unadjusted Clustal alignments for rRNA is rare, and adjusting a Clustal alignment turns the alignment to a manual one. Malign is similarly rarely used, and no longer supported by its authors. In a survey of phylogenetic papers in *Systematic Biology*, *Molecular Biology and Evolution*, and *Cladistics* from the last three years, Kjer et al. (in press) found that 76% of the papers that utilized rRNA were manually aligned. POY is an implementation of direct optimization, or DO (Sankoff et al. 1973; Sankoff 1975; Sankoff & Cedergren 1983; Kruskal 1983), which is explained in detail in Wheeler (1996). Direct optimization is a good idea because homology is tree dependent. There are cases when insertions and deletions have occurred with such frequency that the only way their history could possibly be recovered is through a direct optimi-

zation approach with dense taxon sampling, accurately recovering ancestral states throughout the backbone of a tree. Structural alignments are limited to conserved portions of the molecules, which sometimes leaves the most length variable regions unaligned and discarded, or worse, arbitrarily aligned and left in. So why should anyone favor structural alignments? Because in rRNA, structure is conserved to a greater degree than are nucleotides (e.g., GUTELL et al. 1994). This can be seen by superimposing folded rRNA molecules from exceedingly distant taxa. They all fold into the same basic conserved structure, but when you look at the nucleotides that make up these hydrogen-bonded stems, they may share little or no similarity in their nucleotides. For the most part, ribosomal RNAs function on the basis of their structure, not their nucleotides (see recent review in NOLLER 2005). POY does offer an option to incorporate structural constraints, and this option is disucussed in KJER et al. (in press) and GILLESPIE et al. (2005a). Yet, as it is most commonly used, nucleotides (i.e., transformations) are the only thing that POY "sees," using fixed gap costs and other parameters that are currently undefined, and should vary across sites. The allegation that manual alignments are subjective and unrepeatable applies equally to a system in which trees are determined through a variety of arbitrary parameter searches.

Nucleotide compositional bias (a state in which the four nucleotides deviate substantially from 25% each) presents a severe challenge to computer alignments. Compositional bias reduces character complexity, already low in molecular data. If independent lineages develop similar compositional biases, then aligning a series of non-homologous "A"s together has the same effect as grouping taxa according to overall nucleotide composition. While the analysis may be strictly character-based, the results are phenetic. Another challenge presented by compositional bias that has been shown with empirical data, simulation studies (COLLINS et al. 1994), and mathematical proof (EYRE-WALKER 1998), is that parsimony severely under-represents the rare states in ancestral reconstructions under conditions of compositional bias and/or accelerated substitution rates. In other words, if you measure the percentages of each of the nucleotides in 10 taxa on a tree, you might find that G ranges from 3 to 8 percent across taxa. Then when you look at reconstructed sequences for these same 10 taxa, you would find that G is virtually absent. What happened to all the Gs? The situation is well illustrated in the tree figures presented by COLLINS et al. (1994), which show the reconstructed nucleotide compositions at all the internodes are drastically different from those present in terminal taxa. Counting reconstructed transformations is what POY does, and COLLINS et al. (1994), and EYRE-WALKER (1998) show that this doesn't work when nucleotide composition is biased and/or when rates are elevated. Variable regions in rRNA are defined by elevated substitution rates, and are commonly biased in nucleotide composition (e.g., GILLESPIE et al. 2005b).

Indels in rRNA are also commonly clustered together within variable regions. POY treats indels as independent events. So, if there were a region in which Taxon A had lost five nucleotides, and Taxon B had lost 9, POY would consider a minimum of five independent transformations linking Taxon A and Taxon B together. However, it is more parsimonious to assume a single loss of five nucleotides in the ancestor of A and B, followed by an additional loss of four nucleotides in Taxon B (two transformations). Worse yet, it is equally parsimonious to assume that Taxon A had lost five nucleotides as a single event, and Taxon B had lost nine independently (two transformations). Even with the "extendcost" option, (which allows for a reduced gap cost for the insertion of additional gaps, following an initial gap) any non-zero extendcost will inflate the cost of multiple simultaneous deletions.

To summarize, we would predict POY to fail under conditions of nucleotide compositional bias, and/or when gaps are not uniformly distributed, and/or when gaps are not independent of one-another, and/or in a molecule where the nucleotide composition is less important than the structure with which they are associated. Any one of these conditions would require that we look to the results of a POY analysis with extreme skepticism. All of these conditions are characteristic of rRNA.

These are strong opinions, but they remain untested. How would one compare the repeatability of one method with that of another? The old approach has been one of philosophical proclamation. A better approach would be to send a large number of datasets, with the taxon labels masked, to a variety of investigators. These investigators could be instructed to align that data with secondary structure and compare the results to a similarly blind POY analysis. If gap costs and other parameter selections are indeed arbitrary, we would predict that different investigators would arrive at different parameters and therefore different trees with the POY analyses. We would also predict that if there is an underlying conserved secondary structure to rRNA, then different investigators could find it, and their phylogenetic results would be more similar to one another because they are using a homology criterion that is not arbitrary. We did such an experiment (KJER et al. in press), although on a small scale. The three of us analyzed the entire mitochondrial large subunit rDNA gene (16S rRNA) for 18 mammals. This dataset had the added bonus of having a highly corroborated expected phylogeny: (Monotremes (Marsupials ((Perissodactyls Artiodactyls) (Baboon (Gibbon (Orangutans (Gorilla (Human Chimps)))))))). As predicted,

with the structural alignment, all three of us arrived at the same phylogeny with the exception of the Gorilla/Human/Chimp node. The node grouping Chimps and Humans (excluding Gorillas) was found to have a near zero branch length, was supported by a 39% bootstrap by two of us, and left as a polytomy by the other. Otherwise, the trees we recovered were the expected trees. Also, as predicted, we all selected different parameters for the POY analysis, and arrived upon different trees. None of us converged on the expected tree with POY. We were surprised to find that even when comparing the results of analyses with the same input parameters on the same dataset, none of us recovered the same tree. There are a number of reasons this could have occurred, but the most likely is that we did not perform enough replicates to converge on the best tree, just as an insufficient number of replicates in any heuristic tree search may fail to find the shortest trees. However, we performed between 10 and 100 replicates, which is a standard number, and the search that ran for 100 replicates recovered the longest trees. In this example, we suggest that POY is not an objective means of data analysis unless you have some objective means of selecting the input parameters, and it is not repeatable unless you are given those parameters in advance. Even with the input parameters set in advance, we recovered different trees from each investigator. Manual alignment may be subjective and unrepeatable, as W.C. Wheeler defines these terms, but POY is even more so. As a side note, morphological data are inherently unrepeatable by this strict definition. Decisions still need to be made. Expertise is still required. Under this set of conditions, the more objective approach is that which allows the reader to more readily evaluate the evidence. We suggest you compare the 18S rDNA alignment presented for insects on Kjer's website, from which the results of Kjer (2004) were taken, with the results of Wheeler et al. (2001). Which of the trees from Wheeler et al. (2001) shall we favor? The 1:1:1 tree? The strict consensus? The *discussion* tree? These are decisions best made in light of the data, which is best aligned, visualized and evaluated by its structure.

4. Acknowledgements

We thank Klaus-Dieter Klass for inviting us to participate in this forum and facilitating review. We thank Kenneth (Tripp) MacDonald, Klaus-Dieter Klass, Stephen Cameron and an anonymous reviewer for helpful comments on the manuscript. Financial support from NSF DEB 0423834, and the New Jersey Agricultural Experiment Station, is appreciated.

5. References

Collins, T.M., P.H. Wimberger & G. Naylor 1994. Compositional bias, character-state bias, and character-state reconstruction using parsimony. – Systematic Biology **43**: 482–496.

Eyre-Walker, A. 1998. Problems with parsimony in sequences of biased base composition. – Journal of Molecular Evolution **47**: 686–690.

Farris, J.S. 1969. A successive approximates approach to character weighting. – Systematic Zoology **18**: 374–385.

Farris, J.S., M. Källersjö, A.G. Kluge & C. Bult 1994. Testing significance of incongruence. – Cladistics **10**: 315–319.

Gillespie, J.J. 2004. Characterizing regions of ambiguous alignment caused by the expansion and contraction of hairpin-stem loops in ribosomal RNA molecules. – Molecular Phylogenetics and Evolution **33**: 936–943.

Gillespie, J.J., M.J. Yoder & R.A. Wharton 2005a. Predicted secondary structures for 28S and 18S rRNA from Ichneumonoidea (Insecta: Hymenoptera: Apocrita): Impact on sequence alignment and phylogeny estimation. – Journal of Molecular Evolution **61**: 114–137.

Gillespie, J.J. C.H. McKenna, M.J. Yoder, R.R. Gutell, J.S. Johnston, J. Kathirithamby & A.I. Cognato 2005b. Assessing the odd secondary structural properties of nuclear small subunit ribosomal RNA sequences (18S) of the twisted-wing parasites (Insecta: Strepsiptera). – Insect Molecular Biology **14**: 625–643.

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

Gorodkin, J., R.B. Lyngsø & G.D. Stormo 2001. A minigreedy algorithm for faster structural RNA stem-loop search. – Genome Informatics **12**: 184–193.

Grant, T. & A.G. Kluge 2003. Data exploration in phylogenetic inference; scientific, heuristic, or neither. – Cladistics **19**: 379–418.

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

Hofacker, I.L., S.H.F. Bernhart & P.F. Stadler 2004. Alignment of RNA base pairing probability matrices. – Bioinformatics **20**: 2222–2227.

Holmes, I. 2004. A probabalistic model for the evolution of RNA structure. – BMC Bioinformatics **5**: 166.

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. – Molecular Phylogenetics and Evolution **4**: 314–330.

Kjer, K.M. 2004. Aligned 18S and insect phylogeny. – Systematic Biology **53**: 506–514.

Kjer, K.M., J.J. Gillespie & K.A. Ober in press, 2007. Opinions on multiple sequence alignment, and an empirical comparison of repeatability and accuracy between POY and structural alignments. – Systematic Biology **56**(1).

KJER, K.M., F.L. CARLE, J.A. LITMAN & J. WARE 2006. A molecular phylogeny of Hexapoda. – Arthropod Systematics & Phylogeny **64**(1): 35–44.

KLASS, K.-D. 2001: Morphological evidence on blattarian phylogeny: "phylogenetic histories and stories" (Insecta, Dictyoptera). – Deutsche Entomologische Zeitschrift **48**: 223–265.

KRUSKAL, J.B. 1983. An overview of sequence comparison. Pp. 1–45 *in*: D. SANKOFF & J.B. KRUSKAL (eds.), Time Warps, String Edits, and Macromolecules. – Addison-Wesley, Reading, MA.

MISOF, B. & G. FLECK 2003. Comparative analysis of mt LSU secondary structure of odonates: structural variability and phylogenetic signal. – Insect Molecular Biology **12**: 535–547.

NEEDLEMAN, S.B. & C.D. WUNSCH 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. – Journal of Molecular Biology **48**: 443–453.

NIEHUIS, O., C.M. NAUMANN & B. MISOF 2006. Identification of evolutionary conserved structural elements in the mt SSU rRNA of Zygaenoidea (Lepidoptera): a comparative sequence analysis. – Organisms Diversity and Evolution **6**: 17–32.

NOLLER, H.F. 2005. RNA structure: Reading the ribosome. – Science **309**: 1508–1514.

NOTREDAME, C., E.A. O'BRIEN & D.G. HIGGINS 1997. RAGA: RNA sequence alignment by genetic algorithm. – Nucleic Acids Research **25**: 4570–4580.

OGDEN, T.H., M.F. WHITING & W.C. WHEELER 2005. 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. – Cladistics **21**: 295–302.

REDELINGS, B.D. & M.A. SUCHARD 2005. Joint Bayesian estimation of alignment and phylogeny. – Systematic Biology **54**: 401–418.

SANKOFF, D., C. MOREL & R.J. CEDERGREN 1973. Evolution of 5S RNA and the non-randomness of base replacement. – Nature New Biology **245**: 232–234.

SANKOFF, D. 1975. Minimal mutation trees of sequences. SIAM – Journal of Applied Mathematics **28**: 35–42.

SANKOFF, D. & R.J. CEDERGREN 1983. Simultaneous comparison of three or more sequences related by a tree. Pp. 253–263 *in*: D. SANKOFF & J.B. KRUSKAL (eds.), Time Warps, String Edits, and Macromolecules. – Addison-Wesley, Reading, MA.

SWOFFORD, D.L. 1995. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). – Sinauer Associates, Sunderland, Massachusetts.

THOMPSON, J.D., D.G. HIGGINS & T.J. GIBSON 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. – Nucleic Acids Research **22**: 4673–4680.

WHEELER, W.C. 1995. Sequence alignment, parameter sensitivity, and the phylogenetic analysis of molecular data. – Systematic Biology **44**: 321–331.

WHEELER, W.C. 1996. Optimization alignment: The end of multiple sequence alignment in phylogenetics? – Cladistics **12**: 1–9.

WHEELER, W.C. & D.L. GLADSTEIN 1994. Malign. Computer program. – American Museum of Natural History, New York.

WHEELER, W.C., M.F. WHITING, Q.D. WHEELER & J.C. CARPENTER 2001. The phylogeny of extant insect orders. – Cladistics **17**: 113–169.

WHITING, M.F., J.C. CARPENTER, Q.D. WHEELER & W.C. WHEELER 1997. The Strepsiptera problem: Phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. – Systematic Biology **46**: 1–68.