biological interaction between Cape fur seals *Arctocephalus pusillus pusillus* and the Cape hakes *Merluccius capensis* and *M. paradoxis*. *S. Afr. J. Mar. Sci.* 16, 255–285

31 Yodzis, P. (2000) Diffuse effects in food webs. *Ecology* 81, 261–266

32 Yodzis, P. and Innes, S. (1992) Body size and consumer–resource dynamics. *Am. Nat.* 139, 1151–1175

33 Schmitz, O.I. (1997) Press perturbations and the predictability of ecological interactions in a food web. *Ecology* 78, 55–69

34 Walters, C.J. *et al.* (1997) Structuring dynamic models of exploited ecosystems from trophic mass-balance assessments. *Rev. Fish. Biol. Fish.* 7, 1–34

35 Christensen, V. and Pauly, D. (1992) ECOPATHII: a software for balancing steady-state models and calculating network characteristics. *Ecol. Model.* 61, 169–185

36 Christensen, V. (1998) Fishery-induced changes in a marine ecosystem: insight from models of the Gulf of Thailand. *J. Fish Biol.* 53, (Suppl. A), 128–142

37 Trites, A.W. *et al.* (1999) Ecosystem considerations and the limitations of ecosystem models in fisheries management: insights from the Bering Sea. In *Ecosystem Approaches for Fisheries Management*, pp. 609–620, University of Alaska, Fairbanks

38 Farber, D. A. (1999) *Eco-Pragmatism: Making Sensible Environmental Decisions in an Uncertain World,* University of Chicago Press

---

Trees for bees

Peter J. Lockhart and Sydney A. Cameron

Controversy over the origins and evolution of social behaviour in the major groups of social bees (the corbiculate bees) has fuelled arguments over different approaches for building evolutionary trees. However, the application of different analytical methodologies does not explain why molecular and morphological data suggest strikingly different hypotheses for the evolution of eusociality in bees. Determining the phylogenetic root is expected to help resolve the question of the social evolution of corbiculate bees. However, this requires that the long branch attraction problem is overcome. This phenomenon affects both molecular and morphological data for corbiculate bees.

Peter J. Lockhart*
Institute of Molecular BioSciences, Massey University, Palmerston North, New Zealand.
*e-mail: p.j.lockhart@massey.ac.nz

Sydney A. Cameron
Dept of Entomology, University of Arkansas, Fayetteville, AR 72701, USA.

The corbiculate bees include four tribes (Fig. 1), the highly eusocial honey bees (Apini) and stingless bees (Meliponini), the intermediately eusocial bumble bees (Bombini) and the essentially solitary orchid bees (Euglossini). These tribes include the major groups of highly social bees, and because they present different levels of sociality, they are useful for studying the evolution of social behavior. The phylogenetic relationships among these four tribes have been examined by several authors, using morphological and molecular characters. However, a clear understanding of their phylogeny has not been reached. Phylogenies from molecular data sets[1–4] have been interpreted to suggest independent origins for the two highly eusocial groups (Apini and Meliponini), whereas those obtained from morphological data sets[5] mostly suggest a common origin.

This lack of consensus has raised questions concerning the validity of certain methods of DNA sequence analysis[5]. It has been argued that the application of standard statistical tests for interpreting bee phylogenies[1–4] is invalid, and that an alternative approach, termed 'total evidence' (simultaneous analysis)[5,6], should be used instead, in which all available morphological and molecular data sets are combined, and the global best parsimony fit of the data onto a bifurcating tree is sought. Total evidence is appealing because it is based on the idea that evolutionary hypotheses can be tested most effectively by looking for corroboration among all available character data. However, the method has been criticized on the grounds that not all data in a combined phylogenetic analysis are necessarily equivalent[7–9].

Although the controversy surrounding corbiculate bee origins has been said to stem from a failure to observe the total evidence philosophy[5], this suggestion does not explain why phylogenetic inference from morphology and molecular analysis leads to different conclusions on the evolution of eusociality. Here, we examine the complexity of the corbiculate bee problem. We show that the phylogenetic question studied is a difficult one, that incompatibility exists between the data sets studied by comparative morphologists and molecular systematists, and that the controversy will not be resolved by advocating one evolutionary tree-building philosophy in preference to another.

Building trees for bees: a classic problem

The corbiculate bee phylogeny represents a classic example of an evolutionary tree model in which the juxtaposition of long external branches and a short internal branch (Fig. 2a) makes it difficult to place outgroups correctly[10]. With this tree shape, the root and direction of evolution are difficult to determine. In both morphological and molecular bee data, this problem can be visualized using splits graphs[11], which can represent more of the information contained in the data than is possible using a bifurcating tree representation. Figure 2 illustrates how the problem of identifying the root presents itself as reticulation when using morphological data (Fig. 2b), and sometimes as a collapse of internal tree structure when using molecular data (Fig. 2c). This uncertainty of the root position, although not explaining the disagreement between molecular and morphological results, makes it difficult to know whether highly eusocial behavior was an ancestral characteristic of corbiculate bees before tribal lineages diverged. As we discuss later, knowing the position of the root might help explain the corbiculate bee controversy.

Incongruence between morphological and molecular data forms

Tree-building analyses of molecular and morphological data have suggested strikingly different phylogenetic relationships among

http://tree.trends.com 0169–5347/01/$ – see front matter © 2001 Elsevier Science Ltd. All rights reserved. PII: S0169-5347(00)02054-1

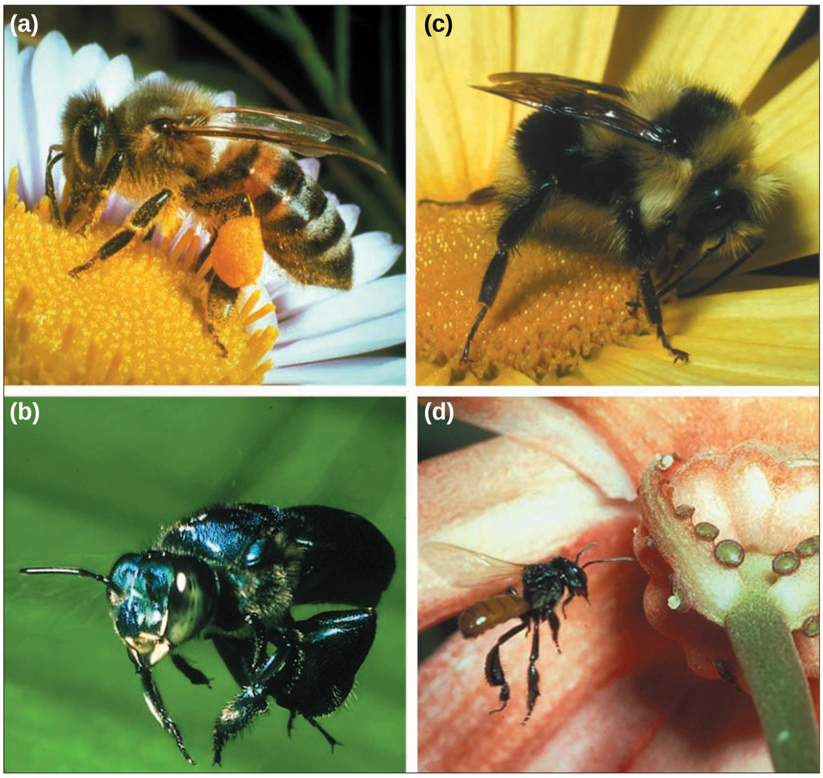
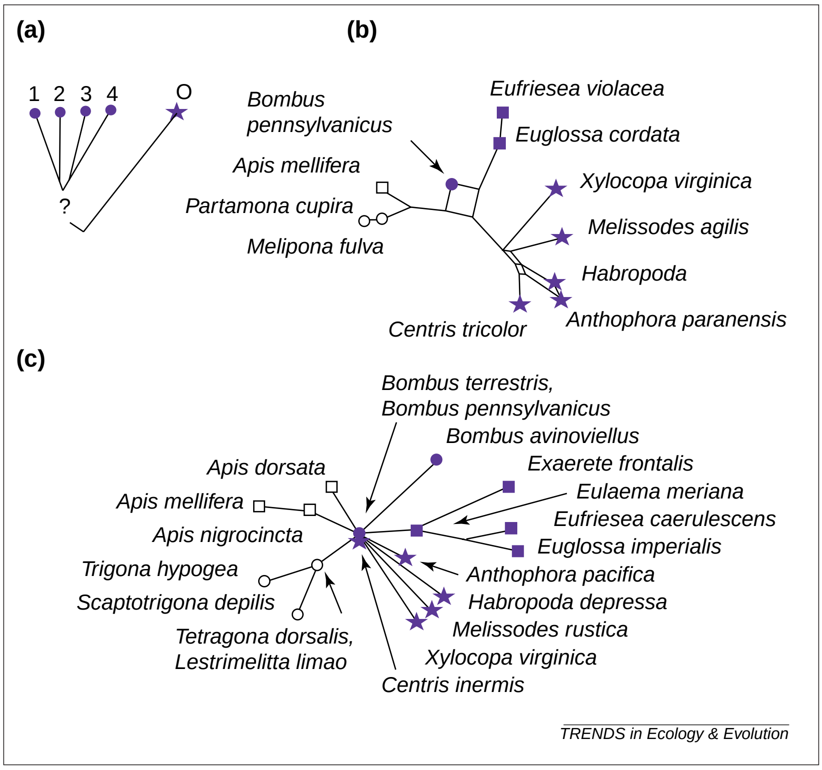


Fig. 1. Representatives of the four tribes of corbiculate bees: the clade encompassing honey bees (*Apis mellifera*, a), orchid bees (*Euglossa mixta*, b), bumble bees (*Bombus edwardsii*, c) and stingless bees (*Trigona fulviventris*, d). *Photographs, reproduced with permission, from E.S. Ross (a–c) and from the late L. Johnson (d).*



TRENDS in Ecology & Evolution

Fig. 2. Postitioning of outgroups in evolutionary models. (a) With the evolutionary tree model studied by Hendy and Penny[10] (i.e. where there are long external branches and short internal branches), it can be difficult to use outgroups to identify the root of the tree, because homoplasy is expected to result in conflicting signals indicating root placement. Irrespective of the true placement, outgroups tend to be drawn towards the long external branches in the tree. (b) An outgroup-rooted network built from morphological characters under split decomposition[11] with parsimony scoring criteria. Signals for conflicting root placement in the data cause a box-like reticulation, suggesting that the root could be either on the Euglossini lineage or between the highly eusocial, and primitively eusocial and solitary tribes. (c) An outgroup-rooted network built using first codon position opsin sequences under split decomposition with parsimony scoring criteria. In this case, conflicting signals for root placement result in an unresolved polytomy. Outgroups are indicated by stars, and in (b) and (c) highly eusocial species are indicated by open circles and squares.

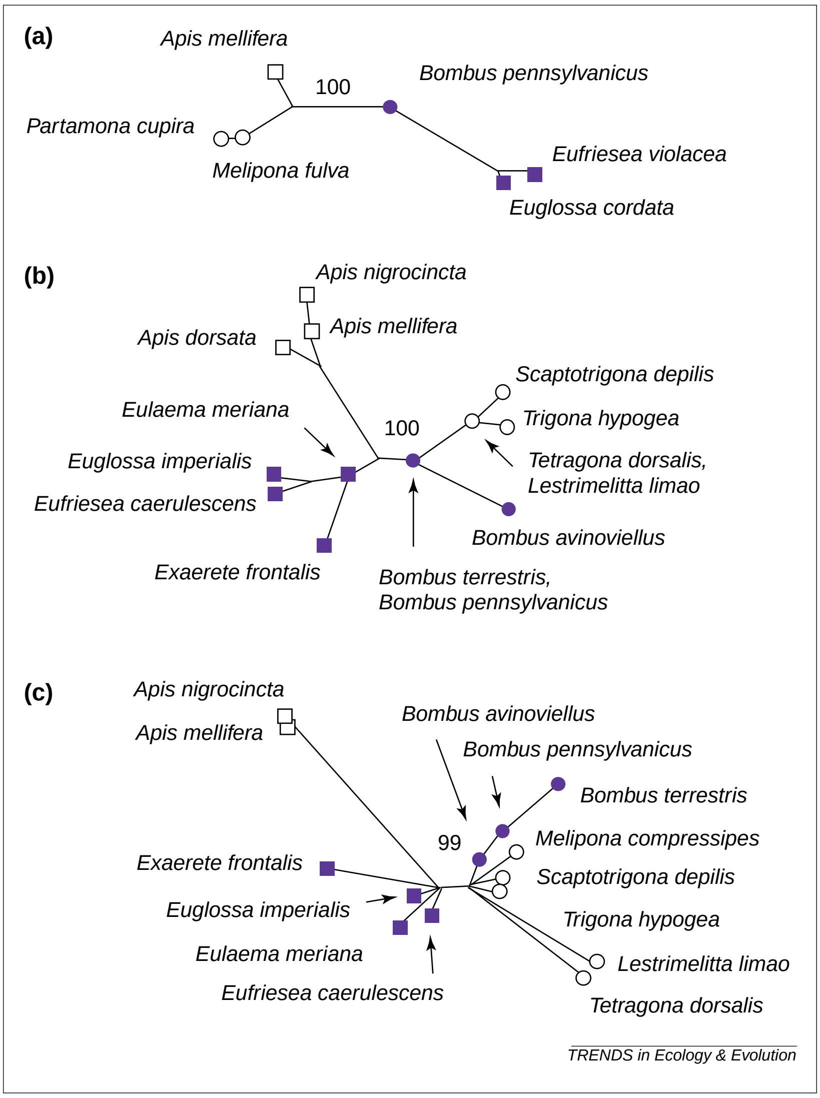
corbiculate bee tribes[1–5]. Parsimony splits graphs[11] illustrate that morphological and protein encoding sequence data sets have strongly incompatible signals. The morphological data group the highly eusocial tribes together and partition them from the Bombini and Euglossini tribes (Fig. 3a). By contrast, information from the first codon positions in the opsin sequences (Fig. 3b) and at the second codon positions in the cytochrome b sequences (Fig. 3c) indicates a split that groups together the Bombini and the highly eusocial Meliponini tribe. The differences between the morphological and sequence splits graphs cannot be explained by error in sequence alignment (a question that was raised[5] when the first rDNA sequences were reported[1]) as the alignments are unambiguous for the protein-encoding sequence data sets.

The evolution of molecular and morphological data

The total evidence philosophy of combining molecular with morphological data has been accepted and applied in diverse studies of both animals[14,15] and plants[16]. The study of corbiculate bees is only one of many examples. An important assumption when using total evidence is that the levels of 'homoplasy' (similarity which does not reflect phylogenetic relationships) are similar in both data forms[5]. In fact, the processes underlying morphological evolution differ greatly from the processes underlying the evolution of most molecular sequences[17–27]. A consequence of this is that the nature and level of homoplasy can differ between the two data forms, and this limits the analysis of combined data.

One of the advantages of using only sequence data for tree building is that it allows the assumption of a common evolutionary process at different nucleotide and/or amino acid positions[28,29]. This assumption enables hypotheses to be tested using more sophisticated model-based approaches[18–30]. These methods cannot be used when morphological and molecular data are combined, and need to be used cautiously when different genes or sequence data partitions are combined[7–9,19,21]. In particular, tree building models might not accommodate heterogeneity in evolutionary rates and/or differences in the processes of character evolution that could occur in joint data sets[7,8].

Empirical observations have shown that sequence evolution is constrained by cellular processes in a way that can give rise to patterns unexpected under the simple bifurcating divergence models used in analyses of morphological data. More commonly recognized complexities of sequence data include irregular nucleotide and amino acid frequencies[18–22], the presence and changing distributions of invariable sites (sites that are inferred to be unable to



TRENDS in Ecology & Evolution

Fig. 3. Groupings of the corbiculate bees. Unrooted parsimony splits graphs for (a) morphological data, (b) first codon position opsin data and (c) second codon position cytochrome *b* data. Quartet puzzle 'reliability values'[12] shown on the figures have been calculated under parsimony criteria[13]. These indicate the high degree of compatible signals within data sets for hypotheses of eusocial origins, and the incongruence between morphological and molecular data forms. Open circles and squares identify highly eusocial species.

change)[23,24], positional substitution rate heterogeneity (the general form of having variable and invariable sites in a single analysis)[25], non-independence of change due to secondary structure interactions[26], and recombination[27]. If sufficient consideration is not given to the biological properties of sequence data, tree building from sequences could be unreliable. This problem will affect not only parsimony tree building, but also more explicitly model-based methods, such as those involving maximum likelihood inference or asymmetric substitution models[19,22]. Problems could also exist in analyses of morphological data, for example, in the form of functionally correlated characters. However, in the absence of a common mechanism underlying the evolution of different morphological characters, these problems are expected to be more difficult to identify and compensate for in the way that can be done with molecular data. Perhaps morphologists do not expect systematic biases to be as pervasive as they are in some molecular data sets, but the issue needs to be considered seriously.

Are the molecular analyses of corbiculate bee data wrong?

An obvious question concerning corbiculate bee sequence analyses is whether systematic biases in sequence data (such as those mentioned in the previous section) might explain the different results from analyses of molecular and morphological data. However, no such biases have been detected. Figure 4 illustrates one approach used to investigate the potentially misleading effect of compositional heterogeneity and invariable sites in bee sequence data, by showing the phylogenetic signal in opsin and cytochrome b sequences for and against the grouping of Bombini and Meliponini tribes. The height of the histograms in this distance Hadamard spectrum[30,31] indicate the support (above the x-axis) and conflict (below the x-axis) in the data for this grouping of tribes. The phylogenetic signal has been corrected for sequence composition irregularity and the presence of invariable sites using a LogDet/invariable sites model[21,22]. The key observation is that the signal in the data for the Bombini and Meliponini grouping always remains much greater than the signal for other possible groupings of tribes. By contrast, one of us[22] recently reported an example using similar analyses. In this earlier study, as invariable sites in an alignment of insect rDNA sequences were removed, the support for a partition grouping 'Strepsiptera and Diptera' became greatly reduced, favoring other possible hypotheses. This observation suggested that the grouping of Strepsiptera and Diptera in tree building analyses was unreliable because it was largely caused by compositional differences and positional rate heterogenity in the molecular sequences[22].

Observations such as those in Fig. 4 suggest that the grouping of Bombini and Meliponini tribes in molecular analyses cannot be easily dismissed. Of course, it remains possible that unrecognized biases might still mislead the phylogenetic interpretation of corbiculate bee sequences. However, the observation that bias in substitution processes would need to similarly affect sequences in both the nuclear (opsin, 28S rDNA sequences)[3] and mitochondrial genomes (16S rDNA, cytochrome b sequences)[3] would seem to make this possibility unlikely.

Resolving the controversy

A better understanding of the evolution of eusociality in corbiculate bees can be expected with the acquisition of more data. Additional molecular sequences might help to identify hidden biases in the current data, test the strength of support for unrooted data partitions, and confirm the position of the root.

The potential value of additional sequence data is indicated from the conservative substitution patterns

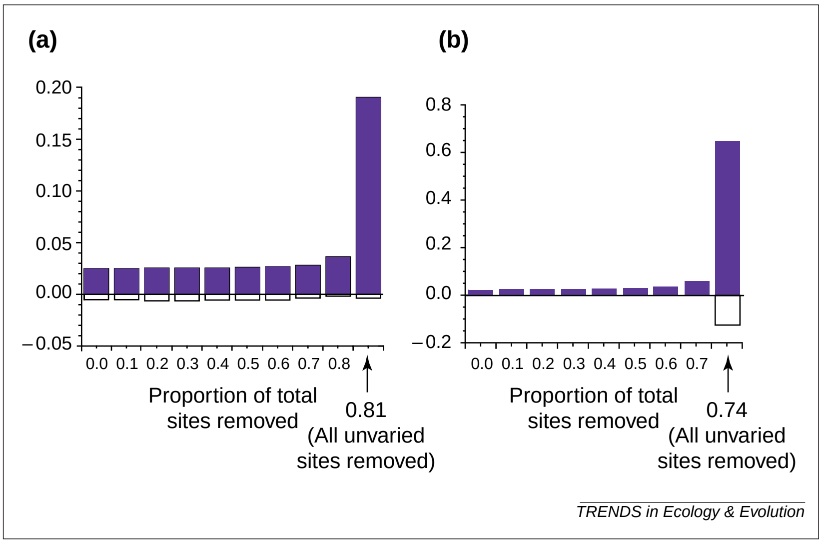


Fig. 4. Distance Hadamard spectra indicating support under the LogDet transformation for the groupings Bombini with Meliponini, and Apini with Euglossini. The analysis used (a) opsin – first codon position and (b) cytochrome b – second codon position data. LogDet values were calculated following the removal of different proportions of unvaried sites (observed not to have changed). Values on the x-axis indicate the proportion of unvaried sites removed. These were removed to test the effect of potential substitution rate differences at different sequence positions on the LogDet path-length estimates[21,22].

observed in the protein genes studied to date. These data show low amounts of contradictory signal in distance Hadamard spectra (Fig. 4). Thus, efforts to obtain complete mitochondrial genome and nuclear sequences seem very worthwhile. With the use of new sequence data there will be debate over the use of homogeneity tests[7–9,21] and whether or not sequences and sequence partitions should be combined when building trees. It could be that not all sequence data can be combined and jointly analysed using only one tree building method or sequence evolution model. In this case, sequences and/or data partitions which have evolved under similar processes (i.e. homogeneous data) will need to be identified, and studied under appropriate evolutionary tree building models. Where data cannot be combined, consensus methods for comparing trees built using different criteria should help evaluate the best phylogenetic estimates for tribal phylogeny[9].

Improvements in the resolution of morphological data could come from several areas. New independent morphological character sets could reduce the impact of any bias or sampling error in character selection, which can be one of the less objective aspects of morphological analyses[32]. In existing analyses, some of the morphological characters are not discrete, but, nevertheless, have been directly coded as discrete characters[33]. These data could be further investigated using methods for converting continuous data to discrete phylogenetic characters[34]. An additional aspect to consider is the degree to which some of the morphological characters are functionally or developmentally, rather than phylogenetically, correlated[35]. With respect to this issue, it is noteworthy that some molecular analyses[3] place outgroups on the highly eusocial lineages. This observation raises the possibility that the last common ancestor of corbiculate bees might have been eusocial; a hypothesis that has not arisen from analyses of morphological data (e.g. Fig. 2b). If additional sequencing studies strongly support the suggestion that the root is on one of the highly eusocial lineages (Apini or Meliponini), there will be a need for closer examination of the morphological data, in case some characters currently used for phylogeny are correlated with an ancestral social condition. Finally, many of the analyses have used supraspecific taxa as terminals in the analysis, and over-simplifications of data could potentially be contributing to bias[33]. In these cases, the individual exemplar taxa might be more carefully coded for their actual morphological character states (this also applies to coding for outgroup taxa).

In summary, efforts to resolve the controversy over the origins of bee social behavior need to be made by both molecular systematists and morphologists. Disagreement exists because analyses of sequences and morphology suggest different hypotheses, and not because researchers have used different criteria for building and testing evolutionary trees[1–5]. Future consideration must be given to identifying more clearly the phylogenetic information contained in different data sources.

Acknowledgements
We thank J. Whitfield, P. Mardulyn and the anonymous referees for helpful comments. The XIII International Congress of the International Union for the Study of Social Insects meeting, which stimulated this research, was supported by National Science Foundation (DEB-9870735) and University of Arkansas grants to S.A.C. The New Zealand Marsden Fund also helped towards some of the attendance costs for P.J.L. and The Alexander von Humboldt Foundation is acknowledged for current support to P.J.L.

References

1 Cameron, S.A. (1993) Multiple origins of advanced eusociality in bees inferred from mitochondrial DNA sequences. *Proc. Natl. Acad. Sci. U. S. A.* 90, 8687–8691
2 Mardulyn, P. and Cameron, S.A. (1999) The major opsin in bees (Insecta: Hymenoptera): a promising nuclear gene for higher level phylogenetics. *Mol. Phylogenet. Evol.* 12, 168–176
3 Cameron, S.A. and Mardulyn, P. Multiple data sets suggest independent origins of highly eusocial behavior in bees (Hymenoptera: Apinae). *Syst. Biol.* (in press)
4 Koulianos, S. *et al.* (1999) Phylogenetic relationships within the apinae (Hymenoptera) and the evolution of eusociality. *J. Evol. Biol.* 12, 380–384
5 Chavarría, G. and Carpenter, J.M. (1994) 'Total evidence' and the evolution of highly social bees. *Cladistics* 10, 229–258
6 Nixon, K.C. and Carpenter, J.M. (1996) On simultaneous analysis. *Cladistics* 12, 221–241
7 Bull, J.J. *et al.* (1993) Partitioning and combining data in phylogenetic analysis. *Syst. Biol.* 42, 384–397
8 Huelsenbeck, J.P. *et al.* (1996) Combining data in phylogenetic analysis. *Trends Ecol. Evol.* 11, 152–157
9 Swofford, D.L. (1991) When are phylogeny estimates from molecular and morphological data incongruent? In *Phylogenetic Analysis of DNA Sequences* (Miyamoto, M.M. *et al.*, eds), pp. 295–333, Oxford University Press
10 Hendy, M.D. and Penny, D. (1989) A framework for the quantitative study of evolutionary trees. *Syst. Zool.* 38, 297–309
11 Huson, D.H. (1998) SplitsTree: analyzing and visualizing evolutionary data. *Bioinformatics* 14, 68–73
12 Strimmer, K. and von Häseler, A. (1996) Quartet puzzling: a quartet maximum likelihood method for reconstructing tree topologies. *Mol. Biol. Evol.* 13, 964–969
13 Swofford, D.L. (1998) PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4. Sinaur Associates
14 Remsen, J. and DeSalle, R. (1998) Character congruence of multiple data partitions and the origin of the Hawaiian Drosophilidae. *Mol. Phylogenet. Evol.* 9, 225–235
15 Springer, M.S. *et al.* (1999) Additional support for Afrotheria and Paenungulata, the performance of mitochondrial versus nuclear genes, and the impact of data partitions with

heterogeneous base composition. *Syst. Biol.* 48, 65–75
16 Pryer, K.M. *et al.* (1995) Phylogenetic relationships of extant ferns based on evidence from morphology and rbcL sequences. *Am. Fern J.* 85, 205–282
17 Kadereit, J.W. (1994) Molecules and morphology, phylogenetics and genetics. *Bot. Acta* 107, 369–373
18 Whitfield, J.B. and Cameron, S.A. (1998) Hierarchical analysis of variation in the mitochondrial 16S rRNA gene among Hymenoptera. *Mol. Biol. Evol.* 15, 1728–1743
19 Lockhart, P.J. *et al.* (1999) Spectral analysis, systematic bias, and the evolution of chloroplasts. *Mol. Biol. Evol.* 16, 573–576
20 Mooers, A. and Holmes, E.C. (2000) The evolution of base composition and phylogenetic inference. *Trends Ecol. Evol.* 15, 365–369
21 Waddell, P.J. *et al.* (1999) Using novel phylogenetic methods to evaluate mammalian mtDNA, including amino acid-invariant sites-LogDet plus site stripping, to detect internal conflicts in the data, with special reference to the positions of Hedgehog, Armadillo and Elephant. *Syst. Biol*. 48, 31–53
22 Steel, M.A. *et al*. (2000) Invariable site models and their use in phylogeny reconstruction. *Syst. Biol.* 49, 225–232
23 Lockhart, P.J. *et al.* (2000) How molecules evolve in eubacteria. *Mol. Biol. Evol.* 17, 835–838
24 Philippe, H. and Germot, A. (2000) Phylogeny of Eukaryotes based on ribosomal RNA: long branch attraction and models of sequence evolution. *Mol. Biol. Evol.* 17, 830–834
25 Van de Peer, Y. *et al.* (1996) Substitution rate calibration of small subunit ribosomal RNA identifies chlorachniophyte endosymbionts as remnants of green algae. *Proc. Natl. Acad. Sci. U. S. A.* 93, 7732–7736
26 Schöniger, M. and von Häseler, A. (1994) A stochastic model for the evolution of autocorrelated DNA sequences. *Mol. Phylogenet. Evol.* 3, 240–247
27 Clegg, M.T. *et al.* (1997) The evolution of plant nuclear genes. *Proc. Natl. Acad. Sci. U. S. A.* 94, 7791–7798
28 Swofford, D.L. *et al.* (1996) Phylogenetic Inference. In *Molecular Systematics* (Hillis, D.M. *et al.*, eds), pp. 407–514, Sinaur
29 Steel, M. and Penny, D. (2000) Parsimony, likelihood, and the role of models in molecular phylogenetics. *Mol. Biol. Evol.* 17, 839–850
30 Penny, D. *et al.* (1996) Corrected parsimony, minimum evolution and Hadamard conjugations. *Syst. Biol.* 45, 594–603
31 Charleston, M.A. and Page, R.D.M. (1999) Spectral analysis – a brief introduction. In *Molecular Systematics and Plant Evolution* (Hollingsworth, P.M. *et al.*, eds), pp. 255–267, Taylor and Francis
32 Poe, S. and Wiens, J. J. (2000) Character selection and the methodology of morphological phylogenetics. *In Phylogenetic Analysis of Morphological Data* (Wiens, J.J. ed.), pp. 20–36, Smithsonian Institution Press
33 Bininda-Emonds, O.R.P. *et al.* (1998) Supraspecific taxa as terminals in cladistic analysis: implicit assumptions of monophyly and a comparison of methods. *Biol. J. Linn. Soc.* 64, 101–133
34 Swiderski, D.L. *et al.* (1998) Why morphometrics is not special: coding quantitative data for phylogenetic analysis. *Syst. Biol.* 47, 508–519
35 Emerson, S.B. and Hastings, P.A. (1998) Morphological correlations in evolution: consequences for phylogenetic analysis. *Q. Rev. Biol.* 73, 141–162