Points of View

Supertrees Are a Necessary Not-So-Evil: A Comment on Gatesy et al.

OLAF R. P. BININDA-EMONDS,[1] KATE E. JONES,[2] SAMANTHA A. PRICE,[2] RICHARD GRENYER,[3]
MARCEL CARDILLO,[3] MICHAEL HABIB,[2] ANDY PURVIS,[3] AND JOHN L. GITTLEMAN[2]

[1]*Lehrstuhl für Tierzucht, Technical University of Munich, Alte Akademie 12, 85354 Freising-Weihenstephan, Germany;*
E-mail: olaf.bininda@tierzucht.tum.de
[2]*Department of Biology, Gilmer Hall, University of Virginia, Charlottesville, Virginia 22904-4328, USA;*
E-mail: kate.jones@virginia.edu (K.E.J.), sprice@virginia.edu (S.A.P.), jlgittleman@virginia.edu (J.L.G.)
[3]*Department of Biological Sciences, Imperial College London, Silwood Park Campus, Ascot SL5 7PY, United Kingdom;*
E-mail: r.grenyer@imperial.ac.uk (R.G.), m.cardillo@imperial.ac.uk (M.C.)

In a recent article, Gatesy et al. (2002) assessed different methods of creating phylogenies of even-toed ungulates (Artiodactyla) and whales (Cetacea; together, the Cetartiodactyla) from compilations of data. Gatesy et al. criticized strongly an analysis by Liu et al. (2001), that combined separate estimates of phylogeny to yield a phylogenetic supertree (sensu Sanderson et al., 1998) rather than using primary character data (e.g., molecular sequences) to infer the phylogeny. The main criticisms that Gatesy et al. raised about the Liu et al. analysis are of data duplication, poor data quality, and data accountability. They claimed that these problems biased the Liu et al. supertree analysis such that "support for any particular node . . . may have little or no empirical basis," meaning that the supertree "should not be used to guide future phylogenetic research or to organize comparative evolutionary studies" (Gatesy et al., 2002:662). Although it was not their intent (J. Gatesy, pers. comm.), it is easy to interpret Gatesy et al.'s criticisms as applying to the supertree approach in general, with the implication that analysis of the primary character data, especially when combined in a supermatrix approach (sensu Sanderson et al., 1998), is the only valid method for phylogeny reconstruction.

Given the growing use of supertree methods for constructing phylogenies of large clades of organisms (reviewed by Bininda-Emonds et al., 2002), it is important to assess this approach in general in light of Gatesy et al.'s comments on the Liu et al. analysis. In our opinion, most of the problems identified can be ameliorated with a well-conceived protocol for selecting the source phylogenies with which to build the supertree. Many of the issues raised by Gatesy et al. have already been addressed by individual supertree studies (e.g., Purvis, 1995a; Bininda-Emonds et al., 1999; Liu et al., 2001; Jones et al., 2002; Kennedy and Page, 2002; Pisani et al., 2002; Stoner et al., 2003). Here, we outline our protocol for selecting source trees as well as point out some analytical advantages that the supertree approach possesses over the supermatrix approach. In so doing, we establish that supertree construction represents a viable and valid alternative to supermatrix analyses.

TREES VERSUS CHARACTERS AS SOURCE DATA

Supertrees differ from conventional phylogenetic analyses in the source data being analyzed. In a supermatrix approach, the source data represent primary character data that can be measured for the taxa in question. In a supertree analysis, the source data are the topologies of the set of source trees, often represented individually by a set of matrix elements. This distinction is key and forms the basis for criticisms of the supertree approach (e.g., Rodrigo, 1993, 1996; Slowinski and Page, 1999; Novacek, 2001; Springer and de Jong, 2001). Below, we discuss several corollaries of the use of source trees as data (see also Bininda-Emonds et al., in press).

Data Duplication

Springer and de Jong (2001) pointed out that a fundamental limitation of the supertree approach is the potential duplication of primary character data among the source trees. These duplicated data will have a disproportionately greater influence on the supertree topology, potentially biasing the end result. Gatesy et al. (2002:661) were of the opinion that for the supertrees of Purvis (1995a) and Bininda-Emonds et al. (1999) "unnecessary repetitions of character evidence negatively influenced both of these supertree data sets." Similar shortcomings were stated to be present in the supertrees of Liu et al. (2001) and Jones et al. (2002).

Given the continual recycling of phylogenetic data, data duplication is difficult to avoid in any phylogenetic

analysis where trees are combined instead of the primary character data. Gatesy et al. nicely summarized the potential for overlap among the source trees in a supertree analysis in their figure 3. Clearly, data duplication exists. This duplication violates a key assumption of phylogenetic analysis, namely that the source data are independent. However, this assumption is routinely violated in analyses based on primary character data. For instance, several characters are often described for a single morphological structure. In molecular studies, secondary structure (e.g., stem regions in tRNAs, protein folding) or codon position in coding DNA mean that nonindependent compensatory mutations may accompany primary ones. The combination of phenotypic and genotypic data (i.e., morphological and molecular, respectively) in a supermatrix approach must represent duplication at some level. Thus, issues of data duplication and nonindependence are not limited to the supertree approach. Instead, it is important to establish how detrimental these issues are in any phylogenetic analysis and whether instances of them can be minimized.

In all the major supertree analyses (Purvis, 1995a; Bininda-Emonds et al., 1999; Liu et al., 2001; Jones et al., 2002; Kennedy and Page, 2002; Pisani et al., 2002; Stoner et al., 2003), the researchers have taken steps to address data duplication, albeit with varying success. We believe that systematic data collection protocols will reduce duplication to such an extent that it would have little effect on the relationships recovered in the supertree. This can be done by creating a formal framework to allow the supertree builder to choose only those source trees that contain what would be considered to be independent data sets for analysis.

Our suggested protocol (presented in detail by Bininda-Emonds et al., in press) operates on the principle of identifying phylogenetic *hypotheses* that can reasonably be viewed as being independent (following Purvis, 1995b). In the extreme, no characters are truly independent given the existence of a common hierarchical set of relationships that has shaped their evolution. However, delimiting pseudoindependent evolutionary "packets" based on genes may be defensible given the recognition of the gene tree/species tree dichotomy (Maddison, 1997). Our protocol attempts to identify these packets through an explicitly defined set of rules formulated according to the ordered criteria of (1) data independence, (2) taxonomic inclusiveness, and (3) (informed) author preference.

Decisions about independence are based solely on the source of the character data and the taxon set. Nonoverlapping data sets (e.g., different genes) are considered to be independent, even if they appear on a single heritable unit like mitochondrial DNA. In contrast, different portions of the same gene are not independent for an overlapping set of taxa, even if these gene portions do not overlap at all. Trees for nonoverlapping taxon sets, even if they are derived from the same set of characters, are independent by practical necessity. These taxon sets specify different regions of the tree as a whole and so do not duplicate data.

The supertree approach combines phylogenetic hypotheses, which we contend are composed of more than the primary character data being analyzed (Purvis, 1995b). Therefore, we would argue that the detrimental effects of data duplication are ameliorated further when the overlap of the primary character data among the source trees is only partial (as is usually the case; see figure 3 of Gatesy et al.). Due to the phenomenon of signal enhancement (sensu de Queiroz et al., 1995), the combination of different data sets within any single study may specify a solution different from those specified by the data sets individually (see Barrett et al., 1991). Thus, despite the partial duplication of the primary character data underlying a set of source trees, our protocol holds each member of the latter to be an independent phylogenetic hypothesis (Purvis, 1995b) and suitable for inclusion in the supertree analysis. A cogent example of the emergent property of the phylogenetic hypothesis presents itself in Cetacea. Messenger and McGuire (1998) showed that the strongly conflicting molecular phylogenies of this group (compare Milinkovitch et al., 1994, 1995, 1996 with Arnason and Gullberg, 1994, 1996), even those based on virtually the same primary character data, could be explained largely by the choice of outgroup altering the root of the tree. The resultant phylogeny was also sensitive to weighting options and sequence alignment decisions in addition to the primary character data.

The final two criteria—taxonomic inclusiveness and author preference—are designed to select single source trees from a set of equally suitable possibilities (e.g., numerous source trees all derived from the same gene). Taxonomic inclusiveness holds that only the most recent and/or comprehensive study (in terms of number of taxa) is used. Author preference is normally restricted to within a single paper, where the same data set may be analyzed using several optimization criteria. If neither of these criteria can be applied, then a supertree of all equally suitable source trees is constructed. This minisupertree then serves as a source tree in the main supertree analysis (Bininda-Emonds et al., in press).

Source Data Quality

Gatesy et al. criticized several supertree analyses for using source trees that included appeals to authority (particularly inappropriate assumptions of monophyly) or with questionable phylogenetic content (e.g., taxonomies or reviews that did not include new data or analyses). Both factors were held to reduce the quality of the source trees and therefore potentially bias the resultant phylogeny.

The use of poor data may compromise the results in any phylogenetic analysis (i.e., including a supermatrix analysis), and researchers should ensure that all data used are of the highest achievable quality. One of the strengths of the supertree approach is the transparency in the source tree selection process. Again, a formal framework (see Bininda-Emonds et al., in press) outlining potential source trees that are or are not appropriate for a supertree analysis will do much to alleviate the

concerns raised by Gatesy et al. Ideally, decisions should be made in an evidence-based fashion, but decisions are presently often a trade-off between choosing only the most rigorous phylogenies and obtaining adequate taxonomic coverage. As pointed out by Gatesy et al., Purvis (1995a) included some source studies that were not rigorous estimates of phylogeny, nor were even intended to represent phylogeny, in order to obtain coverage of all extant species of primate. However, Purvis acknowledged explicitly the variation in data quality in his analysis by heavily downweighting such studies. Differential weighting of source trees was also performed in other supertree analyses (e.g., Bininda-Emonds et al., 1999; Jones et al., 2002; Stoner et al., 2003). In each case, topological changes were minimal, indicating that less robust source trees were either largely congruent with more robust ones (see also Bininda-Emonds, 2000) or had little influence on the analysis.

As the generation of new phylogenies continues at an ever-increasing rate and as rigorous methods of phylogeny reconstruction become the rule rather than the exception, the need to include poorly justified source trees in supertree analyses will diminish. Moreover, these new phylogenies will increasingly include estimates of node support. Future supertree analyses will be able to incorporate this information, which has been shown to improve performance (Bininda-Emonds and Sanderson, 2001). In contrast, including information about differential support among and within source data sets is much more difficult in a supermatrix approach.

Data Accountability

Gatesy et al. contended that supertree construction suffers from a lack of both data accountability and transparency compared with the supermatrix approach. However, their ability to deconstruct the Liu et al. (2001) data set to identify instances of (primary character) data duplication within it and modify it in a reanalysis clearly represents data accountability.

Gatesy et al. argued that the (primary character) data are explicitly presented in supermatrix analyses. There is obviously a more direct connection to the primary character data in a supermatrix approach; however, it is often the case that these data are not available directly. Although web-based resources have increased data availability, few of these resources supply the *aligned* molecular sequences (TreeBASE and the Ribosomal Database Project being notable exceptions), a key factor in data accountability. We note that many of the same web-based resources can archive the source data of both supermatrix and supertree approaches, giving potentially equal source data accountability.

Furthermore, while it is true that use of the primary character data is more transparent in theory, this is not the case in practice. Errors in morphological data are recycled regularly along with the data sets (Jenner, 2001). The same can be expected for any molecular data sets that contain inaccuracies. GenBank database information is known to contain some errors (e.g., vector contamina-

tion; Seluja et al., 1999), and there have been implications that some molecular sequence data are of inferior quality (e.g., Ledje and Arnason, 1996). Given the difficulty of verifying the accuracy of primary character data, be they morphological and especially molecular, any errors are unlikely to be identified (e.g., only when the resultant phylogeny is obviously wrong; Page and Charleston, 1999), much less corrected. Therefore, as pointed out by Jenner (2001), supermatrices may very well suffer from limited data transparency, just like any other phylogenetic analysis.

ADVANTAGES OF THE SUPERTREE APPROACH

We agree that combination of trees rather than the primary character data entails some loss of information. However, available evidence indicates that this loss of information is not detrimental in practice (Bininda-Emonds and Sanderson, 2001; Levasseur and Lapointe, 2001). Moreover, the development of both nonparametric and parametric bootstrapping procedures for supertree construction (Huelsenbeck et al., in prep.; Moore et al., in prep.) will allow a direct connection with the primary character data. Finally, the supermatrix approach suffers from several practical limitations with respect to the supertree approach. These limitations, which involve both the data themselves and especially the analysis of these data, also result in a loss of information and, in our opinion, can potentially outweigh the advantages gained by using the primary character data.

Source Data and the Principle of Total Evidence

The supermatrix approach is limited to compatible primary character data types that can be analyzed using a single optimization criterion (Bininda-Emonds et al., 1999). Until recently, the only criterion available for a broad range of data types was maximum parsimony (MP); however, Lewis's (2001) generalization of the maximum likelihood (ML) framework and Bayesian methods provide other possibilities. Nonetheless, there are still many useful sources of phylogenetic data that cannot be analyzed under any of these frameworks (e.g., distance data, morphometric data) and therefore must be discarded in a supermatrix analysis. This too represents an undesirable loss of information. Supertree construction, in contrast, allows all source studies to be combined (even those that are not based on explicit character data, if desired), which accords with the principle of total evidence and the idea that the best phylogenetic hypothesis is that derived from the greatest number of independent lines of evidence (Mickevich, 1978; Farris, 1983; Penny and Hendy, 1986; Kluge, 1989; Novacek, 1992; de Jong, 1998).

Another issue is the availability of compatible character data, particularly molecular data. As pointed out by Bininda-Emonds et al. (2002), the number of mammals for which homologous molecular sequences are available is comparatively tiny, even though this group is comparatively well studied. Taxonomic coverage is similarly

limited even within more restricted clades. For example, the analysis by Gatesy et al. (2002) includes only 51 of the 293 extant cetartiodactyl species (following Wilson and Reeder, 1993). Further, information for most of these species is available only for cytochrome *b* and partial 12S and 16S rDNA sequences, which together comprise only about 5% of the supermatrix. In a survey of GenBank, we found that 205 cetartiodactyl species had information for at least one of the 19 most commonly sequenced genes for the group, meaning that nearly one-third of the species likely have no (or at least very little) molecular information known for them. The situation will be even worse in clades that are less thoroughly investigated. This consideration leads into the issue of missing data, the associated analytical problems of which were mentioned by Gatesy et al. For a given set of primary character data, the proportion of missing data in the matrix analyzed in the supertree approach will be either equal to or almost always less than that in the supermatrix approach.

In the end, supertrees allow the use of more of the available data sources, thereby providing better, and potentially complete, taxonomic coverage for any given group. Currently, a supermatrix approach cannot produce complete phylogenies except for taxa of restricted size, and then using only a subset of the data.

Analytical Issues and Computational Complexity

A supermatrix analysis is potentially much more computationally complex than a supertree analysis because different phylogenetic data sources are best analyzed under different models of evolution. For molecular data, relevant parameters in the evolutionary model include transition:transversion (ti:tv) ratio, proportion of invariant sites, between site rate variation (gamma correction), differential codon weighting (for coding DNA), and accounting for secondary structure (e.g., stems versus loops in tDNA). Other parameters, which are usually unknown, are relevant for morphological data. Yet, the computational complexity involved in analyzing each data partition under the appropriate evolutionary model means that analyses where such information is included will be largely intractable in a supermatrix approach, especially within a ML framework (Sanderson and Kim, 2000; Daubin et al., 2001; Bininda-Emonds et al., 2002). Only analysis within a MP framework, and possibly within a Bayesian one, is computationally feasible. However, MP cannot accommodate many of the parameters listed above and will have problems accommodating multiple parameters simultaneously (e.g., using weighting to account for both ti:tv ratio and secondary structure).

It is also well appreciated that different data are most informative at different levels in the phylogeny (e.g., slow genes for deep relationships, fast genes for more recent ones; although see Källersjö et al., 1999). Thus, in a supermatrix study of exceedingly broad scope, there will be either alignment or saturation problems among distantly related taxa (molecular data) or the inclusion of many inapplicable characters (morphological data).

Inapplicable data in particular may compromise phylogenetic accuracy and remain an ongoing problem with no adequate solution (Lee and Bryant, 1999; Strong and Lipscomb, 1999).

In contrast, a supertree approach offers analytical efficiency and accuracy. The different data partitions can be analyzed robustly under an appropriate model of evolution in separate analyses (Bininda-Emonds et al., 2002). The resultant trees can then be combined as a supertree, with apparently little loss of accuracy, especially when information about differential node support (e.g., bootstrap values) is included (Bininda-Emonds and Sanderson, 2001). This supertree approach offers two other advantages. First, it is more amenable to the benefits of parallel processing (i.e., for the initial analysis of each data partition) than would be a supermatrix approach. Second, it makes issues involving the relative weight of the different data partitions more transparent. In contrast to the supermatrix approach, the a priori weight of any data partition in a supertree analysis is unaffected by the amount of character data (see Doyle, 1992). This addresses concerns about smaller data sets being "swamped" by larger ones (Miyamoto, 1985; Barrett et al., 1991), be they morphological versus molecular studies or even small versus large genes. Given the widespread recognition of the existence of gene tree/species tree conflicts, the supermatrix approach will eventually have to confront the same questions of weighting data partitions now being tackled in the supertree approach.

SUPERTREES AS SUMMARIES OF PAST AND CURRENT RESEARCH

Gatesy et al. (2002:662) held supertrees to be "*imprecise* summaries of previous work" (emphasis added). It is unclear to us why this might be the case. We contend that supertrees are extremely precise summaries of previous work to the point that they are dependent upon choices of weighting and inclusion of source trees in the same way that conventional parsimony analyses are dependent upon choice and weighting of character partitions. Also, the supertree approach is highly amenable to time series analysis, whereby supertrees can be constructed using only the source trees published during a precise time interval (e.g., Bininda-Emonds, 2000, in press; Jones et al., 2002). This is less true for the supermatrix approach because primary character data, unlike published source trees, are not as obviously time stamped.

We agree with Gatesy et al. (2002:662) that "a supermatrix clearly reviews which *characters* have or have not been scored for particular taxa" (emphasis in original). However, in extending this point to question the overall descriptive value of supertrees, Gatesy et al. ignored the fact that supertrees and supermatrices are summarizing different levels of the systematics database. Because the supertree approach works at the level of the source trees, it can easily compare the phylogenetic stability of a taxon or species with the amount of research effort it has received. Thus, unlike supermatrix approaches, supertrees

can more easily highlight taxa that have been well studied, but the relationships within which remain controversial (e.g., Felidae in Carnivora; Bininda-Emonds et al., 1999). In our opinion, both supertree and supermatrix approaches provide useful summaries of their respective source data.

Conclusions

We contend that many of the problems raised by Gatesy et al. (2002) about the Liu et al. (2001) supertree study can be remedied easily using an explicit and well-justified protocol. Moreover, we indicate that the supertree approach possesses some clear advantages over the supermatrix approach (e.g., historical perspective, possible analytical advantages, ability to include all source data). When *applied properly*, supertree construction has been shown to perform on a par with, if not outperform, supermatrix analyses (Bininda-Emonds and Sanderson, 2001; Levasseur and Lapointe, 2001). Thus, we contend that supertree construction is a valuable, transparent, and valid alternative to the supermatrix approach.

Currently, supertree construction represents a first, and less than ideal, method for producing large, comprehensive phylogenies. However, this indicates the incomplete nature of the systematics database rather than a flaw in the method. We would suggest that in most real-world situations a supertree approach will allow examination of the phylogenetic relationships of a large, complete set of taxa more quickly than will a supermatrix approach, and using the widest possible variety of evidence. It must be remembered that supertrees, like phylogenies derived from supermatrix analyses, are phylogenetic hypotheses. In both cases, these hypotheses are open to falsification and will become increasingly accurate with time in line with the continual refinement of systematic opinion due to the accumulation of data (and different data types) and improvements in methodology. Therefore, like any other phylogenetic hypothesis, we see no reason why supertrees should not be used to analyze comparative evolutionary studies in a phylogenetic framework. The alternative is to use less complete trees, limiting both statistical power and possibly the accuracy of the resultant conclusions (see Bininda-Emonds et al., 2002). In the end, any phylogeny—based on a supertree or supermatrix analysis—really only represents current opinion until the next one is published.

References

Arnason, U., and A. Gullberg. 1994. Relationships of baleen whales established by cytochrome *b* gene sequence comparison. Nature 367:726–728.

Arnason, U., and A. Gullberg. 1996. Cytochrome *b* nucleotide sequences and the classification of five primary lineages of extant cetaceans. Mol. Biol. Evol. 13:407–417.

Barrett, M., M. J. Donoghue, and E. Sober. 1991. Against consensus. Syst. Zool. 40:486–493.

Bininda-Emonds, O. R. P. 2000. Factors influencing phylogenetic inference: A case study using the mammalian carnivores. Mol. Phylogenet. Evol. 16:113–126.

Bininda-Emonds, O. R. P. in press. The phylogenetic position of the giant panda (*Ailuropoda melanoleuca*): A historical consensus through supertree analysis. *In* Pandas: Biology and conservation (D. G. Lindburg and K. Baragona, eds.). Univ. California Press, Berkeley.

Bininda-Emonds, O. R. P., J. L. Gittleman, and A. Purvis. 1999. Building large trees by combining phylogenetic information: A complete phylogeny of the extant Carnivora (Mammalia). Biol. Rev. 74:143–175.

Bininda-Emonds, O. R. P., J. L. Gittleman, and M. A. Steel. 2002. The (super)tree of life: Procedures, problems, and prospects. Annu. Rev. Ecol. Syst. 33:265–289.

Bininda-Emonds, O. R. P., K. E. Jones, S. A. Price, M. Cardillo, R. Grenyer, and A. Purvis. in press. Garbage in, garbage out: Data issues in supertree construction. *In* Phylogenetic supertrees: Combining information to reveal the tree of life (O. R. P. Bininda-Emonds, ed.). Kluwer Academic, Dordrecht, The Netherlands.

Bininda-Emonds, O. R. P., and M. J. Sanderson. 2001. Assessment of the accuracy of matrix representation with parsimony supertree construction. Syst. Biol. 50:565–579.

Daubin, V., M. Gouy, and G. Perrière. 2001. Bacterial molecular phylogeny using supertree approach. Genome Informatics 12:155–164.

de Jong, W. W. 1998. Molecules remodel the mammalian tree. Trends Ecol. Evol. 13:270–275.

de Queiroz, A., M. J. Donoghue, and J. Kim. 1995. Separate versus combined analysis of phylogenetic evidence. Annu. Rev. Ecol. Syst. 26:657–681.

Doyle, J. J. 1992. Gene trees and species trees: Molecular systematics as one-character taxonomy. Syst. Bot. 17:144–163.

Farris, J. S. 1983. The logical basis of phylogenetic analysis. Pages 7–36 *in* Advances in cladistics (N. I. Platnick and V. A. Funk, eds.). Columbia Univ. Press, New York.

Gatesy, J., C. Matthee, R. DeSalle, and C. Hayashi. 2002. Resolution of a supertree/supermatrix paradox. Syst. Biol. 51:652–664.

Jenner, R. A. 2001. Bilaterian phylogeny and uncritical recycling of morphological data sets. Syst. Biol. 50:730–742.

Jones, K. E., A. Purvis, A. MacLarnon, O. R. P. Bininda-Emonds, and N. B. Simmons. 2002. A phylogenetic supertree of the bats (Mammalia: Chiroptera). Biol. Rev. 77:223–259.

Källersjö, M., V. A. Albert, and J. S. Farris. 1999. Homoplasy *increases* phylogenetic structure. Cladistics 15:91–93.

Kennedy, M., and R. D. M. Page. 2002. Seabird supertrees: Combining partial estimates of procellariiform phylogeny. Auk 119:88–108.

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

Ledje, C., and U. Arnason. 1996. Phylogenetic analyses of complete cytochrome *b* genes of the order Carnivora with particular emphasis on the Caniformia. J. Mol. Evol. 42:135–144.

Lee, D.-C., and H. N. Bryant. 1999. A reconsideration of the coding of inapplicable characters: Assumptions and problems. Cladistics 15:373–378.

Levasseur, C., and F.-J. Lapointe. 2001. War and peace in phylogenetics: A rejoinder on total evidence and consensus. Syst. Biol. 50:881–891.

Lewis, P. O. 2001. A likelihood approach to estimating phylogeny from discrete morphological character data. Syst. Biol. 50:913–925.

Liu, F.-G. R., M. M. Miyamoto, N. P. Freire, P. Q. Ong, M. R. Tennant, T. S. Young, and K. F. Gugel. 2001. Molecular and morphological supertrees for eutherian (placental) mammals. Science 291:1786–1789.

MADDISON, W. P. 1997. Gene trees in species trees. Syst. Biol. 46:523–536.

MESSENGER, S. L., AND J. A. McGUIRE. 1998. Morphology, molecules, and the phylogenetics of cetaceans. Syst. Biol. 47:90-124.

MICKEVICH, M. F. 1978. Taxonomic congruence. Syst. Zool. 27:143–158.

MILINKOVITCH, M. C., R. G. LEDUC, J. ADACHI, F. FARNIR, M. GEORGES, AND M. HASEGAWA. 1996. Effects of character weighting and species sampling on phylogeny reconstruction: A case study based on DNA sequence data in cetaceans. Genetics 144:1817–1833.

MILINKOVITCH, M. C., A. MEYER, AND J. R. POWELL. 1994. Phylogeny of all major groups of cetaceans based on DNA sequences from three mitochondrial genes. Mol. Biol. Evol. 11:939–948.

MILINKOVITCH, M. C., G. ORTI, AND A. MEYER. 1995. Novel phylogeny of whales revisited but not revised. Mol. Biol. Evol. 12:518–520.

MIYAMOTO, M. M. 1985. Consensus cladograms and general classifications. Cladistics 1:186–189.

NOVACEK, M. J. 1992. Mammalian phylogeny: Shaking the tree. Nature 356:121–125.

NOVACEK, M. J. 2001. Mammalian phylogeny: Genes and supertrees. Curr. Biol. 11:R573-R575.

PAGE, R. D. M., AND M. A. CHARLESTON. 1999. Comments on Allard and Carpenter (1996), or the "aquatic ape" hypothesis revisited. Cladistics 15:73–74.

PENNY, D., AND M. D. HENDY. 1986. Estimating the reliability of evolutionary trees. Mol. Biol. Evol. 3:403–417.

PISANI, D., A. M. YATES, M. C. LANGER, AND M. J. BENTON. 2002. A genus-level supertree of the Dinosauria. Proc. R. Soc. Lond. B 269:915–921.

PURVIS, A. 1995a. A composite estimate of primate phylogeny. Philos. Trans. R. Soc. Lond. B 348:405–421.

PURVIS, A. 1995b. A modification to Baum and Ragan's method for combining phylogenetic trees. Syst. Biol. 44:251–255.

RODRIGO, A. G. 1993. A comment on Baum's method for combining phylogenetic trees. Taxon 42:631–636.

RODRIGO, A. G. 1996. On combining cladograms. Taxon 45:267–274.

SANDERSON, M. J., AND J. KIM. 2000. Parametric phylogenetics? Syst. Biol. 49:817–829.

SANDERSON, M. J., A. PURVIS, AND C. HENZE. 1998. Phylogenetic supertrees: Assembling the trees of life. Trends Ecol. Evol. 13:105–109.

SELUJA, G. A., A. FARMER, M. MCLEOD, C. HARGER, AND P. A. SCHAD. 1999. Establishing a method of vector contamination identification in database sequences. Bioinformatics 15:106–110.

SLOWINSKI, J. B., AND R. D. M. PAGE. 1999. How should species phylogenies be inferred from sequence data? Syst. Biol. 48:814–825.

SPRINGER, M. S., AND W. W. DE JONG. 2001. Phylogenetics. Which mammalian supertree to bark up? Science 291:1709–1711.

STONER, C. J., O. R. P. BININDA-EMONDS, AND T. M. CARO. 2003. The adaptive significance of coloration in lagomorphs. Biol. J. Linn. Soc. 79:309–328.

STRONG, E. E., AND D. LIPSCOMB. 1999. Character coding and inapplicable data. Cladistics 15:363–371.

WILSON, D. E., AND D. M. REEDER (eds.). 1993. Mammal species of the world: A taxonomic and geographic reference. Smithsonian Institution Press, Washington, D.C.