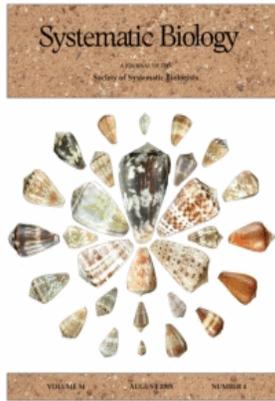


This article was downloaded by:[American Museum of Natural History]
On: 22 July 2008
Access Details: [subscription number 789507793]
Publisher: Taylor & Francis
Informa Ltd Registered in England and Wales Registered Number: 1072954
Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



Systematic Biology

Publication details, including instructions for authors and subscription information:
<http://www.informaworld.com/smpp/title~content=t713658732>

Resolution of a Supertree/Supermatrix Paradox

John Gatesy^a; Conrad Matthee^b; Rob DeSalle^c; Cheryl Hayashi^a

^a Department of Biology, University of California, Riverside, California 92521, USA.

^b Department of Zoology, University of Stellenbosch, Stellenbosch 7602, South Africa.

^c Department of Invertebrates, American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024, USA.

First Published on: 01 July 2002

To cite this Article: Gatesy, John, Matthee, Conrad, DeSalle, Rob and Hayashi, Cheryl (2002) 'Resolution of a Supertree/Supermatrix Paradox', Systematic Biology, 51:4, 652 — 664

To link to this article: DOI: 10.1080/10635150290102311

URL: <http://dx.doi.org/10.1080/10635150290102311>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article maybe used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

information on the history of northeastern Africa because the fauna is composed of different elements with different histories that cannot be shown in a single area cladogram. An analysis of the biotic elements is a more appropriate first step toward understanding the genesis of composite biotas.

REFERENCES

- AXELIUS, B. 1991. Areas of distribution and areas of endemism. *Cladistics* 7:197–199.
- BIRKS, H. J. B. 1987. Recent methodological developments in quantitative descriptive biogeography. *Ann. Zool. Fenn.* 24:165–177.
- DE LATTIN, G. 1957. Die Ausbreitungszentren der holarktischen Landtierwelt. *Zool. Anz. Suppl.* 20: 380–410.
- DENNIS, R. L. H., W. R. WILLIAMS, AND T. G. SHREEVE. 1998. Faunal structures among European butterflies: Evolutionary implications of bias for geography, endemism and taxonomic affiliation. *Ecography* 21: 181–203.
- FREY, J. K. 1992. Response of a mammalian faunal element to climatic changes. *J. Mammal.* 73:43–50.
- HAROLD, A. S., AND R. D. MOOL. 1994. Areas of endemism: Definition and recognition criteria. *Syst. Biol.* 43:261–266.
- HENDERSON, I. M. 1991. Biogeography without area? *Aust. Syst. Bot.* 4:59–71.
- HOLLOWAY, J. D., AND N. JARDINE. 1968. Two approaches to zoogeography: A study based on the distributions of butterflies, birds and bats in the Indo-Australian area. *Proc. Linn. Soc. Lond.* 179:153–188.
- HOVENKAMP, P. 1997. Vicariance events, not areas, should be used in biogeographical analysis. *Cladistics* 13:67–79.
- HUMPHRIES, C. J., AND L. R. PARENTI. 1999. Cladistic biogeography. Oxford Univ. Press, Oxford, U.K.
- JARDINE, N. 1972. Computational methods in the study of plant distributions. Pages 381–393 in *Taxonomy, phytogeography and evolution* (D. H. Valentine, ed.). Academic Press, London.
- KAISER, G. W., L. P. LEFKOVITCH, AND H. F. HOWDEN. 1972. Faunal provinces in Canada as exemplified by mammals and birds: A mathematical consideration. *Can. J. Zool.* 50:1087–1104.
- LINDER, H. P. 2001. On areas of endemism, with an example from the African Restionaceae. *Syst. Biol.* 50:892–912.
- LINDER, H. P., AND D. M. MANN. 1998. The phylogeny and biogeography of *Thamnochortus* (Restionaceae). *Bot. J. Linn. Soc.* 128:319–357.
- MAYR, E. 1944. Wallace's line in the light of recent zoogeographic studies. *Q. Rev. Biol.* 19:1–14.
- MAYR, E. 1965. What is a fauna? *Zool. Jahrb. Syst.* 92:473–486.
- MORRONE, J. J. 1994. On the identification of areas of endemism. *Syst. Biol.* 43:438–441.
- MORRONE, J. J., AND J. V. CRISCI. 1995. Historical Biogeography: Introduction to methods. *Annu. Rev. Ecol. Syst.* 26:373–401.
- NELSON, G., AND N. PLATNICK. 1981. Systematics and biogeography: Cladistics and vicariance. Columbia Univ. Press, New York.
- PALLARY, P. 1909. Catalogue de la faune malacologique de l'Égypte. *Mem. Inst. Égyptien* 6:1–92.
- PETERS, J. A. 1955. Use and misuse of the biotic province concept. *Am. Nat.* 89:21–28.
- PLATNICK, N. I. 1991. On areas of endemism. *Aust. Syst. Bot.* 4:xi–xii.
- PLATNICK, N. I., AND G. NELSON. 1978. A method of analysis for historical biogeography. *Syst. Zool.* 27: 1–16.
- REBEL, H. 1931. Zur Frage der europäischen Faunenelemente. *Ann. Naturhist. Mus. Wien* 46:49–55.
- SIMPSON, G. G. 1977. Too many lines; the limits of the Oriental and Australian zoogeographic regions. *Proc. Am. Philos. Soc.* 121:107–120.
- UDVARDY, M. D. F. 1969. *Dynamic zoogeography*. Van Nostrand Reinhold, New York.
- VANE-WRIGHT, R. I. 1991. Transcending the Wallace line: Do the western edges of the Australian region and the Australian plate coincide? *Aust. Syst. Bot.* 4:183–197.

First submitted 9 May 2001; reviews returned
22 March 2002; final acceptance 2 May 2002
Associate Editor: Peter Linder

Syst. Biol. 51(4):652–664, 2002
DOI: 10.1080/10635150290102311

Resolution of a Supertree/Supermatrix Paradox

JOHN GATESY,¹ CONRAD MATTHEE,² ROB DE SALLE,³ AND CHERYL HAYASHI¹

¹Department of Biology, University of California, Riverside, California 92521, USA

²Department of Zoology, University of Stellenbosch, Stellenbosch 7602, South Africa

³Department of Invertebrates, American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024, USA

In the total evidence approach to systematics, all characters and taxa are merged in a single “supermatrix,” and the data are analyzed simultaneously (Miyamoto, 1985; Kluge, 1989; Nixon and Carpenter, 1996). Recently, however, a plea has been made

for the use of “supertrees” in systematics (Sanderson et al., 1998). In this framework, individual *characters* are not interpreted as phylogenetic evidence. Instead, *topologies* supported by different published studies, that is, source trees, are encoded into

a matrix and used to reconstruct the evolutionary history of a particular group (Baum, 1992; Ragan, 1992; Purvis, 1995b). Such supertrees are considered by some to be excellent summaries of past research and useful frameworks for comparative biologists (Purvis, 1995a; Bininda-Emonds et al., 1999; Liu et al., 2001), but if supertrees are intended to be meaningful phylogenetic tools, the underlying support for these topologies should be accountable. We investigated the consistent support for a monophyletic Artiodactyla (antelopes, pigs, hippos, camels, and close relatives) to the exclusion of Cetacea (whales) in a recent supertree analysis (Liu et al., 2001). We then compared this *topological* support to the strongly conflicting *character* support in a supermatrix analysis of Paraxonia (Artiodactyla + Cetacea) and reconciled the apparently paradoxical conclusions of these two studies.

BACKGROUND

The Artiodactyla historically has been recognized as a monophyletic group based on a small suite of apparently conservative morphological synapomorphies, such as a double-trochleated astragalus and trilobed deciduous lower fourth premolars (Luckett and Hong, 1998). In the past 10 years, a cascade of studies has challenged the traditional view. For example, the supermatrix compiled by Gatesy et al. (1999) contained 3 mitochondrial (mt) gene sequences, 7 nuclear (nu) gene sequences, 5 nu protein sequences, insertions of transposons at 9 loci, and 116 morphological characters; the shortest tree that included a monophyletic Artiodactyla was 125 steps longer than the minimum length (Fig. 1a). The supermatrices of Arnason et al. (2000; 12 mt genes), Matthee et al. (2001; 8 nu genes and 1 mt gene), and Murphy et al. (2001, 15 nu genes and 3 mt genes) gave

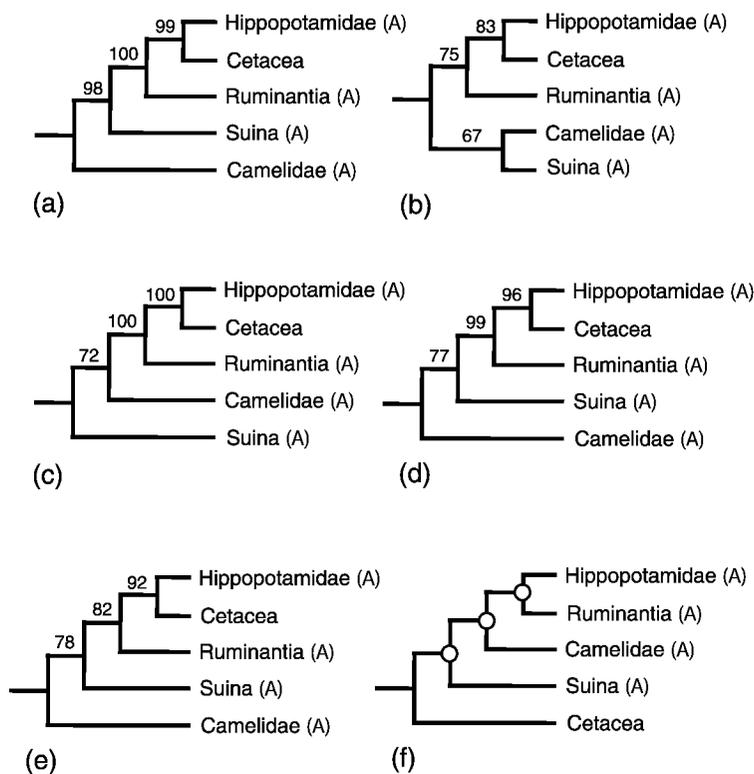


FIGURE 1. Phylogenetic relationships of paraxonians based on explicit character data (a–e) and a combined supertree analysis (f). Data sets are those of Gatesy et al. (1999) (a), Arnason et al. (2000) (b), Matthee et al. (2001) (c), Murphy et al. (2001) (d), Nikaido et al. (1999; also see Hillis, 1999) (e), and Liu et al. (2001) (f). A = artiodactyl subtaxa. Parsimony bootstrap percentages (Felsenstein, 1985) are shown above internodes for character-based studies. A monophyletic Artiodactyla and two other clades in the supertree (f) are inconsistent with the character-based studies (a–e). Circles at internal nodes mark these clades.

similar results (Figs. 1b–d). Each of these large combined data sets included >7,000 characters and robustly supported a paraphyletic Artiodactyla with hippopotamid artiodactylans grouping closer to Cetacea than to ruminant, suine, and camelid artiodactylans (Ruminantia, Suina, and Camelidae, respectively; Fig. 1). Additional combined analyses (Graur and Higgins, 1994; Montgelard et al., 1997; Gatesy, 1998; Madsen et al., 2001; O'Leary, 2001) and specific insertions of transposons (Fig. 1e; Nikaido et al., 1999) also did not support artiodactyl monophyly.

In contrast to these studies, Liu et al. (2001) utilized supertree methods in an analysis of mammalian phylogeny. Four hundred thirty phylogenetic trees from 315 systematic studies were recoded as nodal characters and used to construct a supertree, employing the matrix representation with parsimony method (MRP; Baum, 1992; Ragan, 1992). In part, results from this analysis were consistent with previous supermatrix studies, but there were some controversial discrepancies. For example, monophyly of Artiodactyla was favored by the supertree data set over a range of weighting schemes ("large" data sets weighted from one to four times "small" data sets), and two artiodactyl subclades that have never been endorsed seriously by previous authors were resolved by the giant analysis (Fig. 1f). The 89 morphological source trees in the supertree data set overwhelmingly supported a monophyletic Artiodactyla (Liu et al., 2001).

The supertree results of Liu et al. (2001) contradicted every modern supermatrix analysis of paraxonian phylogeny (Figs. 1a–d), and the robustness of support for a monophyletic Artiodactyla did not seem to be consistent with the limited number of actual characters that have been forwarded as synapomorphies for this group (Prothero, 1993; Luckett and Hong, 1998; O'Leary and Geisler, 1999; Thewissen and Madar, 1999; Geisler, 2001). These discrepancies may be due to the many more data sets accounted for in this supertree analysis relative to past supermatrix studies. Alternatively, the support for Artiodactyla in the study of Liu et al. could be the result of problems in their supertree analysis (Springer and DeJong, 2001).

Previously published mammalian supertrees (Purvis, 1995a; Bininda-Emonds

et al., 1999) have been the basis for a variety of subsequent evolutionary studies (e.g., Purvis et al., 1995; Blackburn et al., 1997; Purvis and Bromham, 1997; Gittleman and Purvis, 1998; Woodroffe and Ginsberg, 1998). Likewise, the supertree of Liu et al. (2001) is the most comprehensive phylogenetic analysis of eutherian mammals at the family level and undoubtedly will be utilized in future comparative studies, so it is important to determine whether this topology has a sound empirical basis. Artiodactylans and close relatives are exceptionally well characterized for diverse character systems (Graur and Higgins, 1994; Montgelard et al., 1997; Gatesy, 1998; Luckett and Hong, 1998; Messenger and McGuire, 1998; O'Leary and Geisler, 1999; Nikaido et al., 1999; Thewissen and Madar, 1999; Arnason et al., 2000; Geisler, 2001; Langer, 2001; Madsen et al., 2001; Matthee et al., 2001; Murphy et al., 2001). We therefore examined the MRP support for artiodactyl relationships in the supertree data set of Liu et al. (2001), and compared this evidence to actual character support in an extensive supermatrix of eutherian mammals that included 75 taxa and >37,000 characters.

PARSIMONY ANALYSIS OF THE SUPERTREE DATA SET

The supertree data set of Liu et al. (2001) was reanalyzed using PAUP* (Swofford, 1998). All 91 operational taxonomic units (OTUs) were included, and the parsimony search was heuristic with 1,000 random taxon addition replicates, TBR branch swapping, the amb- option in effect, all characters unordered, and all character transformations equally weighted. The strict consensus of optimal trees (minimum length = 2,929 steps) was consistent with the results reported by Liu et al. (2001) (Fig. 2). Branch support (Bremer, 1994) for a monophyletic Artiodactyla was 13, estimated using the "constraints" command of PAUP*. Artiodactyla was favored even when nodes derived from "large" data sets, as defined by Liu et al. were weighted five times more than nodes derived from "small" data sets. Source trees that supported artiodactyl monophyly in the equally weighted analysis were identified with the "list of apomorphies" command of PAUP*. Thirty-three nodal characters from 30 source trees provided

Downloaded By: [American Museum of Natural History] At: 01:10 22 July 2008

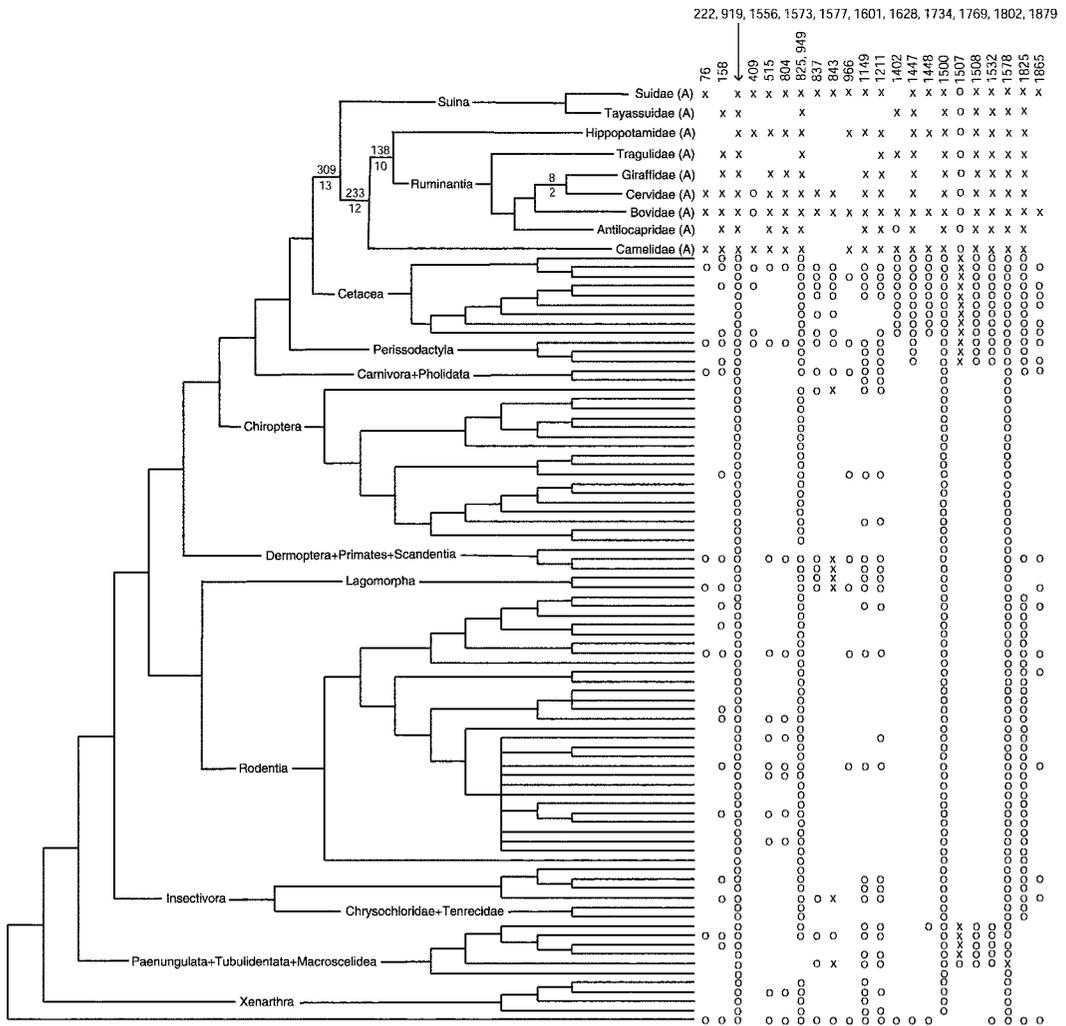


FIGURE 2. Strict consensus of 234 trees (minimum length = 2,929) derived from equally weighted analysis of the supertree data set (Liu et al., 2001). The taxonomic distribution of nodal characters that provided unambiguously optimized synapomorphies for Artiodactyla is shown (character states = x and o; missing = no state). Numbering of nodal characters is as for Liu et al. (2001). Branch support scores (Bremer, 1994) for four artiodactyl clades that contradicted supermatrix results are below internodes. The costs in extra character steps required to recover each of these clades in the supermatrix analysis (Fig. 4) are above internodes. Artiodactyl families (A) are indicated, and higher level taxa are shown at internodes. See Liu et al. (2001) for identities of all OTUs. Branch lengths are not proportional to the numbers of character changes. When dubious nodal characters were deleted (characters 76, 158, 222, 515, 825, 837, 843, 949, 966, 1149, 1211, 1402, 1447, 1448, 1500, 1507, 1508, 1532, 1556, 1573, 1577, 1578, 1601, 1628, 1734, 1769, 1802, 1825, 1865, and 1879 were based on reviews, assumptions of monophyly, redundancies, and poorly justified source trees; see Fig. 3), the supertree data set supported Hippopotamidae + Cetacea and Hippopotamidae + Ruminantia + Cetacea instead of artiodactyl monophyly.

unequivocally optimized synapomorphies for Artiodactyla (Fig. 2).

PROBLEMS IN THE SUPERTREE DATA SET
Duplication of Character Evidence

Although Liu et al. (2001) noted that "... redundancy and nonindependence were

lessened as the 430 source trees were compiled for supertree analysis" (p. 1789), their supermatrix remained highly redundant. Extensive duplications of morphological characters and of certain amino acid sequences were present (red, brown, and blue interconnections, Fig. 3). For example, beta-hemoglobin sequences were used to

construct 6 of the 30 source trees that supported a monophyletic Artiodactyla (Fig. 3). Each of these six source trees was coded as a separate nodal character in the supertree analysis. The morphological character matrix of Novacek (1994) was basically duplicated in three of the source trees that supported Artiodactyla and was partially reproduced in another data set (Shoshani and McKenna, 1998). The analysis of 41 ribonuclease sequences by Fitch and Beintema (1990) was nearly identical to the analysis of 39 sequences by Beintema et al. (1988). In turn, these ribonuclease data sets broadly overlapped with two tandem protein alignments (Miyamoto and Goodman, 1986; Czelusniak et al., 1990; Fig. 3).

The combined data sets of Miyamoto and Goodman (1986) and Czelusniak et al. (1990) included evidence from seven and eight proteins, respectively. These and other large matrices were given a high weight in some analyses by Liu et al. (2001). "This weighting strategy was based on recent empirical and theoretical work in support of the greater power and accuracy of multiple data sets and increased character and taxon sampling in phylogenetic estimation" (Liu et al., 2001, p. 1786). The seven protein sequences in Miyamoto and Goodman (1986), however, were all utilized in the analysis of Czelusniak et al. (1990), and both of these tandem alignments overlapped, at least partially, with nine other data sets that supported Artiodactyla in the supertree analysis (Fig. 3). Upweighting such nonindependent, large data sets only compounded the problem of character redundancy in the supertree data set.

Appeals to Authority

Many of the source trees that supported Artiodactyla in the supertree analysis were simply appeals to authority. An appeal to authority was defined here in three ways.

Studies in which the original author of a source tree assumed that Artiodactyla was monophyletic.—For example, Wyss et al. (1987) examined relationships among extant placental orders. To streamline their analysis, they constrained the monophyly of Artiodactyla, which at the time was not particularly controversial. Liu et al. (2001) translated this assumption into their supertree data set by coding all nine artiodactyl families as composing a monophyletic

group (Figs. 2 and 3), even though four of these nine artiodactyl taxa were not included in the analysis of Wyss et al. (1987). Such appeals to authority applied to 11 of the nodal characters that supported artiodactyl monophyly (blue dots, Fig. 3). Liu et al. (2001) realized that these appeals to authority hindered their analysis but rationalized that assumptions of monophyly were "much less of an issue for the molecular support, because only six of its 314 source phylogenies were based on interfamilial operational taxonomic units" (p. 1789). Three molecular source trees that supported Artiodactyla in the supertree analysis assumed Artiodactyla was monophyletic *a priori*, and two other molecular parsimony analyses, coded by Liu et al. as supporting Artiodactyla, recovered equally short trees that contradicted artiodactyl monophyly (Goodman et al., 1982; Czelusniak et al., 1990; Fig. 3).

Composite trees pieced together by researchers based on results from previously published studies.—These composite trees were supertrees but were not assembled according to specific rules (e.g., the strict supertree method or MRP; see Sanderson et al., 1998). Three appeals to authority of this type supported Artiodactyla in the supertree data set (yellow dots, Fig. 3), and each of these composite trees used reviews as their basis for artiodactyl monophyly (purple connections, Fig. 3).

Reviews of previous studies.—Such reviews (pink dots, Fig. 3) presented no new data or analyses and essentially summarized the opinions of influential mammalogists based on previously published studies. In four cases, reviews coded as independent nodal characters by Liu et al. (2001) might be considered both appeals to authority and duplications of evidence. These reviewers (Janis, 1988; Novacek, 1992; Milinkovitch and Thewissen, 1997; DeJong, 1998) cited other source trees in the supertree data set that supported artiodactyl monophyly (green connections, Fig. 3). One reviewer (Novacek, 1992) cited five studies in the supertree data set, and these five studies in turn shared character data with six other source trees that supported Artiodactyla (Fig. 3). Another set of reviews (Janis, 1988; Novacek, 1992; Prothero, 1993; Milinkovitch and Thewissen, 1997) had overlapping references (pink connections, Fig. 3), but

these common references (Schaeffer, 1947; Prothero et al., 1988) were not included as characters in the supertree data set.

Poorly Justified Source Trees

Several of the source trees that supported Artiodactyla in the supertree analysis were poorly justified. Four of these trees were based on character data from more than 30 taxa, but the original authors did not execute rigorous tree searches. Data sets were analyzed by hand and by eye without the aid of computer programs. This situation was prevalent in some of the older studies that did not even include explicit character matrices. Another problematic study (Thewissen and Madar, 1999) included putative anklebones from fossil whales that were misidentified (see Thewissen et al., 2001). Thus, five of the nodal characters that supported artiodactyl monophyly were derived from dubious source trees (orange dots, Fig. 3).

Miscoding of Source Trees

Two characters that supported artiodactyl monophyly in the supertree analysis were miscoded. When recoded, these nodal characters did not provide unequivocal support for Artiodactyla (green dots, Fig. 3).

Misinterpretation of Relative Support in Different Data Sets

Many nodes supported by different published data sets were treated as if they were equally well supported in the supertree

study of Liu et al. (2001). For example, the grouping of Suidae, Hippopotamidae, and Camelidae favored by the beta-hemoglobin data set analyzed by Gatesy et al. (1999) (character 409, Fig. 2) had a branch support score of 1 and a bootstrap (Felsenstein, 1985) value of 46%. In the supertree analysis of Liu et al. this weakly supported node was given the same weight as robustly supported contradictory nodes from other published analyses (e.g., Cetacea + Hippopotamidae + Ruminantia; branch support = 5, bootstrap = 97%; based on gamma-fibrinogen sequences of Gatesy, 1997). It is impossible to translate accurately the relative character support from different source trees into a supertree data set (Purvis, 1995b; Ronquist, 1996; Bininda-Emonds and Bryant, 1998; Bininda-Emonds et al., 1999). Weighting nodal characters according to branch support scores, relative bootstrap percentages, or branch lengths in the original source trees (see Sanderson et al., 1998) is not a perfect solution to this problem because hidden character support in different data sets that emerges in supermatrix analysis (Barrett et al., 1991) is ignored in a supertree analysis (Pisani and Wilkinson, 2002).

SUPERMATRIX ANALYSIS OF PARAXONIA

All 33 nodal characters of Liu et al. (2001) that supported a monophyletic Artiodactyla were appeals to authority, duplications of data, miscodings, or derivatives of poorly justified source trees (Fig. 3). To determine how this information compared to evidence

FIGURE 4. Minimum length topology of 67,357 steps supported by parsimony analysis of the supermatrix with all characters unordered. Operational taxonomic units (OTUs) are shown to the right. Higher level taxa are in capitals and are delimited by brackets to the right of OTUs. Families of ruminant artiodactylans (A-R), families of suine artiodactylans (A-S), other families of artiodactylans (A), families of cetaceans (C), and outgroup taxa to Paraxonia (O) are indicated. Data sets are shown at the top of the figure. Taxa that were sampled for these data sets are indicated by colored circles (yellow = morphology; red = transposons; green = nu amino acid sequences; blue = nu DNA sequences; purple = mt DNA sequences), and missing data in the supermatrix are colored beige. Branch support scores (Bremer, 1994) for relationships among paraxonian families are above internodes. One thousand random taxon addition replicates were utilized in each constrained heuristic search, but given the complexity of the supermatrix data set these branch support scores may be lower than indicated. Bootstrap values that were >69% are indicated below internodes (7 = 70–79%; 8 = 80–89%; 9 = 90–99%; 10 = 100%). One thousand bootstrap replicates were generated using heuristic searches of informative characters with simple taxon addition and tree bisection-reconnection branch swapping (Swofford, 1998). The statistical assumptions of the bootstrap (Felsenstein, 1985) were violated by the supermatrix data set. Black circles mark clades that were inconsistent with the supertree analysis of Liu et al. (2001). Taxonomic exemplars for higher level groups that were assumed to be monophyletic are given at the *Systematic Biology* web site. The tree is rooted according to the hypotheses of Liu et al. (2001), Madsen et al. (2001), and Murphy et al. (2001). When some morphological characters were ordered as prescribed by Geisler (2001), the same minimum length topology was supported. Branch lengths are not proportional to the number of character changes.

for or against artiodactyl monophyly in supermatrices, we analyzed a revised version of the WHIPPO-2 data set from Gatesy et al. (1999). The original supermatrix included character data from three mt genes and seven nu genes. This data set was augmented by the addition of morphological evidence (Messenger and McGuire, 1998; Geisler, 2001; Langer, 2001), retropositional insertions (Shimamura et al., 1997; Nikaido et al., 1999; 2001), 12 mt genes, segments of 24 nu genes, amino acid sequences from 5 nu proteins, and additional DNA sequence data for mt cytochrome *b*, mt 12S ribosomal DNA (rDNA), nu vWF, and nu IRBP. Sequence data were downloaded from NCBI or were from published alignments (Gatesy et al., 1999; Madsen et al., 2001; Matthee et al., 2001; Murphy et al., 2001).

Mitochondrial cytochrome *b* sequences were available for all 75 OTUs in the supermatrix, and 12S rDNA sequences were published for 73 of these taxa. Each OTU shared at least some characters with every other OTU in the data set and had sequence data for minimally three genes/gene products (Fig. 4). Liu et al. (2001) included 91 taxa in their supertree data set and implicitly assumed that each of these OTUs was monophyletic. For our supermatrix, we did not make assumptions of monophyly that extended to higher taxonomic levels than those of the 91 OTUs considered monophyletic by Liu et al. (see Fig. 4). The artiodactyl supermatrix (SuperArtio-1.dat) is available at the *Systematic Biology* web site (Systematicbiology.org/info/issues.html).

There were extensive missing data in our supermatrix (Fig. 4). In parsimony analysis, these missing data might make long branch attraction effects more pronounced (e.g., Halanych, 1998). In maximum likelihood and phenetic analyses, extensive missing data can impact parameters that are based on summations of all characters (i.e., branch lengths, nucleotide composition biases, corrected distances between taxa, etc.). In turn, such parameters affect the influence of *all characters* in the phylogenetic analysis. In cladistic analyses, missing information for a particular character only affects the influence of *that character* in phylogenetic analysis. To our knowledge, all previously published analyses of supermatrices that included morphological characters, molecular information, and extensive missing data

have utilized a parsimony approach (e.g., O'Leary, 2001). We also have chosen to use standard cladistic methods to search for maximum congruence among diverse characters (Farris, 1983; Kluge, 1989).

Analysis of the 14,124 informative characters in the supermatrix was as described above for the supertree data set. Although the supermatrix was quite large and very incomplete for several component data sets, our heuristic search with 1,000 random taxon additions took only ~200 min on a 450-MHz Macintosh Power PC G4 (128 MB SDRAM). Perhaps because mt cytochrome *b* data linked all included taxa (Bininda-Emonds and Sanderson, 2001), perceived practical and analytical limitations (Liu et al., 2001) of supermatrix studies were avoided. A single optimal topology of 67,357 steps was recovered (Fig. 4). Cetacea was nested two nodes within a paraphyletic Artiodactyla, and relationships among Cetacea, Hippopotamidae, Ruminantia, Suina, and Camelidae were as in the mt genome analysis of Arnason et al. (2000) (see Fig. 1b). The Camelidae + Suina clade, which was contradicted by some previous supermatrix studies (Figs. 1a, 1c, 1d), was very unstable (branch support = 8, bootstrap = 57%) and collapsed with the exclusion of all mt DNA data from our supermatrix.

For the total supermatrix, artiodactyl monophyly demanded an additional 309 character steps relative to the minimum length tree and was strongly denied according to a signed-rank test of character support (two-tailed $P < 0.0001$; Templeton, 1983). Within a strictly cladistic framework (Farris, 1983), such statistical tests are difficult to interpret. Nevertheless, in terms of stability to the addition of new character data (see Nixon and Carpenter, 1996), artiodactyl paraphyly is a robust phylogenetic hypothesis. Even if 308 perfectly congruent synapomorphies for Artiodactyla were added to our supermatrix, the combined data set would not yield a monophyletic Artiodactyla.

SUPERMATRIX/SUPERTREE INCONSISTENCIES

Some clades that were overwhelmingly supported by the supermatrix contradicted the trees supported by the supertree data set of Liu et al. (2001). For example, a sister group relationship between Cetacea and Hippopotamidae had a branch support

score of 103 in the supermatrix analysis (bootstrap = 100%, Fig. 4). In the supertree analysis, Cetacea was separated from Hippopotamidae by three nodes (Fig. 2), and 13 extra steps were required to join these taxa. Instead of grouping Hippopotamidae with Cetacea, the supertree data set linked hippopotamids with ruminant artiodactylans (Fig. 2). The 13 source trees that provided unequivocal synapomorphies for this unorthodox clade were highly redundant; none were completely independent in terms of character data. One source tree (Gatesy et al., 1999) partially overlapped with every other data set that supported this same node! In the original publications (Gatesy et al., 1996, 1999; Gatesy, 1997; Shimamura et al., 1997), 9 of the 13 source trees resolved Hippopotamidae + Ruminantia + Cetacea and *not* Hippopotamidae + Ruminantia. Hippopotamidae + Ruminantia + Cetacea was robustly supported by our supermatrix (Fig. 4; branch support = 97, bootstrap = 100%), but in the supertree analysis appeals to authority and character duplications excluded Cetacea from a position within Artiodactyla (Fig. 3). Thus, Hippopotamidae + Cetacea and Hippopotamidae + Ruminantia + Cetacea were not supported; in their place, three clades that strongly conflicted with actual character data were favored (Fig. 2).

Of eight nodes that defined relationships among artiodactyl families in the supertree (Fig. 2), four disagreed with the supermatrix results (Fig. 4). Four hundred seventy-three extra character steps were required from the supermatrix to enforce all of these topological differences. A 50% match between trees was unimpressive, given the amounts of data summarized in each of these large data sets. Such conflicts apparently had no empirical basis in the characters and taxa examined, however, and were quirky epiphenomena of the source tree miscodings in the supertree data set (Fig. 3). When problematic and redundant nodal characters that supported artiodactyl monophyly were deleted from the supertree data set, Hippopotamidae + Cetacea and Hippopotamidae + Ruminantia + Cetacea were supported (see Fig. 2). Unfortunately, these corrections solved only a fraction of the difficulties in the supertree data set. The paraxonian section of the combined supertree from Liu et al. (2001) is flawed by the methods employed and should not be

utilized in future comparative evolutionary studies.

Inconsistencies in the supertree data set were not limited to Paraxonia. Duplications of evidence and other problems may have skewed other relationships favored by this data set. For example, the alpha- and beta-hemoglobin source trees from Wyss et al. (1987) were both coded as supporting a grouping of golden moles (Chrysochloridae), tenrecs (Tenrecidae), moles, shrews, solenodons, and hedgehogs (Insectivora; see Fig. 2). However, hemoglobins have not been sequenced from golden moles or tenrecs. Insectivoran monophyly simply was assumed in this case and in several other studies coded as supporting this clade. The beta-hemoglobin sequences of Wyss et al. (1987) broadly overlapped with the beta-hemoglobin sequences analyzed by Novacek (1994), which in turn shared morphological characters with trees of Shoshani and McKenna (1998). Wyss et al. (1987) was cited by Novacek (1992), who also cited McKenna (1975). In the supertree data set of Liu et al. (2001), all five of these studies (McKenna, 1975; Wyss et al., 1987; Novacek, 1992; Novacek, 1994; Shoshani and McKenna, 1998) were coded as independent characters that grouped insectivorans. For each node in the supertree that we investigated, such duplications of supporting evidence were present (also see Springer and DeJong, 2001).

Other recent supertrees of mammals (Purvis, 1995a; Bininda-Emonds et al., 1999) showed similar patterns. When different researchers had analyzed the same character set for carnivorans, Bininda-Emonds et al. (1999) "used each tree because differences in the analyses might change the results between the studies" (p. 147). Thus, a data set analyzed by a variety of authors in the past was given a redundantly high weight in their supertree analysis. Purvis (1995a) constrained his analysis of 203 primate species by enforcing the monophyly of 24 traditional higher level taxa and incorporated studies into his supertree data that "were never *intended* to estimate phylogenetic relations," adding that "any information is better than none [!]" (p. 406). Unnecessary repetitions of character evidence negatively influenced both of these supertree data sets, and appeals to authority further distorted the original data. To this point, only four major supertree analyses of mammals have been published

(Purvis, 1995a; Bininda-Emonds et al., 1999; Liu et al., 2001; Jones et al., 2002); each of these studies included similar disappointing inconsistencies.

CONCLUSION

Proponents of mammalian supertrees repeatedly have argued that these topologies are useful summaries of past work and efficient guides to what systematic research is needed (Purvis, 1995a; Bininda-Emonds et al., 1999; Liu et al., 2001; Jones et al., 2002). It would be difficult, however, to argue that a supertree is a better depiction of past research than a supermatrix. A supermatrix clearly reviews which *characters* have or have not been scored for particular taxa. These primary data are explicitly presented, no information is duplicated, and any editing errors or sequence alignment discrepancies are easily accessible to the scrutiny of other researchers. By comparison, recently published mammalian supertrees are imprecise summaries of previous work. These topologies are not necessarily good indicators of what relationships are supported by the current systematic database and should not be used to guide future phylogenetic research or to organize comparative evolutionary studies.

Liu et al. (2001) suggested that "our supertrees advance the field one step closer toward its ultimate goal of a phylogenetic synthesis for eutherian mammals at the level of their individual datum and character" (p. 1789). Based on an extensive supermatrix analysis of approximately 1 megabase of molecular sequence data, transposons, and morphological characters, we do not agree with this statement. The supertree analysis of Liu et al. has perpetuated phylogenetic myths by duplicating character evidence, resurrecting outdated analyses that in a modern context are deficient, confusing appeals to authority with actual evidence for competing phylogenetic hypotheses, and distorting the relative character support in different data sets. The end result of such supertree analyses is that the support for any particular node (e.g., Fig. 3) may have little or no empirical basis.

ACKNOWLEDGMENTS

M. Allard, R. DeBry, A. de Queiroz, C. Simon, M. Springer, J. Sullivan, and an anonymous reviewer made helpful comments. R. Liu and M. Miyamoto made

their supertree data set available for reanalysis, and a variety of scientists contributed molecular data to this study through the NCBI database. J. Geisler allowed use of his morphological data set prior to its publication. This work was funded by NSF grant DEB-9985847.

REFERENCES

- ARNASON, U., A. GULLBERG, S. GRETARSDOTTIR, B. URSING, AND A. JANKE. 2000. The mitochondrial genome of the sperm whale and a new molecular reference for estimating eutherian divergence dates. *J. Mol. Evol.* 50:569–578.
- BARRETT, M., M. J. DONOGHUE, AND E. SOBER. 1991. Against consensus. *Syst. Zool.* 40:486–493.
- BAUM, B. 1992. Combining trees as a way of combining data sets for phylogenetic inference, and the desirability of combining gene trees. *Taxon* 41:3–10.
- BEINTEMA, J., C. SCHULLER, M. IRIE, AND A. CARSONA. 1988. Molecular evolution of the ribonuclease superfamily. *Prog. Biophys. Mol. Biol.* 51:165–192.
- BININDA-EMONDS, O. R. P., AND H. BRYANT. 1998. Properties of matrix representation with parsimony analyses. *Syst. Biol.* 47:497–508.
- BININDA-EMONDS, O., J. 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., AND M. J. SANDERSON. 2001. Assessment of the accuracy of matrix representation with parsimony analysis supertree construction. *Syst. Biol.* 50:565–579.
- BLACKBURN, T., K. GASTON, R. QUINN, H. ARNOLD, AND R. GREGORY. 1997. Of mice and wrens: The relation between abundance and geographic range size in British mammals and birds. *Philos. Trans. R. Soc. Lond. B* 352:419–427.
- BREMER, K. 1994. Branch support and tree stability. *Cladistics* 10:295–304.
- CZELUSNIAK, J., M. GOODMAN, B. KOOP, D. TAGLE, J. SHOSHANI, G. BRAUNITZER, T. KLEINSCHMIDT, W. DEJONG, AND G. MATSUDA. 1990. Perspectives from amino acid sequences on cladistic relationships among higher taxa of Eutheria. *Curr. Mammal.* 2:545–572.
- DEJONG, W. 1998. Molecules remodel the mammalian tree. *Trends Ecol. Evol.* 13:270–275.
- FARRIS, J. 1983. The logical basis of phylogenetic analysis. Pages 7–36 in *Advances in cladistics*, Volume 2 (N. Platnick and V. Funk, eds.). Columbia Univ. Press, New York.
- FELSENSTEIN, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783–791.
- FITCH, W., AND J. BEINTEMA. 1990. Correcting parsimonious trees for unseen nucleotide substitutions: The effect of dense branching as exemplified by ribonuclease. *Mol. Biol. Evol.* 7:438–443.
- GATESY, J. 1997. More DNA support for a Cetacea/Hippopotamidae clade: The blood-clotting protein gene gamma-fibrinogen. *Mol. Biol. Evol.* 14:537–543.
- GATESY, J. 1998. Molecular evidence for the phylogenetic affinities of Cetacea. Pages 63–111 in *The emergence of whales* (J. Thewissen, ed.). Plenum, New York.
- GATESY, J., C. HAYASHI, M. CRONIN, AND P. ARCTANDER. 1996. Evidence from milk casein genes that cetaceans

- are close relatives of hippopotamid artiodactyls. *Mol. Biol. Evol.* 13:954–963.
- GATESY, J., M. MILINKOVITCH, V. WADDELL, AND M. STANHOPE. 1999. Stability of cladistic relationships between Cetacea and higher-level artiodactyl taxa. *Syst. Biol.* 48:6–20.
- GAUDIN, T., J. WIBLE, J. HOPSON, AND W. TURNBULL. 1996. Reexamination of the morphological evidence for the cohort Epitheria (Mammalia, Eutheria). *J. Mammal. Evol.* 3:31–79.
- GEISLER, J. 2001. New morphological evidence for the phylogeny of Artiodactyla, Cetacea, and Mesonychia. *Am. Mus. Novit.* 3344:1–53.
- GITTLEMAN, J., AND A. PURVIS. 1998. Body size and species richness in carnivores and primates. *Philos. Trans. R. Soc. Lond. B* 265:113–119.
- GOODMAN, M., A. ROMERO-HERRERA, H. DENE, J. CZELUSNIAK, AND R. TASHIAN. 1982. Amino acid sequence evidence on the phylogeny of primates and other eutherians. Pages 115–191 in *Macromolecular sequences in systematic and evolutionary biology* (M. Goodman, ed.). Plenum, New York.
- GRAUR, D., AND D. HIGGINS. 1994. Molecular evidence for the inclusion of cetaceans within the order Artiodactyla. *Mol. Biol. Evol.* 11:357–364.
- HALANYCH, K. M. 1998. Lagomorphs misplaced by more characters and fewer taxa. *Syst. Biol.* 47:138–146.
- HILLIS, D. 1999. SINEs of the perfect character. *Proc. Natl. Acad. Sci. USA* 96:9979–9981.
- JANIS, C. 1988. New ideas in ungulate phylogeny and evolution. *Trends Ecol. Evol.* 3:291–297.
- JONES, K., A. PURVIS, A. MACLARNON, O. BININDA-EMONDS, AND N. SIMMONS. 2002. A phylogenetic supertree of bats (Mammalia: Chiroptera). *Biol. Rev.* 77:223–259.
- KLUGE, A. G. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Syst. Zool.* 38:7–25.
- LANGER, P. 2001. Evidence from the digestive tract on phylogenetic relationships in ungulates and whales. *J. Zool. Syst. Evol. Res.* 39:77–90.
- LILOUPIS, A., O. WALLIS, AND M. WALLIS. 1997. Cloning and characterisation of the gene encoding red deer (*Cervus elaphus*) growth hormone: Implications for the molecular evolution of growth hormone in artiodactyls. *J. Mol. Endocrinol.* 19:259–266.
- LIU, F., M. MIYAMOTO, N. FREIRE, P. ONG, M. TENNANT, T. YOUNG, AND K. GUGEL. 2001. Molecular and morphological supertrees for eutherian (placental) mammals. *Science* 291:1786–1789.
- LUCKETT, W., AND N. HONG. 1998. Phylogenetic relationships between the orders Artiodactyla and Cetacea: A combined assessment of morphological and molecular evidence. *J. Mammal. Evol.* 5:127–182.
- LYNCH, M. 1997. Mutation accumulation in nuclear, organelle, and prokaryotic transfer RNA genes. *Mol. Biol. Evol.* 14:914–925.
- MADSEN, O., M. SCALLY, C. DOUADY, D. KAO, R. DEBRY, R. ADKINS, H. AMRINE, M. STANHOPE, W. DEJONG, AND M. SPRINGER. 2001. Parallel adaptive radiations in two major clades of placental mammals. *Nature* 409:610–614.
- MATTHEE, C. A., J. D. BURZLAFF, J. F. TAYLOR, AND S. K. DAVIS. 2001. Mining the mammalian genome for artiodactyl systematics. *Syst. Biol.* 50:367–390.
- MCKENNA, M. 1975. Toward a phylogenetic classification of the Mammalia. Pages 21–46 in *Phylogeny of the primates* (W. Luckett and F. Szalay, eds.). Plenum, New York.
- MCKENNA, M. 1987. Molecular and morphological analysis of high-level mammalian interrelationships. Pages 55–93 in *Molecules and morphology in evolution: Conflict or compromise?* (C. Patterson, ed.). Cambridge Univ. Press, Cambridge, U.K.
- MESSINGER, S. L., AND J. A. MCGUIRE. 1998. Morphology, molecules, and the phylogenetics of cetaceans. *Syst. Biol.* 47:90–124.
- MILINKOVITCH, M., AND J. THEWISSEN. 1997. Even-toed fingerprints on whale ancestry. *Nature* 388:622–624.
- MIYAMOTO, M. 1985. Consensus cladograms and general classifications. *Cladistics* 1:186–189.
- MIYAMOTO, M. M., AND M. GOODMAN. 1986. Biomolecular systematics of eutherian mammals: Phylogenetic patterns and classification. *Syst. Zool.* 35:230–240.
- MONTGELARD, C., F. CATZEFELIS, AND F. DOUZERY. 1997. Phylogenetic relationships of artiodactyls and cetaceans as deduced from the comparison of cytochrome *b* and 12S rRNA mt sequences. *Mol. Biol. Evol.* 14:550–559.
- MURPHY, W., E. EIZRIK, W. JOHNSON, Y. ZHANG, O. RYDER, AND S. O'BRIEN. 2001. Molecular phylogenetics and the origins of placental mammals. *Nature* 409:614–618.
- NIKAIDO, M., F. MATSUNO, H. HAMILTON, R. BROWNELL, Y. CAO, W. DING, Z. ZUOYAN, A. SHEDLOCK, R. FORDYCE, M. HASEGAWA, AND N. OKADA. 2001. Retroposon analysis of major cetacean lineages: The monophyly of toothed whales and the paraphyly of river dolphins. *Proc. Natl. Acad. Sci. USA* 98:7384–7389.
- NIKAIDO, M., A. ROONEY, AND N. OKADA. 1999. Phylogenetic relationships among cetartiodactyls based on insertions of short and long interspersed elements: Hippopotamuses are the closest extant relatives of whales. *Proc. Natl. Acad. Sci. USA* 96:10261–10266.
- NIXON, K., AND J. CARPENTER. 1996. On simultaneous analysis. *Cladistics* 12:221–241.
- NOVACEK, M. 1992. Mammalian phylogeny: Shaking the tree. *Nature* 356:121–125.
- NOVACEK, M. 1994. Morphological and molecular inroads to phylogeny. Pages 85–131 in *Interpreting the hierarchy of nature* (L. Grande and O. Rieppel, eds.). Academic Press, New York.
- O'LEARY, M. 2001. The phylogenetic position of cetaceans: Further combined data analyses, comparisons with the stratigraphic record, and a discussion of character optimization. *Am. Zool.* 41:487–506.
- O'LEARY, M., AND J. GEISLER. 1999. The position of Cetacea within Mammalia: Phylogenetic analysis of morphological data from extinct and extant taxa. *Syst. Biol.* 48:455–490.
- PISANI, D., AND M. WILKINSON. 2002. Matrix representation with parsimony, taxonomic congruence, and total evidence. *Syst. Biol.* 51:151–155.
- PROTHERO, D. 1993. Ungulate phylogeny: Molecular vs. morphological evidence. Pages 173–181 in *Mammal phylogeny. Placentals* (F. Szalay, M. Novacek, and M. McKenna, eds.). Springer-Verlag, New York.
- PROTHERO, D., E. MANNING, AND M. FISCHER. 1988. The phylogeny of the ungulates. Pages 201–234 in *The phylogeny and classification of the tetrapods, Volume 2* (M. Benton, ed.). Oxford Univ. Press, Oxford, U.K.
- 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.
- PURVIS, A., AND L. BROMHAM. 1997. Estimating the transition/transversion ratio from independent pairwise comparisons with an assumed phylogeny. *J. Mol. Evol.* 44:112–119.
- PURVIS, A., S. NEE, AND P. HARVEY. 1995. Macroevolutionary inferences from primate phylogeny. *Proc. R. Soc. Lond. B* 260:329–333.
- RAGAN, M. 1992. Phylogenetic inference based on matrix representation of trees. *Mol. Phylogenet. Evol.* 1:53–58.
- REYNOLDS, P. 1997. Phylogenetic analysis of surface areas of mammals. *J. Mammal.* 78:859–868.
- RONQUIST, F. 1996. Matrix representation of trees, redundancy, and weighting. *Syst. Biol.* 45:247–253.
- SANDERSON, M., A. PURVIS, AND C. HENZE. 1998. Phylogenetic supertrees: Assembling the trees of life. *Trends Ecol. Evol.* 13:105–109.
- SCHAEFFER, B. 1947. Notes on the origin and function of the artiodactyl tarsus. *Am. Mus. Novit.* 1356:1–24.
- SHIMAMURA, M., H. YASUE, K. OHAHIMA, H. ABE, H. KATO, T. KISHIRO, M. GOTO, I. MUNECHEKA, AND N. OKADA. 1997. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. *Nature* 388:666–670.
- SHOSHANI, J. 1986. Mammalian phylogeny: Comparison of morphological and molecular results. *Mol. Biol. Evol.* 3:222–242.
- SHOSHANI, J. 1992. Cuvier vis-à-vis Huxley on the relationship of Hyracoidea and an update on an old controversy. Pages 103–112 in *Ongules/ungulates 91* (F. Spitz, ed.). Societe Francaise pour d'Etude Protection des Mammals, Paris.
- SHOSHANI, J. 1993. Hyracoidea–Tethytheria affinity based on myological data. Pages 235–256 in *Mammal phylogeny. Placentals* (F. Szalay, M. Novacek, and M. McKenna, eds.). Springer-Verlag, New York.
- SHOSHANI, J., AND M. MCKENNA. 1998. Higher taxonomic relationships among extant mammals based on morphology, with selected comparisons of results from molecular data. *Mol. Phylogenet. Evol.* 9:572–584.
- SPRINGER, M., AND W. DEJONG. 2001. Which mammalian supertree to bark up? *Science* 291:1709–1711.
- SWOFFORD, D. L. 1998. PAUP*: Phylogenetic analysis using parsimony (and other methods), Version 4.0b8. Sinauer, Sunderland, Massachusetts.
- TEMPLETON, A. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and the apes. *Evolution* 37:221–244.
- THEWISSEN, J. 1994. Phylogenetic aspects of cetacean origins: A morphological perspective. *J. Mammal. Evol.* 2:157–184.
- THEWISSEN, J., AND S. MADAR. 1999. Ankle morphology of the earliest cetaceans and its implications for the phylogenetic relations among ungulates. *Syst. Biol.* 48:21–30.
- THEWISSEN, J., E. WILLIAMS, L. ROE, AND S. HUSSAIN. 2001. Skeletons of terrestrial cetaceans and the relationship of whales to artiodactyls. *Science* 413:277–281.
- WIBLE, J. 1987. The eutherian stapedial artery: Character analysis and implications for superordinal relationships. *Zool. J. Linn. Soc.* 91:107–136.
- WOODROFFE, R., AND J. GINSBERG. 1998. Edge effects and the extinction of populations inside protected areas. *Science* 280:2126–2128.
- WYSS, A., M. NOVACEK, AND M. MCKENNA. 1987. Amino acid sequence versus morphological data and the interordinal relationships of mammals. *Mol. Biol. Evol.* 4:99–116.

First submitted 26 August 2001; reviews returned
22 January 2002; final acceptance 15 April 2002
Associate Editor: Jack Sullivan

Syst. Biol. 51(4):664–671, 2002
DOI: 10.1080/10635150290102357

Increased Taxon Sampling Is Advantageous for Phylogenetic Inference

DAVID D. POLLOCK,¹ DERRICK J. ZWICKL,² JIMMY A. MCGUIRE,³ AND DAVID M. HILLIS²

¹Department of Biological Sciences and Biological Computation and Visualization Center, Louisiana State University, Baton Rouge, Louisiana 70803, USA; E-mail: dpollock@lsu.edu

²Section of Integrative Biology and Center for Computational Biology and Bioinformatics, University of Texas, Austin, Texas 78712, USA; E-mail: zwickl@mail.utexas.edu (D.J.Z.), dhillis@mail.utexas.edu (D.M.H.)

³Museum of Natural Science and Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana 70803, USA; E-mail: jmcguire@lsu.edu

Until recently, it was believed that complex phylogenies might be extremely difficult to reconstruct due to the phenomenal rate of increase in the number of possible phylogenies as the number of taxa increases. However, Hillis (1996) showed through simulation that, for at least one complex phylogeny

of angiosperms with 228 taxa, reconstruction was far more accurate than expected, even with relatively modest amounts of DNA sequence data. This led to a flurry of papers on the subject of taxon sampling and phylogenetic reconstruction, with focus quickly shifting from the question of whether complex