Letter to the Editor

Support for a Monophyletic Lemuriformes: Overcoming Incongruence Between Data Partitions

K. Stanger-Hall and C. W. Cunningham†*

*Department of Zoology, University of Texas at Austin; and †Zoology Department, Duke University

The Malagasy lemurs (Lemuriformes), represented today by 32 extant species (Mittermeier et al. 1994), are thought to have diverged approximately 60 MYA from their last common ancestor with the Lorisiformes in Africa and Asia (Yoder et al. 1996). Despite a great deal of morphological and molecular data, the family-level relationships—and even the monophyly—of the Lemuriformes remain controversial (e.g., Tattersall and Schwartz 1974; Cartmill 1975; Szalay 1975; Dene et al. 1976; Sarich and Cronin 1976; Dene, Goodman, and Prychodko 1980; Schwartz and Tattersall 1985; MacPhee and Cartmill 1986; Adkins and Honeycutt 1994; Yoder 1994; Del Pero et al. 1995; Porter et al. 1995; Yoder et al. 1996; Yoder, Vilgalys, and Ruvolo 1996; Stanger-Hall 1997). In this paper, we focus on resolving a remarkable pattern of incongruence between codon positions found in two mitochondrial DNA genes, cytochrome *b* (Cyt *b*) (Yoder, Vilgalys, and Ruvolo 1996) and cytochrome *c* oxidase subunit II (COII) (Adkins and Honeycutt 1994).

To test the hypothesis that first and second codon positions of these genes are misleading in their support for a diphyletic Lemuriformes (Yoder, Vilgalys, and Ruvolo 1996), we collected partial sequences of the mitochondrial large ribosomal subunit (16S) from 17 species and subspecies of the Lemuriformes. Our sequences represented the 3′ portions of the gene (33%), which are known to be far more conserved than the 5′ portions (Hillis and Dixon 1991). This study represents the largest taxon sampling to date of the Lemuriformes for any gene. We chose the 16S gene, because, as a mitochondrial gene, it is guaranteed to have had the same history as the two protein-coding genes, but, as a ribosomal gene, it presumably operates under different selective constraints. As many taxa as possible were analyzed in an effort to resolve the long deep branches found in previous analyses with fewer taxa. After analyzing the 16S gene by itself, we used an iterative procedure to investigate whether the incongruence between data partitions is aggravated by inadequate phylogenetic reconstruction methods (Cunningham 1997).

Three different phylogenetic reconstruction models were applied to the data using PAUP* 4.0d63 (written by D. L. Swofford, Smithsonian Institution): the best-fit model under maximum likelihood (ML), equally weighted parsimony, and six-parameter parsimony. The six-parameter parsimony method gives a different weight to each of the six substitution classes (AC, AG, AT, CG, CT, GT), as described by Williams and Fitch (1990). The weights were applied using a posteriori stepmatrices under a generalized parsimony framework (Williams and Fitch 1990; Swofford et al. 1996). These stepmatrices were based on estimates of the substitution frequencies of the six nucleotide substitution classes. These estimates were obtained by ML (using PAUP* 4.0d63), because the general time-reversible ML estimate not only considers unequal base frequencies, but also takes multiple substitutions into account.

To obtain the substitution frequencies, we first found the most parsimonious tree (using equally weighted, unordered parsimony) and then used this tree to estimate a matrix of substitution frequencies (''R'' matrix) under the general time-reversible ML model (Lanave et al. 1984; Tavaré 1986). A Microsoft Excel spreadsheet is available from the authors that takes the six values from the R matrix, converts them to proportions, and takes their natural log. Of the many ways of converting a matrix of substitution frequencies to rates (Wheeler 1990; Rodrigo 1992; Collins, Kraus, and Estabrook 1994; Cunningham 1997), taking the natural log of the substitution frequency has the strongest theoretical justification (Felsenstein 1981; Albert and Mishler 1992), produces stepmatrices that usually obey the triangle inequality, and has performed well when applied to sequence data from strongly corroborated phylogenies (Cunningham 1997). Before model fitting, the sequences were tested to confirm that there was no significant difference in base composition among taxa ($\chi^2 = 11.322075$, df = 30, $P > 0.99$). The best-fit models for the ML analyses were chosen for each data partition separately by adding parameters in a stepwise procedure (Goldman 1993; Yang 1994 [as described by Cunningham, Zhu, and Hillis 1998]). Parameters were only retained if their addition significantly improved the fit between the model and the data.

A phylogenetic analysis of the complete taxon set for the 16S data (fig. 1) contradicts the diphyletic Lemuriformes supported by the combined first/second codon positions of the Cyt *b* and COII genes (fig. 2*A*; see also Yoder, Vilgalys, and Ruvolo 1996). Both equally weighted parsimony and six-parameter parsimony support a monophyletic Lemuriformes, albeit with fairly low bootstrap support (67% and 68% respectively). The best-fit ML model also supports a monophyletic Lemuriformes, but with much weaker bootstrap support (39%).

All of the genera from which we sampled multiple taxa are monophyletic, as are family groups such as the Lemuridae and the Cheirogaleidae (fig. 1). This is consistent with earlier studies that included fewer taxa. As

Key words: Lemuriformes, Lorisiformes, six-parameter parsimony, maximum likelihood, best-fit models, incongruence.

Address for correspondence and reprints: Cliff Cunningham, Zoology Department, Duke University, Durham, North Carolina 27708. E-mail: cliff@acpub.duke.edu.

Mol. Biol. Evol. 15(11):1572–1577. 1998

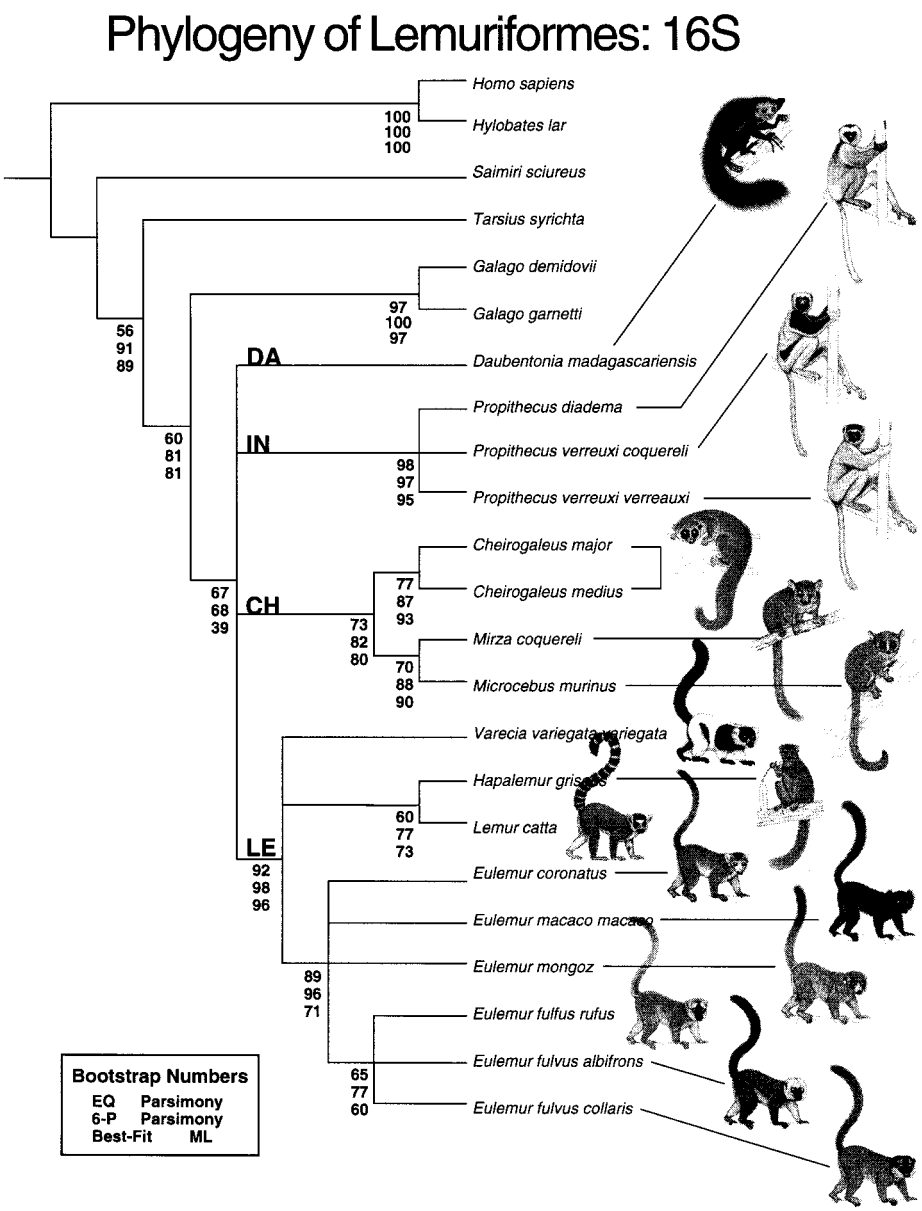

FIG. 1.—Phylogenetic analysis of the 16S data for all taxa studied. The topology shown is the 50% majority-rule consensus tree supported by equally weighted, unordered parsimony, with bootstrap support for 500 pseudoreplicates for this method given as the first of three numbers (Felsenstein 1985). The next two numbers represent bootstrap support for six-parameter stepmatrix parsimony and the best-fit ML model, respectively. Computational intensity of the ML searches reduced the number of bootstraps to 100. All three methods were applied using heuristic searches with tree-bisection-reconnection branch swapping with PAUP* 4.0d63 and with the maximum number of trees saved fixed at 5. The best-fit model for the 16S data was the general time-reversible model (Lanave et al. 1984; Tavaré 1986) assuming equal base frequencies and among-site rate variation by simultaneously estimating the invariable sites and discrete gamma distribution methods. Lemuriform taxa from four family groups were analyzed: Daubentoniidae (DA), Indridae (IN), Cheirogaleidae (CH), and Lemuridae (LE). In addition, several anthropoids (*Homo, Hylobates, Saimiri, Tarsius*) and Lorisiformes (*Galago*) were included. Partial 16S DNA sequences were collected by direct sequencing of polymerase chain reaction products amplified using primers 16Sar and 16Sbr as described in Cunningham, Blackstone, and Buss (1992), except that reactions were carried out using Buffer A from Invitrogen (300 mM Tris HCl; 75 mM $(NH_4)_2SO_4$; 7.5 mM $MgCl_2$, pH 8.5). Both strands of each sequence were obtained by cycle sequencing with ABI prism kits as per the manufacturer's instructions, read on an ABI 373 automated sequencer, and edited using Sequencher 3.0 (Genecodes Corp.). Sequences were aligned using CLUSTAL W, and regions of ambiguous sequence alignment were eliminated using a variant of the protocol proposed by Gatesy, DeSalle, and Wheeler (1993) as described by Cunningham (1997). The final sequence alignment consisted of 461 bp. The sequences have been deposited in GenBank under accession numbers AF072412–AF072430, and the alignment has been deposited at EMBL (ftp://ftp.ebi.ac.uk/pub/databases/emb/align/d535460). For other outgroup sequences please see Horovitz and Meyer (1995) and references therein. The drawings were created by Stephen Nash (and adapted with permission from Mittermeier et al. 1994).

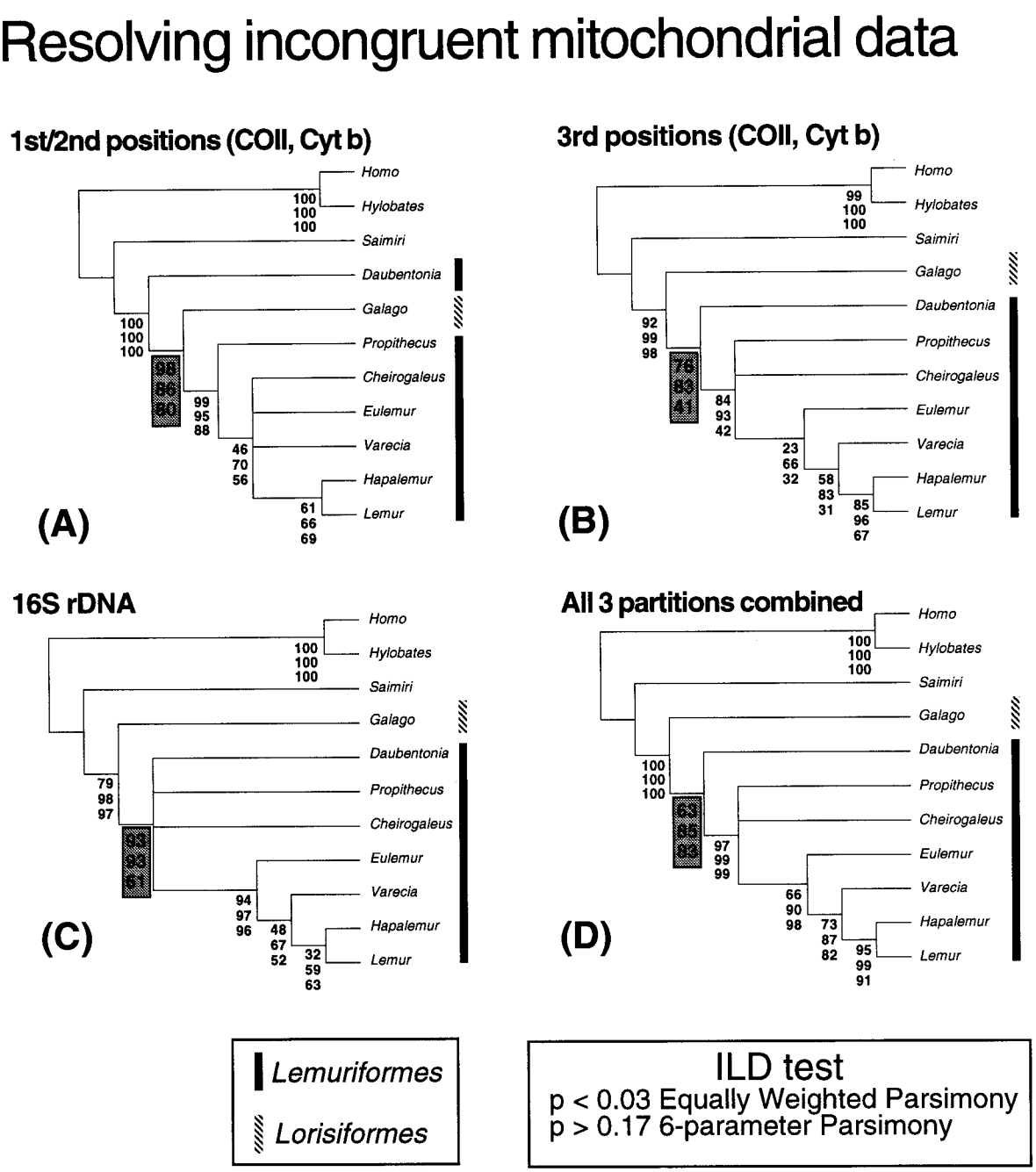

FIG. 2.—Phylogenetic analysis of the subset of the taxa in figure 1 for which data were available for the Cyt *b* and COII genes. The arrangement of bootstrap values corresponds to figure 1. The numbers refer to 1,000 bootstraps for both parsimony methods and 100 bootstraps for the ML method. Bootstrap values in shaded boxes refer to the node relevant to the monophyly or diphyly of the Lemuriformes. Analyses are presented for first and second positions of the combined COII and Cyt *b* genes (*A*), for third positions of those genes (*B*), and for 16S (*C*), and a combined analysis is presented (*D*). The best-fit model for the first and second positions is the general time-reversible model (Lanave et al. 1984; Tavaré 1986) with rate heterogeneity accommodated by simultaneously estimating both the proportion of invariant sites and a five-class discrete gamma distribution model (Yang 1994). For third positions, the best-fit model was a Hasegawa, Kishino, and Yano (1985) model with estimation of the proportion of the invariant sites (for 16S, see fig. 1).

in previous studies, the phylogenetic position of the Indridae and the Cheirogaleidae relative to the other Lemuriformes was inconclusive (Adkins and Honeycutt 1994; Yoder, Vilgalys, and Ruvolo 1996). The position of *Daubentonia* (the aye-aye) is unresolved by the 16S data (but see combined analysis below).

If the 16S data are correct in supporting a monophyletic Lemuriformes, this raises the question of why first/second codon positions of Cyt *b* and COII are incongruent with other independent data sets that also support a monophyletic Lemuriformes. These data sets include third-position transversions from the Cyt *b* and COII genes (Yoder, Vilgalys, and Ruvolo 1996), epsilon-globin gene sequences (Porter et al. 1995), and immunodiffusion (Dene et al. 1976; Dene, Goodman, and Prychodko 1980), DNA hybridization (Bonner, Heine-

mann, and Todaro 1980), karyological (e.g., Rumpler et al. 1988), and biogeographical data (Yoder et al. 1996). Although Yoder, Vilgalys, and Ruvolo (1996) suggest that selective constraints on amino acids may be responsible for the presumably misleading signal in the first and second codon positions, incongruence between data partitions can also arise when inadequate reconstruction methods cause systematic errors in one or more data partitions.

Between-partition incongruence can be explored using an iterative approach to phylogeny reconstruction (Bull et al. 1993; Cunningham 1997). First, the degree of incongruence is tested under a simple reconstruction model, such as equally weighted, unordered parsimony. Then, the incongruence test is reapplied under a more realistic model, such as six-parameter parsimony. If this incongruence is partly caused by the use of an overly simple model, a more appropriate model should cause the various data partitions to converge on support for the same—and presumably correct—tree. This increased support for a common solution should be accompanied by a decreased level of incongruence.

Under equally weighted parsimony, the degree of incongruence between the codon positions of the two protein-coding genes is striking. The first and second positions support a diphyletic Lemuriformes, with 98% bootstrap support (fig. 2*A*), whereas the third positions support a monophyletic Lemuriformes with 76% support (fig. 2*B*). When the 16S data for the corresponding taxa are included as a third partition (supporting a monophyletic Lemuriformes with 93% bootstrap support), the degree of incongruence between the three partitions is significant as detected by the incongruence length difference test ($P < 0.03$; Farris et al. 1994).

When we applied six-parameter parsimony to a combined analysis of the three data partitions (with each partition being assigned its own stepmatrix), six-parameter parsimony dramatically increased support for a monophyletic Lemuriformes relative to equally weighted parsimony (from 63% to 85%, fig. 2*D*). For the presumably misleading first and second positions, bootstrap support for a diphyletic Lemuriformes dropped from 98% to 85% (fig. 2*A*). For third codon positions, the support for a monophyletic Lemuriformes increased from 76% to 83% (fig. 2*B*), and for the 16S data, the support remained at 93% (fig. 2*C*). Not only does six-parameter parsimony increase support for a monophyletic Lemuriformes among all three partitions (albeit rather weakly for first and second positions), but it also causes the degree of incongruence as detected by the incongruence length difference test to drop below significance ($P > 0.17$).

Interestingly, the performance of the most complex model tested in our analysis (i.e., the best-fit model under ML) appears to be somewhat erratic. For the 16S data set with all the taxa included (fig. 1), the best-fit model only very weakly supported a monophyletic Lemuriformes (39%), although support for a diphyletic Lemuriformes was even weaker (7%, not shown in the figure). In contrast, when a much simpler ML model was applied to the 16S data set (one substitutional class with no rate variation; Jukes and Cantor 1969), the support for a monophyletic Lemuriformes increased to 61% (results not shown). This appears to be a case in which a simpler model performs better than the more complex best-fit model (Yang 1997). For the 16S data set, this is most likely because too many parameters are being estimated from a fairly short sequence (461 bp).

When best-fit models were applied to the individual data partitions (fig. 2*A–C*) the results of best-fit ML models were again mixed. When best-fit models were applied to first/second positions, support for the diphyletic Lemuriformes dropped to 80% (fig. 2*A*). This is consistent with the hypothesis that at least part of the misleading signal in the first and second positions can be overcome by applying more appropriate models of DNA sequence evolution. However, the support of best-fit models for a monophyletic Lemuriformes is much lower than that of parsimony for both the third positions (fig. 2*B*; 41%, compared with 80% for six-parameter parsimony) and the comparable taxon set of the 16S data (fig. 2*C*; 61%, compared with 94% for six-parameter parsimony).

Although best-fit models showed decidedly mixed results when applied to the individual data partitions, their support for a monophyletic Lemuriformes in the combined data partition was strong, and compared favorably with that of six-parameter parsimony (fig. 2*D*). This is consistent with the hypothesis that under ML, complex best-fit models may be most appropriate in large data sets, even when the data are unable to reject parameter-rich models. For smaller data sets, simpler models should always be run for comparison.

In summary, the 16S data support a monophyletic Lemuriformes. This is further supported by the combined analysis of the 16S, Cyt *b,* and COII sequence data. These results suggest that the third positions of the mitochondrial protein-coding genes in lemurs may indeed contain a more accurate phylogenetic signal than the structurally more constrained first and second positions, resulting in incongruent data partitions (Yoder, Vilgalys, and Ruvolo 1996). However, our analysis shows that the apparent incongruence between data partitions is at least partly due to inadequate phylogenetic reconstruction methods.

Applying multiparameter models using parsimony and maximum likelihood to the combined data partitions increased support for a monophyletic Lemuriformes (as well as support for other relationships, such as the basal position of *Daubentonia*; see fig 2*D*). These results confirm that in some cases, applying more realistic models to phylogenetic reconstruction can reduce incongruence between data partitions and increase support in combined analyses (Cunningham 1997). Whereas the six-parameter parsimony model seems to outperform best-fit ML models for smaller data partitions, the best-fit ML models seem to perform well in the combined data set.

Acknowledgments

We are grateful to Dr. K. Glander (Director) for the opportunity to use the tissue sample collection of the

Duke University Primate Center (DUPC) for this study. We are also grateful to D. L. Swofford for permission to use a test version of PAUP*. This study was supported by an NSF grant to the DUPC and by NSF grant DEB-9615461 to C.W.C.. This is DUPC publication 662.

LITERATURE CITED

ADKINS, R. M., and R. L. HONEYCUTT. 1994. Evolution of primate cytochrome c oxidase subunit II gene. J. Mol. Evol. **38**:215–231.

ALBERT, V. A., and B. D. MISHLER. 1992. On the rationale and utility of weighting nucleotide sequence data. Cladistics **8**: 73–83.

BONNER, T. I., R. HEINEMANN, and G. J. TODARO. 1980. Evolution of DNA sequences has been retarded in Malagasy primates. Nature **286**:420–423.

BULL, J. J., J. P. HUELSENBECK, C. W. CUNNINGHAM, D. L. SWOFFORD, and P. J. WADELL. 1993. Partitioning and combining data in phylogenetic analysis. Syst. Biol. **42**:384–397.

CARTMILL, M. 1975. Strepsirhine basicranial structures and the affinities of the Cheirogaleidae. Pp. 313–354 *in* W. P. LUCKETT and F. S. SZALAY, eds. Phylogeny of the Primates. Plenum Press, New York.

COLLINS, T., F. KRAUS, and G. ESTABROOK. 1994. Compositional effects and weighting of nucleotide sequences for phylogenetic analysis. Syst. Biol. **43**:449–459.

CUNNINGHAM, C. W. 1997. Is congruence between data partitions a reliable predictor of phylogenetic accuracy? Empirically testing an iterative procedure for choosing among phylogenetic methods. Syst. Biol. **46**:464–478.

CUNNINGHAM, C. W., N. W. BLACKSTONE, and L. W. BUSS. 1992. Evolution of king crabs from hermit crab ancestors. Nature **355**:539–542.

CUNNINGHAM, C. W., H. ZHU, and D. M. HILLIS. 1998. Best-fit maximum likelihood models for phylogenetic inference: empirical tests with known phylogenies. Evolution **52**:978–987.

DEL PERO, M., S. CROVELLA, P. CERVELLA, G. ARDITO, and Y. RUMPLER. 1995. Phylogenetic relationships among Malagasy lemurs as revealed by mitochondrial DNA sequence analysis. Primates **36**:431–440.

DENE, H., M. GOODMAN, and W. PRYCHODKO. 1980. Immunodiffusion systematics of the primates. IV. Lemuriformes. Mammalia **44**:211–223.

DENE, H., M. GOODMAN, W. PRYCHODKO, and G. W. MOORE. 1976. Immunodiffusion systematics of the primates. Folia Primatol. **25**:35–61.

FARRIS, J. S., M. KÄLLERSJÖ, A. G. KLUGE, and C. BULT. 1994. Testing significance of incongruence. Cladistics **10**:315–319.

FELSENSTEIN, J. 1981. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. Biol. J. Linn. Soc. **16**:183–196.

———. 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39**:783–791.

GATESY, J., R. DESALLE, and W. C. WHEELER. 1993. Alignment-ambiguous nucleotide sites and the exclusion of systematic data. Mol. Phylogenet. Evol. **2**:152–157.

GOLDMAN, N. 1993. Statistical tests of models of DNA substitution. J. Mol. Evol. **36**:182–198.

HASEGAWA, M., H. KISHINO, and T. A. YANO. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22**:160–174.

HILLIS, D. M., and M. T. DIXON. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. Q. Rev. Biol. **66**:411–453.

HOROVITZ, I., and A. MEYER. 1995. Systematics of New World monkeys (Platyrrhini, Primates) based on 16S mitochondrial DNA sequences: a comparative analysis of different weighting methods in cladistic analysis. Mol. Phylogenet. Evol. **4**:448–456.

JUKES, T. H., and C. H. CANTOR. 1969. Evolution of protein molecules. Pp. 21–123 *in* H. M. MUNRO, ed. Mammalian protein metabolism. Academic Press, New York.

LANAVE, C., G. PREPARATA, C. SACCONE, and G. SERIO. 1984. A new method for calculating evolutionary substitution rates. J. Mol. Evol. **20**:86–93.

MACPHEE, R. D. E., and M. CARTMILL. 1986. Basicranial structures and primate systematics. Pp. 219–275 *in* D. R. SWINDLER and J. ERWIN, eds. Comparative primate biology. Vol. 1. Systematics, evolution and anatomy. Alan R. Liss, New York.

MITTERMEIER, R. A., I. TATTERSALL, W. R. KONSTANT, D. M. MEYERS, and R. B. MAST. 1994. Lemurs of Madagascar. Conservation International Tropical Field Guide Series. Conservation International, Washington, D.C.

PORTER, C. A., I. SAMPAIO, H. SCHNEIDER, M. P. C. SCHNEIDER, J. CZELUSNIAK, and M. GOODMAN. 1995. Evidence on primate phylogeny from epsilon-globin gene sequences and flanking regions. J. Mol. Evol. **40**:30–55.

RODRIGO, A. G. 1992. A modification to Wheeler's combinatorial weights calculations. Cladistics **8**:165–170.

RUMPLER, Y., S. WARTER, J. J. PETTER, R. ALBIGNAC, and B. DUTRILLAUX. 1988. Chromosomal evolution of Malagasy lemurs. XI. Phylogenetic position of *Daubentonia madagascariensis.* Folia Primatol. **50**:124–129.

SARICH, V. M., and J. E. CRONIN. 1976. Molecular systematics of the primates. Pp. 139–168 *in* M. GOODMAN, R. E. TASHIAN, and J. H. TASHIAN, eds. Molecular anthropology. Academic Press, New York.

SCHWARTZ, J. H., and I. TATTERSALL. 1985. Evolutionary relationships among living lemurs and lorises (Mammalia, Primates) and their potential affinities with European Eocene Adapidae. Anthropol. Pap. Am. Mus. Nat. Hist. **60**:1–100.

STANGER-HALL, K. F. 1997. Phylogenetic affinities among the extant Malagasy lemurs (Lemuriformes) based on morphology and behavior. J. Mamm. Evol. **4**:163–194.

SWOFFORD, D. L., G. J. OLSEN, P. J. WADDELL, and D. M. HILLIS. 1996. Phylogenetic inference. Pp. 407–514 *in* D. M. HILLIS, C. MORITZ, and B. K. MABLE, eds. Molecular systematics. 2nd edition. Sinauer, Sunderland, Mass.

SZALAY, F. S. 1975. Phylogeny of primate higher taxa: the basicranial evidence. Pp. 91–125 *in* W. P. LUCKETT and F. S. SZALAY, eds. Phylogeny of Primates. Plenum Press, New York.

TATTERSALL, I., and J. H. SCHWARTZ. 1974. Craniodental morphology and the systematics of the Malagasy lemurs (Primates, Prosimii). Anthropol. Pap. Am. Mus. Nat. Hist. **52**: 138–192.

TAVARÉ, S. 1986. Some probabilistic and statistical problems on the analysis of DNA sequences. Lect. Math. Life Sci. **17**:57–86.

WHEELER, W. C. 1990. Combinatorial weights in phylogenetic analysis: a statistical parsimony procedure. Cladistics **6**: 269–275.

WILLIAMS, P. L., and W. M. FITCH. 1990. Phylogeny determination using the dynamically weighted parsimony method. Methods Enzymol. **183**:615–626.

YANG, Z. 1994. Estimating the pattern of nucleotide substitution. J. Mol. Evol. **39**:105–111.

———. 1997. How often do wrong models produce better phylogenies? Mol. Biol. Evol. **14**:105–108.

YODER, A. D. 1994. Relative position of the Cheirogaleidae in strepsirhine phylogeny: a comparison of morphological and molecular methods and results. Am. J. Phys. Anthropol. **94**: 25–46.

YODER, A. D., M. CARTMILL, M. RUVOLO, K. SMITH, and R. VILGALYS. 1996. Ancient single origin for Malagasy primates. Proc. Natl. Acad. Sci. USA **93**:5122–5126.

YODER, A. D., R. VILGALYS, and M. RUVOLO. 1996. Molecular evolutionary dynamics of cytochrome beta in strepsirrhine primates: the phylogenetic significance of third-position transversions. Mol. Biol. Evol. **13**:1339–1350.

MANOLO GOUY, reviewing editor

Accepted July 28, 1998

NOTICE!

The *Molecular Biology and Evolution* Editorial Office has moved from the University of Rochester to the Australian National University, Canberra, Australia.

The new Editor is Dr. Simon Easteal.

Effective immediately all Editorial correspondence should be addressed to:

MBE Editorial Office
The John Curtin School of Medical Research
The Australian National University
Canberra, ACT 0200
AUSTRALIA
Telephone: 61 2 6249 2435
E-mail: mbe@anu.edu.au

Contact information for the Editor:
Dr. Simon Easteal
(tel) 61 2 6249 4719
(fax) 61 2 6249 4712
Email Simon.Easteal@anu.edu.au