The Colugo (*Cynocephalus variegatus*, Dermoptera): The Primates' Gliding Sister?

Jürgen Schmitz, Martina Ohme,* Bambang Suryobroto,† and Hans Zischler**

*Primate Genetics, German Primate Center, Göttingen; and †Department of Biology, Faculty of Mathematics and Natural Sciences, Bogor Agricultural University, Bogor, Indonesia

Although a general agreement on the major groups of eutherian orders and their phylogenetic affiliations is emerging, the evolutionary affiliations among the members constituting these groups are still subject to debate. A prominent example is the recently published molecular evidence that challenges the long assumed monophyly of primates, displaying the colugo or flying lemur (*Cynocephalus*, Dermoptera) as a sister to anthropoid primates (Arnason et al. 2002) and positioning them after the prosimian primates (tarsiers and strepsirhines) split off.

The phylogenetic analysis of the complete mitochondrial (mt) genome sequence of *Cynocephalus variegatus* presented in this study first appears to corroborate interpretations of primates as a paraphyletic group. However, more detailed analyses disclosed that mt nucleotide composition and consequently amino acid (AA) composition varied considerably among the species analyzed. This led us to assume that the flying lemur may be incorrectly grouped with anthropoids on the basis of similar mt nucleotide and AA compositions, rather than reflecting the true evolutionary relationship.

To reanalyze the flying lemur's evolutionary association with other eutherian orders from a completely different molecular perspective, a molecular cladistic approach was applied. To this end, we determined the presence/absence pattern of transposable elements that provide a nearly homoplasy-free and copious source of molecular evolutionary markers, with well-defined character polarity. We could identify transposable elements, both on a multilocus and single-locus level, being present in all extant primate infraorders but absent in the flying lemur, thus clearly supporting the monophyly of primates by retropositional evidence.

Introduction

Early primate evolution between the late Cretaceous and the end of the Eocene continues to be an issue of lively debate. Of particular interest are the interrelationships between extant and fossil primates and other eutherians, where questions of phylogeny still largely remain unresolved.

Using molecular evidence, several authors have recently split the eutherian orders into four major groups, displaying primates in a cluster with rodents, lagomorphs, tree shrews, and flying lemurs (Madsen et al. 2001; Murphy et al. 2001*a*, 2001*b*). Although the positioning of the major groups is largely confirmed, both by nuclear and mitochondrial (mt) DNA data sets, the phylogenetic affiliations among the members of the major groups remain, in part, contradictory. This is illustrated by the varying affiliations of colugos to other eutherian taxa, as described in Murphy et al. (2001*a*, 2001*b*). Although it is difficult to compare distinct topologies because of the different taxonomic sampling, it is clear that primate affiliations to other eutherians have not been resolved to a degree that would permit the launch of large-scale comparative projects, e.g., on molecular character evolution.

Another difficulty in establishing a firm phylogenetic framework for linking primates to other eutherians is represented by the scanty early primate fossil record, so that the discovery of new fossils often requires major revisions of phylogenetic hypotheses (Beard 1990, 1993; Kay, Thorington, and Houde 1990; Martin 1993).

The concept of the superorder Archonta, first proposed by Gregory (1910) and revived in the last decades (Novacek 1992), indicates a close relationship between primates and dermopterans. This superorder comprises primates, colugos, bats, and tree shrews. Although the monophyly of the Archonta is partially refuted both by molecular (Schmitz, Ohme, and Zischler 2000; Teeling et al. 2000; Lin and Penny 2001) and by palaeontological (Kay, Thorington, and Houde 1990) evidence, it is commonly agreed that the primate origin lies somewhere among archontan representatives (for an overview see Fleagle and MacPhee 1993, pp. 1–383).

To complement early primate fossil records with molecular evidence, we analyzed the complete mtDNA of *Cynocephalus variegatus*. Results were combined with retropositional data for testing different phylogenetic interpretations in primates and other eutherians (for GenBank accession numbers see *Supplementary Material*).

Materials and Methods

The DNA sequence of the entire mt genome of *C. variegatus* was obtained applying long-range PCRs covering the entire mt genome (see *Supplementary Material*, www.molbiolevol.org). Products from nested PCRs carried out on the long-range PCR products were cloned and sequenced (sequence information of the nested primers is available from the corresponding author on request). The subsequent phylogenetic reconstructions included all protein coding genes located on the mt H-strand from representatives of all primate infraorders and archontan orders as well as additional nonprimate

Key words: *Cynocephalus variegatus*, colugo, mitochondrial genome, mammalian phylogeny.

Address for correspondence and reprints: Jürgen Schmitz, Primate Genetics, German Primate Center, Kellnerweg 4, 37077 Göttingen, Germany. E-mail: jschmitz@dpz.gwdg.de.

Mol. Biol. Evol. 19(12):2308–2312. 2002

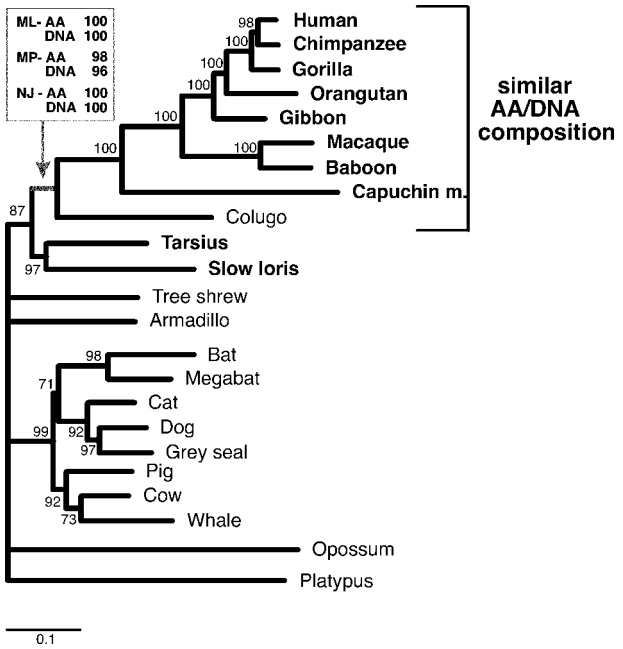

FIG. 1.—ML tree reconstruction of the mammalian relationship performing eight categories of rate heterogeneity and based on the 12 H-strand AA. ML, MP, and distance-based (NJ) bootstrap or quartet puzzling support values corresponding to the artificial sister group relationship of higher primates and colugos indicate a similar AA and nucleotide composition. We excluded rodents and lagomorphs from the analyses because of their unstable position in mt phylogenies (see also Graur, Hide, and Li 1991 and Waddell et al. 1999). Branch lengths represent AA substitutions per site.

mammals (for gene order see *Supplementary Material*, www.molbiolevol.org).

Sequence alignments were carried out by CLUSTAL X (Thompson et al. 1997). Phylogenetic analyses of both DNA and amino acids (AA) were performed using maximum parsimony (MP, heuristic search), maximum likelihood (ML, with eight categories of the discrete, gamma-distribution model of rate heterogeneity over sites, mtREV model of sequence evolution), and distance-based (neighbor-joining, LogDet distance measure, either including or excluding constant sites, or Dayhoff PAM matrix) methods implemented in PAUP* 4.0b8 (Swofford 2000), PHYLIP 3.6 (Felsenstein 1995), and TREE-PUZZLE 5.0 (Strimmer and von Haeseler 1996). All bootstrap and puzzling supports are based on 1,000 replicates.

Results and Discussion

Surprisingly, in phylogenetic reconstructions the colugo clusters inside the primate tree, appearing as a sister taxon to higher primates. This result is obtained independently of the method of reconstruction and with convincing bootstrap or puzzling support values (fig. 1). Recently, Arnason et al. (2002) arrived at a similar conclusion after comparing 60 mammalian mt genomes in a phylogenetic reconstruction.

Based on the AA reconstruction, the log L value of the tree shown in figure 1 is significantly better than the log L value of a user-defined tree with the colugo set as the primates sister group ($\log L = -55159.43$ vs. $\log L = -55183.14$, 23.71 ± 9.65 SE, Kishino and Hasegawa 1989).

Because this result is at odds with previous hypotheses, we used an entirely different class of phylogenetic markers, performing single-locus and multilocus analyses of primate-specific short interspersed nuclear element (SINE) transpositions (for review see Shedlock and Okada 2000). This approach makes use of a peculiar feature of the primate genome, which is the high abundance of dimerized, seven SL-RNA–derived SINEs or Alu-sequences. Because of a wavelike spreading of Alu-SINEs during primate evolution, Alu-SINEs cover more than 10% of the entire human genome (Li et al. 2001) and apparently also considerable parts of the genome of other primates (Hwu et al. 1986). Given the irreversible integration of individual Alu-SINEs at mainly random genomic target sites, they are ideally suited as molecular cladistic markers and landmarks of primate evolution (Schmitz, Ohme, and Zischler 2001; Singer et al. 2003).

In Southern blot experiments using primate Alu elements as a probe, we challenged *Hinf*I-restricted total DNA of humans and lemurs—the two primates representing the descendants of the deepest split in the primate tree—along with tree shrew, colugo, bat, rabbit, hedgehog, and mouse (fig. 2*A*). Hybridization conditions were chosen to detect hybrids with at least 70% sequence similarity between the genomic DNAs and an Alu SINE probe. Only the two primate representatives revealed a signal indicating high copy numbers of the Alu-element. No signal was detected in the colugo, as would be expected with colugos as a sister group of the Anthropoidea, as indicated by the mtDNA-based phylogenetic reconstructions.

This multilocus evidence was further corroborated by a single-locus approach, where the presence/absence pattern of an individual medium reiterated repeat (MER), a transposable element different from Alu-repeats, was also tested. By screening the GenBank database for primate-specific SINEs, we identified the respective MER localized on the human chromosome 11q23 between the zinc finger protein exons 9 and 10 (all-1). Using conserved exon-based primers we were able to PCR amplify the orthologous DNA fragments in higher primates and strepsirhines, comprising the MER element and flanking regions throughout (larger fragments in fig. 2*B* indicate presence). On the other hand, all nonprimate representatives including the colugo harbored the unoccupied retropositional target site and the flanking regions at the orthologous locus only, thus firmly precluding the presence of the MER element (smaller fragments in fig. 2*B* indicate absence). All presence/absence patterns were verified by sequence analyses of PCR products (for GenBank accession numbers see *Supplementary Material*).

Finally, we analyzed three haplorhine-specific Alu SINE markers published earlier (Schmitz, Ohme, and Zischler 2001), which unequivocally cluster together tarsiers and anthropoids as sister groups. All were absent in *C. variegatus* (as an example see fig. 2*C*). *Cynocephalus variegatus* thus displays the ancestral character

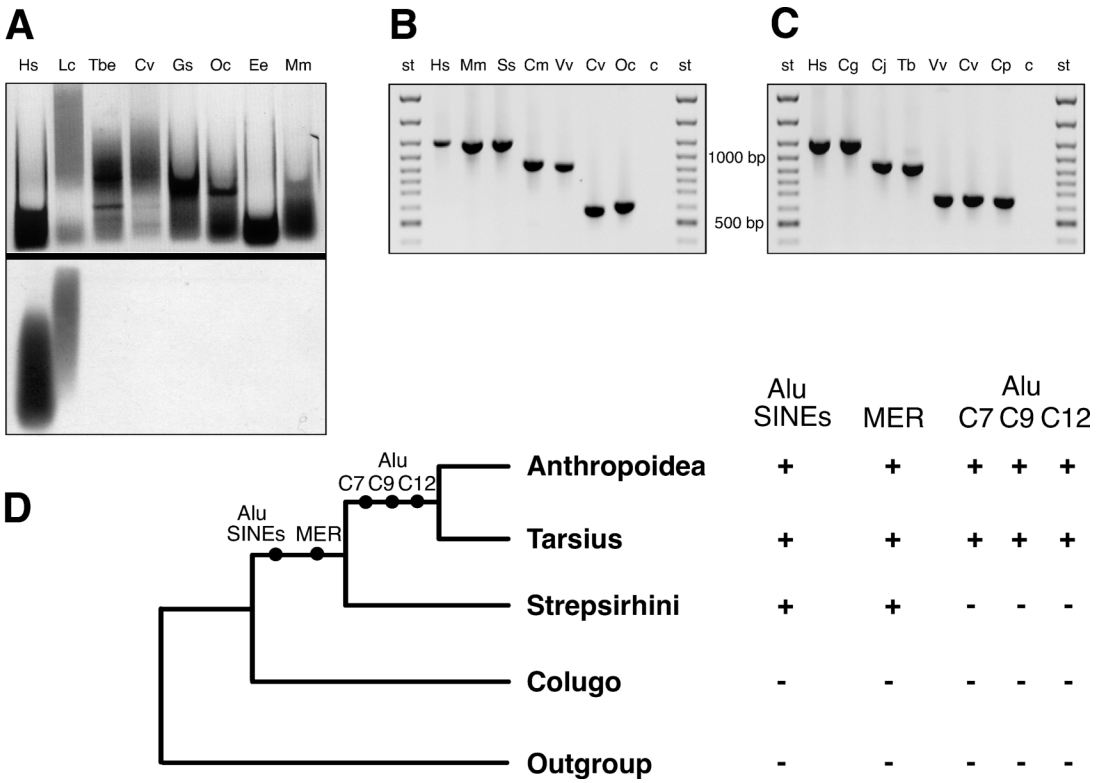

FIG. 2.—Presence/absence analyses of retrotransposable SINEs. Larger fragments indicate the presence of SINE elements and their flanking regions, whereas smaller fragments represent the unoccupied flanking sequences only. (*A*) Multilocus presence/absence analysis of primate-specific Alu SINEs. *Hinf*I-restricted total genomic DNA (upper half) and hybridization signals of an Alu SINE-probe (lower half) indicate the presence of Alu SINEs in primates only. (*B*) Single-locus analysis corresponding to a primate-specific MER present in all primates and absent in nonprimate mammals including colugo. (*C*) Single-locus analysis of an Alu SINE integration (C9 locus) in *Tarsius* and higher primates (Haplorhini), with absence in lemurs, colugo, and guinea-pig. In (*B*) and (*C*) additional Alu SINE integrations took place in the lineage leading to anthropoids and catarrhines, respectively. (*D*) Presence (+) or absence (−) scheme and the evolutionary origin of the multilocus (Alu SINEs) and single-locus (MER, Alu C7, C9, C12) markers. C7, C9, and C12 correspond to three independent chromosomal loci published in Schmitz, Ohme, and Zischler (2001). Abbreviations are: *Homo sapiens* (Hs), *Macaca mulatta* (Mm), *Saimiri sciureus* (Ss), *Colobus guereza* (Cg), *Callithrix jacchus* (Cj), *Tarsius bancanus* (Tb), *Lemur catta* (Lc), *Cheirogaleus medius* (Cm), *Varecia variegata* (Vv), *Tupaia belangeri* (Tbe), *Cynocephalus variegatus* (Cv), *Glossophaga soricina* (Gs), *Oryctolagus cuniculus* (Oc), *Cavia porcellus* (Cp), *Erinaceus europaeus* (Ec), *Mus musculus* (Mm), 100-bp ladder (st), PCR control reaction without DNA (c).

state at all analyzed orthologous loci that became targets for retroposon integration at various times during primate evolution.

The independent retropositional evidence clearly demonstrates that the apparently significant affiliation of colugos to higher primates obtained in mtDNA-based phylogenetic reconstructions is due to a misleading mtDNA signal. We consider this effect sufficiently strong to also suggest a close relationship between colugos and higher primates in data sets composed of both nuclear and mt sequences (see fig. 1 of Murphy et al. 2001*a*).

We recently presented evidence that mt-specific, directional mutational pressure might be responsible for the artifact of tarsiers and strepsirhines appearing as sister groups on account of high bootstrap values. Depending on the algorithm of reconstruction, they even fall together with nonprimate representatives (Schmitz, Ohme, and Zischler 2002). In this study we propose that the same mechanism inversely places the colugo amidst the primates. Directional mutation pressure leads to an alteration of the nucleotide composition; hence, in extreme cases, phylogenetic reconstructions might include an artificial clustering of phylogenetically unrelated taxa due to similar base composition (Tarrio, Rodriguez-Trelles, and Ayala 2001). This is obviously most pronounced where two successive splits occur within a short period of time, leaving behind only weak phylogenetic signals. In an analysis of mt genes and their respective nuclear pseudogenes, we were able to show that in primates the plasticity of nucleotide composition is specific for mtDNA (Schmitz, Ohme, and Zischler 2002).

We compared the nucleotide composition for several mammals (fig. 3*A*). *Cynocephalus variegatus* clearly represents the nucleotide composition exclusively found in higher primates. Using correspondence analysis to examine the AA composition of *C. variegatus* mtDNA, a multivariate method in which observations (AA composition) and variables (taxa) can be jointly displayed in a two-dimensional space, a clustering of *C. variegatus* and higher primates was revealed, disclosing a similar AA preference in these taxa. In contrast, the prosimians (tarsiers and stepsirhines) group with nonprimate eutherians, exhibiting an affinity for AAs encoded by AT-rich codons (circled in fig. 3*B*).

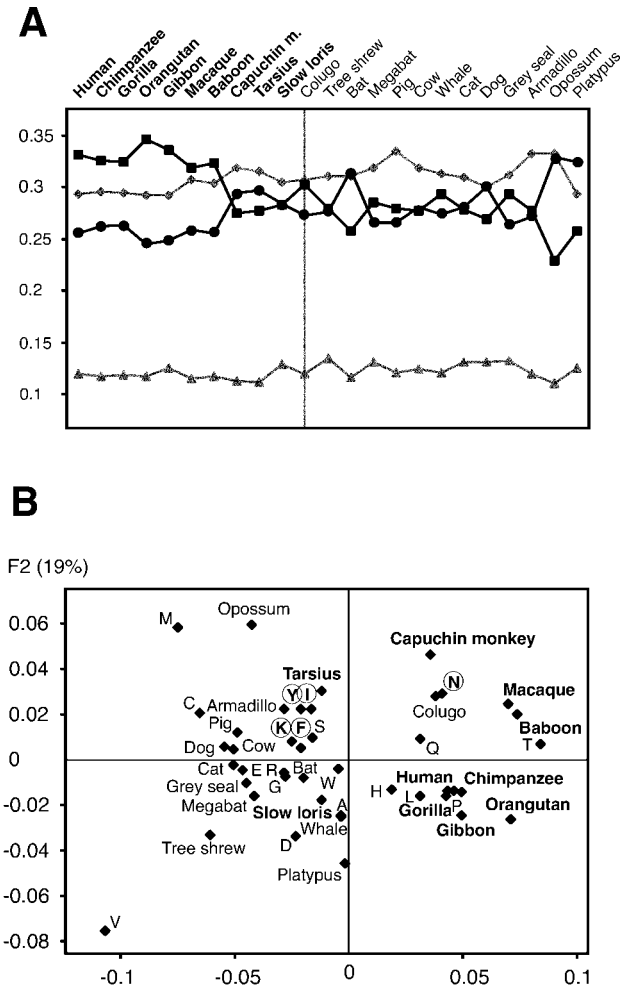

FIG. 3.—(*A*) Nucleotide frequencies of the 23 mammalian mtDNAs as obtained from the 12 H-strand genes. The adenosine base frequency is represented by gray rhombs, cytosine frequency by black rectangles, guanine frequency by gray triangles, and thymine frequency by black circles. (*B*) Correspondence analysis of the 12 H-strand protein-coding genes as a function of AA composition. The multivariate analysis of codon usage was carried out with codonW version 1.4.2 (http://bioweb.pasteur.fr/seqanal/interfaces/codonw.html). Adenosine- and thymine-rich AAs (Foster, Jermiin, and Hickey 1997) are circled. F1 and F2, first and second factorial axes represent 58% and 19% of the total variability in the correspondence analysis, respectively.

In the study by Schmitz, Ohme, and Zischler (2002), we found evidence for an extensive nucleotide compositional plasticity in primates whose AA composition changed in the direction predicted by the underlying nucleotide bias. In the present study we suggest that a similar effect of base composition similarity is responsible for joining the colugo with higher primates in phylogenetic tree reconstructions.

To circumvent the problem of variable base composition among sequences, we applied the LogDet transformation by both including and excluding invariant sites in separate analyses. However, this method did not resolve the conflicting tree shown in figure 1. One potential reason for this could be that LogDet does not compensate for rate heterogeneity across sites (see also Waddell et al. 1999).

Thus, from an ad hoc analysis of the mtDNA of mammalian representatives, we cannot gain conclusive evidence for a sister group relationship of dermopterans and primates. To avoid an inadvertent effect of the nucleotide composition on phylogenetic reconstructions, we excluded all higher primates from further computations, thus taking into account only taxa with similar mt nucleotide composition. As a result, independent of the method of phylogenetic reconstruction applied, the mtDNA sequence of colugo supports a sister taxon relationship to tarsiers and strepsirhines as representatives of the primate order (tree not shown).

In summary, the SINE results clearly refute both the mtDNA-based phylogenetic position of the colugo presented in the study by Arnason et al. (2002) and the interpretations obtained from the composite mtDNA and nuclear data of Murphy et al. (2001*a*). Thus, the newly proposed clade ''Dermosimii'' joining together Dermoptera and higher or anthropoid primates (Arnason et al. 2002) is not supported by SINE-evidence. To tease apart the influence of mtDNA and nuclear DNA information in the composite nuclear and mt data set of Murphy et al. (2001*a*), we reanalyzed this data set by including or excluding mt sequence information. Whereas a separate analysis of nuclear sequence information generated a largely unresolved tree that involved among others the dermopteran representative, the composite data set gives rise to a Dermoptera-Anthropoidea sister group with high support. It therefore seems that the clustering of Dermoptera and higher primates in Murphy et al. (2001*a*) is largely driven by mt sequence information.

In addition, our mt data do not agree with the modified phylogenetic position of dermopterans as presented in Murphy et al. (2001*b*). In this article the colugo shows a strong affiliation to *Tupaia* (up to 100% Bayesian posterior probability). No supporting signals for this affiliation could be found in the mt genome.

Using the presence or absence of retrotransposable elements as evidence, we were essentially able to demonstrate that the nucleotide compositional plasticity of mammalian mt genomes strongly interferes with phylogenetic reconstructions involving primate and primate-related taxa (figs. 1 and 2*D*). Sampling only mammalian species with similar nucleotide composition gives rise to a phylogenetic constellation in which colugos and primates share a common ancestor that is exclusive to all other extant eutherian taxa. Given this unsatisfactory incomplete taxonomic sampling, future comparative sequencing projects that reveal the presence of retrotransposable elements shared by dermopterans and primates but not by other living eutherians are urgently required. For the part of nuclear DNA-based evidence, this could provide a final settlement for the debate of primate origins.

Supplementary Material

Sequences are deposited in GenBank under accession numbers: mt genome—*Cynocephalus variegatus* (AF460846); MER locus—*Homo sapiens* (X83604, positions 7059–8222), *Macaca mulatta* (AF468826), *Sai-*

miri sciureus (AF468827), *Cheirogaleus medius* (AF468828), *Varecia variegata* (AF468829), *Cynocephalus variegatus* (AF468830), and *Oryctolagus cuniculus* (AF468831); Alu-SINE locus—*Homo sapiens* (X54816), *Colobus guereza* (AF278728), *Callithrix jacchus* (AF278730), *Tarsius bancanus* (AF278731), *Varecia variegata* (AF278732), *Cynocephalus variegatus* (AF460847), *Cavia porcellus* (AF278734).

Acknowledgments

We thank H. J. Kuhn and D. Penny for discussion and K. Hodges for support. We are indebted to C. Roos, Y. Rumpler, and members of the German Primate Center (Department of Pathology) for providing DNA and tissue samples and thank K. Gee for editorial assistance.

LITERATURE CITED

ARNASON, U., J. A. ADEGOKE, K. BODIN, E. W. BORN, Y. B. ESA, A. GULLBERG, M. NILSSON, R. V. SHORT, X. F. XU, and A. JANKE. 2002. Mammalian mitogenomic relationships and the root of the eutherian tree. Proc. Natl. Acad. Sci. **99**: 8151–8156.

BEARD, K. C. 1990. Gliding behaviour and palaeoecology of the alleged primate family Paromomyidae (Mammalia, Dermoptera). Nature **345**:340–341.

———. 1993. Origin and evolution of gliding in early Cenozoic Dermoptera (Mammalia, Primatomorpha). Pp. 63–90 *in* J. G. Fleagle and R. D. E. Macphee, eds. Primates and their relatives in phylogenetic perspective. Plenum Press, N.Y.

FELSENSTEIN, J. 1995. PHYLIP (phylogeny inference package). Version 3.6. Distributed by the author, Department of Genetics, University of Washington, Seattle.

FLEAGLE, J. C., and R. D. E. MACPHEE. 1993. Primates and their relatives in phylogenetic perspective. Plenum Press, N.Y.

FOSTER, P. G., L. S. JERMIIN, and D. A. HICKEY. 1997. Nucleotide composition bias affects amino acid content in proteins coded by animal mitochondria. J. Mol. Evol. **44**:282–288.

GRAUR, D., W. A. HIDE, and W.-H. LI. 1991. Is the guinea-pig a rodent? Nature **351**:649–652.

GREGORY, W. K. 1910. The orders of mammals. Bull. Am. Mus. Nat. Hist. **27**:1–524.

HWU, H. R., J. W. ROBERTS, E. H. DAVIDSON, and R. J. BRITTEN. 1986. Insertion and/or deletion of many repeated DNA sequences in human and higher ape evolution. Proc. Natl. Acad. Sci. USA **83**:3875–3879.

KAY, R. F., R. W. THORINGTON, and P. HOUDE. 1990. Eocene plesiadapiform shows affinities with flying lemurs not primates. Nature **345**:342–344.

KISHINO, H., and M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J. Mol. Evol. **29**:170–179.

LI, W.-H., Z. GU, H. WANG, and A. NEKRUTENKO. 2001. Evolutionary analyses of the human genome. Nature **409**:847–849.

LIN, Y.-H., and D. PENNY. 2001. Implications for bat evolution from two new complete mitochondrial genomes. Mol. Biol. Evol. **18**:684–688.

MADSEN, O., M. SCALLY, C. J. DOUADY, D. J. KAO, R. W. DEBRY, R. ADKINS, H. M. AMRINE, M. J. STANHOPE, W. W. DE JONG, and M. S. SPRINGER. 2001. Parallel adaptive radiations in two major clades of placental mammals. Nature **409**:610–614.

MARTIN, R. D. 1993. Primate origins: plugging the gaps. Nature **363**:223–234.

MURPHY, W. J., E. EIZIRIK, W. E. JOHNSON, Y. P. ZHANG, O. A. RYDER, and S. J. O'BRIEN. 2001*a*. Molecular phylogenetics and the origins of placental mammals. Nature **409**: 614–618.

MURPHY, W. J., E. EIZIRIK, S. J. O'BRIEN et al. (11 co-authors). 2001*b*. Resolution of the early placental mammal radiation using bayesian phylogenetics. Science **294**:2348–2351.

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

SCHMITZ, J., M. OHME, and H. ZISCHLER. 2000. The complete mitochondrial genome of *Tupaia belangeri* and the phylogenetic affiliation of Scandentia to other eutherian orders. Mol. Biol. Evol. **17**:1334–1343.

———. 2001. SINE insertions in cladistic analyses and the phylogenetic affiliations of *Tarsius bancanus* to other primates. Genetics **157**:777–784.

———. 2002. The complete mitochondrial sequence of *Tarsius bancanus*: evidence for an extensive nucleotide compositional plasticity of primate mitochondrial DNA. Mol. Biol. Evol. **19**:544–553.

SHEDLOCK, A. M., and N. OKADA. 2000. SINE insertions: powerful tools for molecular systematics. Bioessays **22**:148–160.

SINGER, S. S., J. SCHMITZ, C. SCHWIEGK, and H. ZISCHLER. 2003. Molecular cladistic markers in New World monkey phylogeny (Platyrrhini, Primates). Mol. Phylogenet. Evol. (in press).

STRIMMER, K., and A. VON HAESELER. 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. **13**:964–969.

SWOFFORD, D. L. 2000. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer, Sunderland, Mass.

TARRIO, R., F. RODRIGUEZ-TRELLES, and F. J. AYALA. 2001. Shared nucleotide composition biases among species and their impact on phylogenetic reconstructions of the Drosophilidae. Mol. Biol. Evol. **18**:1464–1473.

TEELING, E. C., M. SCALLY, D. J. KAO, M. L. ROMAGNOLI, M. S. SPRINGER, and M. J. STANHOPE. 2000. Molecular evidence regarding the origin of echolocation and flight in bats. Nature **403**:188–192.

THOMPSON, J. D., T. J. GIBSON, F. PLEWNIAK, F. JEANMOUGIN, and D. G. HIGGINS. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. **25**: 4876–4882.

WADDELL, P. J., Y. CAO, J. HAUF, and M. HASEGAWA. 1999. Using novel phylogenetic methods to evaluate mammalian mtDNA, including amino acid-invariant sites-LogDet plus site stripping, to detect internal conflicts in the data, with special reference to the positions of hedgehog, armadillo, and elephant. Syst. Biol. **48**:31–53.

DAN GRAUR, reviewing editor

Accepted August 26, 2002