

# Estimation of Primate Speciation Dates Using Local Molecular Clocks

Anne D. Yoder,\*† and Ziheng Yang‡

\*Department of Cell and Molecular Biology, Northwestern University Medical School, Chicago, Illinois; †Department of Zoology, Field Museum of Natural History, Chicago, Illinois; and ‡Department of Biology, Galton Laboratory, University College London, London, England

Protein-coding genes of the mitochondrial genomes from 31 mammalian species were analyzed to estimate the speciation dates within primates and also between rats and mice. Three calibration points were used based on paleontological data: one at 20–25 MYA for the hominoid/cercopithecoid divergence, one at 53–57 MYA for the cetacean/artiodactyl divergence, and the third at 110–130 MYA for the metatherian/eutherian divergence. Both the nucleotide and the amino acid sequences were analyzed, producing conflicting results. The global molecular clock was clearly violated for both the nucleotide and the amino acid data. Models of local clocks were implemented using maximum likelihood, allowing different evolutionary rates for some lineages while assuming rate constancy in others. Surprisingly, the highly divergent third codon positions appeared to contain phylogenetic information and produced more sensible estimates of primate divergence dates than did the amino acid sequences. Estimated dates varied considerably depending on the data type, the calibration point, and the substitution model but differed little among the four tree topologies used. We conclude that the calibration derived from the primate fossil record is too recent to be reliable; we also point out a number of problems in date estimation when the molecular clock does not hold. Despite these obstacles, we derived estimates of primate divergence dates that were well supported by the data and were generally consistent with the paleontological record. Estimation of the mouse-rat divergence date, however, was problematic.

## Introduction

Evolutionary biologists are interested in determining the age of speciation events. Aside from the intrinsic fascination for knowing the antiquity of a given species or clade, accurate date estimation is important for advancing evolutionary theory. By calculating the temporal setting for a given divergence event, one can determine the geological and environmental context for that event and consequently gain a better understanding of speciation and dispersal mechanisms. Genetic data are commonly used for these purposes (Bermingham et al. 1992; Da Silva and Patton 1993; Patton, da Silva, and Malcolm 1994; Horai et al. 1995; Hedges et al. 1996; Rassmann 1997; Avise, Walker, and Johns 1998; Bermingham and Moritz 1998; Cooper and Fortey 1998; Voelker 1999). Unfortunately, the results of such studies are often controversial due to their disagreement with the fossil record and/or with each other.

A number of factors may account for such disagreements. For the most part, the arguments have focused on perceived flaws in the fossil record. Some geneticists have dismissed the discrepancies between their studies and the fossil record by claiming that the latter is simply too incomplete to be reliable (Easteal, Collet, and Betty 1995; Hedges et al. 1996; Kumar and Hedges 1998). Paleontologists have been quick to respond with evidence to the contrary (Alroy 1999; Foote et al. 1999). Others have pointed out that by their very nature, genetic and fossil data are destined to give different estimates of organismal ages, because fossil morphological data can only identify crown clades, while genetic data

can only identify the earliest stages of divergence, long before clade-defining morphological synapomorphies were established (Archibald 1999). It is therefore argued that genetic estimates of clade age will always be older than those identified via the fossil record. Molecular evolutionary issues are also important. For example, selective constraints, genetic repair mechanisms, and effective population sizes will affect the relative utility and accuracy of a given genetic marker. Thus, it should not be surprising that nucleotide data may give age estimates different from those of amino acid data, just as a mitochondrial gene may give a different estimate than a nuclear gene.

Finally, the role of statistical analysis should not be underestimated. Increasingly sophisticated models are being applied to reconstruction of molecular phylogenies, and model assumptions appear even more important for date estimation. Simple unrealistic models do not correct for multiple hits properly and cause biased estimates of speciation dates (Yang 1996a). Most important with regard to date estimation is the assumption of the molecular clock, that is, rate constancy among lineages; incorrectly assuming the clock may lead to spurious date estimates (Takezaki, Rzhetsky, and Nei 1995). Recent work has suggested the possibility of estimating dates without assuming a global molecular clock (Sanderson 1997; Rambaut and Bromham 1998; Thorne, Kishino, and Painter 1998; Huelsenbeck, Larget, and Swofford 2000). However, the utility and limitations of such methods are not yet well understood.

We implemented maximum-likelihood (ML) models of local molecular clocks and applied them to protein-coding genes in the mitochondrial genome from 31 mammalian species. The effects of calibration date, data type, and statistical model on estimation of speciation dates were explored. We examined these effects on the recovery of divergence dates within anthropoid primates and between two murine rodents, the mouse and the rat.

Key words: local molecular clock, maximum likelihood, mitochondrial DNA, molecular clock, primates, mammals, speciation dates.

Address for correspondence and reprints: Anne D. Yoder, Department of Cell and Molecular Biology, 303 East Chicago Avenue, Chicago, Illinois 60611. E-mail: ayoder@nwu.edu.

*Mol. Biol. Evol.* 17(7):1081–1090, 2000

© 2000 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

**Table 1**  
**Sources of Complete mtDNA Sequences**

Binomial	Common Name	Genbank Accession No.
<i>Ornithorhynchus anatinus</i>	Platypus	X83427
<i>Didelphis virginiana</i>	Opossum	Z29573
<i>Macropus robustus</i>	Wallaroo	Y10524
<i>Mus musculus</i>	Mouse	J01420
<i>Rattus norvegicus</i>	Rat	X14848
<i>Glis glis</i>	Fat dormouse	AJ001562
<i>Cavia porcellus</i>	Guinea pig	WWW
<i>Oryctolagus cuniculus</i>	Rabbit	AJ001588
<i>Artibeus jamaicensis</i>	Fruit bat	AF061340
<i>Bos taurus</i>	Cow	J01394
<i>Ovis aries</i>	Sheep	AF010406
<i>Sus scrofa</i>	Pig	AJ002189
<i>Erinaceus europaeus</i>	Hedgehog	X88898
<i>Balaenoptera physalus</i>	Fin whale	X61145
<i>Balaenoptera musculus</i>	Blue whale	X72204
<i>Equus caballus</i>	Horse	X79547
<i>Equus asinus</i>	Donkey	X97337
<i>Rhinoceros unicornis</i>	Indian rhinoceros	X97336
<i>Ceratotherium simum</i>	White rhinoceros	Y07726
<i>Felis catus</i>	Domestic cat	U20753
<i>Canis familiaris</i>	Domestic dog	U96639
<i>Phoca vitulina</i>	Harbor seal	X63726
<i>Halichoerus grypus</i>	Gray seal	X72004
<i>Papio hamadryas</i>	Hamadryas baboon	Y18001
<i>Hylobates lar</i>	Gibbon	X99256
<i>Pongo pygmaeus</i>	Bornean orangutan	D38115
<i>Pongo abelii</i>	Sumatran orangutan	X97707
<i>Gorilla gorilla</i>	Gorilla	X93347
<i>Homo sapiens</i>	Human	X93334
<i>Pan troglodytes</i>	Common chimpanzee	X93335
<i>Pan paniscus</i>	Pygmy chimpanzee	D38116

Both groups have received much attention in the literature, the anthropoid primates due primarily to our inherent interest in human evolution, and the rodents due primarily to the large discrepancies between paleontological and genetic estimates of their antiquity. A recent study of primate divergence dates (Arnason, Gullberg, and Janke 1998) explored the effect of fossil dates outside the primate clade for tree calibration. The authors argued that the chosen calibrations (between cetaceans and artiodactyls and between equids and rhinoceroses) were based on more complete fossil records than were any of the fossil calibrations typically employed within

the primate clade. That study estimated primate divergence dates to be much earlier than had previously been determined by genetic studies. On the other hand, genetic studies that have addressed the issue of murine rodent antiquity (O'hUigin and Li 1992; Frye and Hedges 1995; Kumar and Hedges 1998) have repeatedly found the *Mus/Rattus* divergence to be far older than the 12–14 MYA date indicated by the fossil record (Jacobs and Downs 1994). One likely explanation for the discrepancy is that rodents tend to show higher rates of molecular evolution than do other mammals, including primates (Wu and Li 1985; Robinson et al. 1997). By applying statistical methods that allow for variable evolutionary rates among mammalian lineages, we hope to address some of the complications that have handicapped previous studies.

## Materials and Methods

### Alignment

Complete mtDNA sequences for 31 mammalian species were retrieved from GenBank (table 1). The 12 protein-coding genes on the H-strand of the mitochondrial genome were concatenated according to their relative positions (fig. 1). There are several regions in which one gene overlaps another, using differing reading frames. In such cases, overlapping sequence of the 3' end of the 5' gene was removed. Additional bases of the downstream gene were removed in order to preserve the codon structure of that gene (e.g., in the case of the 7-bp overlap between NADH4L and NADH4, 9 bp were removed from NADH4L). The NADH6 gene was not used, as it is coded by the other strand of the mitochondrial genome and has very different codon usage and substitution patterns. Start and stop codons were removed. Sequences are aligned manually at the amino acid level, aided by probabilities of changes between amino acids estimated previously (Yang, Nielsen, and Hasegawa 1998). The aligned sequence has 10,806 nt (3,602 amino acids). The alignment is available from the authors.

**Typical Mammal mtDNA Alignment  
(5' to 3')**

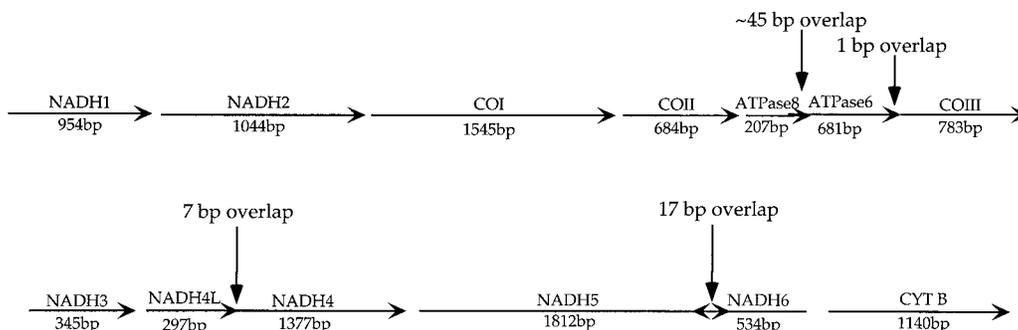


FIG. 1.—Alignment map showing locations of the genes in the mammalian mitochondrial genome.

## Phylogenetic Reconstruction and Calculation of Bootstrap Support

Complete-genome mtDNA data have been analyzed extensively for phylogeny reconstruction and are known to produce unorthodox phylogenies for vertebrates (e.g., Cao et al. 1998; Takezaki and Gojobori 1999). We did not attempt to resolve the phylogeny with certainty, but collected four candidate tree topologies for date estimation. We used the PROTML and NUCML programs in the MOLPHY package (Adachi and Hasegawa 1996b) to perform heuristic tree searches. The HKY85 model (Hasegawa, Kishino, and Yano 1985) was used for nucleotide sequences, and the mtREV24 model was used for amino acid sequences (Adachi and Hasegawa 1996a). Further comparisons among candidate trees were performed using the BASEML and CODEML programs in the PAML package to account for variable rates among sites (Yang 1999). Nucleotide data were analyzed in two ways, one using the third codon positions only and another using data at all three positions, with different substitution rates, transition/transversion rate ratios ( $\kappa$ ), and base frequencies among codon positions assumed (Yang 1996b).

Tree topologies were compared with the REL approximation to the bootstrap (Felsenstein 1985; Kishino and Hasegawa 1989). For models accounting for heterogeneity among codon positions, data from different positions do not follow the same distribution. As pointed out by H. Shimodaira (personal communication), bootstrap replicates for such partitioned data are constructed by stratified sampling. That is, each replicate consists of the same number of sites as in the original data set for each codon position, obtained by sampling with replacement. The calculation was done with the BASEML program. It is noted that this approach produced results very similar to those produced by sampling ignoring the data partitions when each partition contained many sites.

## Calibration Points

Lee (1999) stressed the importance of critically assessing the reliability of fossil calibration points. In this paper, we use three calibration points: one based on the primate fossil record and two taken from outside of the primates. A range of dates was explored for each to reflect the uncertainty concerning fossil identities. The first calibration assumed a range of 20–25 MYA for the hominoid/cercopithecoid divergence ( $C_1$  in the trees of fig. 2), the second assumed a range of 53–57 MYA for the cetacean/artiodactyl divergence ( $C_2$ ), and the third assumed a range of 110–130 MYA for the metatherian/eutherian divergence ( $C_3$ ). The  $C_1$  calibration was based on the estimated ages of the three earliest putative cercopithecoid and hominoid fossils, *Victoriapithecus* and *Prohylobates*, both of which are considered to be early cercopithecoids prior to the divergence of the colobine and cercopithecoine lineages, have been dated to approximately 18–20 MYA (Benefit 1993). *Kamoyapithecus*, the earliest known hominoid, has been dated to 25 MYA (Leakey, Ungar, and Walker 1995). The  $C_1$  paleontological estimate of catarrhine antiquity is thus considerably

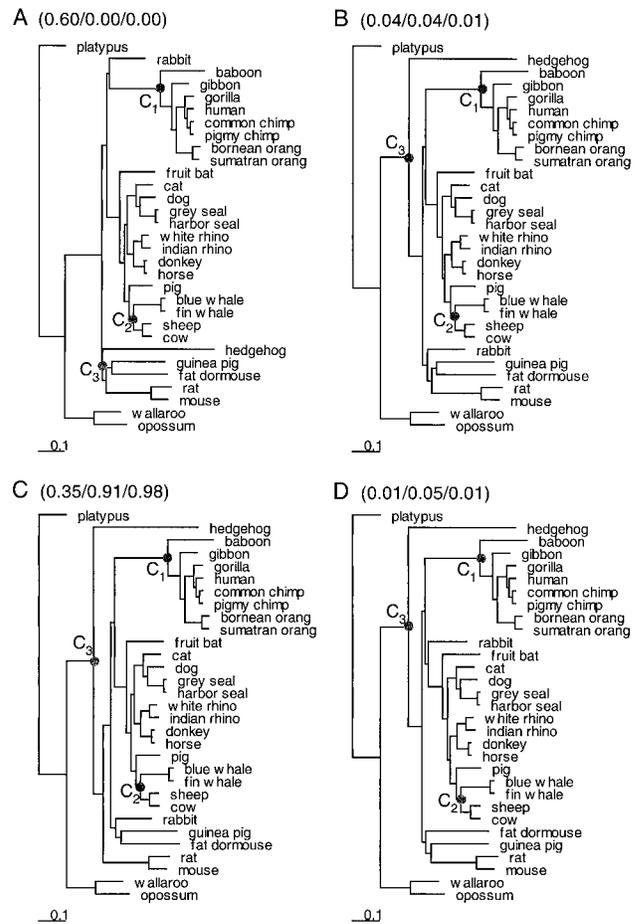


FIG. 2.—Four phylogenies used for calculating speciation dates. Trees A and B were obtained from heuristic tree searches using third-position nucleotide data with tree A being the ML tree and tree B being most consistent with morphological data. Trees C and D were obtained from heuristic tree searches using amino acid data, with tree C being the ML tree and tree D being most consistent with morphological data. Numbers in parentheses, in the format “(3rd/a.a./all),” are the RELL bootstrap proportions calculated using third-position nucleotide sequences, amino acid sequences, and complete nucleotide sequences, respectively. The three calibration points ( $C_1$ ,  $C_2$ , and  $C_3$ ) are indicated on the trees. The branch lengths are estimated from the amino acid sequences under the mtREV24 model (Adachi and Hasegawa 1996a) with gamma rates among sites (Yang 1993, 1994).

younger than the >50 MYA estimated by Arnason, Gullberg, and Janke (1998). The  $C_2$  calibration is based primarily on paleontological estimates for the antiquity of Cetacea (Bajpai and Gingerich 1998; Thewissen and Hussain 1998). Arnason, Gullberg, and Janke (1998) similarly used the cetacean/artiodactyl divergence as a calibration for their genetic analysis, although they employed a slightly earlier date of 60 MYA. The range of dates assumed for the  $C_3$  calibration was based on paleontological estimates of the antiquity of the metatherian and eutherian lineages. *Deltatheridium*, a basal metatherian, is known from the late Cretaceous of Mongolia (Rougier, Wible, and Novacek 1998), and *Prokennalestes*, a basal eutherian, is known from the early Cretaceous, also of Mongolia (Kielan-Jaworowska and Dashzeveg 1989).

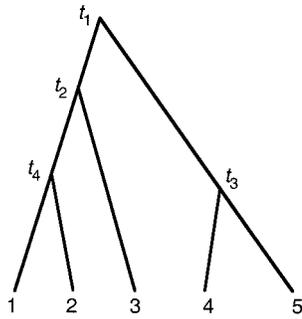


FIG. 3.—An example tree to explain PAML implementation of models of local molecular clocks.

### Models of Local Molecular Clocks

Current phylogenetic packages (such as PAUP, PHYLIP, and PAML) implement two extreme models concerning substitution rates among lineages. One is the global clock, assuming the same rate for all lineages. The other assumes free rates, that is, one independent rate for each branch. Models of local molecular clocks lie between these two extremes and assume that some branches (e.g., those for a closely related group of species) have the same rate, while different parts of the tree may have different rates. Such models are implemented in the PAML package. In the following, we describe the implementation of the global clock first, and then the implementation of local molecular clocks.

The natural parameters for the model of a global molecular clock are the node ages (e.g.,  $t_1$ ,  $t_2$ ,  $t_3$ , and  $t_4$  in fig. 3), measured as the expected number of substitutions per site from the node to the present time. A rooted tree of  $n$  species has  $n - 1$  interior nodes, so there are  $n - 1$  parameters in the model. Maximum-likelihood estimation of node ages requires numerical optimization under inequality constraints, such as  $t_4 \leq t_2 \leq t_1$  and  $t_3 \leq t_1$ . Algorithms currently used in PAML deal with simple bounds on parameters but not general inequality constraints, so we redefined the parameters to facilitate the iteration, according to a suggestion by Jeff Thorne. We used the age of the root as the first parameter. With other node ages, we used the ratio of the age of a node to the age of its immediate ancestor. In the example tree of figure 3, the new parameters are thus  $x_1 = t_1$ ,  $x_2 = t_2/t_1$ ,  $x_3 = t_3/t_1$ , and  $x_4 = t_4/t_2$ . As a result of this transform, only simple bounds are involved (that is,  $0 \leq x_1 < \infty$ ,  $0 \leq x_i \leq 1$  for  $i = 2, 3, 4$ ).

In a model of local molecular clock, we assume that each branch in the phylogeny can take one of  $k$  possible rates. We let  $r_0 = 1$  be the default rate and simply use  $k - 1$  rate multiplication factors as additional parameters. The model then has  $n - k - 2$  parameters. When  $k = 1$ , all branches have the same rate, and the model reduces to the global clock. Note that the number of branches in an unrooted tree topology is  $2n - 3$ , so that  $n + k - 2$  should not exceed  $2n - 3$ . Furthermore, certain specifications of rates for branches may make it impossible to identify all parameters in the model, and such identifiability problems have to be avoided (see *Results and Discussion*).

## Results and Discussion

### Collection of Candidate Phylogenies

We employed two data sets for generating phylogenetic trees: one of amino acid sequences derived from the aligned nucleotide sequences and another of nucleotide sequences composed of third codon positions only. The MOLPHY package was used for a quick search, followed by a proper likelihood evaluation of the top 50 trees in each data set. Substitution rates were assumed to be free to vary among branches in those analyses; that is, no clock was assumed. Two trees were selected from each data set: the maximum-likelihood tree and another tree that appeared most consistent with morphological hypotheses of mammalian relationships (Miyamoto 1996; Shoshani and McKenna 1998; Liu and Miyamoto 1999). The four trees from the two data sets were employed for date estimation (fig. 2).

A number of results are universal to all trees found in the quick search: (1) primates are monophyletic, with interrelationships among them stable (as illustrated in fig. 2); (2) carnivores, artiodactyls, cetaceans, and perissodactyls form a clade (the Ferungulata); (3) cetaceans nest within the artiodactyls, with the pig as the basal taxon; (4) opossum and wallaroo form a clade (Metatheria) that is basal to the Eutheria. Our study did not address the issue of a Marsupionta clade, as platypus was designated a priori as the outgroup. Inconsistencies among the trees found in the quick search included the placement of the rabbit, the hedgehog, and the fruit bat, which was highly variable, although there was a general tendency for the hedgehog to be basal within the Eutheria and for the fruit bat to be sister to the ferungulates (as is shown in all four trees in fig. 2). The placement of the perissodactyls relative to the carnivores and the artiodactyls varied, as did the monophyly of the rodents. The majority of trees did not recover rodent monophyly.

We note that the amino acid data and the third-position nucleotide data gave conflicting support for the phylogeny. The RELL bootstrap proportions are listed for the four trees in figure 2 when only those four trees are compared. The nucleotide data at the third codon positions favored tree 1, while the amino acid sequences, as well as the combined data of all three codon positions, favored tree 3. Notably, the results obtained from the analysis of third codon positions were as reliable as those obtained from the amino acid sequences, despite the very large estimates of tree length for the third position (about 50 or 150 substitutions per site on the tree under the one-rate and gamma-rates models, respectively). This result contradicts the widespread view that rapidly evolving sites should be omitted or down-weighted in phylogenetic analysis (e.g., Waddell et al. 1999; Springer et al. 1999) and is more consistent with observations that these sites can contain phylogenetic signal (Yoder, Vilgalys, and Ruvolo 1996; Yang 1998). While the nucleotide data at the third positions and amino acid data gave inconsistent results, the conflicts between the molecular data and the morphological evidence were even greater, as trees 2 and 4 received little

**Table 2**  
**Estimation of Dates from Amino Acid Sequences (in Myr)**

	ONE RATE			GAMMA		
	C <sub>1</sub> : 25–20	C <sub>2</sub> : 57–53	C <sub>3</sub> : 130–110	C <sub>1</sub> : 25–20	C <sub>2</sub> : 57–53	C <sub>3</sub> : 130–110
Global clock						
Human-chimp (HC) . . . . .	5.48–4.38	16.71–15.53	14.63–12.38	4.88–3.90	15.38–14.30	12.07–10.21
Common-pigmy chimps (C) . . . . .	2.68–2.15	8.18–7.60	7.16–6.06	2.34–1.87	7.39–6.87	5.80–4.91
Gorilla (HC-G) . . . . .	7.57–6.06	23.08–21.46	20.21–17.10	6.76–5.41	21.33–19.84	16.74–14.17
Bornean-sumatran orang . . . . .	5.14–4.11	15.67–14.57	13.73–11.62	4.57–3.65	14.40–13.39	11.31–9.57
Orang (HCG) . . . . .	14.88–11.90	45.35–42.17	39.72–33.61	14.04–11.23	44.30–41.19	34.77–29.42
Gibbon (Gb-HCGO) . . . . .	17.25–13.80	52.58–48.89	46.05–38.96	16.24–12.99	51.24–47.65	40.21–34.03
Rat-mouse . . . . .	15.23–12.19	46.44–43.18	40.67–34.42	14.29–11.43	45.09–41.92	35.39–29.94
C <sub>1</sub> (baboon-hominoid) . . . . .	25–20	76.21–70.86	66.74–56.47	25–20	78.87–73.34	61.90–52.38
C <sub>2</sub> (Cetacean-Artiodactl) . . . . .	18.70–14.96	57–53	49.92–42.24	18.07–14.45	57–53	44.73–37.85
C <sub>3</sub> (Meta-Eutherian) . . . . .	48.69–38.96	148.44–138.02	130–110	52.50–42.00	165.65–154.02	130–110
ℓ . . . . .		–64,966.66			–60,220.47	
Local clock, gamma						
		Model 1 (four rates)			Model 2 (five rates)	
Human-chimp (HC) . . . . .	4.21–3.37	8.39–7.80	7.34–6.21	4.21–3.37	10.12–9.41	7.32–6.19
Common-pigmy chimps (C) . . . . .	2.00–1.60	3.99–3.71	3.50–2.96	2.00–1.60	4.82–4.48	3.48–2.95
Gorilla (HC-G) . . . . .	5.89–4.71	11.73–10.91	10.28–8.69	5.89–4.71	14.15–13.16	10.24–8.66
Bornean-sumatran orang . . . . .	4.00–3.20	7.98–7.42	6.99–5.91	4.01–3.21	9.63–8.95	6.97–5.89
Orang (HCG) . . . . .	12.68–10.15	25.28–23.51	22.14–18.74	12.69–10.16	30.50–28.36	22.06–18.67
Gibbon (Gb-HCGO) . . . . .	14.78–11.82	29.45–27.39	25.79–21.83	14.79–11.83	35.52–33.03	25.70–21.74
Rat-mouse . . . . .	22.25–17.80	44.35–41.24	38.84–32.87	22.46–17.97	53.95–50.17	39.03–33.02
C <sub>1</sub> (baboon-hominoid) . . . . .	25–20	49.83–46.34	43.64–36.93	25–20	60.06–55.85	43.45–36.77
C <sub>2</sub> (Cetacean-Artiodactl) . . . . .	28.59–22.88	57–53	49.92–42.24	23.73–18.98	57–53	41.23–34.89
C <sub>3</sub> (Meta-Eutherian) . . . . .	74.47–59.58	148.45–138.03	130–110	74.80–59.84	179.70–167.09	130–110
ℓ . . . . .		–59,966.46			–59,941.23	
		$r_0 = 1, r_1 = 0.548, r_2 = 0.618, r_3 = 1.923$			$r_0 = 1, r_1 = 0.551, r_2 = 0.620, r_3 = 1.944, r_4 = 1.718$	

NOTE.—Local-clock model 1 specifies four rates:  $r_1$  for platypus,  $r_2$  for wallaroo and opossum,  $r_3$  for hedgehog and primates; all other lineages have rate  $r_0 = 1$ . Local-clock model 2 adds a fourth rate  $r_4$  for whales (see fig. 2).  $\ell$  is the log-likelihood value.

statistical support from either the nucleotide data set or the amino acid data set.

#### Date Estimation

Three data sets (inferred amino acid sequences, complete nucleotide sequences, and third codon positions), four tree topologies (fig. 2), three calibration points (C<sub>1</sub>, C<sub>2</sub>, and C<sub>3</sub>), and a variety of statistical models were employed to estimate divergence dates within the primates and between the mouse and the rat. The estimates are given in tables 2–4. The assumption of a global molecular clock is easily rejected for any data type or tree topology. For example, the likelihood ratio statistic comparing models with and without the clock assumption is  $2\Delta\ell = 751.9$  for the amino acid sequence data and tree 1; this is much greater than the  $\chi^2$  critical value of 49.6 with  $df = 29$ . Branch lengths estimated from amino acid sequences without the clock assumption are shown in the trees of figure 2, where violation of the clock is obvious. Branch lengths estimated using nucleotide data showed similar patterns and are thus not shown.

The effect of tree topology on date estimation was small in all tests. For example, under the global-clock model and using third-position nucleotide sequences, the C<sub>1</sub> calibration at 25 MYA gave estimates of the human-chimpanzee separation in the range 4.1–4.2 MYA among the four trees. With the C<sub>2</sub> calibration at 57 MYA, the estimates ranged from 6.3 to 6.7, and with the C<sub>3</sub> calibration at 130 MYA, the estimates ranged

from 6.2 to 7.0 MYA. Estimates for other divergences were also similar among trees. The patterns were also similar in analysis of amino acid sequences. The uncertainty in the phylogenetic relationship appears to have had little effect on estimation of the primate dates, especially in comparison with other factors considered below. We thus present results for the tree of figure 2A only.

Substitution rates were highly variable among sites in the amino acid sequences, with the gamma shape parameter  $\alpha$  estimated at 0.41 for all trees and models (Yang 1996a). Even the third-position nucleotide sequences showed considerable rate variation, with  $\alpha$  estimated at 0.80. In general, accounting for variable evolutionary rates among sites makes shallow nodes (e.g., human/chimpanzee divergence) more recent and deep nodes (e.g., cercopithecoid/hominoid) older. This is because ignoring the rate variation leads to underestimated distances, with the bias being more serious for large distances than for small distances (Yang 1996a). Among-site rate variation has substantial effects, especially on estimation of recent divergence dates (tables 2–4).

Several local-clock models were implemented to account for variable rates among lineages, with the aid of branch length estimates obtained without the clock assumption (see fig. 2 for estimates obtained from the amino acid sequences). The non-eutherian outgroups had much lower rates, and independent rates were assigned to those lineages. In effect, the marsupial species

**Table 3**  
**Estimation of Dates from Third Codon Positions**

	ONE RATE			GAMMA		
	C <sub>1</sub> : 25–20	C <sub>2</sub> : 57–53	C <sub>3</sub> : 130–110	C <sub>1</sub> : 25–20	C <sub>2</sub> : 57–53	C <sub>3</sub> : 130–110
<b>Global clock</b>						
Human-chimp (HC) . . . . .	4.20–3.36	6.33–5.88	6.97–5.90	2.94–2.36	4.41–4.10	4.50–3.81
Common-pigmy chimps (C) . . . . .	1.41–1.13	2.12–1.97	2.34–1.98	0.57–0.46	0.86–0.80	0.87–0.74
Gorilla (HC-G) . . . . .	5.86–4.68	8.82–8.20	9.72–8.22	5.11–4.09	7.65–7.12	7.81–6.61
Bornean-sumatran orang . . . . .	2.60–2.08	3.92–3.65	4.32–3.66	1.57–1.26	2.36–2.19	2.40–2.03
Orang (HCG) . . . . .	9.98–7.99	15.04–13.98	16.57–14.02	10.46–8.37	15.67–14.57	15.99–13.53
Gibbon (Gb-HCGO) . . . . .	14.04–11.23	21.14–19.66	23.29–19.71	14.66–11.73	21.96–20.42	22.42–18.97
Rat-mouse . . . . .	37.12–29.69	55.91–51.99	61.60–52.12	36.45–29.16	54.60–50.77	55.73–47.15
C <sub>1</sub> (baboon-hominoid) . . . . .	25–20	37.66–35.01	41.49–35.11	25–20	37.45–34.82	38.22–32.34
C <sub>2</sub> (cetacean-artiodactyl) . . . . .	37.84–30.27	57–53	62.80–53.14	38.05–30.44	57–53	58.18–49.23
C <sub>3</sub> (meta-eutherian) . . . . .	78.33–62.67	117.99–109.71	130–110	85.03–68.02	127.37–118.43	130–110
ℓ . . . . .		–95,329,37			–92,738.70	
<b>Local clock (three rates)</b>						
Human-chimp (HC) . . . . .	4.18–3.34	6.30–5.86	6.84–5.79	2.95–2.36	4.43–4.12	4.45–3.77
Common-pigmy chimps (C) . . . . .	1.40–1.12	2.11–1.96	2.29–1.94	0.58–0.46	0.86–0.80	0.87–0.74
Gorilla (HC-G) . . . . .	5.82–4.66	8.79–8.17	9.54–8.07	5.11–4.09	7.67–7.13	7.71–6.52
Bornean-sumatran orang . . . . .	2.59–2.07	3.91–3.63	4.24–3.59	1.58–1.26	2.37–2.20	2.38–2.01
Orang (HCG) . . . . .	9.94–7.95	14.99–13.94	16.27–13.77	10.46–8.37	15.69–14.59	15.77–13.35
Gibbon (Gb-HCGO) . . . . .	13.99–11.19	21.10–19.62	22.91–19.38	14.66–11.73	21.99–20.45	22.11–18.71
Rat-mouse . . . . .	37.26–29.80	56.21–52.26	61.01–51.63	36.56–29.25	54.85–51.00	55.14–46.65
C <sub>1</sub> (baboon-hominoid) . . . . .	25–20	37.72–35.07	40.94–34.64	25–20	37.51–34.88	37.71–31.91
C <sub>2</sub> (cetacean-artiodactyl) . . . . .	37.78–30.22	57–53	61.88–52.36	37.99–30.39	57–53	37.30–48.48
C <sub>3</sub> (meta-eutherian) . . . . .	79.38–63.50	119.76–111.35	130–110	86.19–68.95	129.33–120.25	130–110
ℓ . . . . .		–95, 306.59			–92,731.01	
		$r_0 = 1, r_1 = 0.694, r_2 = 0.686$			$r_0 = 1, r_1 = 0.789, r_2 = 0.750$	

NOTE.—The local-clock model specifies three rates;  $r_1$  for platypus,  $r_2$  for wallaroo and opossum, and  $r_0$  for all other lineages. Estimate of the gamma shape parameter  $\alpha = 0.80$  under both the global- and the local-clock models.

**Table 4**  
**Date Estimation from All Three Codon Positions**

	ONE RATE			GAMMA		
	C <sub>1</sub> : 25–20	C <sub>2</sub> : 57–53	C <sub>3</sub> : 130–110	C <sub>1</sub> : 25–20	C <sub>2</sub> : 57–53	C <sub>3</sub> : 130–110
<b>Global clock</b>						
Human-chimp (HC) . . . . .	6.16–4.93	13.89–12.92	14.31–12.11	3.65–2.92	8.04–7.48	7.17–6.07
Common-pigmy chimps (C) . . . . .	2.32–1.85	5.23–4.86	5.39–4.56	0.72–0.58	1.60–1.48	1.42–1.20
Gorilla (HC-G) . . . . .	8.60–6.88	19.39–8.03	19.98–6.91	6.04–4.83	13.32–12.38	11.87–10.04
Bornean-sumatran orang . . . . .	4.33–3.46	9.77–9.08	10.06–8.51	2.68–2.15	5.92–5.50	5.28–4.47
Orang (HCG) . . . . .	14.42–11.53	32.52–30.24	33.50–28.35	12.38–9.91	27.31–25.40	24.35–20.61
Gibbon (Gb-HCGO) . . . . .	17.07–13.65	38.50–35.80	39.67–33.56	15.00–12.00	33.08–30.76	29.49–24.95
Rat-mouse . . . . .	22.44–17.95	50.61–47.06	52.15–44.12	23.29–18.63	51.38–47.78	45.81–38.76
C <sub>1</sub> (baboon-hominoid) . . . . .	25–20	56.39–52.43	58.10–49.16	25–20	55.15–51.28	49.17–41.60
C <sub>2</sub> (cetacean-artiodactyl) . . . . .	25.27–20.22	57–53	58.73–49.69	25.84–20.67	57–53	50.82–43.00
C <sub>3</sub> (meta-eutherian) . . . . .	55.94–44.75	126.17–117.32	130–110	66.10–52.88	145.82–135.59	130–110
ℓ . . . . .		–169,578.34			–160,531.18	
<b>Local clock, gamma</b>						
		Model 1 (four rates)		Model 2 (five rates)		
Human-chimp (HC) . . . . .	3.36–2.68	4.92–4.58	4.75–4.02	3.36–2.69	5.49–5.11	4.74–4.01
Common-pigmy chimps (C) . . . . .	0.66–0.53	0.97–0.91	0.94–0.80	0.66–0.53	1.09–1.01	0.94–0.79
Gorilla (HC-G) . . . . .	5.58–4.46	8.18–7.61	7.89–6.68	5.58–4.46	9.13–8.49	7.87–6.66
Bornean-sumatran orang . . . . .	2.48–1.98	3.64–3.38	3.51–2.97	2.48–1.99	4.06–3.78	3.50–2.96
Orang (HCG) . . . . .	11.62–9.29	17.05–15.86	16.44–13.91	11.62–9.30	19.01–17.68	16.40–13.87
Gibbon (Gb-HCGO) . . . . .	14.15–11.32	20.77–19.31	20.03–16.94	14.16–11.33	23.16–21.53	19.97–16.90
Rat-mouse . . . . .	34.59–27.68	50.78–47.21	48.96–41.42	34.82–27.85	56.96–52.96	49.12–41.56
C <sub>1</sub> (baboon-hominoid) . . . . .	25–20	36.69–34.12	35.38–29.94	25–20	40.90–38.03	35.27–29.84
C <sub>2</sub> (cetacean-artiodactyl) . . . . .	38.83–31.07	57–53	54.96–46.50	34.84–27.87	57–53	49.16–41.59
C <sub>3</sub> (meta-eutherian) . . . . .	91.86–73.49	134.84–125.37	130–110	92.14–73.72	150.75–140.17	130–110
ℓ . . . . .		–160,307.64			–160,291.04	
		$r_0 = 1, r_1 = 0.264, r_2 = 0.701, r_3 = 1.669$		$r_0 = 1, r_1 = 0.395, r_2 = 0.702, r_3 = 1.681, r_4 = 1.363$		

NOTE.—See note for table 2 for specifications of local clock models 1 and 2. Substitution rates at the three codon positions are in the proportions 1:0.40:8.10 without the gamma and 1:0.37:39 under the gamma models.

root the tree but do not appear to affect date estimation in primates. While the local-clock models did not seem to matter much when the primate calibration point  $C_1$  was used, probably due to the relative rate homogeneity within primates, the global and local clock models produced very different estimates when calibration points outside the primates ( $C_2$  and  $C_3$ ) were used. Take, for example, the gamma model analyses of the amino acid sequences in table 2. The global clock gave 15.4–14.3 MYA for the human-chimpanzee divergence under calibration point  $C_2$ , while local-clock model 1 gave dates of almost half that, at 8.4–7.8 MYA, much closer to dates expected from conventional wisdom. Similar patterns are seen in the analysis of the nucleotide data of all three codon positions. The different local-clock models most often produced similar results, presumably because local-clock model 1 accounts for most of the rate variation in the data, and other models, all more complex than model 1, were not very different from local-clock model 1 (see note to table 2). Date estimates obtained using a few other local-clock models we implemented were also similar to those obtained under local-clock model 1 and are thus not presented.

The most important factor affecting date estimation is the calibration point. For all three data sets analyzed (tables 2–4), dates estimated from the  $C_1$  calibration were considerably more recent than those estimated using the  $C_2$  and  $C_3$  calibrations, usually unrealistically so. The unrealistic nature of the  $C_1$  calibration is most forcibly demonstrated by comparing reciprocal calibrations. In all cases where the  $C_1$  calibration was used to estimate dates for the  $C_2$  and  $C_3$  nodes, the estimates were unequivocally too recent and thus falsified by the fossil record. This was particularly true for the  $C_2$  divergence, as the cetacean fossil record is so thoroughly understood (Bajpai and Gingerich 1998; Thewissen and Hussain 1998). For example, among all of the data-model combinations, the  $C_2$  dates estimated using the  $C_1$  calibration range from only 15 MYA (table 2: global clock without gamma for the amino acid data) to a maximum of 38 MYA (table 3: global clock with gamma for the third-position data). Even this maximum is at least 15 Myr too recent when judged against the cetacean fossil record (Bajpai and Gingerich 1998). These results therefore bolster the argument of Arnason, Gullberg, and Janke (1998) that primate fossil calibrations tend to be too recent. Arnason, Gullberg, and Janke (1998), however, determined the divergence between humans and chimpanzees to be 10–13 MYA, that between gorilla and the human/chimpanzee clade to be approximately 17 MYA, and that between hominoids and cercopithecoids to be >50 MYA. Our study found primate divergence dates of this antiquity only by analyzing the amino acid data (table 2) or the complete nucleotide data (table 4) with the assumption of a global molecular clock, an assumption that is clearly violated.

The above discussion leaves open the question of whether this study has accurately estimated primate divergence dates. At this point, we can at least rule out a subset of the results as inaccurate. Clearly, the global-clock assumption for the amino acid and complete nu-

cleotide sequences is invalid, as are all dates estimated via the  $C_1$  calibration. This leaves us with dates estimated via the local-clock models for the amino acid and complete nucleotide sequence data, as well as dates from the third-position sites. We examine these results in turn. For the amino acid data (table 2), the local-clock estimates for primate divergences are more compatible with conventional wisdom than are the global-clock estimates, but there are some troubling inconsistencies when the  $C_2$  and  $C_3$  reciprocal calibrations are compared. For example, in one analysis, the  $C_2$  prediction of  $C_1$  is as old as 60 MYA, whereas the  $C_3$  prediction is as recent as 36 MYA. These results raise doubts as to the reliability of the amino acid data, the appropriateness of the  $C_2$  and  $C_3$  calibrations, or both. Examination of the third-position data (table 3) and complete nucleotide data (table 4) suggest that the discrepancy is due less to inconsistencies between the fossil calibrations than to intrinsic properties of the amino acid data. Estimates of primate divergence dates using calibrations  $C_2$  and  $C_3$  are remarkably consistent under the local-clock model for the complete nucleotide sequences and under both global- and local-clock models for the third-position data. In the six analyses, divergence dates range from approximately 4–6 MYA for humans and chimpanzees, 7–9 MYA between gorillas and the human/chimpanzee clade, and 30–40 MYA between hominoids and cercopithecoids. We propose that these dates are those best supported by the data, given appropriate statistical analysis and proper fossil calibration. Moreover, these dates are generally compatible with the known primate fossil record (Benefit 1993; Leakey, Ungar, and Walker 1995; Ward et al. 1999) or recent molecular studies (Martin 1993; Takahata and Satta 1997).

Estimates of the mouse-rat divergence date do not present such a convergence of genetic and paleontological data. On the contrary, the six analyses that we deemed to be most reliable for the primate data yielded dates that were far too ancient to be realistic for the rodents. On the other hand, those analyses that we dismissed as irrelevant for estimating primate divergence dates tended to give more recent and, thus, more plausible estimates for the mouse-rat divergence. This suggests that there must be molecular-evolutionary properties of the rodent sequences that differ markedly from those of the primate sequences. Given that the discrepancies are most remarkable in the comparison of the amino acid data and the third-position data, we postulate that the two clades may have experienced different selection constraints and have different substitution rates that are not accounted for by our statistical models. The inclusion of only two rodent species in the data also appears to create problems in identifying rate changes within and immediately ancestral to the two rodent lineages (see below).

#### Difficulties of Date Estimation Without a Molecular Clock

The first application of local molecular-clock models to date estimation appears to be that of Kishino and

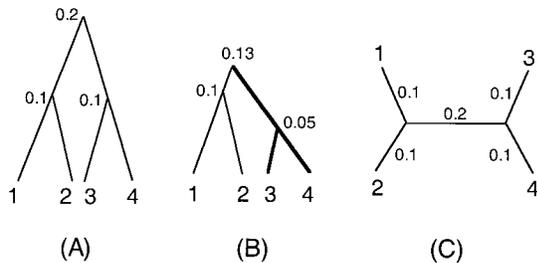


FIG. 4.—Three trees indistinguishable by molecular data, illustrating the identifiability problem when the molecular clock does not hold. Numbers beside nodes in *A* and *B* are node ages, proportional to the number of substitutions per site from that node to the present time. In tree *A*, a global clock holds, and all branches have the same rate (1). In tree *B*, branches on the left side of the tree have the same rate (1) as in tree *A*, while those on the right side have a rate twice as high. Both trees give the same branch lengths in the unrooted tree (*C*).

Hasegawa (1990), who estimated dates within hominoids using models of variable transition and transversion rates along lineages. The authors used a multivariate normal distribution to approximate the observed numbers of transitional and transversional differences between each pair of sequences. Rambaut and Bromham (1998) discussed a maximum-likelihood model with two rates on a tree of four species (quartets). The likelihood models implemented here are similar to those of previous studies. It should be pointed out that several problems are shared by those methods. The first is that the likelihood ratio test of hypotheses concerning rates among lineages is valid only if the null hypothesis is specified beforehand. This may be the case when there are extrinsic reasons to assign rates to branches. For example, two groups of species may be expected to have different evolutionary rates. However, if the hypothesis about rates is derived from the sequence data and tested using the same data, the significance values suggested by the  $\chi^2$  distribution will not be reliable, with too high a probability of incorrectly finding a difference. In this regard, the Bayes approach (Thorne, Kishino, and Painter 1998; Huelsenbeck, Larget, and Swofford 2000) may have advantages, as it does not require prior specification of rates for branches. This problem should be borne in mind if a likelihood ratio test is used to compare local-clock models implemented in this paper. For date estimation, this does not pose a serious concern, as incorrectly rejecting a hypothesis of equal rates and thus using additional rate parameters in the model may not be expected to lead to biased date estimates.

A serious problem attendant upon any attempt to estimate dates without a global clock is the well-known confounding effect between date and rate in molecular data. The likelihood of the tree/model depends solely on the branch lengths in the unrooted tree. As a result, global- and local-clock models that generate identical branch lengths in the unrooted tree are indistinguishable by the data. For example, the tree in figure 4*A* conforms to a global clock, while the tree in figure 4*B* has two rates for branches on the two sides of the root. Both trees give the same branch lengths in the unrooted tree (fig. 4*C*) and will have exactly the same likelihood. Either one can be the correct tree with the other being the es-

timiate, or neither may be correct. If tree *B* is true, in which case the global clock is violated, a test of the molecular clock will not detect the rate difference; by sliding the root along the branch of the unrooted tree, the tree becomes clocklike (tree *A*). Date estimation in such cases will be grossly misleading. For example, the root is 100% older than the node ancestral to species 1 and 2 in tree *A*, but the root is only 33% older in tree *B*. Arbitrary date estimates can be obtained by assuming different relative rates on the two sides of the root. A similar problem exists when an independent rate is assigned to one of the two branches around the root. The problem is particularly acute when only four species are analyzed and exists in Bayes methods allowing for variable evolutionary rates among lineages (Thorne, Kishino, and Painter 1998; Huelsenbeck, Larget, and Swofford 2000). It is possible that estimation of the mouse-rat divergence date in this study suffers from this problem. If the two rodent lineages have accelerated rates in the recent past, the data will be compatible with much more recent divergence dates for the two species than those in tables 2–4. In this regard, increased sampling of species may allow reliable estimation of within-group rates.

#### Acknowledgments

We are grateful to Chris Beard, Marian Dagosto, and Hans Thewissen for sharing their expert knowledge of the mammalian fossil record. Z.Y. thanks Nick Goldman and Bruce Rannala for discussions on models of local molecular clocks and Hidetoshi Shimodaira for discussions on bootstrap resampling of partitioned data. Ulfur Arnason and Axel Janke kindly allowed us to use their mtDNA alignment for purposes of comparison. We thank Axel Janke and two anonymous referees for comments. This study was supported by BBSRC grants 31/MMI09806 and 31/G10434 to Z.Y., and by NSF grant BCS-9905614 to A.D.Y.

#### LITERATURE CITED

- ADACHI, J., and M. HASEGAWA. 1996*a*. Model of amino acid substitution in proteins encoded by mitochondrial DNA. *J. Mol. Evol.* **42**:459–468.
- . 1996*b*. MOLPHY version 2.3: programs for molecular phylogenetics based on maximum likelihood. *Comput. Sci. Monogr.* **28**:1–150.
- ALROY, J. 1999. The fossil record of North American mammals: evidence for a Paleocene evolutionary radiation. *Syst. Biol.* **48**:107–118.
- ARCHIBALD, J. D. 1999. Molecular dates and the mammalian radiation. *Trends Ecol. Evol.* **14**:278.
- ARNASON, U., A. GULLBERG, and A. JANKE. 1998. Molecular timing of primate divergences as estimated by two nonprimate calibration points. *J. Mol. Evol.* **47**:718–727.
- AVISE, J. C., D. WALKER, and G. C. JOHNS. 1998. Speciation durations and Pleistocene effects on vertebrate phylogeography. *Proc. R. Soc. Lond. B Biol. Sci.* **265**:1707–1712.
- BAJPAI, S., and P. D. GINGERICH. 1998. A new Eocene archaeocete (Mammalia, Cetacea) from India and the time of origin of whales. *Proc. Natl. Acad. Sci. USA* **95**:15464–15468.

- BENEFIT, B. R. 1993. The permanent dentition and phylogenetic position of *Victoriapithecus* from Maboko Island, Kenya. *J. Hum. Evol.* **25**:83–172.
- BERMINGHAM, E., and C. MORITZ. 1998. Comparative phylogeography: concepts and applications. *Mol. Ecol.* **7**:367–369.
- BERMINGHAM, E., S. ROHWER, S. FREEMAN, and C. WOOD. 1992. Vicariance biogeography in the Pleistocene and speciation in North American wood warblers, a test of Menger's model. *Proc. Natl. Acad. Sci. USA* **89**:6624–6628.
- CAO, Y., A. JANKE, P. J. WADDELL, M. WESTERMAN, O. TAKENAKA, S. MURATA, N. OKADA, S. PAABO, and M. HASEGAWA. 1998. Conflict among individual mitochondrial proteins in resolving the phylogeny of eutherian orders. *J. Mol. Evol.* **47**:307–322.
- COOPER, A., and R. FORTEY. 1998. Evolutionary explosions and the phylogenetic fuse. *Trends Ecol. Evol.* **13**:151–156.
- DA SILVA, M. N. F., and J. L. PATTON. 1993. Amazonian phylogeography: mtDNA sequence variation in arboreal echimyid rodents (Caviomorpha). *Mol. Phylogenet. Evol.* **2**:243–255.
- EASTEAL, S., C. COLLET, and D. BETTY. 1995. The mammalian molecular clock. R. G. Landes, Austin, Tex.
- FELSENSTEIN, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**:783–791.
- FOOTE, M., J. P. HUNTER, C. M. JANIS, and J. J. SEPKOSKI. 1999. Evolutionary and preservational constraints on origins of biologic groups: divergence times of eutherian mammals. *Science* **283**:1310–1314.
- FRYE, M. S., and S. B. HEDGES. 1995. Monophyly of the order Rodentia inferred from mitochondrial DNA sequences of the genes for 12S rRNA, 16S rRNA, and tRNA-Valine. *Mol. Biol. Evol.* **12**:168–176.
- HASEGAWA, M., H. KISHINO, and T. YANO. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**:160–174.
- HEDGES, S. B., P. H. PARKER, C. G. SIBLEY, and S. KUMAR. 1996. Continental breakup and the ordinal diversification of birds and mammals. *Nature* **381**:226–229.
- HORAI, S., K. HAYASAKA, R. KONDO, K. TSUGANE, and N. TAKAHATA. 1995. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. *Proc. Natl. Acad. Sci. USA* **92**:532–536.
- HUELSENBECK, J. P., B. LARGET, and D. SWOFFORD. 2000. A compound Poisson process for relaxing the molecular clock. *Genetics* **154**:1879–1892.
- JACOBS, L. L., and W. R. DOWNS. 1994. The evolution of murine rodents in Asia. Pp. 149–156 in Y. TOMIDA, C. LI, and T. SETOGUCHI, eds. *Rodent and lagomorph families of Asian origin and diversification*. National Science Museum Monograph, Tokyo.
- KIELAN-JAWOROWSKA, Z., and D. DASHZEVEG. 1989. Eutherian mammals from the early Cretaceous of Mongolia. *Zool. Scripta* **18**:347–355.
- 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.
- . 1990. Converting distance to time: application to human evolution. *Methods Enzymol.* **183**:550–570.
- KUMAR, S., and S. B. HEDGES. 1998. A molecular timescale for vertebrate evolution. *Nature* **392**:917–920.
- LEAKEY, M. G., P. S. UNGAR, and A. WALKER. 1995. A new genus of large primate from the late Oligocene of Lothidok, Turkana District, Kenya. *J. Hum. Evol.* **28**:519–531.
- LEE, M. S. Y. 1999. Molecular clock calibrations and metazoan divergence dates. *J. Mol. Evol.* **49**:385–391.
- LIU, F.-G. R., and M. M. MIYAMOTO. 1999. Phylogenetic assessment of molecular and morphological data for eutherian mammals. *Syst. Biol.* **48**:54–64.
- MARTIN, R. D. 1993. Primate origins, plugging the gaps. *Nature* **376**:565–571.
- MIYAMOTO, M. M. 1996. A congruence study of molecular and morphological data for eutherian mammals. *Mol. Phylogenet. Evol.* **6**:373–390.
- O'HUIGIN, C., and W.-H. LI. 1992. The molecular clock ticks regularly in murid rodents and hamsters. *J. Mol. Evol.* **35**:377–384.
- PATTON, J. L., M. N. F. DA SILVA, and J. R. MALCOLM. 1994. Gene genealogy and differentiation among arboreal spiny rats (Rodentia: Echimyidae) of the Amazon Basin: a test of the riverine barrier hypothesis. *Evolution* **48**:1314–1323.
- RAMBAUT, A., and L. BROMHAM. 1998. Estimating divergence dates from molecular sequences. *Mol. Biol. Evol.* **15**:442–448.
- RASSMANN, K. 1997. Evolutionary age of the Galapagos iguanas predates the age of the present Galapagos islands. *Mol. Phylogenet. Evol.* **7**:158–172.
- ROBINSON, M., F. CATZEFLIS, J. BRIOLAY, and D. MOUCHIROUD. 1997. Molecular phylogeny of rodents, with special emphasis on murids: evidence from nuclear DNA evidence. *Mol. Phylogenet. Evol.* **8**:423–434.
- ROUGIER, G. W., J. R. WIBLE, and M. J. NOVACEK. 1998. Implications of *Deltatherium* specimens for early marsupial history. *Nature* **396**:459–463.
- SANDERSON, M. J. 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. *Mol. Biol. Evol.* **14**:1218–1232.
- SHOSHANI, J., and M. C. 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. S., H. M. AMRINE, A. BURK, and M. J. STANHOPE. 1999. Additional support for Afrotheria and Paenungulata, the performance of mitochondrial versus nuclear genes, and the impact of data partitions with heterogeneous base composition. *Syst. Biol.* **48**:65–75.
- TAKAHATA, N., and Y. SATTA. 1997. Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences. *Proc. Natl. Acad. Sci. USA* **94**:4811–4815.
- TAKEZAKI, N., and T. GOJOBORI. 1999. Correct and incorrect vertebrate phylogenies obtained by the entire mitochondrial DNA sequences. *Mol. Biol. Evol.* **16**:590–601.
- TAKEZAKI, N., A. RZHETSKY, and M. NEI. 1995. Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol.* **12**:823–833.
- THEWISSEN, J. G. M., and S. T. HUSSAIN. 1998. Systematic review of the Pakicetidae, early and middle Eocene Cetacea (Mammalia) from Pakistan and India. Pp. 220–238 in C. BEARD and M. R. DAWSON, eds. *Dawn of the age of mammals in Asia*. Carnegie Museum of Natural History, Pittsburgh, Pa.
- THORNE, J. L., H. KISHINO, and I. S. PAINTER. 1998. Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol. Evol.* **15**:1647–1657.
- VOELKER, G. 1999. Dispersal, vicariance, and clocks: historical biogeography and speciation in cosmopolitan passerine genus (*Anthus*: Motacillidae). *Evolution* **53**:1536–1552.
- 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.
- WARD, S., B. BROWN, A. HILL, J. KELLEY, and W. DOWNS. 1999. *Equatorius*: a new hominoid genus from the Middle Miocene of Kenya. *Science* **285**:1382–1386.
- WU, C.-L., and W.-H. LI. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc. Natl. Acad. Sci. USA* **82**:1741–1745.
- YANG, Z. 1993. Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. *Mol. Biol. Evol.* **10**:1396–1401.
- . 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* **39**:306–314.
- . 1996a. Among-site rate variation and its impact on phylogenetic analyses. *TREE* **11**:367–372.
- . 1996b. Maximum-likelihood models for combined analyses of multiple sequence data. *J. Mol. Evol.* **42**:587–596.
- . 1998. On the best evolutionary rate for phylogenetic analysis. *Syst. Biol.* **47**:125–133.
- . 1999. Phylogenetic analysis by maximum likelihood (PAML) (<http://abacus.gene.ucl.ac.uk/software/paml.html>). University College London, London.
- YANG, Z., R. NIELSEN, and M. HASEGAWA. 1998. Models of amino acid substitution and applications to mitochondrial protein evolution. *Mol. Biol. Evol.* **15**:1600–1611.
- YODER, A. D., R. VILGALYS, and M. RUVOLO. 1996. Molecular evolutionary dynamics of cytochrome b in strepsirrhine primates: the phylogenetic significance of third-position transversions. *Mol. Biol. Evol.* **13**:1339–1350.

NARUYA SAITOU, reviewing editor

Accepted March 29, 2000