The Platypus Is in Its Place: Nuclear Genes and Indels Confirm the Sister Group Relation of Monotremes and Therians

Teun van Rheede,[1] Trijntje Bastiaans,* David N. Boone,† S. Blair Hedges,† Wilfried W. de Jong,*‡ and Ole Madsen**

*Department of Biochemistry, Radboud University Nijmegen, Nijmegen, The Netherlands; †Department of Biology, The Pennsylvania State University; and ‡Institute for Biodiversity and Ecosystem Dynamics, University of Amsterdam, Amsterdam, The Netherlands

Morphological data supports monotremes as the sister group of Theria (extant marsupials + eutherians), but phylogenetic analyses of 12 mitochondrial protein-coding genes have strongly supported the grouping of monotremes with marsupials: the Marsupionta hypothesis. Various nuclear genes tend to support Theria, but a comprehensive study of long concatenated sequences and broad taxon sampling is lacking. We therefore determined sequences from six nuclear genes and obtained additional sequences from the databases to create two large and independent nuclear data sets. One (data set I) emphasized taxon sampling and comprised five genes, with a concatenated length of 2,793 bp, from 21 species (two monotremes, six marsupials, nine placentals, and four outgroups). The other (data set II) emphasized gene sampling and comprised eight genes and three proteins, with a concatenated length of 10,773 bp or 3,669 amino acids, from five taxa (a monotreme, a marsupial, a rodent, human, and chicken). Both data sets were analyzed by parsimony, minimum evolution, maximum likelihood, and Bayesian methods using various models and data partitions. Data set I gave bootstrap support values for Theria between 55% and 100%, while support for Marsupionta was at most 12.3%. Taking base compositional bias into account generally increased the support for Theria. Data set II exclusively supported Theria, with the highest possible values and significantly rejected Marsupionta. Independent phylogenetic evidence in support of Theria was obtained from two single amino acid deletions and one insertion, while no supporting insertions and deletions were found for Marsupionta. On the basis of our data sets, the time of divergence between Monotremata and Theria was estimated at 231–217 MYA and between Marsupialia and Eutheria at 193–186 MYA. The morphological evidence for a basal position of Monotremata, well separated from Theria, is thus fully supported by the available molecular data from nuclear genes.

Introduction

In 1997, a News and Views feature in Nature announced "The platypus put in its place" (Penny and Hasegawa 1997). It commented on the exciting finding that sequence analyses of mitochondrial genomes strongly joined platypus with the marsupial lineage (Janke, Xu, and Arnason 1997). Such a sister group relation of monotremes and marsupials would restore a variant of the Marsupionta hypothesis of Gregory (1947) and contradict the traditional Theria hypothesis, which holds that monotremes diverged before the separation of marsupials and placentals. On the basis of additional mitochondrial and 18S rRNA sequences, Janke et al. (2002) concluded that "the currently unambiguous support for the Marsupionta hypothesis from molecular data can no longer be ignored." However, hardly any other substantial molecular data have been reported in support of the Marsupionta hypothesis. In fact, a recent reanalysis of mitochondrial genomes, in which nucleotide recoding (RY) and data partitioning was applied, favored Theria over Marsupionta (Phillips and Penny 2003). To definitively decide whether monotremes (e.g., the platypus) can stay in their position at the base of the mammal tree, it is important to find further decisive molecular evidence. Moreover, such an evaluation of the Marsupionta case is of broader interest for understanding how valid data and analyses can sometimes be positively misleading in phylogeny reconstruction.

Extant monotremes, the platypus and echidnas, are morphologically unique mammals found only in Australia and New Guinea. Their mosaic appearance, with reptilian (e.g., egg laying) and mammalian (e.g., mammary glands) characteristics, coined them as remnants of an ancestral stock of mammals, classified into their own subclass Prototheria, basal to Theria. The fossil record of monotremes is sparse, but recently discovered Mesozoic fossils provide considerable information about their evolution and indicate an ancient origin with support for a Prototheria-Theria resolution (e.g., Q. Ji, Luo, and S. A. Ji 1999; Luo, Crompton, and Sun 2001; see Musser 2003 for a review).

The original Marsupionta hypothesis, as proposed by Gregory (1947), placed monotremes within marsupials, as closest relatives of the Australian Diprotodontia. Additional support for Marsupionta was found by Kühne (1973), based on resemblances in the pattern of tooth replacements between marsupials and monotremes. Subsequent morphological analyses invalidated their arguments and marginalized the Marsupionta concept (Musser 2003 and references therein), until a series of papers based on protein-coding genes of complete mitochondrial genomes reanimated this hypothesis, with generally high statistical support (Janke et al. 1996, 2001, 2002; Janke, Xu, and Arnason 1997; Cao et al. 1998; Zardoya and Meyer 1998; Kumazawa and Nishida 1999; Nilsson et al. 2004). In contrast, only a few studies of nuclear genes favored Marsupionta (Kirsch and Mayer 1998; Toyosawa et al. 1998; Janke et al. 2002; Vernersson et al. 2002; Nowak et al. 2004), whereas the vast majority have supported Theria (e.g., Retief, Winkfein, and Dixon 1993; Kullander, Carlson, and Hallbook 1997; Messer et al. 1997; Killian et al. 2001; Miska et al. 2002; Vernersson et al. 2002; Belov and Hellman 2003; Miska, Hellman, and Miller 2003; Baker et al. 2004; Vernersson,

[1] Deceased May 21, 2003.

Key words: Marsupionta, Theria, phylogeny, base composition, divergence time.

E-mail: o.madsen@ncmls.ru.nl.

Mol. Biol. Evol. 23(3):587–597. 2006
doi:10.1093/molbev/msj064
Advance Access publication November 16, 2005

Aveskogh, and Hellman 2004). However, these studies provide variable support and are generally based on single genes or short sequences or suffer from limited taxon sampling and rooting with distant outgroups or paralogs. It is not clear in all cases whether the genes are orthologs and have the same functions in the different taxonomic groups, while phylogenetic analyses may not always be optimal (see Phillips and Penny 2003 for details). Molecular evidence for the phylogenetic position of the monotremes remains therefore ambiguous.

The usefulness of mitochondrial genomes in deeper phylogeny has been questioned (Curole and Kocker 1999; García-Moreno, Sorenson, and Mindell 2003), and nuclear genes perform better at this level (Springer et al. 2001). Analyses of large data sets of concatenated nuclear genes have recently indeed been used to resolve the major relationships among eutherians (e.g., Murphy et al. 2001) and estimate their divergence times (Springer et al. 2003). A similar approach is the logical way to elucidate the base of the mammalian tree and establish whether Marsupionta or Theria is eventually favored by molecular data. We therefore generated two large and independent data sets, one with a broad taxon sampling (21 taxa) and a shorter concatenated length (931 amino acids) and the other with only five taxa but a greater length (3,669 amino acids). Data set I consisted of mostly new sequences from five nuclear genes and comprised 17 mammalian ingroup and four outgroup taxa. Data set II was composed mainly of sequences retrieved in the databases, comprising eight nuclear genes and three proteins from four mammalian taxa and one outgroup. The two independent data sets were used for comprehensive phylogenetic and molecular dating analyses.

This is the first study of monotreme phylogeny applying data from many concatenated nuclear genes with different functions and using a broad taxon sampling in both in- and outgroups. The results unambiguously place monotremes in their traditional position at the base of the mammalian tree. Such a knowledge of the order and times of divergences is essential to understand major aspects of early mammalian evolution, such as the morphological transition from a reptilian to a mammal-like form and to determine the ancestral architecture of the mammalian genome, including the evolution of the sex-determining system (Grützner et al. 2004; Rens et al. 2004)

Materials and Methods

Genes and Taxa

Full species names, with taxonomic groupings, and accession numbers of all sequences used in this study are given in Supplementary Material tables 1 and 2 (Supplementary Material online) for data sets I and II, respectively. Data set I comprised 2,793 bp or 931 amino acids for 21 taxa, and data set II had a concatenated length of 10,773 bp or 3,669 amino acids. For data set I, we sequenced segments of five nuclear genes, coding for proteins with widely different functions, three G-protein coupled receptors: the acetylcholinergic receptor M4 (*chrm4*), dopamine receptor type 1A (*drd1a*), and α-2B adrenergic receptor (*adra2b*) and two transcription factors: the proto-oncogene C-MOS (*c-mos*) and sex-determining transcription factor SOX9 (*sox9*). Taxon sampling includes two monotremes, six marsupials from the orders Didelphimorpha, Paucituberculata, Microbiotheria, Dasyuromorphia, Peramelemorphia, and Diprotodontia (Graves and Westerman 2002) and nine eutherians from the orders Primates, Rodentia, Cetartiodactyla, Carnivora, Chiroptera, Eulipotyphla, Sirenia, Tubulidentata, and Xenarthra, representing the major superordinal eutherian clades (Murphy et al. 2001). As outgroup taxa, we obtained sequences from a lizard, a turtle, a crocodile, and a bird (see Supplementary Material table 1 for names, Supplementary Material online). Data set II consists mainly of sequences from the databases, with the following taxon sampling: Primates (*Homo sapiens*), Rodentia (*Mus musculus* or *Rattus norvegicus*), a representative of marsupials, a monotreme, and as outgroup the chicken (*Gallus gallus*). We could retrieve suitable nucleotide sequences for the following genes, *ldha*, *hprt*, *bdnf*, *nt3*, *ngfb*, *m6p/igf2r* receptor, and *rag1*, and amino acid sequences for insulin (*ins*), myoglobin (*mb*), and alpha lactalbumin (*lalba*). Other monotreme sequences were found in the databases but were excluded from our analysis because of a known or suspected evolutionary history of gene conversion, duplication, or concerted evolution, that is, α- and β-hemoglobin (Lee et al. 1999), amelogenin (Toyosawa et al. 1998), olfactory receptor (Glusman et al. 2000), immunoglobulins (e.g., Belov and Hellman 2003), and major histocompatibility cells (e.g., Miska et al. 2002), so that orthology of these genes cannot be established. Data set II also includes newly determined sequences for the αB-crystallin gene (*cryab*) for a marsupial and a monotreme and their human, rat, and chicken orthologs from the database.

Polymerase Chain Reaction Amplication and Sequencing

Amplification of segments of 500–1,200 bp was performed on the genes coding for *chrm4*, *c-mos*, *drd1a*, *sox9*, *adra2b*, and *cryab*. Primers (Supplementary Material table 3, Supplementary Material online) were based on alignments of known human, rat/mouse, and chicken sequences and additional tetrapod sequences when available. All polymerase chain reactions (PCRs) were performed with a polymerase mix (Expand HF system, Roche Molecular Biochemicals, Alameda, Calif.) and contained approximately 50–100 ng genomic DNA or ~10 ng cDNA (for *cryab* only; reverse trancribed with Superscript II RT, Invitrogen, Carlsbad, Calif.). PCRs were typically performed in 50 μl, with 20–100 pmol of primers. Gel-extracted PCR fragments (GFX PCR gel extraction kit, Amersham Biosciences, Piscataway, N.J.) were sequenced directly using Big Dye fluorescent technology on an ABI 3700 96-capillary sequencer.

Phylogenetic Analysis

Sequence data were assembled using the STADEN package programs PreGAP4 and GAP4 (http://www.staden.sourceforge.net). The χ^2 tests of compositional homogeneity of nucleotide data were performed with PAUP* 4.0b10 (Swofford 2003). Nucleotide and amino acid alignments were produced using ClustalW and adjusted manually. All alignments were inspected by eye for insertions and deletions (indels) that could provide support for

Table 1
Probability Scores (*P*) of χ^2 Test of Homogeneity on Different Codon Partitions for All Positions of Data Set I

		Codon Partition						
Taxon		123	1	2	12	3	123^{RY}	3^{RY}
Eutherians	$\chi^2(P)$	0.003	**>0.999**	**>0.999**	**0.999**	<0.001	**0.969**	**0.921**
Marsupials	$\chi^2(P)$	**0.976**	**>0.999**	**>0.999**	**>0.999**	**0.411**	**>0.999**	**0.942**
Monotremes	$\chi^2(P)$	**0.969**	**0.949**	**0.967**	**0.958**	**0.815**	**0.936**	**0.745**
Outgroups	$\chi^2(P)$	<0.001	**0.211**	**0.701**	**0.098**	<0.001	0.012	**0.781**
Mammals	$\chi^2(P)$	<0.001	**0.989**	**>0.999**	**0.954**	<0.001	**0.584**	**0.901**
Mammals + outgroups	$\chi^2(P)$	<0.001	0.002	**0.998**	0.001	<0.001	<0.001	**0.811**
Mammals + bird	$\chi^2(P)$	<0.001	**0.936**	**>0.999**	**0.828**	<0.001	**0.218**	**0.800**

NOTE.—Bold numbers indicate $\chi^2(P) > 0.05$.

Theria, Marsupionta, or a Monotremata-Eutheria clade. For phylogenetic analysis, ambiguous positions in the concatenated alignment were excluded. For some taxa in data sets I and II, chimeric concatenations of sequences from related species were used, as specified in Supplementary Material tables 1 and 2 (Supplementary Material online). Analysis of nucleotide data was performed using PAUP* 4.0b10 (Swofford 2003) on the following data partitions (see *Results* for details): all codon positions unweighted (partition 123); first and second codon positions unweighted

Table 2
Bootstrap Support Values and Bayesian Posterior Probabilities for Theria, Marsupionta, and Monotremata-Eutheria Based on Data Sets I and II

			Support					
Analyses	"Model"	Partition	Theria (node a)		Marsupionta		Monotremata-Eutheria	
Data set I								
MP		12	76.5	(98)	5.2	(1.5)	18.3	(0.1)
		123	75	(80)	1.8	(4.1)	23.2	(15.6)
		123^{RY}	87.6	(85.6)	3.3	(12.3)	9.1	(2.0)
		aa	89.1		5.7		5.2	
ME/NJ	Logdet + I	12	99.5	(100)	0.2	(0)	0.3	(0)
	Logdet + I	123	55	(100)	0	(0)	45	(0)
	Logdet + I	123^{RY}	97	(100)	0.1	(0)	2.9	(0)
	JTT	aa	98.4		0.8		0.8	
ML	GTR + Γ + I	12	95.4	(94.4)	2.6	(2.8)	2	(2.8)
	$HKY^{(TVM)}$ +Γ + I	123	76.8	(94.2)	1.6	(4.4)	21.6	(1.4)
	GTR + Γ + I	123^{RY}	94.8	(89.4)	2.2	(8.2)	3	(2.4)
	JTT	aa	99.2		0.6		0.2	
Bayes	Mixed	12	1.00	(1.00)	0	(0)	0	(0)
	Mixed	123	1.00	(1.00)	0	(0)	0	(0)
	Mixed	123^{RY}	1.00	(1.00)	0	(0)	0	(0)
	GTR + Γ + I	aa	1.00		0		0	
Data set II								
MP		12	100		0		0	
		123	100		0		0	
		123^{RY}	100		0		0	
		aa	100		0		0	
ME/NJ	Logdet + I	12	100		0		0	
	Logdet + I	123	100		0		0	
	Logdet + I	123^{RY}	100		0		0	
	JTT	aa	100		0		0	
ML	HKY + Γ + I	12	100		0		0	
	TrN + Γ + I	123	100		0		0	
	HKY + Γ + I	123^{RY}	100		0		0	
	JTT	aa	100		0		0	
Bayes	Mixed	12	1.00		0		0	
	Mixed	123	1.00		0		0	
	Mixed	123^{RY}	1.00		0		0	
	GTR + Γ + I	aa	1.00		0		0	

NOTE.—aa, amino acids. Analyses are: MP (maximum parsimony); ME/NJ, minimum evolution/neighbor joining; and ML (maximum likelihood). "Mixed" model indicates that each gene segment has its own optimal model of sequence evolution. Models and partition schemes are described in *Materials and Methods*. Support values without parentheses were obtained when using all four outgroups, whereas values in parentheses were obtained with bird only as outgroup (data set I, DNA only). Model parameters with superscript in parentheses indicate that deviating parameters were used with the bird only outgroup analyses.

Table 3
Statistical Results of Log-Likelihood Scores and KH and SH Tests to Compare the Three Possible Prior Hypotheses About Monotreme Relationships

Partition	Relationship	−ln L	Δ −ln L	P_{KH}	P_{SH}
Data set I					
12	Theria	9691.044 (8387.936)			
	Marsupionta	9699.670 (8396.339)	8.626 (8.403)	0.109 (0.133)	0.065 (0.069)
	Monetreme-Eutheria	9699.935 (8397.250)	8.891 (9.314)	0.101 (0.090)	0.052 (0.042*)
123	Theria	23625.937 (20143.473)			
	Marsupionta	23635.377 (20154.327)	9.440 (10.854)	0.072 (0.111)	0.044* (0.062)
	Monetreme-Eutheria	23631.223 (20155.121)	5.287 (11.648)	0.432 (0.066)	0.238 (0.039*)
123^{RY}	Theria	15180.136 (13076.344)			
	Marsupionta	15189.760 (13083.147)	9.624 (6.803)	0.066 (0.194)	0.039* (0.090)
	Monetreme-Eutheria	15189.565 (13084.173)	9.429 (7.829)	0.073 (0.107)	0.037* (0.054)
Data set II					
12	Theria	23119.320			
	Marsupionta	23339.394	214.074	<0.001*	<0.001*
	Monetreme-Eutheria	23448.122	328.802	<0.001*	<0.001*
123	Theria	44774.548			
	Marsupionta	45131.928	357.380	<0.001*	<0.001*
	Monetreme-Eutheria	45259.470	484.922	<0.001*	<0.001*
123^{RY}	Theria	32098.164			
	Marsupionta	32380.839	282.675	<0.001*	<0.001*
	Monetreme-Eutheria	32511.908	413.744	<0.001*	<0.001*

NOTE.—* significantly rejected ($P < 0.05$). Numbers in parentheses are with bird only as outgroup (data set I).

(partition 12); and first and second codon positions unweighted and third codon positions transversions only (123^{RY}). Phylogenetic criteria were: maximum parsimony, minimum evolution with LogDet + I distances, and maximum likelihood (ML) with a single model of sequence evolution for each of the above mentioned partitions, selected using the hierarchial likelihood ratio test or akaike information criterion in Modeltest 3.06/3.5 (Posada and Crandall 1998). In parsimony analyses stepwise additions with 100 random input orders of sequences were used, and in ML analyses a neighbor-joining (NJ) tree was used as starting tree. In all PAUP* analyses, the Tree Bisection-Reconnection option was used to swap branches. ML and NJ distance analyses of amino acid sequences was performed using the PROML, PROTDIST, NEIGHBOR, SEQBOOT, and CONSENSE programs of the PHYLIP 3.6a3 package (Felsenstein 2002) with the Jones, Taylor, and Thornton (JTT) model of sequence evolution (Jones, Taylor, and Thornton 1992). Bootstrapping included 1,000 replicates for parsimony and minimum evolution analyses and 500 replicates for ML and NJ amino acid analyses.

Bayesian phylogenetic analysis of both nucleotides and amino acids was performed with MrBayes 2.1/3.1.1 (Huelsenbeck and Ronquist 2001). For the nucleotide data sets and partitions mixed models were used, giving each gene its own best fitting model of sequence evolution as determined by Modeltest. For the amino acid data, a single model of sequence evolution was used for the concatenated data sets, namely, the general time reversible model (GTR + Γ_4 + I). A Metropolis-coupled Markov chain Monte Carlo sampling approach was used to calculate posterior probabilities with initial equal probabilities for all trees and random starting trees. Four Markov chains were run simultaneously two times for 1,000,000 generations to check if stationary posterior probabilities had been reached. Tree sampling was done each 20 generations, and burn-in values were determined from the likelihood values.

Statistical Tests

Kishino-Hasegawa (1989; KH) and Shimodaira-Hasegawa (1999; SH) statistical tests were performed in PAUP* 4.0b10 (Swofford 2003) with the resampling of estimated log likelihood optimization and 1,000 bootstrap replicates to evaluate the a priori hypotheses about monotreme relationships: Theria, Marsupionta, or a Monotremata-Eutheria clade. For each hypothesis the best ML tree and likelihood score were calculated and used for the statistical tests.

Time Estimation

For seven genes in data set II (*rag1*, *ngfb*, *ldha*, *nt3*, *m6p/igf2r*, *bdnf*, *hprt*), an actinopterygian fish sequence was available as outgroup (see Supplementary Material table 2, Supplementary Material online), resulting in an amino acid alignment of 1,662 sites and nucleotide alignments of 4,986 sites (all codon positions) and 3,324 sites (first and second codon positions). For this data set, only one reliable fossil calibration point is available, namely, the divergence of the lineages leading to birds and mammals, ~310 MYA (Hedges et al. 1996; Benton 2000; Hedges and Kumar 2004). We used a Bayesian method (Divtime5b and MULTIDIVTIME) (Kishino, Thorne, and Bruno 2001) for time estimations. Like in the Bayesian phylogenetic analyses, the nucleotide data were partitioned by gene with ML branch lengths calculated for each gene under a F84 + G model (Felsenstein 1984). For the amino acid data, branch lengths were calculated using a single JTT + G model (Jones, Taylor, and Thornton 1992). The means of the prior distributions ("priors") for the rate parameter and the root time (rt and t, respectively) were calculated for the nucleotide and amino acid data sets. The root prior was set to 310 MYA and was constrained with a lower and upper value of 310 and 370 MYA, respectively (Hedges and Kumar 2004; Reisz and Muller 2004). The

Markov chain Monte Carlo analyses were run for 1,000,000 generations with sampling each 100 generation after a "burnin" of 100,000 generations. Each analysis was run twice to test for consistency of results.

For data set I, no reliable fossil calibration points are present for the root or for any of the ingroup relationships. We therefore used molecular calibration point constraints for this data set, being well aware that such secondary calibration points should be selected and used with special care (for discussion see Graur and Martin 2004; Hedges and Kumar 2004). Three molecular calibration points were taken from Springer et al. (2003), who probably calculated the most reliable eutherian molecular time estimates available at present, being based on a relaxed molecular clock approach and an extensive data set in terms of length (~16,000 bp) and species diversity (42 species). Importantly, the three selected molecular constraints are independently supported by divergence time estimates based on mitochondrial amino acid data (3,392 amino acid) (Hasegawa, Thorne, and Kishino 2003). To account for the uncertainty of the molecular calibration points, we used in our analyses the 95% credibility intervals as lower and upper constraints. This resulted in the following constraints, as indicated in figure 2: 86–74 MYA for the earliest divergence of Afrotheria, 94–81 MYA for the earliest divergence of Euarchontoglires, and 90–80 MYA for the earliest divergence of Laurasiatheria. No real consensus for the age of the mammalian root (divergence of Theria and Monotremata) has been obtained from molecular data. We therefore used three different priors for the age of the root: (1) 237 MYA based on Woodburne, Rich, and Springer (2003); (2) 170 MYA representing the mean of 170 MYA (Belov, Hellman and Cooper 2002), 186–163 MYA (Messer et al. 1997), >167 MYA (Flynn et al. 1999) and 180–160 MYA (Phillips and Penny 2003); and (3) 140 MYA as the mean of >144 MYA (Lou et al. 2003) and 143–130 MYA (Janke, Xu, and Arnason 1997). We used this same procedure for time calculation with data set II.

Results

Base Compositional Heterogeneity of Data Sets

Variation in base composition often occurs between taxa and/or codon partitions and can result in erroneous phylogenetic trees (for review see Mooers and Holmes 2000). The support for Marsupionta from mitochondrial data is a good example of this problem because the same data support Theria if base compositional heterogeneity is taken into account (Phillips and Penny 2003). We therefore were cautious to assess the base compositions in our nuclear sequence data. The mean base frequencies of different taxon and codon partitions for data set I are presented in figure 1. At first and second codon positions the base frequencies are quite similar for the different taxon partitions, the greatest difference being observed for base A and C at first codon position in the outgroup partition. In contrast, the third codon position displays a pronounced compositional heterogeneity, with monotremes being notably low in A and T and high in C and G, while marsupials are high in A/T and low in C/G. As a consequence, the GC contents ranges from 90% in monotremes to 69% in marsupials. Conspicuous, too, is the generally high level of variation within eutherians and outgroups, as evident from the large standard deviations. The compositional heterogeneity at third codon positions, both between and within taxon partitions, vanishes after recoding to purines and pyrimidines (R and Y in fig. 1). This indicates that the third codon position heterogeneity is caused by a bias in both purines and pyrimidines, in contrast to the bias in whole mitochondrial genomes which is most prominent in pyrimidines (Phillips and Penny 2003; Gibson et al. 2005).

We also noticed that, among the outgroup taxa, birds were relatively G rich and reptiles relatively A rich at first and second codon position (data not shown). As a consequence, the base composition of birds is more similar to that of mammals than of reptiles (Supplementary Material table 4, Supplementary Material online). This observation is corroborated by χ^2 test for homogeneity of base composition (table 1). For the mammals + outgroups partition, the χ^2 test only supports homogeneity for second codon position and for third codon position recoded as transversions only (3^{RY}); when bird alone is used as outgroup, the only partitions failing the χ^2 test are the all (123) and third codon positions. However, it should be noted that χ^2 tests combining one of the other outgroups with mammals gave similar results in terms of passing the base homogeneity criterion, but with lower probability scores (data not shown).

Base compositional analyses of data set II yielded mutatis mutandis similar results. Considering these findings and to account for any possible effects of base compositional heterogeneity, we decided to use the following data partitions for phylogenetic analyses: deleting third codon positions (12), coding third codon position as purines and pyrimidines (123^{RY}), using all codon position (123), and using amino acids, and to use for data set I reptiles + bird as well as bird only as outgroup.

Phylogenetic Analyses and Support for Theria and Marsupionta

Phylogenetic analyses of data set I, applying a range of methods on different partitions and using reptiles + bird or bird only as outgroup, provided strong and consistent support for Theria (table 2). The support for Theria was in general lower if reptiles + bird were used as outgroup, especially under parsimony and distance criteria (table 2). In contrast, if base composition was taken into account (partition 12 and 123^{RY}), the support for Theria increased in ML analyses. In no case was there any meaningful support for Marsupionta, the highest bootstrap value being 12.3%, which is even lower than the support for the never-postulated Monotremata-Eutheria clade. Phylogenetic analysis of the much longer 5-taxon data set II were even more straigthforward and unequivocal: all types of analysis favor Theria with maximum support values (table 2).

To compare the three possible a priori hypotheses about monotreme relationships—the Theria, Marsupionta, and Monotremata-Eutheria hypotheses—we calculated log-likelihood scores and used KH and SH tests (table 3). With data set II, the Marsupionta and the Monotremata-Eutheria hypotheses could be significantly rejected with

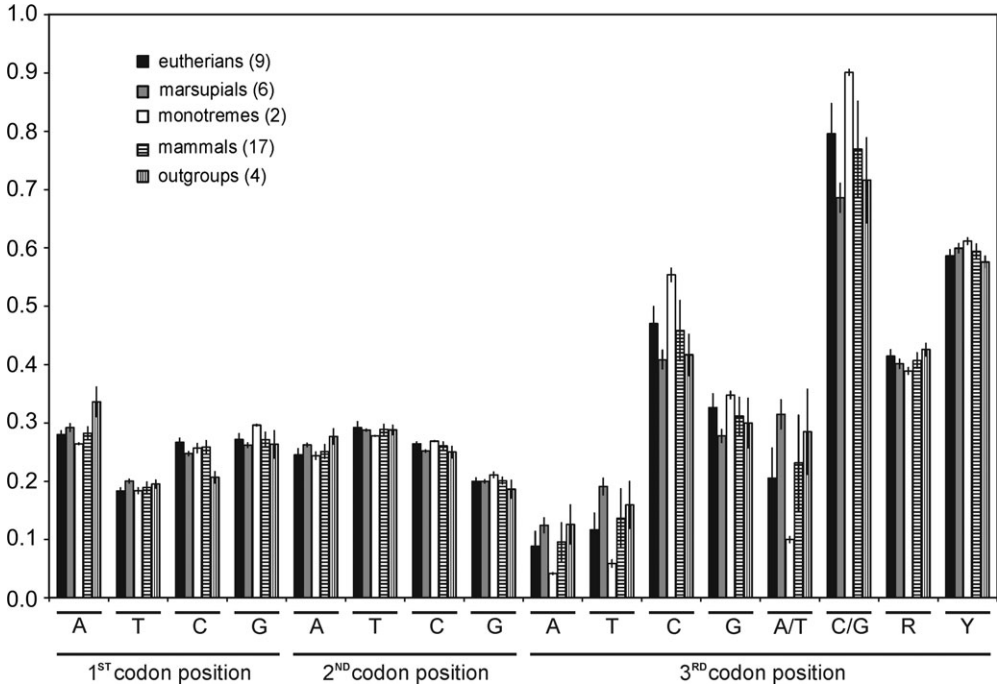

FIG. 1.—Mean base frequencies with standard deviation of different taxon and codon partitions of data set I. Numbers in parentheses indicate numbers of species in taxon partition. R and Y are the sum of purine (A, G) and pyrimidine (C, T) frequencies, respectively.

all data partitions ($P < 0.001$), whereas with data set I, the Marsupionta and Monotremata-Eutheria hypotheses could not always be rejected. However, for all partitions with data set I, Theria is the best solution, and the Marsupionta and Monotremata-Eutheria hypotheses are in general equally unlikely and close to significant rejection.

Figure 2 presents the ML tree obtained from data set I, partition 123, using reptiles + chicken as outgroup. It may be noticed that the eutherian species are correctly grouped into their respective basal clades: aardvark and manatee in Afrotheria; human and murids in Euarchontoglires; and seal, bat, pig, and eulipotyphlan insectivores in Laurasiatheria (Murphy et al. 2001). However, some relationships are deviating from well-supported molecular evidence such as, for example, the nesting of opossums within the other marsupials rather than being their outgroup and the topology inside Laurasiatheria (e.g., Amrine-Madsen et al. 2003; Murphy et al. 2001, respectively). But none of the deviating nodes is well supported, and features like long-branch attraction and inadequate species sampling may account for

Table 4
Estimates of Divergence Time (MYA ± standard error) Among Mammals from Data Sets I and II

Data Type	Model	Marsupials Versus Eutherians	Monotremes Versus Therians
Data set I			
DNA—all codons			
140 MYA[a]		185 ± 19 (165 ± 15)	195 ± 20 (185 ± 20)
170 MYA[a]	F84 + G	190 ± 20 (170 ± 16)	201 ± 21 (193 ± 21)
237 MYA[a]		200 ± 21 (181 ± 18)	212 ± 22 (207 ± 24)
DNA—first and second codon			
140 MYA		185 ± 24 (183 ± 22)	197 ± 26 (204 ± 27)
170 MYA	F84 + G	193 ± 26 (191 ± 23)	208 ± 28 (215 ± 29)
237 MYA		211 ± 29 (208 ± 27)	228 ± 32 (237 ± 35)
Amino acids			
140 MYA		189 ± 29 (191 ± 27)	220 ± 37 (223 ± 35)
170 MYA	JTT + G	195 ± 32 (201 ± 30)	230 ± 40 (238 ± 40)
237 MYA		211 ± 37 (219 ± 36)	252 ± 48 (264 ± 48)
Mean Data set I		193 ± 28	217 ± 31
Data set II			
DNA—all codons	F84 + G	197 ± 22	244 ± 21
DNA—first and second codon	F84 + G	175 ± 26	219 ± 26
Amino acids	JTT + G	187 ± 27	230 ± 24
Mean Data set II		186 ± 25	231 ± 24

NOTE.—Numbers in parentheses indicate when bird only is used as outgroup.
[a] Root prior.

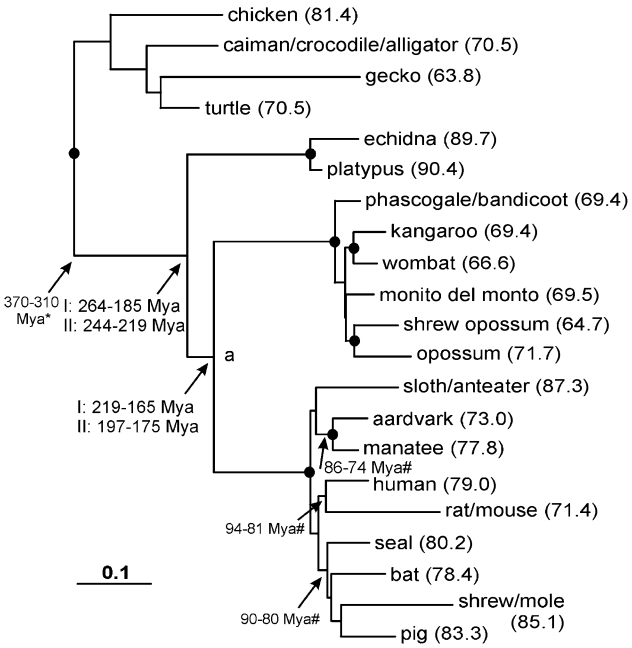

FIG. 2.—ML tree showing mammalian relationships as based on data set I, using all codon positions and a HKY + Γ_4 + I model of sequence evolution (ML partition 123 in table 2; −ln L = 23625.937). Branch lengths are proportional to evolutionary distance (bar = 0.1 base substitution per site). Numbers in parentheses indicate GC_3, which is highest in monotremes, lowest in marsupials, and intermediate but variable in eutherians and outgroups. Support values for node "a" (Theria) are given in table 2 and black circles indicate nodes with >75% bootstrap support and >0.95 posterior probabilities in ML or Bayesian analyses, respectively. The ranges of divergence times, as estimated from data sets I and II, are indicated for the three subclasses; the actual estimates are presented in table 4. The fossil calibration point constraint of 370–310 MYA (*) was used for data set II and the intraeutherian molecular-based calibration point constraints (#) for data set I.

these observations. Moreover, many sites that are informative within a mammalian subclass or superorder had to be scored as ambiguous and were thus removed from subsequent analyses when distant outgroup sequences were added to the alignment. This could also contribute to erroneous resolution within that subclass or superorder.

Indel Support for Theria

Indels can provide an additional source of phylogenetic information, independent of the base substitutions in a sequence data set. Indels in protein-coding sequences are more constrained than those in noncoding sequences and can be identified with reasonable certainty if they occur in conserved regions of proteins and provided that in- and outgroup taxon sampling is adequate (de Jong et al. 2003). We therefore searched all available protein-coding sequence data for the presence of indels that could be informative for resolving the trichotomy Monotremata-Marsupialia-Eutheria. Three single amino acid indels were found in support of Theria: one deletion in exon 3 of *sox9* and one deletion and one insertion in the M6P/IGF2 receptor (fig. 3). All three occur in sequence regions which can be aligned with confidence. No indels supporting the two possible alternative relationships could be retrieved.

Time Estimations

Molecular divergence time estimations have been considerably improved in recent years by introducing methods applying a relaxed molecular clock and allowing calibration constraints (e.g., a time window) instead of fixed calibration points (for reviews on time estimations see e.g., Bromham and Penny 2003; Hedges and Kumar 2003). Divergence times were estimated from data sets I and II using different partitions, methods, and models (table 4). Using data set I, the estimates for the marsupial/eutherian split showed less variation than the estimates for the monotreme/therian split, and the estimates including all codon positions were in general younger than those obtained on first and second codon positions and amino acid sequences. The opposite was true for data set II, where estimates based on all codon positions were the oldest. Such fluctuations in estimated divergence times are to be expected between different data sets and data types (nucleotides vs. amino acids). This is especially true for data sets with limited length (data set I) or number of taxa (data set II) and when the calibration points are few in number (only one in data set II) or "uncertain" (data set I, see *Materials and Methods*). Furthermore, the base heterogeneity at the third codon position (fig. 1) implies the possibility of saturation in the nucleotide data, which could influence the results and be one of the reasons for the observed differences. Nevertheless, the time estimates remain within reasonably narrow ranges.

If the mean time estimates are taken as the most likely divergence times, the molecular clock analyses indicate that marsupials diverged from eutherians in the Early Jurassic (193–186 MYA) and that monotremes diverged from therians in the Middle or Late Triassic (231–217 MYA).

Discussion

Base compositional bias is the reason that mitochondrial protein-coding genes support Marsupionta rather than Theria (Phillips and Penny 2003). We therefore paid special attention to the variation in base composition in our nuclear data sets. At first and second codon positions the base compositions were relatively homogeneous between different taxon partitions, but at third codon positions a pronounced heterogeneity was observed (fig. 1). Most notable is the strong bias in GC content at third codon position (GC_3), with monotremes being GC_3 rich (90%) and marsupials being GC_3 poor (69%), while eutherians (80%) and outgroups (72%) are intermediate. This nuclear GC_3 bias results from heterogeneity in both purines and pyrimidines, while the base compositional discrepancies among mammalian mitochondrial genomes are mainly caused by a skew in pyrimidines (Phillips and Penny 2003; Gibson et al. 2005). Such a difference might reflect that the base compositions of protein-coding genes in the nuclear and mitochondrial genomes of mammals are differentially affected by evolutionary constraints and processes.

GC richness has been related to the body temperatures of vertebrates. Warm-blooded birds and mammals have similar compositional heterogeneities in genomic isochore pattern, with some regions being very GC rich, in contrast to cold-blooded vertebrates which have a more homogeneous isochore pattern with lower average GC contents

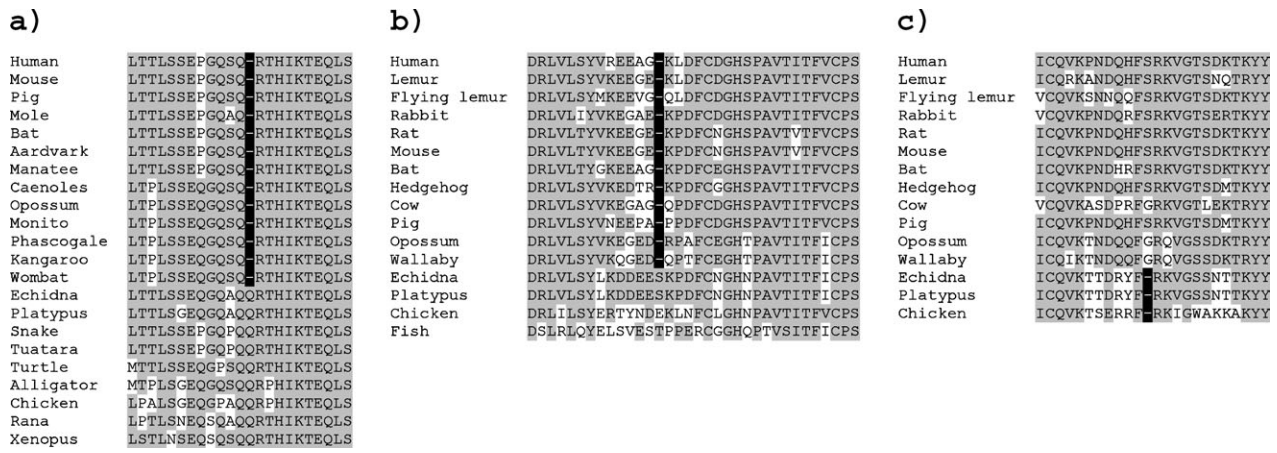

FIG. 3.—Support for Theria from indels in the deduced protein sequences of *sox9* (*a*) and *m6p/igf2r* (*b*, *c*). Gray shading shows majority consensus sequence with R = K, S = T, and D = E. Deletions are in black. Additional *sox9* sequences are consistent with the therian deletion: five primates (Patel et al. 2001), two bats, golden mole, hyrax, and four reptiles (our unpublished data).

(Bernardi 2000). A higher proportion of GC might be advantageous in homeothermic vertebrates by stabilizing DNA and RNA structures (e.g., G. Bernardi and G. Bernardi 1986; Clay et al. 2003; Jabbari, Clay, and Bernardi 2003). The greater similarity in base composition of the chicken genes in our data set to those of mammals than to those of the three poikilotherm reptiles (Supplementary Material table 4, Supplementary Material online) seems to agree with the relation between GC content and body temperature. However, it remains debated whether body temperature does indeed relate to genomic GC contents (Hughes, Zelus, and Mouchiroud 1999; Belle, Smith, and Eyre-Walker 2002; Ream, Johns, and Somero 2003), and doubt about the validity of the isochore concept is raising (e.g., Cohen et al. 2005). Also the high GC content of the monotreme genes in our study, as in the monotreme genome in general (Jabbari and Bernardi 2004; Margulies et al. 2005), does not correlate with body temperature; monotremes can regulate their temperature, but it never exceeds ~33°C, and thus remains lower than that of birds and eutherians.

Belle et al. (2004) recently suggested that the ancestral therian genome was GC rich and underwent major declines in GC contents in early eutherian and marsupial lineages and subsequent minor declines in primates, rodents, and carnivores. The high GC content in monotremes indicates that the ancestral mammalian genome was GC rich as well and that little or no decline in GC contents occurred in the monotreme lineage. The mechanisms responsible for the decrease in genomic GC content among therians are unknown but seem not to have been operating during monotreme evolution. In our data set, the GC_3 levels of primate, rodent, and carnivore genes are all in the lower eutherian range (fig. 2), supporting the minor decline in GC contents in these lineages (Belle et al. 2004). However, the GC_3 contents in manatee, aardvark, and bat are as low or lower, indicating that GC decline is more general among eutherians.

Bias in base composition is known to influence phylogenetic reconstruction, particularly when the model of evolution implicit in the applied method does not fit the data. In our phylogenetic analyses of data set I, the most noticeable differences in support were observed when partition 123 (all codon positions) was used with all four outgroups or with bird as only outgroup (table 2). With bird as outgroup, a much higher support for Theria was persistently obtained. Thus, as expected, base compositional noise introduced by outgroup taxa can have an effect on the statistical support of the deepest nodes. The greatest improvement between the two partitions, from 55% to 100% bootstrap support, was observed in distance analyses. This is quite surprising because we used the LogDet method with removal of invariable sites, which should be the most suitable for distance analyses when base heterogeneity is observed (Lockhart et al. 1996). Parsimony is known to be sensitive to bias in base composition (Eyre-Walker 1998), but with the most heterogeneous partition (123 and all outgroups) it performed better than the LogDet + I-based distance analysis (75% and 55% support, respectively). This indicates that LogDet has trouble dealing with the nonstationary base composition in our data set, as observed in various other cases (e.g., Waddell et al. 1999; Phillips and Penny 2003).

Despite the base heterogeneity in some partitions, we always recovered Theria, no matter which analyses were applied (table 2), indicating that the phylogenetic signal for the deeper mammalian nodes is higher than the noise in our data. This is in contrast to mitochondrial data which at the level of deeper eutherian and mammalian relationships often provide deviating tree topologies because of a systematic bias in base composition (e.g., Phillips and Penny 2003; Gibson et al. 2005). The better performance of nuclear genes than mitochondrial genes in deeper mammalian phylogeny (Springer et al. 2001) might well be due to a better fit of the commonly used models in phylogeny reconstruction to the ratio between phylogenetic signal and homoplastic noise in nuclear genes. Analyses of mitochondrial data in which better fitting models take base composition or convergent noise into account indeed manifest in general a better congruency to nuclear based trees (Phillips and Penny 2003; Gibson et al. 2005; Kitazoe et al. 2005).

The congruent outcome of all phylogenetic analyses on the large and independent data sets I and II is convincing

evidence for the Theria hypothesis and against Marsupionta. Additional independent support for Theria is provided by three indels that we detected in two unrelated genes (fig. 3). No indels in our data sets were found to support the alternative hypotheses. Also Phillips and Penny (2003) presented a highly conserved indel in the tRNA-serine (UCN), between the acceptor and D arms, in support of Theria. On the other hand, Janke et al. (2002) presented an insertion in the 18S rRNA gene of monotremes and marsupials as evidence for Marsupionta. However, in that case, the indel boundaries are not located at precisely the same position in placental mammals and outgroup species, hampering an unambiguous interpretation (Phillips and Penny 2003). It may be noticed that indels in noncoding DNA, such as the genes for tRNA and rRNA, are generally less constrained than indels in protein-coding genes as the latter must leave the reading frame intact. Indels in protein sequences are therefore phylogenetically more informative, although they, too, are certainly not free of homoplasy and ambiguity (de Jong et al. 2003). It is the finding of three independent indels in the *sox9* and *m6p/igf2* genes (fig. 3), together with the regular phylogenetic analyses, that makes the support for Theria from our data sets compelling.

Although the branching order of the three mammalian subclasses is now well established, the times of their divergences remain for the time being a matter of further investigation. We estimate from our data that marsupials and eutherians diverged in the Early Jurassic (193–186 MYA), while monotremes separated from therians in the Middle or Late Triassic (231–217 MYA). Our dating of the marsupial-eutherian divergence is concordant with other molecular estimations based on nuclear data (173 MYA, Kumar and Hedges 1998; 190–182 MYA, Woodburne, Rich, and Springer 2003) and mitochondrial data (176 MYA, Penny et al. 1999). Paleontological evidence places the marsupial-eutherian divergence at 125 MYA or earlier if based on the age of the oldest eutherian (Ji et al. 2002) and oldest marsupial (Lou et al. 2003). However, some recent fossil analyses would bring this split back to at least 167 MYA (Flynn et al. 1999; Woodburne, Rich, and Springer 2003), coming close to the molecular date. The divergence of marsupials from eutherians, in the Early or Middle Jurassic, corresponds roughly with the initial breakup of Pangaea into Laurasia and Gondwana. Although this raises the possibility of an origin by vicariance, the fossil record of mammals at this early time period is currently too fragmentary to either support or refute that hypothesis.

Our estimate for the divergence time of theria and monotremes (231–217 MYA) is older than most other molecular estimates, which are around 170 MYA (Belov, Hellman, and Cooper 2002), 183–163 MYA (Messer et al. 1997), and 180–160 MYA (Phillips and Penny 2003), but is close to a recent dating of 236 MYA by Woodburne, Rich, and Springer (2003). The oldest established monotreme fossil is from the early Cretaceous (~110–120 MYA, Archer et al. 1985; Rich et al. 2001), but late Triassic/early Jurassic fragmentated fossil remains (~205 MYA) may have monotreme affinities (Musser 2003). Such an earlier appearance of monotremes would be compatible with our datings and those of Woodburne, Rich, and Springer (2003). Phillips and Penny (2003) suggested on basis of their analyses of mitochondrial data a very short time period between the monotreme/therian split and the marsupial/eutherian split. Our datings, in agreement with those of Woodburne, Rich, and Springer (2003), would rather indicate a considerable time span of about 24–45 Myr between the two successive divergence events.

The brief and transient revival of the Marsupionta hypothesis has stimulated the molecular analysis of early mammal phylogeny and has provided further insights in the problems and pitfalls of molecular phylogeny. Rejection of Marsupionta avoids the unlikely scenarios that would have been required to explain the reptilian characteristics of monotremes, from egg laying, bone, and sperm morphology to the recently discovered sex-determining system (Grützner et al. 2004; Rens et al. 2004). But many challenges remain to fully understand the morphological transition from a reptile to a mammal-like form and in determining the ancestral architecture of the mammalian genome.

Supplementary Material

Supplementary material tables 1–4 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

W.W. de J. was supported by grants from the Netherlands Organization for Scientific Research (NWO-ALW 805.41.044) and the European Commission (CT98-0221) and S.B.H. by grants from the U.S. National Aeronautics and Space Administration (NCC2-1057 and NNA04CC06A) and National Science Foundation (DBI-0112670). We thank Jaime E. Blair for assistance with analyses and a reviewer and Mark Springer for useful comments on a previous version of the paper.

Literature Cited

Amrine-Madsen, H., M. Scally, M. Westerman, M. J. Stanhope, C. Krajewski, and M. S. Springer. 2003. Nuclear gene sequences provide evidence for the monophyly of australidelphian marsupials. Mol. Phylogenet. Evol. **28**:186–196.

Archer, M., T. F. Flannery, A. Ritchie, and R. E. Molnar. 1985. First Mesozoic mammal from Australia—an early Cretaceous monotreme. Nature **318**:363–366.

Baker, M. L., J. P. Wares, G. A. Harrison, and R. D. Miller. 2004. Relationships among the families and orders of marsupials and the major mammalian lineages based on recombination activating gene-1. J. Mamm. Evol. **11**:1–16.

Belle, E. M., L. Duret, N. Galtier, and A. Eyre-Walker. 2004. The decline of isochores in mammals: an assessment of the GC content variation along the mammalian phylogeny. J. Mol. Evol. **58**:653–660.

Belle, E. M., N. Smith, and A. Eyre-Walker. 2002. Analysis of the phylogenetic distribution of isochores in vertebrates and a test of the thermal stability hypothesis. J. Mol. Evol. **55**:356–363.

Belov, K., and L. Hellman. 2003. Immunoglobulin genetics of *Ornithorhynchus anatinus* (platypus) and *Tachyglossus aculeatus* (short-beaked echidna). Comp. Biochem. Physiol. A Mol. Integr. Physiol. **136**:811–819.

Belov, K., L. Hellman, and D. W. Cooper. 2002. Characterisation of echidna IgM provides insights into the time of divergence of extant mammals. Dev. Comp. Immunol. **26**:831–839.

Benton, M. J. 2000. Vertebrate palaeontology. Blackwell Science, Oxford.
Bernardi, G. 2000. The compositional evolution of vertebrate genomes. Gene **259**:31–43.
Bernardi, G., and G., Bernardi. 1986. Compositional constraints and genome evolution. J. Mol. Evol. **24**:1–11.
Bromham, L., and D. Penny. 2003. The modern molecular clock. Nat. Rev. Genet. **4**:216–224.
Cao, Y., P. J. Waddell, N. Okada, and M. Hasegawa. 1998. The complete mitochondrial DNA sequence of the shark *Mustelus manazo*: evaluating rooting contradictions to living bony vertebrates. Mol. Biol. Evol. **15**:1637–1646.
Clay, O., S. Arhondakis, G. D'Onofrio, and G. Bernardi. 2003. LDH-A and alpha-actin as tools to assess the effects of temperature on the vertebrate genome: some problems. Gene **317**:157–160.
Cohen, N., T. Dagan, L. Stone, and D. Graur. 2005. GC composition of the human genome: in search of isochores. Mol. Biol. Evol. **22**:1260–1272.
Curole, J. P., and T. D. Kocher. 1999. Mitogenomics: digging deeper with complete mitochondrial genomes. Trends Ecol. Evol. **14**:394–398.
de Jong, W. W., M. A. van Dijk, C. Poux, G. Kappe, T. van Rheede, and O. Madsen. 2003. Indels in protein-coding sequences of Euarchontoglires constrain the rooting of the eutherian tree. Mol. Phylogenet. Evol. **28**:328–340.
Eyre-Walker, A. 1998. Problems with parsimony in sequences of biased base composition. J. Mol. Evol. **47**:686–690.
Felsenstein, J. 1984. Distance methods for inferring phylogenies: a justification. Evolution **38**:16–24.
———. 2002. PHYLIP (phylogeny inference package) version 3.6a3. Department of Genetics, University of Washington, Seattle.
Flynn, J. J., J. M. Parrish, B. Rakotosamimanana, W. F. Simpson, and A. R. Wyss. 1999. A middle Jurassic mammal from Madagascar. Nature **401**:57–60.
García-Moreno, J., M. D. Sorenson, and D. P. Mindell. 2003. Congruent avian phylogenies inferred from mitochondrial and nuclear DNA sequences. J. Mol. Evol. **57**:27–37.
Gibson, A., V. Gowri-Shankar, P. G. Higgs, and M. Rattray. 2005. A comprehensive analysis of mammalian mitochondrial genome base composition and improved phylogenetic methods. Mol. Biol. Evol. **22**:251–264.
Glusman, G., A. Bahar, D. Sharon, Y. Pilpel, J. White, and D. Lancet. 2000. The olfactory receptor gene superfamily: data mining, classification, and nomenclature. Mamm. Genome **11**:1016–1023.
Graur, D., and W. Martin. 2004. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. Trends Genet. **20**:80–86.
Graves, J. A., and M. Westerman. 2002. Marsupial genetics and genomics. Trends Genet. **18**:517–521.
Gregory, W. K. 1947. The monotremes and the palimpsest theory. Bull. Am. Mus. Nat. Hist. **88**:1–52.
Grützner, F., W. Rens, E. Tsend-Ayush, N. El-Mogharbel, P. C. O'Brien, R. C. Jones, M. A. Ferguson-Smith, and J. A. Marshall Graves. 2004. In the platypus a meiotic chain of ten sex chromosomes shares genes with the bird Z and mammal X chromosomes. Nature **432**:913–917.
Hasegawa, M., J. L. Thorne, and H. Kishino. 2003. Time scale of eutherian evolution estimated without assuming a constant rate of molecular evolution. Genes Genet. Syst. **78**:267–283.
Hedges, S. B., and S. Kumar. 2003. Genomic clocks and evolutionary timescales. Trends Genet. **19**:200–206.
———. 2004. Precision of molecular time estimates. Trends Genet. **20**:242–247.
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.
Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics **17**:754–755.
Hughes, S., D. Zelus, and D. Mouchiroud. 1999. Warm-blooded isochore structure in Nile crocodile and turtle. Mol. Biol. Evol. **16**:1521–1527.
Jabbari, K., and G. Bernardi. 2004. Cytosine methylation and CpG, TpG (CpA) and TpA frequencies. Gene **333**:143–149.
Jabbari, K., O. Clay, and G. Bernardi. 2003. GC3 heterogeneity and body temperature in vertebrates. Gene **317**:161–163.
Janke, A., D. Erpenbeck, M. Nilsson, and U. Arnason. 2001. The mitochondrial genomes of the iguana (*Iguana iguana*) and the caiman (*Caiman crocodylus*): implications for amniote phylogeny. Proc. R. Soc. Lond. B Biol. Sci. **268**:623–631.
Janke, A., N. J. Gemmell, G. Feldmaier-Fuchs, A. von Haeseler, and S. Paabo. 1996. The mitochondrial genome of a monotreme—the platypus (*Ornithorhynchus anatinus*). J. Mol. Evol. **42**:153–159.
Janke, A., O. Magnell, G. Wieczorek, M. Westerman, and U. Arnason. 2002. Phylogenetic analysis of 18S rRNA and the mitochondrial genomes of the wombat, *Vombatus ursinus*, and the spiny anteater, *Tachyglossus aculeatus*: increased support for the Marsupionta hypothesis. J. Mol. Evol. **54**:71–80.
Janke, A., X. Xu, and U. Arnason. 1997. The complete mitochondrial genome of the wallaroo (*Macropus robustus*) and the phylogenetic relationship among Monotremata, Marsupialia, and Eutheria. Proc. Natl. Acad. Sci. USA **94**:1276–1281.
Ji, Q., Z. X. Luo, and S. A. Ji. 1999. A Chinese triconodont mammal and mosaic evolution of the mammalian skeleton. Nature. **398**:326–330.
Ji, Q., X. Luo, C. X. Yuan, J. R. Wible, J. P. Zhang, and J. A. Georgi. 2002. The earliest known placental mammal. Nature **416**:816–822.
Jones, D. T., W. R. Taylor, and J. M. Thornton. 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Appl. Biosci. **8**:275–282.
Killian, J. K., T. R. Buckley, N. Stewart, B. L. Munday, and R. J. Jirtle. 2001. Marsupials and Eutherians reunited: genetic evidence for the Theria hypothesis of mammalian evolution. Mamm. Genome **12**:513–517.
Kirsch, J. A., and G. C. Mayer. 1998. The platypus is not a rodent: DNA hybridization, amniote phylogeny and the palimpsest theory. Phil. Trans. R. Soc. Lond. B Biol. Sci. **353**:1221–1237.
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.
Kishino, H., J. L. Thorne, and W. J. Bruno. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. Mol. Biol. Evol. **18**:352–361.
Kitazoe, Y., H. Kishino, T. Okabayashi, T. Watabe, N. Nakajima, Y. Okuhara, and Y. Kurihara. 2005. Multidimensional vector space representation for convergent evolution and molecular phylogeny. Mol. Biol. Evol. **22**:704–715.
Kühne, W. G. 1973. The systematic position of monotremes reconsidered (Mammalia). Z. Morphol. Tiere **75**:59–64.
Kullander, K., B. Carlson, and F. Hallbook. 1997. Molecular phylogeny and evolution of the neurotrophins from monotremes and marsupials. J. Mol. Evol. **45**:311–321.
Kumar, S., and S. B. Hedges. 1998. A molecular timescale for vertebrate evolution. Nature **392**:917–920.
Kumazawa, Y., and M. Nishida. 1999. Complete mitochondrial DNA sequences of the green turtle and blue-tailed mole skink: statistical evidence for archosaurian affinity of turtles. Mol. Biol. Evol. **16**:784–792.
Lee, M. H., R. Shroff, S. J. Cooper, and R. Hope. 1999. Evolution and molecular characterization of a beta-globin gene from the

Australian Echidna *Tachyglossus aculeatus* (Monotremata). Mol. Phylogenet. Evol. **12**:205–214.

Lockhart, P. J., A. W. Larkum, M. Steel, P. J. Waddell, and D. Penny. 1996. Evolution of chlorophyll and bacteriochlorophyll: the problem of invariant sites in sequence analysis. Proc. Natl. Acad. Sci. USA **93**:1930–1934.

Luo, Z. X., A. W. Crompton, and A. L. Sun. 2001. A new mammaliaform from the early Jurassic and evolution of mammalian characteristics. Science **292**:1535–1540.

Luo, Z. X., Q. Ji, J. R. Wible, and C. X. Yuan. 2003. An Early Cretaceous tribosphenic mammal and metatherian evolution. Science **302**:1934–1940.

Margulies, E. H., V. V. Maduro, P. J. Thomas, J. P. Tomkins, C. T. Amemiya, M. Luo, E. D. Green, and NISC Comparative Sequencing Program. 2005. Comparative sequencing provides insights about the structure and conservation of marsupial and monotreme genomes. Proc. Natl. Acad. Sci. USA **102**:3354–3359.

Messer, M., M. Griffiths, P. D. Rismiller, and D. C. Shaw. 1997. Evolution of the monotremes: phylogenetic relationship to marsupials and eutherians, and estimation of divergence dates based on α-lactalbumin amino acid sequences. Comp. Biochem. Physiol. B Biochem. Mol. Biol. **118**:403–410.

Miska, K. B., G. A. Harrison, L. Hellman, and R. D. Miller. 2002. The major histocompatibility complex in monotremes: an analysis of the evolution of Mhc class I genes across all three mammalian subclasses. Immunogenetics **54**:381–393.

Miska, K. B., L. Hellman, and R. D. Miller. 2003. Characterization of β_2-microglobulin coding sequence from three non-placental mammals: the duckbill platypus, the short-beaked echidna, and the grey short-tailed opossum. Dev. Comp. Immunol. **27**:247–256.

Mooers, A. O., and E. C. Holmes. 2000. The evolution of base composition and phylogenetic inference. Trends Ecol. Evol. **15**:365–369.

Murphy, W. J., E. Eizirik, S. J. O'Brien. et al. (11 co-authors). 2001. Resolution of the early placental mammal radiation using Bayesian phylogenetics. Science **294**:2348–2351.

Musser, A. M. 2003. Review of the monotreme fossil record and comparison of palaeontological and molecular data. Comp. Biochem. Physiol. A Mol. Integr. Physiol. **136**:927–942.

Nilsson, M. A., U. Arnason, P. B. Spencer, and A. Janke. 2004. Marsupial relationships and a timeline for marsupial radiation in South Gondwana. Gene **340**:189–196.

Nowak, M. A., Z. E. Parra, L. Hellman, and R. D. Miller. 2004. The complexity of expressed kappa light chains in egg-laying mammals. Immunogenetics **56**:555–563.

Patel, M., K. S. Dorman, Y. H. Zhang, B. L. Huang, A. P. Arnold, J. S. Sinsheimer, E. Vilain, and E. R. McCabe. 2001. Primate DAX1, SRY, and SOX9: evolutionary stratification of sex-determination pathway. Am. J. Hum. Genet. **68**:275–280.

Penny, D., and M. Hasegawa. 1997. Molecular systematics. The platypus put in its place. Nature **387**:549–550.

Penny, D., M. Hasegawa, P. J. Waddell, and M. D. Hendy. 1999. Mammalian evolution: timing and implications from using the LogDeterminant transform for proteins of differing amino acid composition. Syst. Biol. **48**:76–93.

Phillips, M. J., and D. Penny. 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. Mol. Phylogenet. Evol. **28**:171–185.

Posada, D., and K. A. Crandall. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14**:817–818.

Ream, R. A., G. C. Johns, and G. N. Somero. 2003. Base compositions of genes encoding alpha-actin and lactate dehydrogenase-A from differently adapted vertebrates show no temperature-adaptive variation in G + C content. Mol. Biol. Evol. **20**:105–110.

Reisz, R. R., and J. Muller. 2004. Molecular timescales and the fossil record: a paleontological perspective. Trends Genet. **20**:237–241.

Rens, W., F. Grutzner, P. C. O'Brien, H. Fairclough, J. A. Graves, and M. A. Ferguson-Smith. 2004. Resolution and evolution of the duck-billed platypus karyotype with an X1Y1X2Y2X3-Y3X4Y4X5Y5 male sex chromosome constitution. Proc. Natl. Acad. Sci. USA. **101**:16257–16261.

Retief, J. D., R. J. Winkfein, and G. D. Dixon. 1993. Evolution of the monotremes. The sequences of the protamine P1 genes of platypus and echidna. Eur. J. Biochem. **218**:457–461.

Rich, T. H., P. Vickers-Rich, P. Trusler, T. F. Flannery, R. L. Cifelli, and A. Constantine. 2001. Monotreme nature of the Australian Early Cretaceous mammal *Teinolophos trusleri*. Acta Palaeontol. Pol. **46**:113–118.

Shimodaira, H., and M. Hasegawa. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. **16**:1114–1116.

Springer, M. S., R. W. DeBry, C. Douady, H. M. Amrine, O. Madsen, W. W. de Jong, and M. J. Stanhope. 2001. Mitochondrial versus nuclear gene sequences in deep-level mammalian phylogeny reconstruction. Mol. Biol. Evol. **18**:132–143.

Springer, M. S., W. J. Murphy, E. Eizirik, and S. J. O'Brien. 2003. Placental mammal diversification and the Cretaceous-Tertiary boundary. Proc. Natl. Acad. Sci. USA **100**:1056–1061.

Swofford, D. L. 2003. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4.0b10. Sinauer Associates, Sunderland, Mass.

Toyosawa, S., C. O'hUigin, F. Figueroa, H. Tichy, and J. Klein. 1998. Identification and characterization of amelogenin genes in monotremes, reptiles, and amphibians. Proc. Natl. Acad. Sci. USA **95**:13056–13061.

Vernersson, M., M. Aveskogh, and L. Hellman. 2004. Cloning of IgE from the echidna (*Tachyglossus aculeatus*) and a comparative analysis of epsilon chains from all three extant mammalian lineages. Dev. Comp. Immunol. **28**:61–75.

Vernersson, M., M. Aveskogh, B. Munday, and L. Hellman. 2002. Evidence for an early appearance of modern post-switch immunoglobulin isotypes in mammalian evolution (II); cloning of IgE, IgG1 and IgG2 from a monotreme, the duck-billed platypus, *Ornithorhynchus anatinus*. Eur. J. Immunol. **32**:2145–2155.

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.

Woodburne, M. O., T. H. Rich, and M. S. Springer. 2003. The evolution of tribospheny and the antiquity of mammalian clades. Mol. Phylogenet. Evol. **28**:360–385.

Zardoya, R., and A. Meyer. 1998. Complete mitochondrial genome suggests diapsid affinities of turtles. Proc. Natl. Acad. Sci. USA **95**:14226–14231.

Yang, Z. 1997. PAML: a program for package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13**:555–556.

Mark Springer, Associate Editor

Accepted November 14, 2005