ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Molecular Phylogenetics and Evolution 28 (2003) 171–185

MOLECULAR PHYLOGENETICS AND EVOLUTION

www.elsevier.com/locate/ympev

The root of the mammalian tree inferred from whole mitochondrial genomes

Matthew J. Phillips* and David Penny

Allan Wilson Center for Molecular Ecology and Evolution, Institute of Molecular BioSciences, P.O. Box 11222, Massey University, Palmerston North, New Zealand

Received 8 October 2002; revised 12 December 2002

Abstract

Morphological and molecular data are currently contradictory over the position of monotremes with respect to marsupial and placental mammals. As part of a re-evaluation of both forms of data we examine complete mitochondrial genomes in more detail. There is a particularly large discrepancy in the frequencies of thymine and cytosine (T–C) between mitochondrial genomes that appears to affect some deep divergences in the mammalian tree. We report that recoding nucleotides to RY-characters, and partitioning maximum-likelihood analyses among subsets of data reduces such biases, and improves the fit of models to the data, respectively. RY-coding also increases the signal on the internal branches relative to external, and thus increases the phylogenetic signal. In contrast to previous analyses of mitochondrial data, our analyses favor Theria (marsupials plus placentals) over Marsupionta (monotremes plus marsupials). However, a short therian stem lineage is inferred, which is at variance with the traditionally deep placement of monotremes on morphological data.

Keywords: Marsupionta; Mitochondrial genomes; Monotremes; Partitioned likelihood; Pyrimidine bias; RY-coding; Theria

1. Introduction

The relationship of monotremes (platypus and echidnas) to other mammals has a long history of uncertainty. The comparison of modern mammals with their Mesozoic relatives (for example, Hu et al., 1997; Lou et al., 2002; Marshall, 1979; Rowe, 1988) has provided much evidence for monotreme affinities lying outside the therian crown group (metatherians, including marsupials and eutherians, including placentals). Studies of reproduction (e.g., Carrick and Hughes, 1982; Renfree, 1993) and soft tissue/cytology (Griffiths, 1978; Selwood, 1994; Tsujii et al., 1992) also provide substantial support for excluding monotremes from Theria.

Prior to the discovery of near-complete Mesozoic mammalian fossils during the latter half of the 20th century, Gregory (1947) provided an alternative interpretation of monotreme relationships. Gregory relied on his palimpsest theory, for which monotreme characters were considered as either adaptive for the current niche (and largely ignored), or as reflective of phylogenetic and past adaptive history (and used for studying relationships). He inferred that monotreme affinities lay within marsupials, his Marsupionta hypothesis. Kühne (1973) advocated Marsupionta, citing similarities in tooth replacement between marsupials and platypuses (ancestrally at least). However, Luckett and Zeller (1989) showed that Kühne (1973) misinterpreted the dental series, while Parrington (1974) argues that most of the similarities between marsupials and monotremes that Gregory cited are, in fact, primitive for Mammalia.

Janke et al. (1996) revived a version of the Marsupionta hypothesis, finding that the platypus grouped with the Virginia opossum to the exclusion of eutherians on the basis of analysis of 12 mitochondrial (mt) protein-coding genes. Addition of another marsupial (Wallaroo, Janke et al., 1997) and another monotreme (short-beaked echidna, Janke et al., 2002) still finds the monotreme/marsupial grouping. Results from other phylogenetic analyses of complete mt-genomes

* Corresponding author. Fax: +64-6-350-5626.
E-mail address: m.j.phillips@massey.ac.nz (M.J. Phillips).

1055-7903/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S1055-7903(03)00057-5

employing an array of analytical methods (e.g., Kumazawa et al., 1998; Penny and Hasegawa, 1997; Zardoya and Meyer, 1998) are similar, with bootstrap support often above 95%. However, as noted by Janke et al. (1996), strong support for Marsupionta is limited to the protein-coding genes. Phylogenetic analyses of the mitochondrial (mt) transfer RNA (tRNA) and/or ribosomal RNA (rRNA) sequences find reduced support for Marsupionta (e.g., Waddell et al., 1999; Zardoya and Meyer, 2001), or even weak support for Theria (Gemmell and Westerman, 1994). Janke et al. (2002) found some support for Marsupionta in nuclear-encoded 18S rRNA genes. However, this largely relied on a putative insertion shared by therians, but more than one indel is required because the putative insert sequences are of different length in the marsupials and the platypus. The data is explained equally well by an insertion in stem mammals with a subsequent deletion along the placental stem lineage.

Most of the nuclear genes for which monotreme sequences are available favor Theria over Marsupionta, although with limited statistical support. These genes include various globins (e.g., Lee et al., 1999; McKenna, 1987), protamine P1 (Retief et al., 1993), the neurotrophins BDNF, NT-3, and NGF (Kullander et al., 1997), and immunoglobulin γ1 (Belov et al., 2002). Gilbert and Labuda (2000) report a SINE family (Ther-2) to be present in marsupials and placentals, but not in monotremes. However, this requires further investigation because Southern analysis only showed substantial hybridization of the probe to the marsupial genomic samples and Ther-2 in placentals was only confirmed by searches through genomic regions that have not been sequenced in monotremes. Using SINEs reliably requires the flanking regions to be sequenced in order to show that the insertions are equivalent (see Shimamura et al., 1997).

Analysis of α-lactalbumin (LA) that involved rooting with lysozyme (LZ) sequences provided the first robust support for Theria based on a molecular sequence (Messer et al., 1998). Nonetheless, the 100% parsimony bootstrap support that Messer et al. (1998) found for Theria is anomalous in that the sequences are only a little over 120 amino acid residues long and average almost one change per site between monotremes and the therians (even before taking constant sites into account). Even placental monophyly, which is strongly supported with longer sequences was only resolved for in 63% of bootstrap replicates. The possibility that the monotreme and therian LAs are paralogous deserves attention. Nevertheless, a period of adaptive LA evolution in stem therians may be a simpler explanation for a much stronger than expected phylogenetic signal for Theria.

Killian et al. (2001) analyzed the $\approx$2250 amino acid residues of the mannose 6-phosphate/insulin-like growth factor II receptor (MP6/IGF2R) gene for 15 mammals and a chicken. Theria gained $\geqslant 97\%$ bootstrap support in each parsimony (MP), minimum-evolution (ME) and maximum-likelihood (ML) analysis, and the inter-ordinal relationships within placentals were resolved as in recent studies of nuclear genes (Madsen et al., 2001; Murphy et al., 2001a). This is the strongest molecular evidence for Theria yet. However, Killian et al. (2000) had earlier shown that in marsupials and placentals (but not in monotremes and chicken), MP6/IGF2R binds IGF2 and is imprinted (expression depends on the parent of origin) so there could be differences in functional constraints on the evolution of the proteins.

The growing number of studies of independent (unlinked) genes that favor Theria has not been matched for Marsupionta. Toyosawa et al. (1998) analyzed the enamel matrix protein amelogenin ($\approx$200 amino acid residues) and did find that monotremes, and the only marsupial in the study (*Didelphis*), grouped together weakly in an uncorrected distance NJ (neighbor-joining) analysis. Kirsch and Mayer (1998) found some support for Marsupionta from DNA to DNA hybridization data, though with a lizard (rather than placentals) as sister to Marsupionta in the better sampled tree, the value of these data at deep levels is questionable. The mt protein-coding sequences remain the only molecular data that have provided statistically robust support for Marsupionta.

Phillips et al. (2001) suggest the possibility that composition bias may be contributing to the mt signal supporting Marsupionta. Bias resulting from differences in the relative frequency of the two pyrimidines (T–C bias, = pyrimidine bias) favored a sister-relationship between the bandicoot, *Isoodon,* and the opossum, *Didelphis*. Both of these mt-genomes have a predominance of T (as opposed to C) relative to the other mammals. This bias was most extreme at the 3rd codon positions (PTN3) and with these included, a bandicoot-opossum clade is weakly favored (unpublished data) over the bandicoot associating with the other australidelphian taxa (which was favored using the more conservative PTN12 and RNA data). The platypus had the next highest thymine content. It is therefore important to determine whether nucleotide composition bias might contribute to the signal for Marsupionta. If there is a strong bias which does not favor the correct phylogeny, then most phylogenetic reconstruction methods are not expected to select the correct tree (see Foster et al., 1997; Lockhart et al., 1994; Moores and Holmes, 2000). In this study, phylogenetic signal for monotreme placement is compared between standard methods, and two methods that are expected to be robust to pyrimidine bias. These are, LogDet (paralinear) distances (Lockhart et al., 1994), and RY-coding, which pools purines (adenine and guanine: R) and pyrimidines (cytosine and thymine: Y) into 2-state categories (R, Y).

Whether a poor fit between substitution models and data contributes to the support for Marsupionta requires further testing. Numerous studies have shown that poorly-fitting models can result in misleading support; for example, Sullivan and Swofford (1997) showed that neglecting rate variation among sites led to the rejection of rodent monophyly. Suspicion that support for Marsupionta may arise from poorly-fitting models is aroused by the observation that MP analyses (with 'no common mechanism among sites,' Steel and Penny, 2000) provide less support for Marsupionta than ML and ME analyses (Janke et al., 2002; Zardoya and Meyer, 1998). On the other hand, ML and corrected-distance approaches are typically more robust to heterogeneity in substitution processes across sites and across the tree than is parsimony (Lockhart et al., 1996; Sullivan and Swofford, 2001). Models approximate concatenated data poorly when substitution processes differ between partitions (Yang, 1996). Evolutionary processes have been shown to differ substantially between vertebrate mt-genomes, such as between protein-coding codon positions (Janke et al., 1996; Schmitz et al., 2002), and between RNA loops and stems (Springer and Douzery, 1996). The influence of this source of phylogenetic inaccuracy has yet to be examined for monotreme affinities with mitochondrial data. However, DeBry (1999) and Cao et al. (2000a) have shown that partitioning likelihood analyses (between codon positions or genes) substantially increased the fit between ML models and mammalian mt protein-coding data.

Here we study both the effect of RY-coding on reducing bias from differences in nucleotide composition, and the effect of partitioning genes into similarly evolving subsets. We have recently analyzed a large set of mammalian mitochondrial genomes (Lin et al., 2002) but in the present study we only require a smaller subset of taxa relevant to the deeper mammalian divergences. We find that both protein-coding and RNA-coding data favor a Theria clade; but the relationship between monotremes and therians is closer than accepted by traditional interpretations of morphological evolution.

2. Materials and methods

2.1. Data

Twelve mammalian ingroup taxa and six outgroup vertebrates (Table 1) were selected to maximize the potential for tracking transformations across the mammalian root, and for providing low taxon variance ratios (determined within RASA 3.0.2 Turbo, Lyons-Weiler, 2000). The exclusion of sequences (such as amphibians, crocodilians and rodents among available mt-genomes) that greatly inflate the taxon variance ratios (Phillips, 2002) ensures that the monotreme relationship found cannot in part be related to the inclusion of such "long-branch taxa." Conservative taxa give greater certainty of alignment, allow more data to be included for phylogenetic analysis, and are expected to retain more

Table 1
Taxon variance ratios (calculated in RASA) and base frequency differences (T–C), (A–G), and (Y–R) for the 10,268 nucleotide PTN12 + RNArt dataset

	GenBank Accession Nos.	Taxon variance ratios	Differences (T–C), (A–G), and (Y–R)		
			T–C	A–G	Y–R
Dogfish (*Mustelus manzano*)	AB015962	2.39175	576	802	1020
Trout (*Oncorhynchus mykiss*)	L29771	2.49268	26	366	960
Mole Skink (*Eumeces egregius*)	AB016606	2.05102	104	814	900
Iguana (*Iguana iguana*)	AJ278511	2.15168	−108	1062	944
Green Turtle (*Chelonia mydas*)	AB012104	2.02396	215	1211	974
Greater Rhea (*Rhea americana*)	AF090339	2.90120	−42	766	1140
Platypus (*Ornithorhyncus anatinus*)	X83427	1.50787	668	1076	1020
Echidna (*Tachyglossus aculeatus*)	AJ303116	1.50978	610	1066	1004
Wallaroo (*Macropus robustus*)	Y10524	1.69779	509	1119	990
Brushtail Possum (*Trichosurus vulpecula*)	AF357238	1.67074	493	1221	962
Bandicoot (*Isoodon macrourus*)	AF358864	1.63988	815	1173	978
Opossum (*Didelphis virginiana*)	Z29573	1.59061	824	1298	916
Hippopotamus (*Hippopotamus amphibius*)	AJ010957	1.65229	296	1090	932
Fin Whale (*Balaenoptera physalus*)	X61145	1.60505	461	1075	994
Horseshoe Bat (*Rhinolophus monoceros*)	AF406806	1.63715	388	1006	992
Flying Fox (*Pteropus scapulatus*)	AF321050	1.68232	444	1064	900
Elephant (*Loxodonta africana*)	AJ224821	1.48384	479	1179	930
Aardvark (*Orycteropus afer*)	Y18475	1.59787	592	1158	1032
Mean for mammals			548.3	1127.1	970.8
SD for mammals			160.8	81.4	42.6

phylogenetic signal. Recent molecular phylogenetic studies (e.g., Cao et al., 2000b; Janke et al., 2001; Murphy et al., 2001b; Phillips et al., 2001; Waddell et al., 2001) indicate that the only uncertainty among the relationships of the 18 selected taxa is the placement of the mammalian root. Hence, our analysis reduces to a comparison between the three hypotheses for the mammalian root, which are referred to by the alternative mammalian sub-class sister groupings they support: 1, Theria; 2, Marsupionta; and 3, Monotremata–Eutheria.

Sequences were aligned manually within Se-Al v1.0a1 (Rambaut, 1996). Ribosomal RNA (12S and 16S) structure was based on the models of Gutell et al. (1993) and Springer and Douzery (1996). Secondary structures were designed for each tRNA. With gaps and ambiguous sites removed, the concatenated data set contained 13,856 nucleotide sites, 10,764 for the 13 protein-coding genes and 3092 for RNA-coding genes. NADH6 and eight of the 22 tRNA genes are coded on the mt L-strand, which has a different nucleotide composition from the H-strand. Using the H-strand for all sites avoids these differences. However, this disrupts the relationship between the DNA code and selection pressures and it is unclear whether this should be maintained in preference to the strand-specific coding relationship, or, whether to exclude the L-strand genes altogether. However, including or excluding L-strand genes alters bootstrap support for ME distance and ML analyses by less than 2% for rooting Mammalia. The L-strand genes have therefore been included, and coded as complements.

The protein-coding DNA sequences are called PTN123, PTN12, PTN1, PTN2, or PTN3, depending on codon positions included. The concatenation of the ribosomal and transfer RNA data is referred to as RNArt. Standard coding of the four DNA nucleotides (A, C, G, and T) is NT-coding; RY-coding reduces them to 2-state categories. The amino acid translation of the protein-coding genes is PTNaa. An alternative treatment (PTNfg) has leucine, isoleucine and valine lumped as one functional group. This follows suggestions that variation in these mid-sized, neutral and hydrophobic amino acids accounts for much of the composition bias among the mt-encoded proteins of vertebrates (Naylor and Brown, 1998), and mammals particularly (DeBry, 1999).

Concatenating protein and RNA-coding sequences allows only one substitution model for characters that are not evolving under a homogenous evolutionary process. Partitioning into more-homogeneous subsets is offset by parameters being estimated from smaller data subsets, so reducing the signal-to-variance ratio. This appears to occur (mildly) for nodes that are well supported regardless of partitioning or concatenation (Cao et al., 2000a; Krajewski et al., 1999; but see Krajewski et al., 2000). In order to balance homogeneity of processes with the loss of precision, the 12S/16S rRNA and tRNA data was divided into just two partitions—RNA stems and RNA loops. For the protein-coding sites, Previous studies have emphasized differences between the codon positions and/or between the genes themselves (e.g. Amrine and Springer, 1999; Yang, 1996). However, DeBry (1999) showed that certain partitioning schemes based on amino acid categories improved likelihood more efficiently (with fewer parameters) than gene-based partitions. Consequently, genes were only partitioned where their properties differed markedly. The five protein-coding gene partitions are based on the proportion of constant sites in the RY data, and the average frequency of purine bases. The five protein-coding partitions were further partitioned into 1st and 2nd codon sites for the ML analysis.

2.2. Composition heterogeneity

If the ingroup is essentially homogeneous for nucleotide and amino acid composition, then alternative positions of the mammalian root will be affected approximately equally by any heterogeneity in the outgroup. As such, we focus on composition homogeneity in the mammalian ingroup because the position of the mammalian root is our main concern. PAUP* 4.0b8 (Swofford, 1998) was used for χ^2 tests of composition homogeneity with nucleotide data, and Microsoft Excel for amino acid data. However, such tests are difficult to compare between data treatments because their statistical power depends on both the number of sites and character states. Amino acid coding provides one third as many sites as nucleotides but has 19, rather than 3, degrees of freedom. Even if composition homogeneity is rejected, this gives little indication of the potential for any bias to actually influence phylogeny reconstruction.

In order to compare the relative potential for heterogeneity to bias phylogeny reconstruction between data treatments, we use a simple statistic, treeness divided by relative composition variability (treeness/RCV). Treeness (or stemminess, Lanyon, 1988) is the proportion of tree distance on internal branches. If we consider substitutions attributed to branches that support taxonomic groupings (i.e., internal branches) as phylogenetic signal and consider other substitutions as noise, then the treeness value is one way of defining a signal-to-noise ratio. Thus, where data treatments are compared for the same tree (taxa and relationships), higher treeness indicates a higher signal-to-noise ratio (see Lanyon, 1988). The unrooted trees in Fig. 1 give the differences in treeness between the NT-coded PTN3, NT-coded PTN12 + RNArt, and RY-coded PTN12 + RNArt datasets. The values are determined on the same tree and

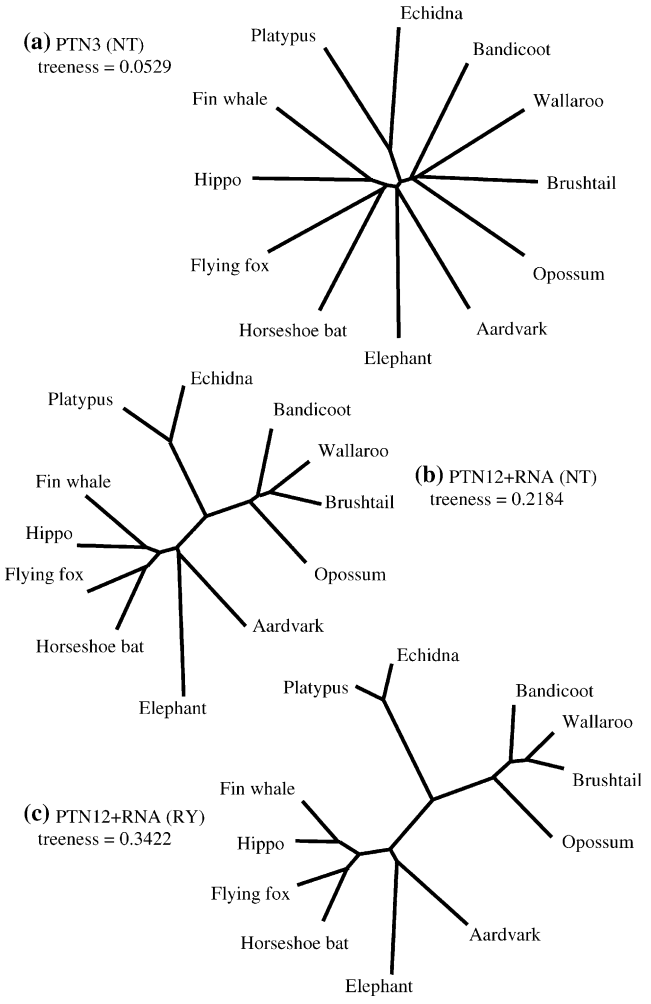

Fig. 1. Minimum-evolution (uncorrected distance) trees illustrating the increase in treeness (the proportion of changes on internal branches) in going from NT-coded 3rd codon positions (a) to the 1st and 2nd codon positions plus RNA data, both NT-coded (b), and RY-coded (c.).

are thus comparable. Relative composition variability (RCV) is the average variability in composition between taxa; for nucleotides this is:

$$\mathrm{RCV} = \sum_{i=1}^{n}(|\mathrm{A}_i - \mathrm{A}^*| + |\mathrm{T}_i - \mathrm{T}^*| + |\mathrm{C}_i - \mathrm{C}^*| + |\mathrm{G}_i - \mathrm{G}^*|)/n \cdot t,$$

where A_i, T_i, C_i, and G_i are the numbers of each nucleotide for the ith taxon. A^*, T^*, C^*, and G^* are averages across the n taxa, and t is the number of sites. Constant sites were excluded for all χ^2 and RCV calculations.

In summary, for data (or data treatments) that are compared for the same tree, lower RCV and higher treeness values, respectively, indicate a lower magnitude of composition bias and a lower potential for bias (composition or other non-phylogenetic signals) to influence phylogeny reconstruction. Hence phylogeny estimates from the data treatments (such as coding methods) and partitions that have the highest treeness/RCV values are expected to be the least susceptible to composition bias. Exclusion of PTN3 (because of saturation and/or composition bias) has been preferred in many phylogenetic analyses. Here we exclude the PTN3 data for NT and RY-coded treatments because of its high relative composition variability (RCV, Table 2), and low signal-to-noise ratio (inferred from low treeness, Fig. 1).

2.3. Phylogenetic analysis

ML analysis of PTNaa was carried out using ProtML within MOLPHY 2.3 (Adachi and Hasegawa, 1996). Other phylogenetic analyses were performed within PAUP* 4.0b8 (Swofford, 1998). TN93 (Tamura and Nei, 1993) and F81 (Felsenstein, 1981) distance corrections were used for ME analyses of the NT and RY-coded data, respectively. These were compared with LogDet (Lockhart et al., 1994) ME analyses, which unlike the former corrections, are intended to be robust to composition heterogeneity among taxa. For ML

Table 2
Composition variability and treeness of variable sites for NT and RY coding, and for PTNaa and PTNfg. χ^2 p values are shown for the full dataset and the 12 mammalian ingroup taxa

Sites	Data	χ^2 (p)		RCV	Treeness
		All taxa	Ingroup	(Ingroup)	(Ingroup)
PTN1	NT	<0.0001	<0.0001	0.0734	0.1765
	RY	0.0123	**0.2583**	**0.0269**	**0.3330**
PTN2	NT	<0.0001	**0.961**	**0.0342**	**0.2558**
	RY	**0.2315**	**0.579**	**0.0392**	**0.3535**
PTN3	NT	<0.0001	<0.0001	0.1430	0.0529
	RY	<0.0001	<0.0001	0.0528	0.1257
RNA stems	NT	<0.0001	**0.0738**	0.0602	**0.3169**
	RY	**>0.9999**	**0.9906**	**0.0302**	**0.4259**
RNA loops	NT	<0.0001	**0.1684**	0.0667	0.1870
	RY	**0.9797**	**0.7017**	**0.0449**	**0.3116**
Amino acids	aa	<0.0001	**0.3963**	0.0819	0.2218
	fg	<0.0002	**>0.9999**	0.0612	0.2425

Relative composition variability (RCV) and treeness values are shown for the ingroup. χ^2 (p) > 0.05, RCV <0.05 and treeness >0.25 are in bold.

analyses, the TN93 (Tamura and Nei, 1993) and CF87 (Cavender and Felsenstein, 1987) models were used with NT and RY-coded data, respectively. Both models incorporate a single transversion category. The Kishino and Hasegawa (1989) test was used to examine the three a priori hypotheses (Theria, Marsupionta, and Monotremata–Eutheria). Two-tailed probabilities of tree rejection are given. For partitioned analyses, $-\ln L$ site values from each partition were transferred to a Microsoft Excel file for the Kishino–Hasegawa test. To further examine composition bias, ME trees based on pairwise base frequency differences between taxa were constructed. Base frequency difference data are shown in Table 1. A–G distance (representing purine bias) between the iguana and the rhea is $|(\text{A–G})_{\text{Rhea}} - (\text{A–G})_{\text{Iguana}}|$, where A–G is the difference between the number of adenine (A) sites and guanine (G) sites for the taxon.

3. Results

3.1. Data examination

The first point is that standard ML analysis of the mt data (both as amino acid and nucleotide sequences) gives the standard result of Marsupionta (Fig. 2). The next step is to analyze the data for compositional heterogeneity, assess its affects on inferring trees, and examine effects of reducing the data to two character-states (RY-coding). Table 1 shows differences in frequencies between pairs of nucleotides (T–C, A–G, and Y–R). The 12 mammals are less variable than the 6 outgroup taxa. Even though our selection of taxa reduces heterogeneity, the differences are still significant. χ^2 Tests (Table 2) indicate that with NT-coding the heterogeneity among the dataset as a whole is highly significant ($p < 0.0002$) for all nucleotide datasets. Similarly, amino acid composition heterogeneity is also highly significant ($p < 0.0002$), irrespective of whether or not Leu, Ile, and Val are combined. With RY-coding heterogeneity is much reduced, and homogeneity cannot be rejected even at $p < 0.20$ for RNA stems, RNA loops, and PTN2.

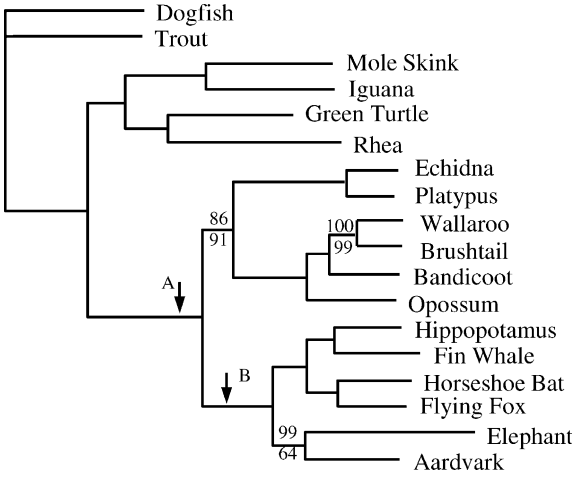

Fig. 2. ML (F81) tree and bootstrap support for the NT-coded PTN12 + RNArt data (above) and RELL-bootstrap support for ML analysis (mt-REV24F) of the amino acid (PTNaa) sequence (below). Values are not shown for nodes that received >99% support in both analyses. Marsupionta is favored. Arrows indicate alternative monotreme placements, giving (A) Theria, and (B) Monotremata–Eutheria.

Heterogeneity within the ingroup (mammals) is lower, though significant at $p < 0.0001$ for PTN3 (with both NT and RY-coding) and for NT-coded PTN1. It is likely that composition heterogeneity among mammals for NT-coded PTN1 has been masked in previous analyses by the inclusion of constant sites and closely related taxa (such as numerous apes). For the ingroup, homogeneity is not rejected (at $p < 0.05$, Table 2) for the amino acid datasets (PTNaa and PTNfg) and the PTN2 and RNA datasets (for both NT and RY-coding), nor for PTN1 with RY-coding. However, considerable variation in the magnitude of heterogeneity among the ingroup exists between the NT, RY, and amino acid treatments. The ingroup RCV values for the amino acids (PTNaa and PTNfg) are higher than for any of the RY-coded data—including PTN3. Values for PTNaa and PTNfg are also higher than for NT-coded PTN2, while the value for PTNaa was higher even than for NT-coded PTN1. Hence any advantage of using the amino acid translation rather than the nucleotides (minus PTN3) is not apparent in terms of the magnitude of compositional variability. However for PTN1 and PTN3, RCV values for RY-coding are approximately 2.7 times less than for NT-coding. The situation is similar, though less extreme, for the RNA stems, and loops.

For NT-coding, most variability comes from differences in frequencies of the pyrimidines. For the 10,268 sites in the PTN12 + RNArt dataset, the standard deviation among the mammals is 160.8 for T–C. It is 81.4 for A–G and 42.6 for Y–R. Across these ingroup taxa, correlation (r^2-values) between PTN12 + RNArt and PTN3 is 0.8815 for T–C, 0. 6376 for A–G, and 0.5629 for Y–R. The treeness values in Table 2 show for PTN1, PTN2, PTN3, RNA stems, and RNA loops that RY-coded values are 1.34–2.38 times higher than the NT-coded equivalents. Furthermore, with the exception of PTN3, the RY-coded data has higher treeness than either treatment of the amino acid translation (PTNaa and PTNfg). For PTN12 + RNArt, RY-coding (Fig. 1c) provides 1.57 times higher treeness than NT-coding (Fig. 1b). These results are important in that lower RCV and higher treeness, respectively, indicate a lower magnitude of compositional non-stationarity and a lower potential for non-phylogenetic signals to influence phylogeny reconstruction. On this basis, results in Fig. 3

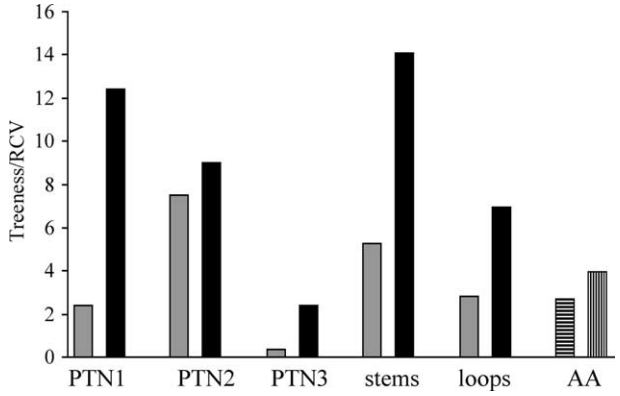

Fig. 3. Treeness/relative composition variability (RCV) for nucleotide codon positions (PTN1, PTN2, and PTN3), RNA stems and loops, and the translated amino acid sequence (AA). Nucleotide sequences are treated as standard NT (grey bars) and RY-coded (black bars). The protein sequence is treated as standard amino acids (horizontal lines) and with Leu, Ile, and Val lumped as a functional group (vertical lines).

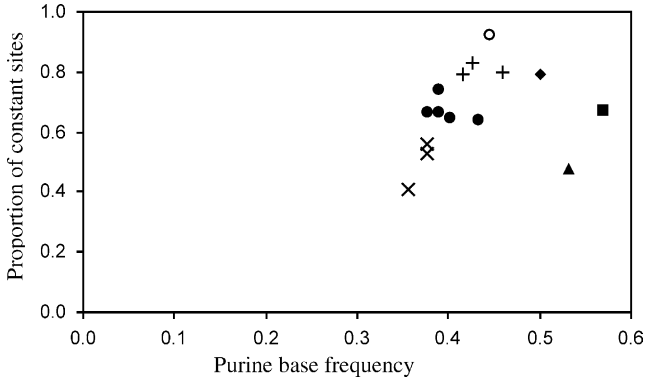

Fig. 4. Grouping of protein-coding (PTN12) and RNA-coding genes, based on the purine frequencies and observed proportion of constant sites with RY-coding. Groupings are: RNA loops (■); RNA stems (◆); COI (○); NADH6 (▲); low constant sites, low purine protein-coding genes (ATPase8, NADH2, NADH4L, and ×); moderate constant sites, low-moderate purine protein-coding genes (ATPase6, NADH1, NADH3, NADH4, NADH5, ●); high constant sites, moderate purine protein-coding genes (COII, COIII, Cyt*b*, and +). The purine base frequency for the L-strand protein, NADH6, is from its complement.

show the relative potential for compositional bias with the different data codings, and between partitions. Prominent is the improvement in treeness/RCV values that RY-coding provides over NT-coding (except for PTN2, where the improvement is only slight). Furthermore, values for PTNaa and PTNfg are low compared with the RY-coded PTN1 and PTN2 datasets.

Frequency differences were determined for A–G (purine bias), T–C (pyrimidine bias) and Y–R (bias between purines and pyrimidines). The pyrimidine bias is the largest affecting the mammalian root on this data (see Table 1). The averages for T–C among the monotremes and the marsupials are much higher (639.0 and 660.3) than among the placentals and outgroup taxa (443.3 and 128.5, respectively).

A way of assessing how strong the biases are in practice comes from constructing ME trees from pairwise distances derived from differences in base frequencies (see Lockhart et al., 1994). To make these results relevant to the ML analyses the constraint trees (Marsupionta, Theria, and Monotremata–Eutheria) were evaluated (Table 3). The T–C distance tree favors Marsupionta, the alternatives require 84.3 and 65.9 more changes. In contrast, Theria is favored in the A–G distance ME analysis, though only by 13.7 and 4.7 A–G changes. The bias for the Y–R distance data is smaller still, with Monotremata–Eutheria favored, but with Theria and Marsupionta requiring only 2.5 and 1.5 further changes. These results were from trees that allowed negative branch-lengths, if negative distances are not permitted the results are essentially the same. However, the small difference is erased between the three hypotheses in the Y–R distance analysis.

Janke et al. (1996) noticed in the platypus mt-genome that the proposed secondary structure for its tRNA serine (UCN) was similar to that of non-therians. With the increase in the number of mt-genomes available, the conserved nature of the deletion in therians stands out. All available amniote mt-genomes reveal two base pairs between the acceptor-arm and D-arm of tRNA-serine for all non-therians, but these are missing in marsupials and eutherians. The lengths of both the acceptor and D-arm stems are conserved among the taxa in this study, and the purine/pyrimidine nature of each of these stem pairs is also maintained (except for the first D-arm pair). The conservation of the acceptor and D-arm stems throughout amniotes is complemented by the conservation (among non-therians) of the two bases that join the stems. These two nucleotides are both likely to have been adenine in the last common ancestor of mammals

Table 3
Minimum-evolution scores for the three positions of the mammal root, from pairwise distances between taxa for base frequency differences

Distance	Theria	Marsupionta	Monotreme–Eutheria
T–C	+84.3 (+67.0)	**2485.2** (**1771.6**)	+65.9 (+62.0)
A–G	**1923.0** (**1452.2**)	+13.7 (+3.3)	+4.7 (+3.3)
Y–R	+2.5 (**532.0**)	+1.5 (**532.0**)	**682.8** (**532.0**)

Sequence data is not used directly, only differences in nucleotide composition. The best score is shown in bold, with differences shown for the alternative hypotheses. Values with negative branch lengths not allowed are in brackets.

and reptiles. This A–A combination occurs in all non-therian amniotes included in this study.

3.2. Maximum-likelihood analysis

The next step uses ML to examine the effects: NT vs. RY-coding; allowing rate heterogeneity (±PAUP* estimates for I and Γ_8); and concatenated vs. partitioned analyses, on support for the three positions for monotremes. The results for the combined protein-coding and RNA-coding data are illustrated in Fig. 5 and Table 4. Of the three monotreme placements, only one is consistently more likely among pairwise comparisons—Theria is more likely than Monotremata–Eutheria in each of the eight treatments (a–h). As shown in Fig. 5, the incorporation of $I + \Gamma_8$ has little impact on the probability of accepting Theria, Marsupionta, or Monotremata–Eutheria. The two factors that have a substantial effect on the Kishino–Hasegawa test p values are coding and partitioning (Fig. 5). The probability of accepting Theria (relative to Marsupionta) is greater with the RY-coding models (a–d) than with the corresponding NT-coding models (e–h). Similarly, partitioning the analyses (c,d,g,h) increases the probability of accepting Theria (relative to Marsupionta). Considered together, RY-coding and partitioning the analyses (among RNA stems, loops, and PTN1, and PTN2 for the 5 protein groupings) results in a substantial turnaround in support for the position of the mammalian root. The concatenated NT-coding models favor Marsupionta; Theria being rejected by approximately 0.8 standard deviations. In contrast, the partitioned RY-coding models favor Theria (with Marsupionta being rejected by approximately 1.1 standard deviations).

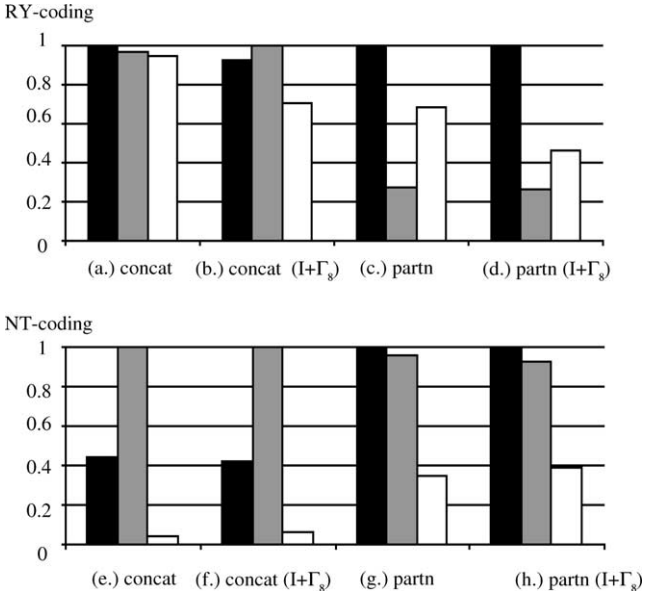

Fig. 5. Kishino–Hasegawa test p values for Theria (black), Marsupionta (grey), and Monotremata–Eutheria (open bars) with the ML tree allotted a p value of 1.0, under RY-coding (above) and NT-coding (below). The data is PTN12 + RNArt and is either concatenated (concat) or partitioned (partn) into RNA stems; RNA loops; and first and second codon positions for five groups of protein-coding genes. Inclusion of PAUP* estimates of invariable sites and an 8 category gamma distribution is denoted $(I + \Gamma_8)$.

The Akaike Information Criterion (AIC, $-2\ln L + 2 \times$ # free parameters) is used to evaluate improvement in adding parameters to the ML models. TN93 + I + Γ_8(con) has 40 free parameters: 33 branch-lengths, three base frequencies, two ti/tv rates and two rate parameters (I and α). For TN93 + I + Γ_8(part) this is multiplied by the number of partitions $(12 \times 40 = 480)$. ML models are compared this way (see Adachi et al., 2000; Cao et al., 2000a) because of concerns that the χ^2 distribution is a poor approximation for likelihood ratio tests (see

Table 4
Log-likelihood scores ($-\ln L$) for the three hypotheses with the PTN12 + RNArt data concatenated and partitioned, and with and without invariable sites and an 8-category gamma rate distribution $(I + \Gamma_8)$

Treatment	Hypothesis	NT-coding (TN93 models a–d)			RY-coding (CF87 models e–h)		
		$-\ln L$	t	p	$-\ln L$	t	p
Concatenated	Theria	87741.29	0.7630	0.4455	36108.02	ML	ML
	Marsupionta	87722.80	ML	ML	36108.54	0.0344	0.9726
	Mon + Euth	87767.18	2.0170	0.0437	36108.93	0.0603	0.9519
Concatenated $(I + \Gamma_8)$	Theria	81090.26	0.8029	0.4221	33052.15	0.0866	0.9310
	Marsupionta	81081.24	ML	ML	33051.94	ML	ML
	Mon + Euth	81099.51	1.8819	0.0599	33052.69	0.3764	0.7066
Partitioned	Theria	84002.95	ML	ML	34062.08	ML	ML
	Marsupionta	84004.12	0.0525	0.9582	34078.51	1.1063	0.2686
	Mon + Euth	84022.80	0.9490	0.3426	34068.86	0.4123	0.6801
Partitioned $(I + \Gamma_8)$	Theria	78547.16	ML	ML	31809.89	ML	ML
	Marsupionta	78548.30	0.0949	0.9244	31816.34	1.1082	0.2678
	Mon + Euth	78556.63	0.8660	0.3865	31814.52	0.7329	0.4636

The number of standard deviations (t) from the most likely tree (denoted ML), and the two-tailed significance (p) of this difference are shown. Variance estimates and significance values were determined by the Kishino and Hasegawa (1989) test.

Table 5
Pairwise comparisons of Akaike Information Criterion (AIC) values for the ML substitution models (a–h) that were used for the 18-taxon PTN12 + RNArt dataset (for the Theria tree)

	Free parameters	AIC value		Free parameters	AIC value
Model 1 vs. Model 2					
a. TN93-con	38	175558.6	b. TN93 + I + Γ_8-con	40	**162260.5**
a. TN93-con	38	175558.6	c. TN93-par	456	**168917.9**
a. TN93-con	38	175558.6	d. TN93 + I + Γ_8-par	480	**158054.3**
b. TN93 + I + Γ_8-con	40	**162260.5**	c. TN93-par	456	168917.9
b. TN93 + I + Γ_8-con	40	162260.5	d. TN93 + I + Γ_8-par	480	**158054.3**
c. TN93-par	456	168917.9	d. TN93 + I + Γ_8-par	480	**158054.3**
e. CF87-con	34	72284.0	f. CF87 + I + Γ_8-con	36	**66176.3**
e. CF87-con	34	72284.0	g. CF87-par	408	**68940.2**
e. CF87-con	34	72284.0	h. CF87 + I + Γ_8-par	432	**64483.8**
f. CF87 + I + Γ_8-con	36	**66176.3**	g. CF87-par	408	68940.2
f. CF87 + I + Γ_8-con	36	66176.3	h. CF87 + I + Γ_8-par	432	**64483.8**
g. CF87-par	408	68940.2	h. CF87 + I + Γ_8-par	432	**64483.8**

The model that better approximates the evolutionary process (according the AIC) for each pairwise (across the table) model-comparison has the lower value, which is bolded. TN93 models were used for NT-coded data, and CF87 models were used for RY-coded data. Models that vary across the 12 partitions are denoted par (= partitioned), while the single models are denoted con (= concatenated). The AIC value is $-2\ln L + 2 \times$ (# free parameters).

Whelan and Goldman, 1999). The conclusion from the AIC scores is that the gain from adding parameters by partitioning is more than offset by the improved likelihood values (Table 5).

3.3. Corrected distance analyses

Fig. 6 shows the results of ME analyses with corrected distance models for which compositional stationarity is either assumed (symmetrical models, F81 and TN93) or not assumed (LogDet). Beginning with PTN12 + RNArt (Fig. 6a), support for Marsupionta decreases as the proportion of RY-coded sites assumed to be invariable increases, while support for Theria increases (≈20% with all sites included, up to ≈75% with all constant sites excluded). In contrast, support for Marsupionta remains above 95% in each treatment of the NT-coded data. By comparison with their respective symmetrical distance corrections, LogDet correction makes little difference to the pattern of support for the mammalian root in Fig. 6a. Nevertheless, as constant sites are excluded from the NT-coded data, more support (still below 5%) for Theria emerges under LogDet than with the Tamura–Nei correction. Separating the protein and RNA data (Fig. 6b) and comparing these with the equivalent PTN12 + RNArt treatment (RY-coded, F81) indicates that the overall signal is dominated by PTN12.

4. Discussion

4.1. Mitochondrial protein and RNA-coding data are consistent with the Theria hypothesis

This study has explored two non-phylogenetic explanations for analyses of mt-genome data supporting Marsupionta; nucleotide composition bias, and non-homogenous processes between data partitions. The position of murid rodents and hedgehog (and its relative, moonrat) are controversial (see Lin et al., 2002) but are not included in this study and so cannot have affected the position of the root with respect to monotremes. In fact, alternative taxon sampling from reptiles and placentals also does not effect the outcomes (Phillips, 2002). Rather, it is in RY-coding and model partitioning for data that this study differs from earlier ones of the position of monotremes. As expected, ML analysis of the concatenated NT-coded PTN12 + RNArt dataset supports Marsupionta. Either RY-coding or allowing the substitution model to vary between the 12 partitions, erodes this support such that effectively Marsupionta and Theria become equally likely. Combining both partitioning and RY-coding results in Theria emerging as the favored hypothesis. However, Marsupionta and Monotremata–Eutheria cannot be rejected at $p < 0.20$ even for the best-fitting (according to AIC) of the models (CF87 + I + Γ_8). Nonetheless, the net change from RY-coding and partitioning on relative support for the mammal rooting hypotheses is substantial. As shown in Table 4, Marsupionta is approximately 0.8 standard deviations ahead of Theria for the standard analyses, while Theria is approximately 1.1 standard deviations ahead of Marsupionta with PTN12 + RNArt being RY-coded and partitioned.

The most important result of the current study is that the mt-genome data favors Theria under the most conservative treatment (RY-coding), and best fitting model (partitioning with or without rate heterogeneity). Further, RY-coded and partitioned ML(F81) favors Theria over Marsupionta for both the protein and RNA data (by 11.4 and 5.0 ln L U, respectively), as do the ME analyses (Fig. 6b). The minimum that can be concluded

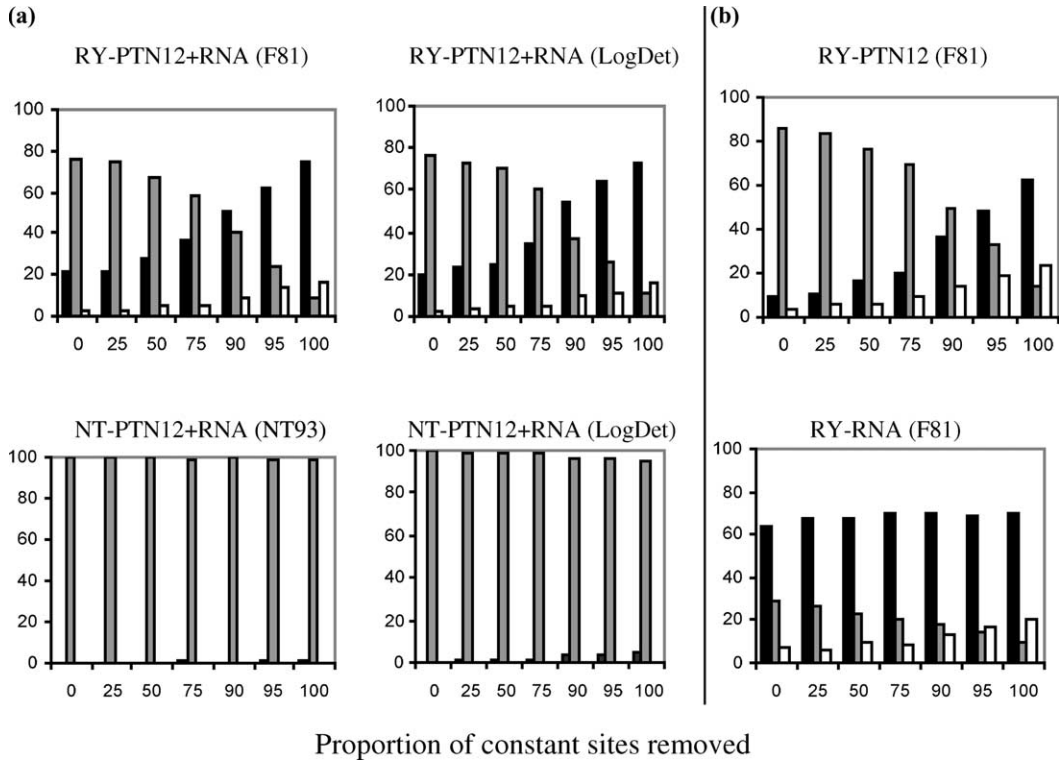

Fig. 6. Minmum-evolution bootstrap support for Theria (black), Marsupionta (grey), and Monotremata–Eutheria (open bars), under proportional removal of constant sites, for symmetrical (F81 and TN93) and LogDet distance models. In (a) RY-coding (above) and NT-coding (below) are compared for the PTN12 + RNArt data. In (b) the protein and RNA data are treated separately, and are RY-coded.

from this is that, contrary to earlier results (e.g., Janke et al., 1996, 2002), whole mt-genome sequence data are not inconsistent with the data from nuclear genes (e.g., α-lactalbumin: Messer et al., 1998; MP6/IGF2R: Killian et al., 2001). Moreover, the mt data do not conflict with the multitude of data that has traditionally defined Theria, such as from the pectoral girdle (Sereno and McKenna, 1995), the basicranium (Rougier et al., 1996), and numerous aspects of reproduction (see Carrick and Hughes, 1982; Renfree, 1993).

Consistency with other data is not the criterion by which the alternative treatments of the mt data are assessed. It must reasonably be expected that RY-coding and partitioning the likelihood analysis provides a better estimate of the mammalian root than does the NT-coded, concatenated data. At least two departures from the assumptions of standard substitution models appear to have resulted in biases that have led to Marsupionta being favored in previous analyses; pyrimidine (T–C) bias, and substitution regime heterogeneity between process partitions.

4.2. Composition heterogeneity

The current study supports the expectation of Phillips et al. (2001) that RY-coding will be more robust to compositional biases than NT-coding. NT-coding provides more chance for compositional heterogeneity to provide a misleading signal, the compositional variability (RCV) results (Table 2) illustrate this. Secondly, the proportion of (true) phylogenetic signal for deep-level relationships that is overwritten is substantially higher for the NT-coded data than for the RY-coded data. This is inferred from the treeness values (Table 2).

Given high composition variability and low phylogenetic signal retention, any method of phylogenetic inference that assumes compositional stationarity can be unreliable (Waddell, 1995). Hence, the treeness/RCV ratio is an indication of the potential for non-stationarity to mislead phylogenetic inference. Fig. 3 indicates that non-stationarity is a severe problem for NT-coded PTN3. This provides further empirical support for the common practice (e.g., Amrine and Springer, 1999; Phillips et al., 2001) of excluding PTN3 (or at least PTN3 transitions) on grounds of compositional non-stationarity or substitution saturation. RY-coding provides an approximately 2.4-fold advantage over NT-coding in terms of the treeness/RCV ratio for PTN12 + RNArt. This advantage is almost 6.5-fold for PTN1, which Janke et al. (1996) stated that most of the information supporting Marsupionta was derived.

The key finding from composition variability for the NT-coded PTN12 + RNArt data is that non-stationarity favors Marsupionta and is dominated by a pyrimidine

bias (the difference between T and C frequencies). This is shown by ME trees (Table 3) that are derived from pairwise distances between taxa for pyrimidine bias (T–C distances). The difference between support for Marsupionta and Theria in terms of overall distance (number of changes), is more than four-fold greater for the T–C distance trees than for ME trees based on absolute distances between the NT-coded PTN12 + RNArt data (which also favors Marsupionta). This indicates the relative strength of the frequency bias among the pyrimidines. By comparison, the biases in purines (A–G) and between purines and pyrimidines (Y–R) are smaller in terms of contribution to the overall composition variability and apparent influence on monotreme placement. Indeed Table 3 shows that if negative branch-lengths are not allowed, and as expected if there is no bias, the Y–R distances equally favor Theria, Marsupionta, and Monotremata–Eutheria.

As suggested by many authors (e.g., Cao et al., 2000a; Waddell et al., 2001; but see Simmons et al., 2002) phylogenetic inference may be more robust to non-stationarity when amino acids are used rather than all three codon positions of DNA. However, the treeness/RCV values in Fig. 3 do not support this when PTN3 is excluded. χ^2 Tests for compositional non-stationarity (including those of this study; Table 2) do give the impression that PTNaa is less affected by composition variability than NT-coded PTN12. However, Table 2 shows that the magnitude of composition variability among the mammals is higher for the amino acid treatment than for NT-coded PTN1. Moreover, NT-coded PTN2 has lower RCV than the amino acid sequence, even when Leu, Ile, and Val are lumped together (PTNfg). The nature of the χ^2 test tends to buffer against compositional stationarity being rejected for amino acid data; many of the amino acids contribute only a small proportion of the characters but elevate the degrees of freedom. Examination of amino acid composition data for vertebrate mt proteins provided in Penny et al. (1999) show this. Preliminary examination of amino acid frequencies in our dataset is consistent with a growing body of evidence (Foster and Hickey, 1999; Schmitz et al., 2002; Singer and Hickey, 2000) that amino acid compositional variability is largely explained by mutational bias. Among the 18 taxa of this study there is a strong negative correlation ($r^2 = 0.85$) between the frequency of MIFVYWC (amino acids with a T but no C at PTN1 or PTN2) and of TAPRQH (amino acids with a C but no T at PTN1 or PTN2). Strong correlation between PTN3, PTN12, and RNArt, for the pyrimidine (T–C) bias further supports the hypothesis that the T–C bias largely derives from an underlying mutation bias.

Reyes et al. (1998a) found two major mutation biases in mammalian whole mt-genomes, G–A, and T–C (both with respect to the L-strand) and suggested these resulted during replication from spontaneous deamination on the H-strand of cytosine to uracil, and adenine to hypoxanthine. As with human mtDNA (Tanaka and Ozawa, 1994), the former deamination occurred at a somewhat higher rate than the latter. The higher cytosine deamination would explain the low representation of guanine on the L-strand (which codes for the mt proteins except for NADH6). Guanine averages 4% at PTN3 among the mammals included in this study, see also Ota and Penny (2003). This helps explain the relatively low purine (A–G) bias. Most of the guanine that is free to vary has been lost and, furthermore, only limited variability can be built upon the low remaining frequency of guanine. In contrast, if H-strand deamination of adenine is the major mutation affecting the relative frequency of cytosine and thymine on the L-strand, then this process (or its proofreading correction) must differ considerably across the tree. For example, the T:C frequency ratio at PTN3 is 1.61 for the opossum, compared to 0.37 for the hippopotamus. The finding that variation in the frequencies of pyrimidines is the dominant bias in mt protein-coding and RNA-coding genes is consistent with previous studies (e.g., Phillips et al., 2001; Reyes et al., 1998b; Springer and Douzery, 1996).

The importance of pyrimidine bias invalidates the conclusion that the main compositional bias is GC content (Sueoka, 1962, 1995; see also Moores and Holmes, 2000 for discussion). This assumption has resulted in studies of directional mutation pressure focusing on the relationship between GC content and codon usage, even for mt-genomes (e.g., Foster et al., 1997; Schmitz et al., 2002). Furthermore, the non-homogenous ML (NHML) of Galtier and Gouy (1998) allows only for GC/AT bias and relies on the assumption that the equilibrium frequency of $A = T$ and $G = C$. In fact, for PTN12 + RNArt, A, and T (or G and C) covary little between the taxa, the respective r^2 values being 0.28 and 0.33. Instead, strong negative correlations exist between the frequencies of T and C ($r^2 = 0.96$) and A and G ($r^2 = 0.94$). These are consistent with transitions dominating substitution among mt genes (e.g., Springer and Douzery, 1996; Yang, 1996), as well as with the more general assertion of Reyes et al. (1998a) that adenine and cytosine deaminations are the major mutations that bias mt nucleotide composition.

LogDet distances (see Lockhart et al., 1994) do not assume compositional stationarity between any of the base frequencies. It is interesting that these did not recover the putatively correct tree (Theria) for the NT-coded data, with the results (Fig. 6a) being almost the same as those for Tamura–Nei distances (which assume stationarity). Lockhart et al. (1996) cautioned that removal of invariant sites may be required to effectively correct for composition variability. By doing this, Haddrath and Baker (2001) recovered a bird phylogeny that contradicted composition bias. However, other authors (e.g., Tarrio et al., 2001; Waddell et al., 1999)

have questioned the ability of LogDet as currently implemented to correct non-stationarity when strong rate variation exists among variable sites (more slowly-evolving sites may hide non-stationarity at the faster-evolving sites). ML estimates of the gamma shape parameter (α) indicated that, even with constant sites removed, there is strong rate heterogeneity among sites in both the RNArt and PTN12 datasets. However, RY-coding will counter biases within pyrimidines and within purines. This is because transitions are "invisible" to RY-coding and it does not distinguish between the four transversion categories.

If the principle compositional bias among DNA sequences involves the relative frequency of purines (R) vs. pyrimidines (Y), RY-coding may be no more (if not less) reliable than either NT-coding or amino acids. It is likely that loss of information from transitions and not distinguishing transversion categories will be a more usual drawback of RY-coding. Currently this is not evident at deep taxonomic levels with whole mt-genomes. RY-coded PTN12 recovers all relationships in this dataset with similar or greater efficiency than NT-coding, or use of PTNaa. RY-coding may be less useful where transitions are less saturated such as at shallower taxonomic levels, or between nuclear sequences. Other relationships for which mt data have provided results that contradict the majority of nuclear and morphological data include the placements of the hedgehog (Cao et al., 2000a; Mouchaty et al., 2000), tarsier (Schmitz et al., 2002), and perching birds (Härlid and Arnason, 1999; Mindell et al., 1999).

Another issue is the nature of the amino acid compositional bias and its relationship with the nucleotide frequency bias. The present study indicates that bias relative to signal retention (inferred from treeness/ RCV) is no less for amino acid than for NT-coded PTN12 data. The strong negative correlation between the frequency of amino acids coded for by T-rich PTN12 codons and the C-rich codons is evidence for the inter-dependence of the aa and DNA biases. Whether PTNaa should provide more reliable phylogeny reconstruction than NT-coded PTN12 is unimportant here, both support Marsupionta. That distance analyses from nucleotide frequency data show that composition (pyrimidine) bias strongly favors Marsupionta (Table 3) is a warning about the NT-coded DNA and PTNaa results. Lower composition variability and higher phylogenetic signal retention should result in phylogeny reconstruction with RY-coding being less susceptible to the effects of non-stationarity.

4.3. Partitioned likelihood

Table 4 shows that partitioning into RNA stems, RNA loops, and 5 protein partitions for each of PTN1 and PTN2 reduced $-\ln L$ scores by between 3.1 and 5.7% for PTN12 + RNArt (depending on coding and whether rate heterogeneity was incorporated in the models). This "partitioning advantage" is slightly less than that found by Amrine and Springer (1999) when all substitution parameters varied between partitions (≈5.8–7.2%). This indicates less heterogeneity in substitution processes between the partitions in the current study, probably because we excluded third codon positions. Akaike Information Criterion (AIC) scores indicate that partitioning the likelihood analyses significantly improved the fit between the data and both NT-coding and RY-coding models. One result does caution against assuming that a better AIC score necessarily means that a result is more reliable. Incorporation of rate variation across sites ($I + \Gamma_8$) provides a much better fitting model than does partitioning, with both TN93 and CF87. However, as shown in Fig. 6, it is partitioning (c and g) rather than $I + \Gamma_8$ (b and f) that allows recovery of the tree (Theria) favored by the models with the overall best fit (partitioning and $I + \Gamma_8$; d and h).

Fig. 4 reveals considerable variability in the proportion of constant sites and base frequencies between different proteins and between RNA stems and loops. Differences between first and second codon positions are even more marked, while relative branch-lengths also differ substantially between partitions (data not shown). Substantial differences in base frequencies, relative branch-length estimation, relative rates across substitution types and rate heterogeneity among sites all justify the partitioning. The partitioning analysis adds considerable weight to the Theria hypothesis, which is favored in all of the partitioned analyses (Fig. 5). Because partitioning, with both NT and RY-coding, increases the signal for Theria relative to Marsupionta, non-homogeneity of substitutions between the PTN12 + RNArt partitions apparently provides a signal favoring Marsupionta. Whether or not different trees are favored, partitioning substantially influences the distribution of phylogenetic signal between hypotheses in a number of likelihood and distance analyses (e.g., Cao et al., 2000b; Caterino et al., 2001; Krajewski et al., 2000). The mechanisms of the biases have not been elaborated on, except to say that analyses of concatenated sequences do not account for heterogeneity in substitution processes, and that the advantage differs by both the partitioning scheme (DeBry, 1999) and the parameters being partitioned (Yang, 1996). As noted by DeBry (1999), the development of comprehensive structural models for the mt protein-coding genes is crucial for partitioning the data. The five protein gene-based partitions used in this study only act as surrogates for structure-based differences in substitution processes.

5. Conclusions

The present finding is that with more conservative treatment (RY-coding), and better fitting ML models

(partitioned likelihood), the mt data favors Theria. The highly conserved nature of an indel between the acceptor and D arms of the tRNA-Serine (UCN) further supports this. Springer et al. (2001) were, without explaining the conflicting signals, critical of the use of mt data for inferring deep-level relationships among mammals. However, mt sequences have some valuable properties, such as confidence in orthology, lack of recombination, and ease of collection from a broad range of organisms. It is important to ascertain whether ML model partitioning for RY-coded data will more generally increase the reliability with which deep-level phylogenies can be reconstructed.

Despite the conclusion that mt data favors Theria, the current study provides some vindication for supporters of Marsupionta (Gregory, 1947; Janke et al., 1996, 2002; Kühne, 1973) and others who argued against the orthodoxy of monotremes being far removed from the therian crown group. An auxiliary aim (detailed elsewhere, Phillips, 2002) was to estimate divergence dates between the mammalian sub-classes, from the ML tree that was derived from the RY-coded PTN12 + RNArt dataset. The monotreme/therian divergence estimate of ≈160–180 million years before present is not controversial (see Kielan-Jaworowska et al., 2002; Rowe, 1993), but it was inferred that monotremes diverged from therians only 10–20 million years before marsupials and placentals split. Such a close relationship between monotremes and therians is consistent with recent questioning of some of the "text book" therian synapomorphies, such as oviparity (Zeller, 1999), tribosphenic (shearing and grinding) molars (see Lou et al., 2001) and a highly mobile shoulder girdle (Ji et al., 1999). However, further sequencing of nuclear genes from monotremes and non-mammals is desirable, to provide independent confirmation of our current conclusion that monotremes are closely related to marsupials and placentals; intermediate between the classic morphological hypothesis of monotremes being a very ancient divergence, and the Marsupionta hypothesis.

Acknowledgments

We thank Axel Janke for a pre-release copy of the echidna mt-genome sequence and Kerryn Slack for useful comments on the manuscript. This work was supported by the Marsden Fund.

References

Adachi, J., Hasegawa, M., 1996. In: MOLPHY Version 2.3: Programs for Molecular Phylogenetics based on Maximum Likelihood, vol. 28. Computer Science Monographs of the Institute of statistical Mathematics, Tokyo, pp. 1–150.

Adachi, J., Waddell, P.J., Martin, W., Hasegawa, M., 2000. Plastid genome phylogeny and a model of amino acid substitution for proteins encoded by chloroplast DNA. J. Mol. Evol. 50, 348–358.

Amrine, H.M., Springer, M.S., 1999. Maximum likelihood analysis of the tethythere hypothesis based on a multigene data set and a comparison of different models of sequence evolution. J. Mammal. Evol. 6, 161–176.

Belov, K., Hellman, L., Cooper, D.W., 2002. Characterization of immunoglobin γ1 from a monotreme, *Tachyglossus aculeatus*. Immunogenetics 53, 1065–1071.

Cao, Y., Fujiwara, M., Nikaido, M., Okada, N., Hasegawa, M., 2000a. Interordinal relationships and timescale of eutherian evolution as inferred from mitochondrial genome data. Gene 259, 149–158.

Cao, Y., Sorenson, M.D., Kumazawa, Y., Mindell, D.P., Hasegawa, M., 2000b. Phylogenetic position of turtles among amniotes: evidence from mitochondrial and nuclear genes. Gene 259, 139–148.

Carrick, F.N., Hughes, R.L., 1982. Aspects of the structure and development of monotreme spermatozoa and their relevance to the evolution of mammalian sperm morphology. Cell Tissue Res. 222, 127–141.

Caterino, M.S., Reed, R.D., Kuo, M.M., Sperling, F.A.H., 2001. A partitioned likelihood analysis of swallowtail butterfly phylogeny (Lepidoptera: Papilionidae). Syst. Biol. 50, 106–127.

Cavender, J.A., Felsenstein, J., 1987. Invariants of phylogenies in a simple case with discrete states. J. Classif. 4, 57–71.

DeBry, R.W., 1999. Maximum likelihood analysis of gene-based and structure-based process partitions, using mammalian mitochondrial genomes. Syst. Biol. 48, 286–299.

Felsenstein, J., 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17, 368–376.

Foster, P.G., Jermiin, L.S., Hickey, D.A., 1997. Nucleotide composition bias affects amino acid content in proteins coded by animal mitochondria. J. Mol. Evol. 44, 282–288.

Foster, P.G., Hickey, D.A., 1999. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. J. Mol. Evol. 48, 284–290.

Galtier, N., Gouy, M., 1998. Inferring pattern and process: maximum likelihood implementation of a nonhomogeneous model of DNA sequence evolutionfor phylogenetic analysis. Mol. Biol. Evol. 15.

Gemmell, N.J., Westerman, M., 1994. Phylogenetic relationships within the class Mammalia: a study using mitochondrial 12S RNA sequences. J. Mammal. Evol. 2, 3–23.

Gilbert, N., Labuda, D., 2000. Evolutionary inventions and continuity of CORE-SINEs in mammals. J. Mol. Biol. 298, 365–377.

Gregory, W.K., 1947. The monotremes and the palimpsest theory. Bull. Am. Mus. Nat. Hist. 88, 1–52.

Griffiths, M., 1978. The Biology of the Monotremes. Academic Press, New York.

Gutell, R., Grey, M., Schnare, M., 1993. A compilation of large subunit (23S and 23S-like) ribosomal RNA structures. Nucleic Acids Res. 21, 3055–3074.

Haddrath, O., Baker, A.J., 2001. Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. Proc. R. Soc. Lond. B 268, 939–945.

Härlid, A., Arnason, U., 1999. Analysis of mitochondrial DNA nest ratite birds within the Neognathae—supporting a neotenous origin of ratite morphological characters. Proc. R. Soc. Lond. B 266, 1–5.

Hu, Y., Wang, Y., Lou, Z.-X., Li, C., 1997. A new symmetrodont mammal from China and its implications for mammalian evolution. Nature 390, 137–142.

Janke, A., Gemmell, N.J., Feldmaier-Fuchs, G., von Haeseler, A., Pääbo, S., 1996. The mitochondrial genome of a monotreme—the platypus (*Ornithorhynchus anatinus*). J. Mol. Evol. 42, 153–159.

Janke, A., Xu, X., Arnason, U., 1997. The complete mitochondrial genome of the wallaroo (*Macropus robustus*) and the phylogenetic relationship of among Monotremata, Marsupialia and Eutheria. Proc. Natl. Acad. Sci. USA 94, 1276–1281.

Janke, A., Erpenbeck, D., Nilsson, M., Arnason, U., 2001. The mitochondrial genomes of the iguana (*Iguana iguana*) and the caiman (*Caiman crocodylus*): implications for amniote phylogeny. Proc. R. Soc. Lond. B 268, 623–631.

Janke, A., Magnell, O., Weiczorek, G., Westerman, M., Arnason, U., 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.

Ji, Q., Lou, Z.-X., Ji, S., 1999. A Chinese triconodont mammal and mosaic evolution of the mammalian skeleton. Nature 398, 326–330.

Kielan-Jaworowska, Z., Cifelli, R., Lou, Z.-X., 2002. Dentition and relationships of the Jurassic mammal *Shuotherium*. Acta Palaeontol. Pol. 47, 479–486.

Killian, J.K., Byrd, J.C., Jirtle, J.V., Munday, B.L., Stoskopf, M.K., Macdonald, R.G., Jirtle, R.L., 2000. M6P/IGF2R imprinting evolution in mammals. Mol. Cell 5, 707–716.

Killian, J.K., Buckley, T.R., Stewart, N., Munday, B.L., Jirtle, R.L., 2001. Marsupials and eutherians reunited: genetic evidence for the Theria hypothesis of mammalian evolution. Mammal. Genome 12, 513–517.

Kirsch, J.A.W., Mayer, G.C., 1998. The platypus is not a rodent: DNA hybridization, amniote phylogeny and the palimpsest theory. Philos. Trans. R. Soc. Lond. B 353, 1221–1237.

Kishino, H., Hasegawa, M., 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence dataand the branching order Hominidae. J. Mol. Evol. 29, 170–179.

Krajewski, C., Fain, M.G., Buckley, L., King, D.G., 1999. Dynamically heterogenous partitions and phylogenetic inference: an evaluation of analytical strategies with cytochrome *b* and ND6 gene sequences in cranes. Mol. Phylogenet. Evol. 13, 302–313.

Krajewski, C., Blacket, M.J., Westerman, M., 2000. DNA sequence analysis of familial relationships among dasyuromorphian marsupials. J. Mammal. Evol. 7, 95–108.

Kullander, K., Carlson, B., Hallböök, F., 1997. Molecular phylogeny and evolution of the neurotrophins of monotremes and marsupials. J. Mol. Evol. 45, 311–321.

Kumazawa, Y., Ota, H., Nishida, M., Ozawa, T., 1998. The complete nucleotide sequence of a snake (*Dinodon semicarinatus*) mitochondrial genome with two identical control regions. Genetics 150, 313–329.

Kühne, W.G., 1973. The systematic position of monotremes reconsidered (Mammalia). Zeitschrift für Morphologie der Tiere 75, 59–64.

Lanyon, S.M., 1988. The stochastic mode of molecular evolution: what consequences for systematic investigators? Auk 105, 565–573.

Lee, M.-H., Shroff, R., Cooper, S.J.B., Hope, R., 1999. Evolution and molecular characterization of a b-globin gene from the Australian echidna *Tachyglossus aculeatus* (Monotremata). Mol. Phylogenet. Evol. 12, 205–214.

Lin, Y.-H., McLenachan, P.A., Gore, A.R., Phillips, M.J., Ota, R., Hendy, M., Penny, D., 2002. Four new mitochondrial genomes, and the increased stability of evolutionary trees of mammals from improved taxon sampling. Mol. Biol. Evol. 19, 2060–2070.

Lockhart, P.J., Steel, M., Hendy, M., Penny, D., 1994. Recovering evolutionary trees under a more realistic model of sequence evolution. Mol. Biol. Evol. 11, 605–612.

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

Lou, Z.-X., Cifelli, R., Kielan-Jaworowska, Z., 2001. Dual origin of tribosphenic mammals. Nature 409, 53–57.

Lou, Z.-X., Kielan-Jaworowska, Z., Cifelli, R., 2002. In quest for a phylogeny of Mesozoic mammals. Acta Palaeontol. Pol. 47, 1–78.

Luckett, W.P., Zeller, U., 1989. Developmental evidence for dental homologies in the monotreme *Ornithorhynchus* and its systematic implications. Zeitschrift für Säugetierekunde 54, 193–204.

Lyons-Weiler, J., 2000. RASA 3.0.2 Turbo for Macintosh and manual. Available from <http://test1.bio.psu.edu/LW/rasatext.html>.

Madsen, O., Scally, M., Douady, C.J., Kao, D., DeBry, R.W., Adkins, R., Amrine, H.M., Stanhope, M.J., de Jong, W.W., Springer, M.S., 2001. Parallel adaptive radiations in two major clades of placental mammals. Nature 409.

Marshall, L.G., 1979. Evolution of metatherian and eutherian characters: a review based on cladistic methodology. Zool. J. Linn. Soc. (Lond). 66, 369–410.

McKenna, M.C., 1987. Molecular and morphological analysis of high-level mammalisn interrelationships. In: Patterson, C. (Ed.), Molecules and Morphology in Evolution: Conflict or Compromise? Columbia University Press, New York, pp. 55–93.

Messer, M., Weiss, A.S., Shaw, D.C., Westerman, M., 1998. Evolution of monotremes: phylogenetic relationship to marsupials and eutherians, and estimation of divergence dates on α-Lactalbumin amino acid sequences. J. Mammal. Evol. 5, 95–105.

Mindell, D.P., Sorenson, M.D., Dimcheff, D.E., Hasegawa, M., Ast, J.C., Yuri, T., 1999. Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. Syst. Biol. 48, 138–152.

Moores, A.Ø., Holmes, E.C., 2000. The evolution of base composition and phylogenetic inference. Trends Ecol. Evol. 15, 365–369.

Mouchaty, S.K., Gullberg, A., Janke, A., Arnason, U., 2000. The phylogenetic position of the Talpidae within Eutheria based on analysis of complete mitochondrial sequences. Mol. Biol. Evol. 17, 60–67.

Murphy, W.J., Eizirik, E., Johnson, W.E., Zhang, Y.P., Ryder, O.A., O'Brien, S.J., 2001a. Molecular phylogenetics and the origins of placental mammals. Nature 409.

Murphy, W.J., Eizirik, E., O'Brien, S.J., Madsen, O., Scally, M., Douady, C.J., Teeling, E.C., Ryder, O.A., Stanhope, M.J., de Jong, W., Springer, M.S., 2001b. Resolution of the early placental mammal radiation using bayesian phylogenetics. Science 294, 2348–2351.

Naylor, G.J.P., Brown, W.M., 1998. Amphioxus mitochondrial DNA, Chordate phylogeny, and the limits of inference based on comparisons of sequences. Syst. Biol. 47, 61–76.

Ota, R., Penny, D., 2003. Estimating changes in mutational mechanisms of evolution. J. Mol. Evol. (in press).

Parrington, F.R., 1974. The problem of the origin of the monotremes. J. Nat. Hist. 8, 421–426.

Penny, D., Hasegawa, M., 1997. The platypus put in its place. Nature 387, 549–550.

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

Phillips, M.J., Lin, Y.-H., Harrison, G.L., Penny, D., 2001. Mitochondrial genomes of a bandicoot and a brushtail possum confirm the monophyly of Australiadelphian marsupials. Proc. R. Soc. Lond. B 268, 1533–1538.

Phillips, M.J., 2002. Monotremata, Marsupialia and Placentalia: Inferring Phylogenetic Relationships from Molecular and Morphological Data Institute of Molecular BioSciences. Massey University, Palmerston North.

Rambaut, A., 1996. Sequence alignment Editor. Available at: http://evolve.zps.ox.ac.uk/software/index.html. Oxford University.

Renfree, M.B., 1993. Ontogeny, genetic control, and phylogeny of female reproduction in monotreme and therian mammals. In: Szalay, F.S., Novacek, M.J., McKenna, M.C. (Eds.), Mammalian Phylogeny: Mesozoic Differentiation, Multituberculates, Mono-

tremes, Early Therians and Marsupials. Springer-Verlag, New York, pp. 4–20.
Retief, J.D., Winkfein, R.J., Dixon, G.H., 1993. Evolution of monotremes: the sequences of the protamine P1 genes of platypus and echidna. Eur. J. Biochem. 218, 457–461.
Reyes, A., Gissi, C., Pesole, G., Saccone, C., 1998a. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. Mol. Biol. Evol. 15, 957–966.
Reyes, A., Pesole, G., Saccone, C., 1998b. Complete mitochondrial DNA sequence of of the fat dormouse, *Glis glis*: further evidence of rodent paraphyly. Mol. Biol. Evol. 15, 499–505.
Rougier, G.W., Wible, J.R., Hopson, J.A., 1996. Basicranial anatomy of *Priacodon fruitaensis* (Triconodontidae, Mammalia) from the Late Jurassic of Colorado, and a reappraisal of mammalia form interrelationships. Am. Mus. Nov. 3183, 1–38.
Rowe, T., 1988. Definition, diagnosis, and origin of Mammalia. J. Vertebr. Paleontol. 8, 241–264.
Rowe, T., 1993. Phylogenetic systematics and the early history of mammals. In: Szalay, F.S., Novacek, M.J., McKenna, M.C. (Eds.), Mammalian Phylogeny: Mesozoic Differentiation, Multituberculates, Monotremes, Early Therians and Marsupials. Springer-Verlag, New York, pp. 129–145.
Schmitz, J., Ohme, M., Zischler, H., 2002. The complete mitochondrial sequence of *Tarsius bancanus*: evidence for extensive nucleotide compositional plasticity of primate mitochondrial DNA. Mol. Biol. Evol. 19, 544–553.
Selwood, L., 1994. Development of early cell lineages in marsupial embryos: and overview. Reprod. Fert. Dev. 6, 507–527.
Sereno, P.C., McKenna, M.C., 1995. Cretaceous multituberculate skeleton and the early evolution of the mammalian shoulder girdle. Nature 377, 144–147.
Simmons, M.P., Ochoterena, H., Freudenstein, J.V., 2002. Amino acid vs. nucleotide characters: challenging preconceived notions. Mol. Phylogenet. Evol. 24, 78–90.
Singer, G.A.C., Hickey, D.A., 2000. Nucleotide bias causes genome-wide bias in the amino acid composition of proteins. Mol. Biol. Evol. 17, 1581–1588.
Shimamura, M., Yasue, H., Ohshima, K., Abe, H., Kato, H., Kishiro, T., Goto, M., Munechika, I., Okada, N., 1997. Molecular evidence from retroposons that whales form a clade within even-toed ungulates. Nature 388, 666–670.
Springer, M.S., Douzery, E., 1996. Secondary structure and patterns of evolution among mammalian 12S rRNA molecules. J. Mol. Evol. 43, 357–373.
Springer, M.S., DeBry, R.W., Douady, C.J., Amrine, H.M., Madsen, O., de Jong, W.W., Stanhope, M.J., 2001. Mitochondrial versus nuclear gene sequences in deep-level phylogeny reconstruction. Mol. Biol. Evol. 18, 132–143.
Steel, M., Penny, D., 2000. Parsimony, likelihood, and the role of models in molecular phylogenetics. Mol. Biol. Evol. 17, 839–850.
Sueoka, N., 1962. On the genetic basis of variation and heterogeneity of DNA base composition. Proc. Natl. Acad. Sci. USA 48, 582–592.
Sueoka, N., 1995. Intrastrand parity rules of DNA base composition and usage biases of synonymous codons. J. Mol. Evol. 40, 318–325.
Sullivan, J., Swofford, D.L., 1997. Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. J. Mammal. Evol. 4, 77–86.
Sullivan, J., Swofford, D.L., 2001. Should we use model-based methods for phylogenetic inference when we know that assumptions about among-site rate variation and nucleotide substitution pattern are violated. Syst. Biol. 50, 723–729.
Swofford, D.L., 1998. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinaur Associates, Sunderland, MA.
Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10, 512–526.
Tanaka, M., Ozawa, T., 1994. Strand asymmetry in human mitochondrial DNA mutations. Genomics 22, 327–335.
Tarrio, R., Rodriguez-Trelles, F., Ayala, F.J., 2001. Shared nucleotide composition biases among species and their impact on phylogenetic reconstructions of the Drosophilidae. Mol. Biol. Evol. 18, 1464–1473.
Toyosawa, S., O'hUigin, C., Figueroa, F., Klein, J., 1998. Identification and characterization of amelogenin genes in monotremes, reptiles, and amphibians. Proc. Natl. Acad. Sci. USA 95, 13056–13061.
Tsujii, T., Inoue, S., Takamiya, H., Liszczynsky, H.R., Naora, H., Seno, S., 1992. Morphology of the kidney of the platypus (*Ornithorhynchus anatinus*: Monotremata). Anat. Rec. 234, 348–358.
Waddell, P.J., 1995. Statistical methods of phylogenetic analysis: Including Hadamard conjugations, LogDet transforms, and maximum likelihood. Massey University, Palmerston North, New Zealand.
Waddell, P.J., Cao, Y., Hauf, J., Hasegawa, M., 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.
Waddell, P.J., Kishino, H., Ota, R., 2001. A phylogenetic foundation for comparative mammalian genomics. Genome Inform. 12, 141–154.
Whelan, S., Goldman, N., 1999. Distributions of statistics used for the comparison of models of sequence evolution in phylogenetics. Mol. Biol. Evol. 16, 1292–1299.
Yang, Z., 1996. Maximum-likelihood models for combined analyses of multiple sequence data. J. Mol. Evol. 42, 587–596.
Zardoya, R., Meyer, A., 1998. Complete mitochondrial genome suggests diapsid affinities of turtles. Proc. Natl. Acad. Sci. USA 95, 14226–14231.
Zardoya, R., Meyer, A., 2001. On the origin of and phylogenetic relationships among living amphibians. Proc. Natl. Acad. Sci. USA 98, 7380–7383.
Zeller, U., 1999. Mammalian reproduction: origin and evolutionary transformations. Zool. Anz. 238, 117–130.