Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular Phylogenetics and Evolution 28 (2003) 261–275
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

# Molecular systematics of armadillos (Xenarthra, Dasypodidae): contribution of maximum likelihood and Bayesian analyses of mitochondrial and nuclear genes

Frédéric Delsuc,[a,1] Michael J. Stanhope,[b,2] and Emmanuel J.P. Douzery[a,*]

[a] *Laboratoire de Paléontologie, Paléobiologie et Phylogénie, Institut des Sciences de l'Evolution, Université Montpellier II, Montpellier, France*
[b] *Queen's University of Belfast, Biology and Biochemistry, 97 Lisburn Road, Belfast BT9 7BL, UK*

Received 26 September 2002; revised 3 March 2003

**Abstract**

The 30 living species of armadillos, anteaters, and sloths (Mammalia: Xenarthra) represent one of the three major clades of placentals. Armadillos (Cingulata: Dasypodidae) are the earliest and most speciose xenarthran lineage with 21 described species. The question of their tricky phylogeny was here studied by adding two mitochondrial genes (NADH dehydrogenase subunit 1 [ND1] and 12S ribosomal RNA [12S rRNA]) to the three protein-coding nuclear genes (α2B adrenergic receptor [ADRA2B], breast cancer susceptibility exon 11 [BRCA1], and von Willebrand factor exon 28 [VWF]) yielding a total of 6869 aligned nucleotide sites for thirteen xenarthran species. The two mitochondrial genes were characterized by marked excesses of transitions over transversions—with a strong bias toward CT transitions for the 12S rRNA—and exhibited two- to fivefold faster evolutionary rates than the fastest nuclear gene (ADRA2B). Maximum likelihood and Bayesian phylogenetic analyses supported the monophyly of Dasypodinae, Tolypeutinae, and Euphractinae, with the latter two armadillo subfamilies strongly clustering together. Conflicting branching points between individual genes involved relationships within the subfamilies Tolypeutinae and Euphractinae. Owing to a greater number of informative sites, the overall concatenation favored the mitochondrial topology with the classical grouping of *Cabassous* and *Priodontes* within Tolypeutinae, and a close relationship between *Euphractus* and *Chaetophractus* within Euphractinae. However, low statistical support values associated with almost equal distributions of apomorphies among alternatives suggested that two parallel events of rapid speciation occurred within these two armadillo subfamilies.


*Keywords:* Molecular systematics; Xenarthra; Armadillos; Evolutionary dynamics; Mitochondrial and nuclear markers; Maximum likelihood and Bayesian phylogenetics

## 1. Introduction

The mammalian order Xenarthra (armadillos, anteaters, and sloths) has long been of special interest in attempts at understanding mammalian phylogenetics (Gregory, 1910). Primarily based on seemingly primitive features of the reproductive anatomy and physiology, this enigmatic placental group has been thought to represent the earliest offshoot of the placental tree (McKenna, 1975; Novacek, 1992; Shoshani and McKenna, 1998). Despite this longstanding interest, modern molecular systematic techniques have only recently been applied to the phylogeny of extant xenarthrans (Delsuc et al., 2001, 2002; van Dijk et al., 1999). Moreover, large scale molecular studies have shown that this undoubtedly monophyletic order, is the sole constituent of one of the four major placental clades (Madsen et al., 2001; Murphy et al., 2001a). Also, molecular surveys have emphasized Xenarthra's fundamental importance for understanding the origins of placental mammals, by showing that it is among the oldest placental groups with living representatives

* Corresponding author. Fax: +33-4-67-14-36-10.
*E-mail address:* douzery@isem.univ-montp2.fr (E.J.P. Douzery).
[1] Present address: The Allan Wilson Centre for Molecular Ecology and Evolution, Massey University, Palmerston North, New Zealand.
[2] Present address: Bioinformatics, GlaxoSmithKline, 1250 South Collegeville Road, UP1345, Collegeville, PA 19426, USA.

1055-7903/03/$ - see front matter © 2003 Elsevier Science (USA). All rights reserved.
doi:10.1016/S1055-7903(03)00111-8

(Delsuc et al., 2002; Madsen et al., 2001; Murphy et al., 2001a,b). This order has its origins in South America where it experienced a radiation that gave birth to a prodigious diversity of fossil forms during the Tertiary (McKenna and Bell, 1997; Patterson and Pascual, 1972). Thirty contemporaneous xenarthran species represent the vestiges from this past diversity. Almost confined to the Neotropics, their diversity is composed of 21 armadillo, four anteater, and five sloth species (Vizcaíno, 1995; Wetzel, 1981).

Both morphological (Engelmann, 1985; Patterson et al., 1989, 1992) and molecular (Delsuc et al., 2001, 2002; Madsen et al., 2001; Murphy et al., 2001a,b; van Dijk et al., 1999) studies unambiguously support the division of Xenarthra into the suborders Cingulata, i.e., armadillos (Dasypoda: Dasypodidae) and Pilosa, i.e., anteaters (Vermilingua: Myrmecophagidae) plus sloths (Folivora: Megalonychidae and Bradypodidae). The phylogenetic relationships and taxonomy of living pilosans are well understood. The two modern tree-sloth's genera *Bradypus* (three-toed sloths) and *Choloepus* (two-toed sloths) are classified in the two distinct families Bradypodidae and Megalonychidae, respectively, reflecting their numerous morphological differences and a possible diphyletic origin from two separate fossil groups (Greenwood et al., 2001; Höss et al., 1996; Webb, 1985). Regarding anteaters, the giant anteater (*Myrmecophaga tridactyla*) and the tamanduas (*Tamandua*) are classically associated to the exclusion of the arboreal pygmy anteater (*Cyclopes didactylus*). Such a relationship, first proposed on the basis of myological (Reiss, 1997) and morphological (Gaudin and Branham, 1998) characters, has been confirmed by molecular studies which supported an early emergence of the pygmy anteater within Vermilingua (Delsuc et al., 2001, 2002).

Of particular interest are the relationships among extant armadillos. Cingulata represents the earliest and most speciose xenarthran lineage with 21 living species classified into eight genera (Vizcaíno, 1995; Wetzel, 1985). This ecologically and morphologically diverse group is currently divided into the five tribes Chlamyphorini, Euphractini, Priodontini, Tolypeutini, and Dasypodini (McKenna and Bell, 1997; Wetzel, 1985). The tribe Chlamyphorini contains two species of fairy armadillos (genus *Chlamyphorus*) restricted to the sandy plains and pampas of Northern Argentina, Paraguay, and Bolivia (Wetzel, 1985). Little is known about the biology and conservation status of these cryptic animals, at least partly because of their mainly nocturnal and subterranean way of life (Meritt, 1985). Hairy armadillos (Euphractini) form a zoogeographically homogeneous group centered in dry open habitats from the high puna of Bolivia to pampas and savannas of Brazil, Paraguay, and Argentina (Wetzel, 1985). The three genera *Chaetophractus*, *Euphractus*, and *Zaedyus* are morphologically very close and their interrelationships remain puzzling (Engelmann, 1985). The tribe Priodontini groups the endangered giant armadillo (*Priodontes maximus*) with naked-tailed armadillos from the genus *Cabassous*. These two fossorial genera possess characteristically shaped carapaces and enlarged manus claws specialized for digging (Wetzel, 1985). They also share the particularity of having unusual spoon-shaped spermatozoon that are among the largest that can be found in mammals (Cetica et al., 1998). The tribe Tolypeutini includes only two species of three-banded armadillos (*Tolypeutes*) which are the only members of the family capable of entirely rolling into a ball to escape predators (Wetzel, 1985). Finally, the tribe Dasypodini comprises seven species of long-nosed armadillos classified in the single genus *Dasypus* (Vizcaíno, 1995; Wetzel and Mondolfi, 1979) which are unique in being the only vertebrates to reproduce by obligate monozygotic polyembryony (Galbreath, 1985; Loughry et al., 1998). Among them, the nine-banded armadillo (*Dasypus novemcinctus*) has the largest distribution, as a consequence of its recent and ongoing invasion of the Southern United States (Taulman and Robbins, 1996). Interestingly, at least three species from this genus (*D. novemcinctus*, *D. hybridus*, and *D. sabanicola*) are the only known animals, with the exception of human, in which the causative agent of leprosy (*Mycobacterium leprae*) can develop both naturally or experimentally (Storrs and Burchfield, 1985). This unforeseen attribute, coupled with the systematic production of clonal sibships, has conferred armadillos with great promise for medical research, and the nine-banded armadillo was established early on as a model for leprosy studies (Storrs et al., 1974). The establishment of a well defined taxonomic framework for armadillos might thus provide directions for investigating the potential of other species as biomedical models, since the development of an anti-leprosy vaccine has proven difficult using the nine-banded armadillo (Storrs, 1999).

A previous study based on nuclear markers (Delsuc et al., 2002) failed to fully resolve the phylogeny of armadillos but identified three major lineages which correspond to the subfamilies Dasypodinae (*Dasypus*), Tolypeutinae (*Tolypeutes*, *Cabassous*, *Priodontes*), and Euphractinae (*Euphractus*, *Chaetophractus*, *Zaedyus*), previously defined on morphological grounds by McKenna and Bell (1997). The lack of resolution within the subfamilies Tolypeutinae and Euphractinae might stand in the relatively slow evolutionary rate of these nuclear genes (Delsuc et al., 2002). Therefore, we turned to two faster evolving mitochondrial genes with the hope of collecting more informative sites to discern between alternative phylogenetic hypotheses within these two clades. Two mitochondrial markers undergoing potentially contrasted mutational and selective constraints were chosen: the protein-coding NADH dehydrogenase subunit 1 (ND1), and the ribosomal

RNA-coding 12S rRNA. These two mitochondrial genes have been previously used for example to examine the relationships between xenarthrans and pangolins (Cao et al., 1998; Delsuc et al., 2001). Complete sequences of these genes were obtained for 13 xenarthran species representing all living genera except the rare *Chlamyphorus*. The new mitochondrial sequences were added to the data set of Delsuc et al. (2002) composed of the three protein-coding nuclear genes: α2B adrenergic receptor (AD RA2B), breast cancer susceptibility exon 11 (BRCA1), and von Willebrand factor exon 28 (VWF), to obtain a combined data set of 6869 aligned nucleotide sites. We compared the evolutionary dynamics of these five genes in regards to pattern and rate of nucleotide substitutions, and employed maximum likelihood (ML; Felsenstein, 1981) and Bayesian phylogenetics (Huelsenbeck et al., 2001), to reconstruct the phylogeny of armadillos, with special reference to the subfamilies Tolypeutinae and Euphractinae.

## 2. Materials and methods

### 2.1. Taxon sampling and data acquisition

We sampled 13 xenarthran species spanning 12 of the 13 extant genera (Table 1). Total DNAs were extracted from 95% ethanol preserved tissue samples stored in the mammalian tissue collection of the Institut des Sciences de l'Evolution de Montpellier (Catzeflis, 1991). Complete mitochondrial 12S rRNA sequences (12S rRNA) were amplified using primers R1 and S2 (Douzery and Catzeflis, 1995). Complete sequences of the NADH dehydrogenase 1 gene (ND1) were amplified using primers R11 = 5′-TTTCTCCCAGTACGAAAGGAC-3′ (forward) and N1 = 5′-CTATTATTTACTCTATCAAAGTAA-3′ (reverse) located in the 16S rRNA and Ile tRNA genes, respectively. PCR products were purified from 1% agarose gels using Amicon Ultrafree-DA columns (Millipore). Automatic sequencing (Big Dye Terminator cycle kit) of purified PCR products was performed on both strands on an ABI 310 (PE Applied Biosystems) using PCR primers plus M1 = 5′-GGTAATTGCRTAARAC TTAAACCTTT-3′ (forward) located in the Leu tRNA and the ND1 internal primer N2 = 5′-TCRGCTADR AARAATAGGGC-3′ (reverse). The 14 new xenarthran sequences have been deposited in the EMBL databank. Taxonomy and accession numbers referring to all sequences used in this study are indicated in Table 1.

### 2.2. Sequence alignment

The ND1 and 12S rRNA new sequences were manually aligned with existing sequences obtained by Cao et al. (1998) and Delsuc et al. (2001), respectively, using the ED editor of the MUST package (Philippe, 1993). We excluded a glutamic acid repeat region of ADRA2B, a 21 nucleotide long repeated region of BRCA1 (Madsen et al., 2001), and also hypervariable regions of 12S rRNA (Springer and Douzery, 1996). All introduced gaps were treated as missing data in subsequent analyses. The five individual data sets of 13 xenarthran taxa

Table 1
Taxonomy and accession numbers for sequences used in this study

| Taxonomy | Species | Common name | ADRA2B | BRCA1 | VWF | ND1 | 12S rRNA |
|---|---|---|---|---|---|---|---|
| Xenarthra | | | | | | | |
| *Pilosa* | | | | | | | |
| Bradypodidae | *Bradypus tridactylus* | Pale-throated three-toed sloth | AJ251179 | AF284002 | U31603 | AB011218[a] | AF038022 |
| Megalonychidae | *Choloepus didactylus* | Southern two-toed sloth | AJ427375 | AF484229 | AJ278160 | AJ505830[b] | AJ278152 |
| Myrmecophagidae | *Myrmecophaga tridactyla* | Giant anteater | AJ427373 | AF484232 | AJ278157 | AJ505831[b] | AJ278153 |
| | *Tamandua tetradactyla* | Collared anteater | AJ427374 | AF284001 | AJ278161 | AB011216 | AJ278154 |
| | *Cyclopes didactylus* | Pygmy anteater | AJ315943 | AF484231 | AJ278156 | AJ505832[b] | AJ278155 |
| *Cingulata* | | | | | | | |
| Dasypodidae | | | | | | | |
| *Dasypodinae* | *Dasypus novemcinctus* | Nine-banded armadillo | AJ427366 | AF484222 | AJ278158 | Y11832 | Y11832 |
| | *Dasypus kappleri* | Great long-nosed armadillo | AJ427367 | AF484223 | AJ427361 | AJ505833[b] | AJ505825[b] |
| *Euphractinae* | *Euphractus sexcinctus* | Six-banded armadillo | AJ427368 | AF484224 | AJ427364 | AJ505834[b] | AJ505826[b] |
| | *Chaetophractus villosus* | Larger hairy armadillo | AJ315935 | AF284000 | AF076480 | AJ505835[b] | U61080 |
| | *Zaedyus pichiy* | Pichi | AJ427369 | AF484226 | AJ427365 | AJ505836[b] | AJ505827[b] |
| *Tolypeutinae* | *Tolypeutes matacus* | Southern three-banded armadillo | AJ427370 | AF484227 | AJ427362 | AJ505837[b] | AJ505828[b] |
| | *Cabassous unicinctus* | Southern naked-tailed armadillo | AJ427371 | AF484228 | AJ278159 | AB011217 | AJ278151 |
| | *Priodontes maximus* | Giant armadillo | AJ427372 | AF484225 | AJ427363 | AJ505838[b] | AJ505829[b] |

[a] Species *B. variegatus* (Brown-throated three-toed Sloth).
[b] Sequences new to this study.

are: ADRA2B (1158 sites), BRCA1 (2793 sites), VWF (1162 sites), ND1 (957 sites), and 12S rRNA (898 sites). We also considered three additional data sets: the nuclear concatenation of ADRA2B + BRCA1 + VWF (NUC: 5113 sites), the mitochondrial concatenation of ND1 and 12S rRNA (MITO: 1855 sites), and the complete combined data set (COMB: 6968 sites). The five pilosan taxa (anteaters and sloths) were used as outgroups for rooting the cingulate (armadillos) subtree in all subsequent phylogenetic analyses. Alignments are available upon request.

### 2.3. Phylogenetic analyses

Phylogenetic analyses were conducted using maximum likelihood (ML) and Bayesian approaches. These probabilistic methods based upon explicit models of sequence evolution were favored because they are known to be robust to a number of systematic biases of phylogenetic reconstruction (Huelsenbeck et al., 2001; Sullivan and Swofford, 2001; Swofford et al., 2001). A likelihood framework also provides us with the ability to statistically compare competing hypotheses and topologies (Goldman et al., 2000; Huelsenbeck and Rannala, 1997; Whelan et al., 2001).

#### 2.3.1. Maximum likelihood

We first determined the best fitting model of sequence evolution for each data set using Modeltest 3.06 (Posada and Crandall, 1998). The Akaike Information Criterion (AIC; Akaike, 1974) indicated that different models were appropriate to best describe the underlying evolutionary process of these eight datasets: TrN + Γ for ADRA2B, TrN + I + Γ for ND1, TVM + Γ for BRCA1 and NUC, GTR + Γ for 12S rRNA, and the GTR + I + Γ model for VWF, MITO, and COMB (see Posada and Crandall, 2001 for a detailed description of these models). However, to render ML analyses comparable between PAUP* version 4.0b10 (Swofford, 2002) and PAML version 3.11 (Yang, 1997), the latter of which does not allow the use of invariant sites, we chose to use the General Time Reversible model (GTR; Yang, 1994) plus a gamma (Γ) distribution with eight categories (GTR + $\Gamma_8$; Yang, 1996a). Using the same model for each data set also afforded the ability to adequately compare the evolutionary dynamics of the five genes in terms of substitution pattern and rate of evolution. We verified that using this GTR + $\Gamma_8$ model did not change the results of ML searches, in terms of the highest likelihood topologies for any of the datasets. This sensitivity analysis suggested that our results were relatively insensitive to subtle deviations between the GTR + $\Gamma_8$ model and the best fitting model for a particular data set (Buckley and Cunningham, 2002; Buckley et al., 2001).

ML base composition stationarity assumptions were evaluated for each of the eight data sets by the $\chi^2$ test implemented in TREE-PUZZLE 5.0 (Strimmer and von Haeseler, 1996). ML phylogenetic analyses consisted of heuristic searches with tree bisection reconnection (TBR) branch swapping, performed with a neighbor-joining (NJ) starting tree. Optimal base composition, substitution rate matrix, and among site substitution rate heterogeneity parameters, were simultaneously estimated during the ML heuristic searches. Reliability of nodes was estimated by ML bootstrap percentages ($BP_{ML}$) (Felsenstein, 1985) obtained after 500 pseudo-replications, using the previously estimated ML parameters, with NJ starting trees, and TBR branch swapping.

#### 2.3.2. Bayesian approach

The Bayesian approach to phylogenetic reconstruction (Rannala and Yang, 1996; Yang and Rannala, 1997) was implemented using MrBayes 2.01 (Huelsenbeck and Ronquist, 2001). Metropolis-coupled Markov chain Monte Carlo (MCMCMC) sampling was performed with five incrementally heated chains that were simultaneously run for 1,000,000 generations, using the program default priors as starting values for GTR + $\Gamma_8$ model parameters. To check for potentially poor mixing of MCMCMC, each analysis was repeated twice. The convergence of MCMCMC was monitored by examining the value of the marginal likelihood through generations. Convergence of substitution rate and rate heterogeneity model parameters was also checked. As previously observed by Miller et al. (2002), the chains reached convergence more rapidly for the likelihood function than for model parameters. However, the high numbers of generations (1,000,000) run here permitted to reach stationarity of likelihood model parameters for our relatively small data sets. Bayesian posterior probabilities (PP) were obtained from the 50% majority rule consensus of 25,000 trees sampled every 20 generations after removing the 25,000 first trees as the "burn-in" stage. We verified that PP did not undergo significant variations between independently repeated runs even when performing only 200,000 generations instead of 1,000,000 (i.e., PP variations were less than 0.05).

Since Bayesian posterior probabilities have recently been shown to provide overestimates of phylogenetic support (Suzuki et al., 2002), we also computed bootstrapped Bayesian posterior probabilities ($BP_{Bay}$) following the new approach of Douady et al. (2003a), as a more conservative measure of Bayesian support. First, 100 bootstrap pseudo-replicates drawn irrespective of codon positions were generated for each data set using the program CODONBOOTSTRAP 3.0 (available from Jonathan P. Bollback at http://brahms.ucsd.edu/software.html). Second, for each pseudo-replicate, MCMCMC sampling of trees was performed as previously described, except that the five chains were run for only 100,000 generations. A conservative one-half of the 5000 trees sampled from the posterior probability

distribution were systematically discarded as "burn-in," to maximize the probability that the chains reached stationarity in each bootstrap resampled data set. Finally, we picked $BP_{Bay}$ from the overall 50% majority rule consensus of the $100 \times 2500 = 250,000$ saved trees.

### 2.4. Statistical tests of alternative hypotheses

Since our genes belong to different cellular compartments (nuclear and mitochondrial) and markedly differ in function, it is likely they evolve under different selective pressures. To accommodate expected differences in evolutionary dynamics between genes, and especially between nuclear and mitochondrial loci, we followed the approach of Yang (1996b) for analyzing multiple genes. This approach, which allows each gene to evolve under its own model in a "partitioned likelihood model," resulted in a general and significant increase in log-likelihood. For example, in the case of the combined data set, the parameter rich model incorporating five independent sets of 31 free parameters—three base frequencies, five GTR rates, one $\Gamma$ shape, and 22 branch lengths for a 13 taxon tree with a basal trifurcation—more appropriately described the underlying evolutionary processes than a single $GTR + \Gamma_8$ model. The AIC decreased from 58,085.86 ($= 2 \times 29011.93 + 2 \times 31$) to 56208.06 ($= 2 \times 27949.03 + 2 \times 31 \times 5$) when an independent set of base, $GTR + \Gamma_8$, and branch length parameters was attributed to each gene. Partitioned ML models were used in PAML to compute log-likelihoods and confidence probability values for alternative phylogenetic hypotheses.

A posteriori selected topologies (Goldman et al., 2000) were statistically compared to the best ML topology by the test of Kishino and Hasegawa (1989) applying the Shimodaira and Hasegawa (1999) correction for multiple comparisons (SH test). One potential problem in using the SH test is that the results of the test are in part dependent on the number of competing topologies examined (Goldman et al., 2000). As an alternative, the use of the parametric SOWH test (Swofford et al., 1996) has been advocated (Goldman et al., 2000). However, this test was not used here, because it is associated with high computational burden and has been shown to be misleading in some cases, being prone to Type I errors, even when using the most sophisticated models of sequence evolution (Buckley, 2002). Moreover, the SH test is thought to be accurate when the number of candidate trees is small (Shimodaira, 2002).

## 3. Results and discussion

### 3.1. Evolutionary properties of the five genes

Striking differences in base composition are displayed by the five genes. Within Xenarthra, the nuclear genes ADRA2B and VWF appear G + C-rich (72.1 and 65.4%, respectively), whereas BRCA1 is rather A + T-rich (61.9%). The two mitochondrial markers 12S rRNA and ND1 present exactly the same excess of A + T (58%) relative to G + C (42%). Although the frequencies of adenine and thymine are nearly identical for these two genes (37/38% for A and 20/21% for T), ND1 exhibits a notably small proportion (7.0%) of guanine. Base composition stationarity assumptions were evaluated by a $\chi^2$ test which compares the nucleotide composition of each sequence to the frequency distribution assumed in the maximum likelihood model. All gene sets appeared homogenous in terms of base composition with only *Cyclopes* deviating from stationarity for ND1 ($P = 0.04$).

Comparative plots of ML estimated $GTR + \Gamma_8$ parameters provides an easily visualized means of contrasting the evolutionary dynamics of the five genes (Fig. 1). This graphical representation highlights the fundamental differences between nuclear and mitochondrial genes for Xenarthra in terms of pattern and rate of nucleotide substitutions. The two mitochondrial genes ND1 and 12S rRNA are characterized by marked excesses of transitions over transversions and exhibit elevated evolutionary rates, approximately five- and twofold faster, respectively, than the fastest nuclear gene (ADRA2B) (Fig. 1). These faster rates of molecular evolution are coupled with strong among site rate heterogeneity, as indicated by low values of the $\Gamma$ shape parameter ($\alpha = 0.22$ and 0.29, respectively) (Fig. 1). The strong rate heterogeneity disclosed in the ND1 gene might find its origin in contrasted patterns of substitution and evolutionary rates among its three codon positions. Also observing such a strong rate heterogeneity in the 12S rRNA was not unexpected as it has been previously reported in the case of sigmodontine rodents (Sullivan et al., 1995). It likely reflects the mosaic structure of this gene, consisting of slowly evolving stems, and faster evolving loops, related to secondary base-pairing interactions (Springer and Douzery, 1996). A strong bias toward C–T transitions is detected here for this gene (Fig. 1). Thus, saturation might be reached at fast evolving positions, particularly those within loops, which are prone to this type of change (Springer and Douzery, 1996). Among nuclear genes, BRCA1 was the slowest evolving gene and exhibited the lowest among site rate heterogeneity ($\alpha = 1.41$; Fig. 1). This relative among site rate homogeneity is explained by the atypical evolutionary pattern of this gene, with first, second and third codon positions evolving in a strikingly similar manner (Adkins et al., 2001; Delsuc et al., 2002). These differences in evolutionary dynamics between genes are reflected in the number and percentage of variable sites observed in each individual data set. For example, about 49.6% of the ND1 sites appeared variable whereas only 24.3% in BRCA1. Given such

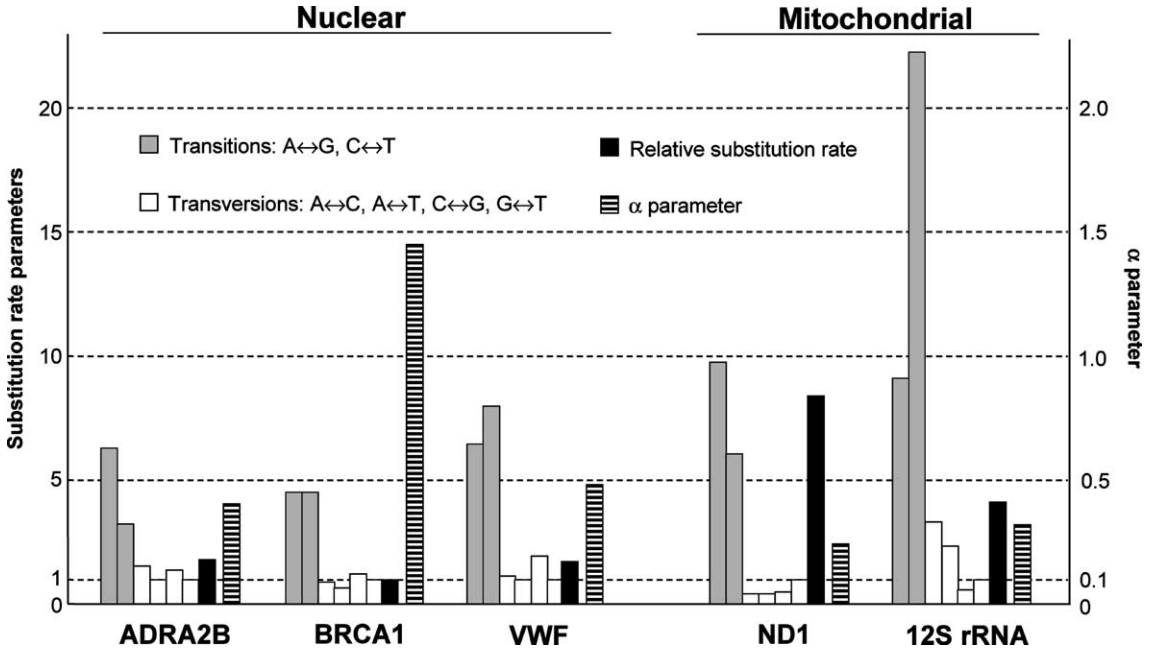


Fig. 1. Comparative evolutionary dynamics of the five genes within Xenarthra. Maximum likelihood estimates of substitution rate and among site heterogeneity parameters under the GTR + $\Gamma_8$ model are graphically represented. From left to right: GTR rate matrix parameters for transitions A ↔ G and C ↔ T (gray bars) and transversions A ↔ C, A ↔ T, C ↔ G, and G ↔ T = 1.00 (white bars), rate between genes relative to the slowest one, i.e., BRCA1 (black bars), and $\alpha$ parameter of the $\Gamma_8$ distribution (horizontally hatched bars).

differences in nucleotide variability, these genes are expected to markedly differ in their phylogenetic resolving power, mitochondrial genes providing more informative sites at taxonomic level such as the one of family Dasypodidae.

### 3.2. Phylogenetic results

#### 3.2.1. Results from individual genes

The maximum likelihood phylograms inferred from each of the five individual genes are presented in Fig. 2. At first glance, it is striking to note that the analysis of each individual gene yielded phylogenies which are topologically incompatible with each other (Fig. 2). However, each gene supported the monophyly of the three armadillo subfamilies: Dasypodinae, Tolypeutinae, and Euphractinae with the noteworthy exception of the 12S rRNA (Fig. 2; Table 2). This gene produced a topology different from the other four by placing the three-banded armadillo *Tolypeutes* as the sister group of all other Dasypodidae, and thus breaking the otherwise strongly supported monophyly of the subfamily Tolypeutinae (Fig. 2). However, this alternative topology is poorly supported and this gene tends towards a general lack of resolution, except for relationships among euphractine armadillos. The topological instability and general lack of resolution exhibited by this gene may be a consequence of nucleotide saturation related to the strong mutational bias toward C–T transitions (Fig. 1). Moreover, model sensitivity analyses revealed that this topology is unstable and largely dependent upon the likelihood model used. Notably, phylogenetic reconstruction with this gene appeared to be sensitive to the number of discrete rate categories used to describe the $\Gamma$ distribution (either 4, 5, 6, 7, or 8; data not shown). These observations underline difficulties to accommodate the strong rate heterogeneity induced by the stem-loop structure and suggest to be cautious when interpreting phylogenetic results issued from the analyses of this gene (Sullivan et al., 1995). The incorporation of structural information in substitution models specific to the evolution of base-paired regions of mammalian mitochondrial RNAs seems promising (Jow et al., 2002). However, frequent and clustered site-specific rate variation has been detected in the mitochondrial 16S rRNA gene of insects (Misof et al., 2002), suggesting that this gene—and presumably other ribosomal molecules—evolves under a covarion-like process. It is thus suggested that phylogenetic analyses of the 12S rRNA xenarthran data set under a covarion-like model might help to recover a topology compatible with other genes. As an alternative explanation for the basal position of *Tolypeutes* among armadillos, we cannot rule out the possibility that the corresponding 12S rRNA sequence of the three-banded armadillo we obtained is a unwittingly amplified divergent mitochondrial-derived nuclear pseudogene. Indeed, even if the use of secondary structure features might help to detect rRNA nuclear paralogs, it has been shown that the identification of 12S rRNA nuclear insertions may prove to be very difficult (Olson and Yoder, 2002).

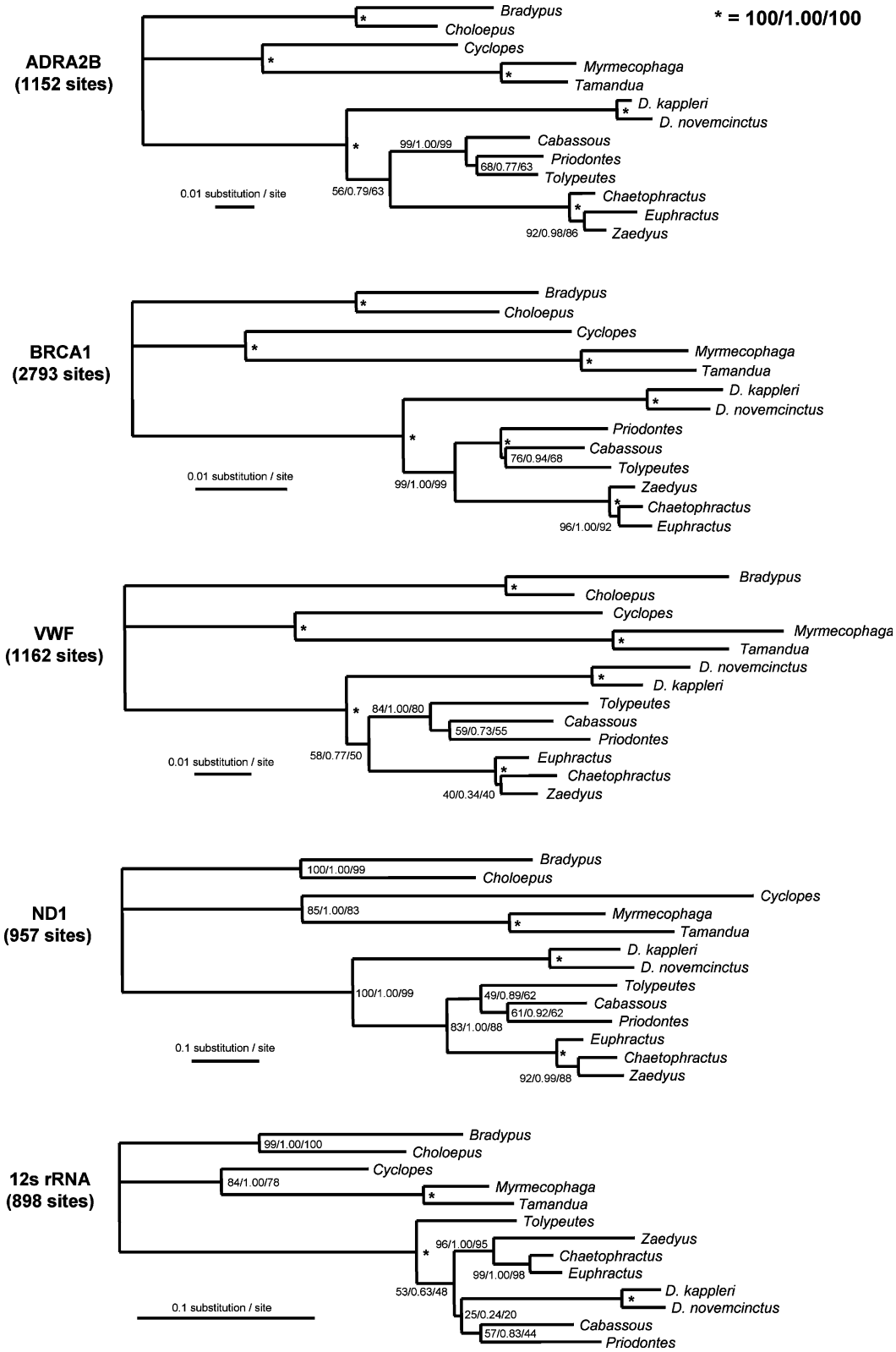


Fig. 2. Maximum likelihood topologies obtained from the five individual data sets under the GTR + $\Gamma_8$ model. Values at nodes indicate maximum likelihood and Bayesian support expressed by maximum likelihood bootstrap ($BP_{ML}$)/Bayesian posterior probability (PP)/bootstrapped posterior probability ($BP_{Bay}$). Nodes marked with a star (*) are supported by 100/1.00/100. Note that different scales are used for nuclear (ADRA2B, BRCA1, and VWF) and mitochondrial (ND1 and 12S rRNA) genes illustrating the large difference in evolutionary rate between these markers. *D.* is an abbreviation for *Dasypus*.

Table 2
Maximum likelihood bootstrap support ($BP_{ML}$), Bayesian posterior probabilities (PP), bootstrapped Bayesian posterior probabilities ($BP_{Bay}$), and ML inferred number of apomorphies (Apo) supporting different alternative phylogenetic relationships between and within the three armadillo subfamilies

| | $BP_{ML}$ | PP | $BP_{Bay}$ | Apo | $BP_{ML}$ | PP | $BP_{Bay}$ | Apo | $BP_{ML}$ | PP | $BP_{Bay}$ | Apo |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ADRA2B | | | | BRCA1 | | | | VWF | | | |
| Dasypodidae | | | | | | | | | | | | |
| Euphractinae + Tolypeutinae | 56 | 0.79 | 63 | 9 | 99 | 1.00 | 100 | 15 | 58 | 0.77 | 50 | 3 |
| Euphractinae + Dasypodinae | 42 | 0.21 | 35 | 13 | — | — | — | 0 | 23 | 0.09 | 26 | 0 |
| Tolypeutinae + Dasypodinae | — | — | — | 0 | — | — | — | 5 | 17 | 0.14 | 20 | 0 |
| Tolypeutinae | | | | | | | | | | | | |
| *Cabassous* + *Priodontes* | — | 0.07 | 10 | 0 | 17 | — | 25 | 1 | 59 | 0.73 | 55 | 5 |
| *Cabassous* + *Tolypeutes* | 26 | 0.16 | 26 | 1 | 76 | 0.94 | 68 | 2 | 16 | 0.10 | 22 | 3 |
| *Priodontes* + *Tolypeutes* | 68 | 0.77 | 63 | 3 | — | — | 7 | 0 | 22 | 0.17 | 17 | 4 |
| Euphractinae | | | | | | | | | | | | |
| *Euphractus* + *Chaetophractus* | — | — | — | 0 | 96 | 1.00 | 92 | 3 | 18 | 0.22 | 27 | 1 |
| *Euphractus* + *Zaedyus* | 92 | 0.98 | 86 | 3 | — | — | — | 0 | 35 | 0.44 | 33 | 1 |
| *Chaetophractus* + *Zaedyus* | 6 | — | 9 | 0 | — | — | — | 0 | 40 | 0.34 | 40 | 1 |
| | ND1 | | | | 12S rRNA | | | | | | | |
| Dasypodidae | | | | | | | | | | | | |
| Euphractinae + Tolypeutinae | 83 | 1.00 | 99 | 25 | 37 | 0.32 | 39 | 13 | | | | |
| Euphractinae + Dasypodinae | 9 | — | — | 16 | 22 | 0.26 | 17 | 2 | | | | |
| Tolypeutinae + Dasypodinae | 6 | — | — | 12 | — | — | — | 0 | | | | |
| Tolypeutinae | | | | | | | | | | | | |
| *Cabassous* + *Priodontes* | 61 | 0.92 | 62 | 13 | 57 | 0.83 | 44 | 11 | | | | |
| *Cabassous* + *Tolypeutes* | 23 | 0.07 | 22 | 6 | 21 | 0.07 | 22 | 14 | | | | |
| *Priodontes* + *Tolypeutes* | — | — | — | 0 | — | — | 8 | 6 | | | | |
| Euphractinae | | | | | | | | | | | | |
| *Euphractus* + *Chaetophractus* | — | — | — | 1 | 99 | 1.00 | 98 | 14 | | | | |
| *Euphractus* + *Zaedyus* | 5 | — | 8 | 2 | — | — | — | 0 | | | | |
| *Chaetophractus* + *Zaedyus* | 92 | 0.99 | 88 | 20 | — | — | — | 0 | | | | |
| | Nuclear | | | | Mitochondrial | | | | Combination | | | |
| Dasypodidae | | | | | | | | | | | | |
| Euphractinae + Tolypeutinae | 100 | 1.00 | 100 | 31 | 86 | 1.00 | 88 | 40 | 100 | 1.00 | 100 | 62 |
| Euphractinae + Dasypodinae | — | — | — | 9 | — | — | — | 12 | — | — | — | 18 |
| Dasypodinae + Tolypeutinae | — | — | — | 11 | 11 | — | 10 | 3 | — | — | — | 9 |
| Tolypeutinae | | | | | | | | | | | | |
| *Cabassous* + *Priodontes* | 23 | 0.12 | 26 | 6 | 69 | 0.99 | 68 | 29 | 54 | 0.87 | 52 | 24 |
| *Cabassous* + *Tolypeutes* | 52 | 0.70 | 50 | 6 | 26 | — | 27 | 20 | 41 | 0.13 | 41 | 20 |
| *Tolypeutes* + *Priodontes* | 25 | 0.18 | 24 | 6 | — | — | — | 6 | 5 | — | — | 18 |
| Euphractinae | | | | | | | | | | | | |
| *Euphractus* + *Chaetophractus* | 35 | 0.13 | 33 | 4 | 54 | 0.83 | 50 | 12 | 60 | 0.97 | 54 | 21 |
| *Euphractus* + *Zaedyus* | 56 | 0.85 | 58 | 4 | — | — | — | 7 | 21 | — | 21 | 15 |
| *Chaetophractus* + *Zaedyus* | 9 | — | 9 | 2 | 43 | 0.17 | 47 | 15 | 19 | — | 25 | 15 |

Apart from the peculiar 12S rRNA topology, all other topologies agreed in grouping the subfamilies Tolypeutinae and Euphractinae with moderate (ADRA2B and VWF) to strong (BRCA1 and ND1) support (Fig. 2; Table 2). However, conflicts exist between genes about relationships within these two subfamilies (Fig. 2; Table 2). Indeed, within tolypeutines, the VWF, 12S rRNA, and ND1 genes favor a *Cabassous* + *Priodontes* clade, whereas ADRA2B prefers *Priodontes* + *Tolypeutes*, and BRCA1 *Cabassous* + *Tolypeutes*, each with almost equivalent moderate support (Fig. 2; Table 2). The situation is even more complicated in the case of the euphractines where the three possible alternatives appear strongly supported by different genes: ADRA2B supports *Euphractus* + *Zaedyus*, BRCA1 and 12S rRNA prefers *Euphractus* + *Chaetophractus*, and VWF and ND1 *Chaetophractus* + *Zaedyus* (Fig. 2; Table 2).

The statistical significance of the topological incongruence between the five individual genes was evaluated by performing crossed SH tests (Delsuc et al., 2002; Huchon et al., 2002). In these tests, the highest-likelihood topologies obtained with individual genes were compared against each other under each of the five character matrices (Table 3). The results indicate that none of the individual data sets significantly rejects the ML topology of the other four, with the exception of the atypical 12S rRNA ML topology under the ADRA2B, BRCA1, and ND1 data sets (Table 3). This reinforces the earlier observation regarding the peculiar nature of the 12S rRNA topology, and the fact that it is likely the result of a phylogenetic artifact related to the particular evolutionary behavior of this gene (see also Sullivan et al., 1995). However, these tests conclude that despite apparently high topological incongruence, each individual gene leads to a phylogenetic estimate compatible with the signal provided by the others. This conclusion is also corroborated by the fact that seemingly incongruent regions between trees are restricted to nodes involving short internal edges, despite being strongly supported in certain cases (Fig. 2). Thus, with the exception of the 12S rRNA tree, individual trees appear to be "different" but are nonetheless "similar" (Penny et al., 1993), based on the results of the conservative SH test.

### *3.2.2. Results from the nuclear, mitochondrial, and combined data sets*

Combining signal from different data sets in phylogenetic analyses has long been a matter of controversy (Huelsenbeck et al., 1996). Here we adopted the "conditional combination" approach which proposes to statistically test the homogeneity of a priori defined data partitions before combining them in an overall approach (Bull et al., 1993). The homogeneity of the signal carried by each individual gene was statistically evaluated by crossed SH tests in which the highest-likelihood topologies obtained with individual genes were compared against those obtained from nuclear (ADRA2B + BRCA1 + VWF), mitochondrial (12S rRNA + ND1) and all gene combinations. These tests indicated that none of the five individual data sets significantly rejects the ML topology resulting from either the nuclear, mitochondrial, or all gene combinations (Table 3). This suggests

Table 3
Results from standard Shimodaira and Hasegawa (1999) crossed tests of topological congruence between the eight data sets

| ML topologies | Data sets | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | ADRA2B | BRCA1 | VWF | ND1 | 12S rRNA | Nuclear | Mitochondrial | Combination |
| ADRA2B | **−4354.35** | 7.98 ($P = 0.41$) | 1.21 ($P = 0.71$) | 11.09 ($P = 0.12$) | 13.84 ($P = 0.06$) | 1.63 ($P = 0.79$) | 13.48 ($P = 0.13$) | 11.21 ($P = 0.43$) |
| BRCA1 | 6.40 ($P = 0.37$) | **−8511.16** | 2.11 ($P = 0.61$) | 9.95 ($P = 0.17$) | 1.96 ($P = 0.68$) | 1.67 ($P = 0.80$) | 4.63 ($P = 0.55$) | 1.87 ($P = 0.82$) |
| VWF | 6.77 ($P = 0.36$) | 7.83 ($P = 0.41$) | **−4364.26** | 0.00 ($P = 1.00$) | 11.29 ($P = 0.11$) | 5.61 ($P = 0.60$) | 1.35 ($P = 0.77$) | 4.16 ($P = 0.73$) |
| ND1 | 6.77 ($P = 0.36$) | 7.83 ($P = 0.41$) | 0.00 ($P = 1.00$) | **−6123.37** | 11.29 ($P = 0.11$) | 5.61 ($P = 0.60$) | 1.35 ($P = 0.77$) | 4.16 ($P = 0.73$) |
| 12S rRNA | 40.07 ($P < 0.05$*) | 69.04 ($P < 0.001$*) | 10.07 ($P = 0.15$) | 16.64 ($P < 0.05$*) | **−4413.94** | 132.71 ($P < 0.001$*) | 13.15 ($P = 0.19$) | 113.80 ($P < 0.001$*) |
| Nuclear | 1.07 ($P = 0.77$) | 5.26 ($P = 0.49$) | 1.48 ($P = 0.69$) | 9.86 ($P = 0.18$) | 12.16 ($P = 0.11$) | **−17691.46** | 10.34 ($P = 0.23$) | 6.22 ($P = 0.59$) |
| Mitochondrial | 6.77 ($P = 0.36$) | 2.57 ($P = 0.71$) | 0.65 ($P = 0.86$) | 5.92 ($P = 0.34$) | 0.78 ($P = 0.86$) | 3.57 ($P = 0.71$) | **−10648.82** | 0.00 ($P = 1.00$) |
| Combination | 6.77 ($P = 0.36$) | 2.57 ($P = 0.71$) | 0.65 ($P = 0.86$) | 5.92 ($P = 0.34$) | 0.78 ($P = 0.86$) | 3.57 ($P = 0.71$) | 0.00 ($P = 1.00$) | **−29011.65** |

*Note.* Log-likelihood values of ML topologies inferred from each data set were computed using each of the eight data matrices and then compared to the corresponding highest log-likelihood value (in bold). The difference in log-likelihood derived from these crossed comparisons and corresponding confidence *P* values of the SH test (between brackets) are also indicated. A * indicates that the tested topology is significantly worse at the 5% level than the best ML topology for the corresponding data set.

that combining individual genes in a priori defined data partitions leads to a phylogenetic estimate compatible with the signal contributed by each individual gene.

The ML phylograms obtained from the nuclear and mitochondrial combinations are compared in Fig. 3. There are clear topological differences between the two phylogenies. Once again the conflicting areas involve relationships within the subfamilies Tolypeutinae and Euphractinae. The nuclear concatenation favors the grouping of *Cabassous* and *Tolypeutes* ($BP_{ML} = 52$/ $PP = 0.70$/$BP_{Bay} = 50$) within Tolypeutinae, whereas the mitochondrial genes support a *Cabassous* + *Priodontes* clade (69/0.99/68). Within Euphractinae, the nuclear partition groups *Euphractus* and *Zaedyus* (56/0.85/58), whereas the mitochondrial combination prefers *Euphractus* + *Chaetophractus* (54/0.83/50). It is, however, noteworthy that these topological conflicts implicate short and statistically poorly supported internal branches, except the possible relationship between *Cabassous* and *Priodontes*, which is moderately supported by the mitochondrial data set (Fig. 3; Table 2). As a consequence of the lack of resolution within these two subfamilies, partitioned SH tests failed to statistically distinguish between alternative hypotheses (Table 4). Apart from these discrepancies, both the nuclear and mitochondrial combinations unequivocally supported the monophyly of the three armadillo subfamilies and strongly favored the early emergence of Dasypodinae within Dasypodidae (Fig. 3; Table 2). Alternatives breaking the monophyly of the Tolypeutinae + Euphractinae clade were statistically rejected by the nuclear data set (Table 4).

The maximum likelihood phylogram obtained from the combined data set is presented in Fig. 4. This topology is fully congruent with the one yielded by the analysis of the mitochondrial data set (see Fig. 3). Examination of the relative contribution of each gene to total branch lengths reveals the preponderant role of the two mitochondrial genes. Predictably, the contribution of these faster evolving genes is particularly relevant in the definition of short internal branch lengths within Dasypodidae (Fig. 4). The concatenation of the five genes adds support to the respective monophyly of the three subfamilies and to the close relationship between Tolypeutinae and Euphractinae (Fig. 4; Table 2). Partitioned SH tests indicated that alternatives to such a

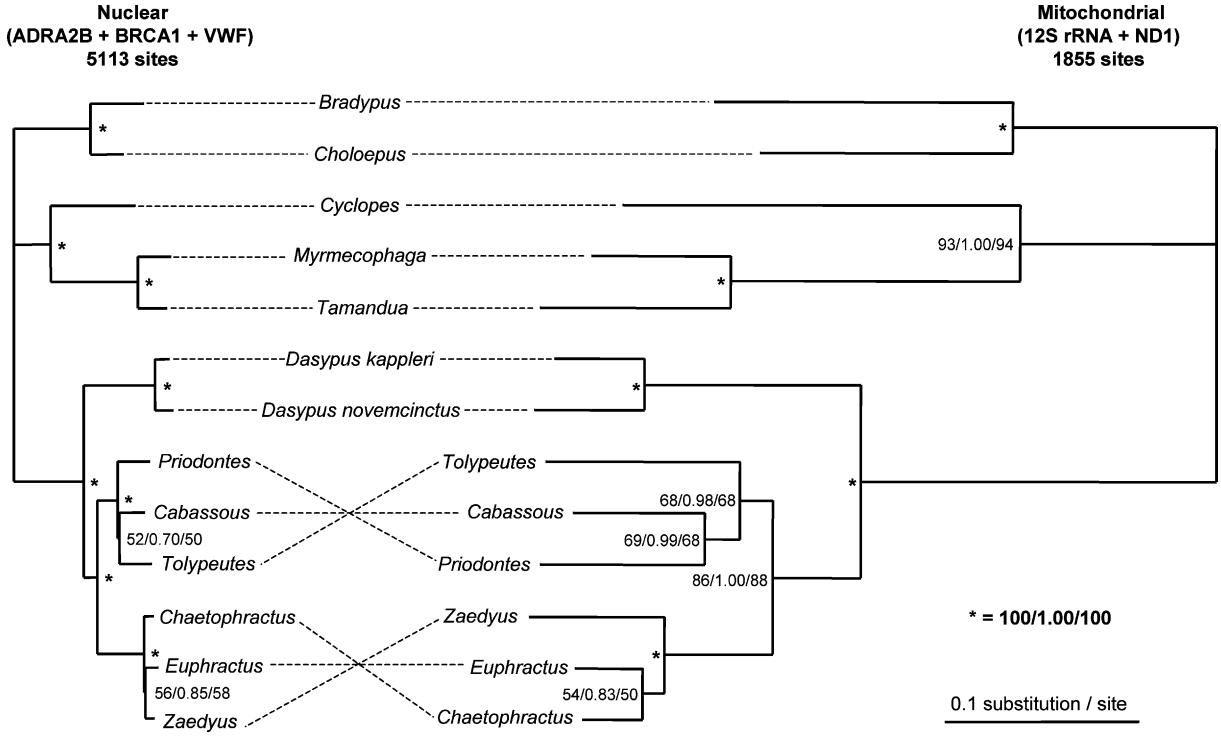


Fig. 3. Comparison of maximum likelihood topologies obtained from the nuclear (5113 sites) and mitochondrial (1855 sites) data sets under the GTR + $\Gamma_8$ model. Maximum likelihood estimates of model parameters were: A = 0.28, C = 0.26, G = 0.25, T = 0.21; A ↔ C = 0.91, A ↔ G = 4.69, A ↔ T = 0.71, C ↔ G = 1.96, C ↔ T = 4.73, G ↔ T = 1.00; $\alpha = 0.60$ for the nuclear data set; and A = 0.37, C = 0.29, G = 0.13, T = 0.21; A ↔ C = 2.16, A ↔ G = 9.47, A ↔ T = 1.56, C ↔ G = 0.57, C ↔ T = 15.89, G ↔ T = 1.00; $\alpha = 0.29$ for the mitochondrial data set. Values at nodes indicate maximum likelihood and Bayesian support expressed by maximum likelihood bootstrap ($BP_{ML}$)/Bayesian posterior probability (PP)/bootstrapped posterior probability ($BP_{Bay}$). Nodes marked with a star (*) are supported by 100/1.00/100. Topological conflicts are indicated by crossed broken lines. Note that the scale is the same for the two topologies illustrating the contrast in substitution rate between nuclear and mitochondrial genes.

Table 4
Results of partitioned Shimodaira and Hasegawa (1999) tests of alternative hypotheses depicting phylogenetic relationships between and within the three armadillos subfamilies

| Topologies | Nuclear | | Mitochondrial | | Combination | |
|---|---|---|---|---|---|---|
| | $\Delta \ln L/SE$ | $P_{SH}$ | $\Delta \ln L/SE$ | $P_{SH}$ | $\Delta \ln L/SE$ | $P_{SH}$ |
| Dasypodidae | | | | | | |
| Euphractinae + Tolypeutinae | **[−17272.22]** | **Best** | **[−10617.99]** | **Best** | **[−27949.03]** | **Best** |
| Euphractinae + Dasypodinae | 24.34/9.42$^{KH}$ | <0.05* | 10.46/6.65 | 0.24 | 32.96/11.39$^{KH}$ | <0.05* |
| Tolypeutinae + Dasypodinae | 23.31/9.66$^{KH}$ | <0.05* | 12.16/6.13$^{KH}$ | 0.17 | 33.28/11.35$^{KH}$ | <0.05* |
| Tolypeutinae | | | | | | |
| *Cabassous* + *Priodontes* | 2.46/3.60 | 0.67 | **[−10617.99]** | **Best** | **[−27949.03]** | **Best** |
| *Cabassous* + *Tolypeutes* | **[−17272.22]** | **Best** | 4.29/6.27 | 0.55 | 0.24/6.09 | 0.82 |
| *Priodontes* + *Tolypeutes* | 2.50/3.57 | 0.67 | 7.26/5.21 | 0.37 | 4.41/4.48 | 0.66 |
| Euphractinae | | | | | | |
| *Euphractus* + *Chaetophractus* | 0.79/4.99 | 0.77 | **[−10617.99]** | **Best** | **[−27949.03]** | **Best** |
| *Euphractus* + *Zaedyus* | **[−17272.22]** | **Best** | 6.83/5.03 | 0.41 | 4.52/6.23 | 0.61 |
| *Chaetophractus* + *Zaedyus* | 3.88/3.67 | 0.59 | 3.00/6.33 | 0.64 | 6.26/5.67 | 0.53 |

*Note.* Log-likelihoods of selected topologies were compared with PAML using independent ML models (GTR + $\Gamma_8$) for each gene. The highest likelihood is given between square brackets. $\Delta \ln L/SE$ = ratio between the log-likelihood difference of a phylogenetic alternative relative to the best hypothesis and its standard error. $P_{SH} = P$ value of the SH test with multiple-comparison correction; a * indicates that the tested hypothesis is significantly rejected at the 5% level. $^{KH}$ indicates that the tested hypothesis is significantly rejected at the 5% level by the Kishino and Hasegawa (1989) test.

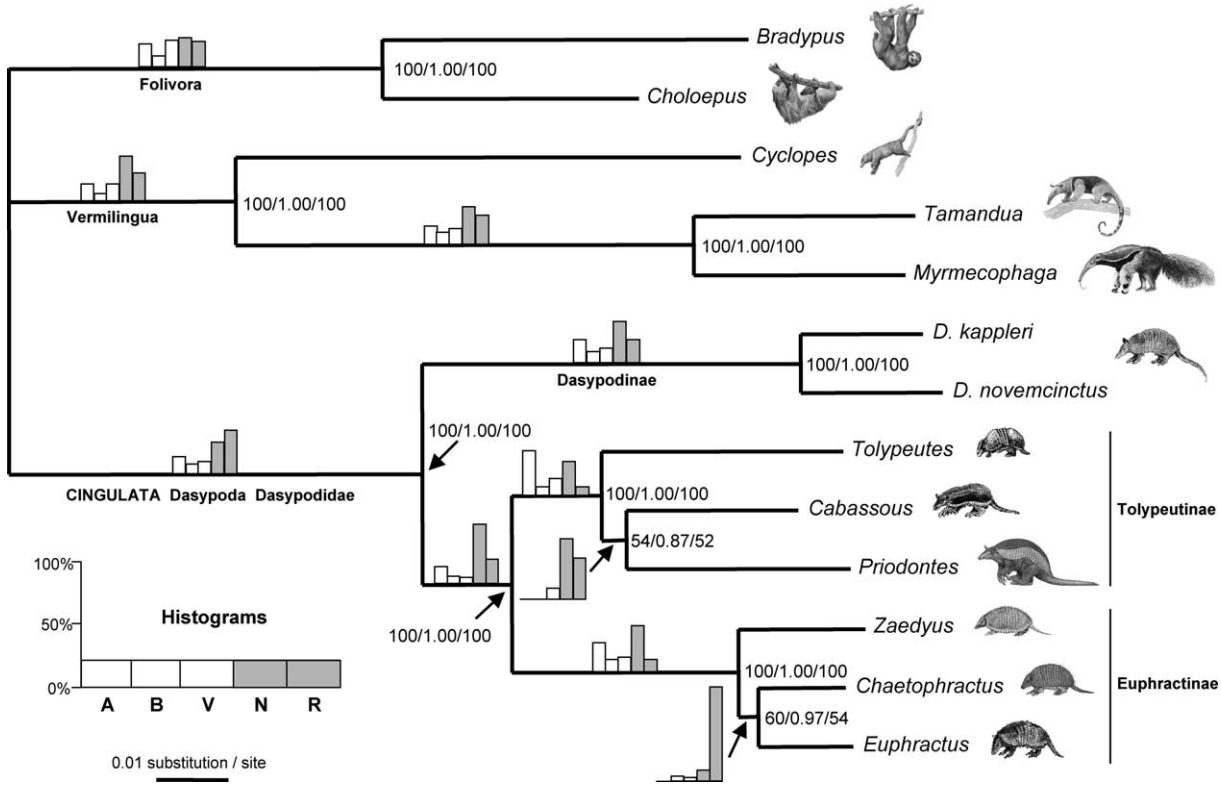


Fig. 4. Maximum likelihood tree obtained from the combined data set (6968 sites) under the GTR + $\Gamma_8$ model. Maximum likelihood estimates of model parameters were: A = 0.29, C = 0.27, G = 10.23, T = 0.21; A ↔ C = 2.13, A ↔ G = 5.73, A ↔ T = 1.55, C ↔ G = 1.73, C ↔ T = 10.01, G ↔ T = 1.00; $\alpha = 0.34$. Values at nodes indicate maximum likelihood and Bayesian support expressed by maximum likelihood bootstrap ($BP_{ML}$)/ Bayesian posterior probability (PP)/bootstrapped posterior probability ($BP_{Bay}$). The histograms at internal branches depict the contribution of each gene to total branch length in the following order (see box): ADRA2B (A), BRCA1 (B), VWF (V), ND1 (N), and 12S rRNA (R).

relationship are significantly worse when using the combined data set (Table 4). However, irresolution persists within these two subfamilies. The combination of the five genes favors the classical grouping of *Cabassous* + *Priodontes* within tolypeutines but with low statistical support (54/0.87/52). It equally poorly

supports a *Chaetophractus* + *Euphractus* clade within euphractines (60/0.97/54). Once again, the partitioned likelihood SH tests are inconclusive regarding the alternatives to the best topology with each hypothesis remaining almost equally likely (Table 4).

### 3.3. Topological incongruence and rapid cladogenesis events

It is clear from our results that Bayesian posterior probabilities (PP) were systematically higher than ML bootstrap proportions ($BP_{ML}$) (Table 2; Figs. 2–4). This propensity has been noted by other authors, who tried to compare the two measures of phylogenetic support (Leaché and Reeder, 2002; Whittingham et al., 2002; Wilcox et al., 2002). In our case, and others (Buckley et al., 2002; Douady et al., 2003b), this tendency led to a situation where high PP were obtained for mutually exclusive nodes inferred from different data sets (Table 2). Thus, based solely on this estimator, one might conclude that there was incongruence between our different data sets. However, some of these conflicts were not so strongly supported by $BP_{ML}$. Furthermore, the use of bootstrap resampling under the Bayesian approach ($BP_{Bay}$)—as suggested by Douady et al. (2003a)—rendered the two indices comparable by lowering PP (Table 2). These observations seem to support the views that: (i) PP tends to provide overestimates of phylogenetic support as recently shown in simulation studies by Suzuki et al. (2002); (ii) a more conservative approach might be to incorporate bootstrap in Bayesian analyses (Douady et al., 2003a; Waddell et al., 2002).

The use of bootstrap resampling under Bayesian analyses was able to eliminate conflicting support within the subfamily Tolypeutinae, but strong incongruence persisted within Euphractinae (Table 2). Also, apart from VWF which is uninformative for this question, the four remaining markers exclusively supported one of the three alternative hypotheses: ADRA2B supports *Euphractus* + *Zaedyus* ($BP_{ML} = 92/BP_{Bay} = 86$), BRCA1 ($BP_{ML} = 96/BP_{Bay} = 92$) and 12S rRNA ($BP_{ML} = 99/BP_{Bay} = 98$) supports *Euphractus* + *Chaetophractus*, and ND1 supports *Chaetophractus* + *Zaedyus* ($BP_{ML} = 92/BP_{Bay} = 88$). Further understanding regarding the origin of these conflicting signals emerged from the examination of the ML inferred number of apomorphies supporting each alternative (Table 2). Nuclear genes contain too few relevant sites to draw firm conclusions for relationships within subfamilies. Indeed, their combination provides only $18 + 10 = 28$ relevant sites over 5113 to distinguish between alternatives within the subfamilies Tolypeutinae and Euphractinae, respectively (Table 2). The strongly supported conflict between ADRA2B and BRCA1 within Euphractinae rests on only three incompatible apomorphies. Thus, incompatibility between these two genes may be the result of stochastic effects associated with a low number of pertinent informative sites. By contrast, mitochondrial genes bring a greater number of informative sites and apomorphies relevant for reconstructing relationships within subfamilies. Hence, 20 of the 23 relevant apomorphies in ND1 supported *Chaetophractus* + *Zaedyus* whereas the 14 relevant 12S rRNA apomorphies exclusively supported *Chaetophractus* + *Euphractus*. The causes of incongruence between these two mitochondrial loci are not obvious. One possible reason might lie in the previously mentioned particular evolutionary behavior of the 12S rRNA gene in Xenarthra. In addition, *Zaedyus* appeared as a very fast evolving taxon for this gene (Fig. 2) and its phylogenetic position might be influenced by a potential long branch attraction artifact (Felsenstein, 1978; Hendy and Penny, 1989).

Owing to the high number of incompatible apomorphies observed between genes, it is not surprising that the combination of the five genes failed to resolve the phylogenetic relationships among Tolypeutinae and Euphractinae (Table 2; Fig. 4). Thus, despite the inclusion of more rapidly evolving mitochondrial genes, we were not able to resolve the two remaining uncertainties in Xenarthra phylogeny. The low bootstrap support values associated with almost equal distributions of apomorphies among alternative hypotheses suggest the occurrence of rapid cladogenesis in Tolypeutinae and Euphractinae, leaving only short time windows for molecular synapomorphies to accumulate. Certainly, relative incongruence between different genes is not unexpected when lineages likely diverged within a short period of time. Hybridization, incomplete lineage sorting or assortment of ancestral polymorphism are commonly cited as potential causes of incongruence between gene and species trees (Maddison, 1997). The persistence of ancestral polymorphism coupled with the differential survival of alleles has been shown to significantly account for the difficulties encountered in accurately resolving the hominoid trifurcation between human, chimpanzee and gorilla (O'hUigin et al., 2002; Satta et al., 2000). Our results, thus, tend to confirm the nuclear based hypothesis (Delsuc et al., 2002) of two parallel episodes of fast diversification during the evolutionary history of extant armadillos, which have led to two difficult-to-resolve trifurcations within the subfamilies Tolypeutinae and Euphractinae. The reconstruction of a reliable molecular time-scale for the chronicle of xenarthran evolution might help to identify the grounds of these events.

## Acknowledgments

We thank the associated editors Wilfried W. de Jong, Mark S. Springer, and Michael J. Stanhope for inviting us to contribute to this special issue on mammalian

phylogeny. We are grateful to François Catzeflis for giving access to tissue samples and laboratory facilities. This work would not have been possible without the help of the following individuals and institutions in providing us with xenarthran samples: Tammie L. Bettinger (Cleveland Metroparks Zoo, US), Pablo Carmanchahi, Pablo Cetica, Jorgue Omar García and Rodolpho Rearte (Complejo Ecológico Municipal de Presidencia Roque Sàenz Peña, Chaco, Argentina), Guillermo Anibal Lemus, Jesus Mavarez, Guillermo Perez Jimeno, Mariella Superina, and Jean-Christophe Vié and his team "Faune Sauvage" in French Guiana. Wilfried W. de Jong and two anonymous reviewers provided helpful comments on the manuscript. This work has been supported by the TMR Network "Mammalian phylogeny" (contract FMRX—CT98—022) of the European Community and by the Génopôle Montpellier Languedoc-Roussillon. FD acknowledges the financial support of a MENRT doctoral grant (contract 99075) and a grant from the British Systematics Association. This is the contribution ISEM 2002-080 of the Institut des Sciences de l'Evolution de Montpellier (UMR 5554-CNRS).

## References

Adkins, R.M., Gelke, E.L., Rowe, D., Honeycutt, R.L., 2001. Molecular phylogeny and divergence time estimates for major rodent groups: evidence from multiple genes. Mol. Biol. Evol. 18, 777–791.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Contr. AC-19, 716–723.

Buckley, T.R., 2002. Model misspecification and probabilistic tests of topology: evidence from empirical data sets. Syst. Biol. 51, 509–523.

Buckley, T.R., Cunningham, C.W., 2002. The effects of nucleotide substitution model assumptions on estimates of nonparametric bootstrap support. Mol. Biol. Evol. 19, 394–405.

Buckley, T.R., Arensburger, P., Simon, C., Chambers, G.K., 2002. Combined data, Bayesian phylogenetics, and the origin of the New Zealand cicada genera. Syst. Biol. 51, 4–18.

Buckley, T.R., Simon, C., Chambers, G.K., 2001. Exploring among-site rate variation models in a maximum likelihood framework using empirical data: effects of model assumptions on estimates of topology, branch lengths, and bootstrap support. Syst. Biol. 50, 67–86.

Bull, J.J., Huelsenbeck, J.P., Cunningham, C.W., Swofford, D.L., Waddell, P.J., 1993. Partitioning and combining data in phylogenetic analysis. Syst. Biol. 42, 384–397.

Cao, Y., Janke, A., Waddell, P.J., Westerman, M., Takenaka, O., Murata, S., Okada, N., Pääbo, S., Hasegawa, M., 1998. Conflict among individual mitochondrial proteins in resolving the phylogeny of eutherian orders. J. Mol. Evol. 47, 307–322.

Catzeflis, F., 1991. Animal tissue collections for molecular genetics and systematics. Trends Ecol. Evol. 6, 168.

Cetica, P.D., Solari, A.J., Merani, M.S., De Rosas, J.C., Burgos, M.H., 1998. Evolutionary sperm morphology and morphometry in armadillos. J. Submicroscop. Cytol. Pathol. 30, 309–314.

Delsuc, F., Catzeflis, F.M., Stanhope, M.J., Douzery, E.J.P., 2001. The evolution of armadillos, anteaters and sloths depicted by nuclear and mitochondrial phylogenies: Implications for the status of the enigmatic fossil *Eurotamandua*. Proc. R. Soc. Lond. B.S. 268, 1605–1615.

Delsuc, F., Scally, M., Madsen, O., Stanhope, M.J., de Jong, W.W., Catzeflis, F.M., Springer, M.S., Douzery, E.J.P., 2002. Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. Mol. Biol. Evol. 19, 1656–1671.

Douady, C.J., Delsuc, F., Boucher, Y., Doolittle, W.F., Douzery, E.J.P., 2003a. Comparison of Bayesian and maximum likelihood bootstrap measures of phylogenetic reliability. Mol. Biol. Evol. 20, 248–254.

Douady, C.J., Dosay, M., Shivji, M.S., Stanhope, M.J., 2003b. Molecular phylogenetic evidence refuting the hypothesis of Batoidea (rays and skates) as derived sharks. Mol. Phylogenet. Evol. 26, 215–221.

Douzery, E., Catzeflis, F.M., 1995. Molecular evolution of the mitochondrial 12S rRNA in Ungulata (Mammalia). J. Mol. Evol. 41, 622–636.

Engelmann, G.F., 1985. The phylogeny of the Xenarthra. In: Montgomery, G.G. (Ed.), The Evolution and Ecology of Armadillos, Sloths and Vermilinguas. Smithsonian Institution Press, Washington, pp. 51–63.

Felsenstein, J., 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. 27, 401–410.

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

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Gaudin, T.J., Branham, D.G., 1998. The phylogeny of the Myrmecophagidae (Mammalia, Xenarthra, Vermilingua) and the relationship of *Eurotamandua* to the Vermilingua. J. Mammal. Evol. 5, 237–265.

Galbreath, G.J., 1985. The evolution of monozygotic polyembryony in Dasypus. In: Montgomery, G.G. (Ed.), The Evolution and Ecology of Armadillos, Sloths and Vermilinguas. Smithsonian Institution Press, Washington, pp. 243–246.

Goldman, N., Anderson, J.P., Rodrigo, A.G., 2000. Likelihood-based tests of topologies in phylogenetics. Syst. Biol. 49, 652–670.

Greenwood, A.D., Castresana, J., Feldmaeir-Fuchs, G., Pääbo, S., 2001. A molecular phylogeny of two extinct sloths. Mol. Phylogenet. Evol. 18, 94–103.

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

Hendy, M.D., Penny, D., 1989. A framework for the quantitative study of evolutionary trees. Syst. Zool. 38, 297–309.

Höss, M., Dilling, A., Currant, A., Pääbo, S., 1996. Molecular phylogeny of the extinct ground sloth *Mylodon darwinii*. Proc. Natl. Acad. Sci. USA 93, 181–185.

Huchon, D., Madsen, O., Sibbald, M.J., Ament, K., Stanhope, M.J., Catzeflis, F., de Jong, W.W., Douzery, E.J.P., 2002. Rodent phylogeny and a timescale for the evolution of Glires: evidence from an extensive taxon sampling using three nuclear genes. Mol. Biol. Evol. 19, 1053–1065.

Huelsenbeck, J.P., Rannala, B., 1997. Phylogenetic methods come of age: testing hypotheses in an evolutionary context. Science 276, 227–232.

Huelsenbeck, J.P., Ronquist, F., 2001. Mrbayes: Bayesian inference of phylogenetic trees. Bioinformatics 17, 754–755.

Huelsenbeck, J.P., Bull, J.J., Cunningham, C.W., 1996. Combining data in phylogenetic analysis. Trends Ecol. Evol. 11, 152–158.

Huelsenbeck, J.P., Ronquist, F., Nielsen, R., Bollback, J.P., 2001. Bayesian inference of phylogeny and its impact on evolutionary biology. Science 294, 2310–2314.

Jow, H., Hudelot, C., Rattray, M., Higgs, P.G., 2002. Bayesian phylogenetics using an RNA substitution model applied to early mammalian evolution. Mol. Biol. Evol. 19, 1591–1601.

Kishino, H., Hasegawa, M., 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.

Leaché, A.D., Reeder, T.W., 2002. Molecular systematics of the Eastern Fence Lizard (*Sceloporus undulatus*): a comparison of parsimony, likelihood, and Bayesian approaches. Syst. Biol. 51, 44–68.

Loughry, W.J., Prodöhl, P.A., McDonough, C.M., Avise, J.C., 1998. Polyembryony in armadillos. Am. Scient. 86, 274–279.

Maddison, W.P., 1997. Gene trees in species trees. Syst. Biol. 46, 523–536.

Madsen, O., Scally, M., Douady, C.J., Kao, D.J., DeBry, R.W., Adkins, R., Amrine, H.M., Stanhope, M.J., de Jong, W.W., Springer, M.S., 2001. Parallel adaptative radiations in two major clades of placental mammals. Nature 409, 610–614.

McKenna, M.C., 1975. Toward a phylogenetic classification of the Mammalia. In: Luckett, W.P., Szalay, F.S. (Eds.), Phylogeny of the Primates. Plenum Press, New York, pp. 21–46.

McKenna, M.C., Bell, S.K., 1997. Classification of Mammals Above the Species Level. Columbia University Press, New York.

Meritt Jr., D.A., 1985. The fairy armadillo *Chlamyphorus truncatus* Harlan. In: Montgomery, G.G. (Ed.), The Evolution and Ecology of Armadillos, Sloths and Vermilinguas. Smithsonian Institution Press, Washington, pp. 393–395.

Miller, R.E., Buckley, T.R., Manos, P.S., 2002. An examination of the monophyly of morning glory taxa using Bayesian phylogenetic inference. Syst. Biol. 51, 740–753.

Misof, B., Anderson, C.L., Buckley, T.R., Erpenbeck, D., Rickert, A., Misof, K., 2002. An empirical analysis of mt 16S rRNA covarion-like evolution in insects: site-specific rate variation is clustered and frequently detected. J. Mol. Evol. 55, 460–469.

Murphy, W.J., Eizirick, E., Jonhson, W.E., Zhang, Y.P., Ryder, O.A., O'Brien, S.J., 2001a. Molecular phylogenetics and the origins of placental mammals. Nature 409, 614–618.

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

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

O'hUigin, C., Satta, Y., Takahata, N., Klein, J., 2002. Contribution of homoplasy and of ancestral polymorphism to the evolution of genes in anthropoid primates. Mol. Biol. Evol. 19, 1501–1513.

Olson, L.E., Yoder, A.D., 2002. Using secondary structure to identify ribosomal numts: cautionary examples from the human genome. Mol. Biol. Evol. 19, 93–100.

Patterson, B., Pascual, R., 1972. The fossil mammal fauna of South-America. In: Keast, A., Erk, F.C., Glass, B.P. (Eds.), Evolution, Mammals and Southern Continents. State University of New York Press, Albany, pp. 247–309.

Patterson, B., Segall, W., Turnbull, W.D., 1989. The ear region in xenarthrans (= Edentata, Mammalia). Part I. Cingulates. Fieldiana: Geology, New Series 18, 1–46.

Patterson, B., Segall, W., Turnbull, W.D., Gaudin, T.J., 1992. The ear region in xenarthrans (= Edentata, Mammalia). Part II. Sloths, anteaters, palaeanodonts, and a miscellany. Fieldiana: Geology, New Series 24, 1–79.

Penny, D., Watson, E.E., Steel, M.A., 1993. Trees from languages and genes are very similar. Syst. Biol. 42, 382–384.

Philippe, H., 1993. MUST: A computer package of management utilities for sequences and trees. Nucleic Acids Res. 21, 5264–5272.

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

Posada, D., Crandall, K.A., 2001. Selecting the best-fit model of nucleotide substitution. Syst. Biol. 50, 580–601.

Rannala, B., Yang, Z., 1996. Probability distribution of molecular evolutionary trees: a new method of phylogenetic inference. J. Mol. Evol. 43, 304–311.

Reiss, K.Z., 1997. Myology of the feeding apparatus of myrmecophagid anteaters (Xenarthra: Myrmecophagidae). J. Mammal. Evol. 4, 87–117.

Satta, Y., Klein, J., Takahata, N., 2000. DNA archives from our nearest relative: the trichotomy problem revisited. Mol. Phylogenet. Evol. 14, 259–275.

Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51, 492–508.

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

Shoshani, J., McKenna, M.C., 1998. Higher taxonomic relationships among extant mammals based on morphology, with selected comparisons of results from molecular data. Mol. Phylogenet. Evol. 9, 572–584.

Springer, M.S., Douzery, E.J.P., 1996. Secondary structure and patterns of evolution among mammalian mitochondrial 12s rRNA molecules. J. Mol. Evol. 43, 357–373.

Storrs, E.E., 1999. A lost talisman: catastrophic decline in yields of leprosy bacilli from armadillos used for vaccine production. Int. J. Lepr. Other Mycobact. Dis. 67, 67–70.

Storrs, E.E., Burchfield, H.P., 1985. Leprosy in wild common long-nosed armadillos *Dasypus novemcinctus*. In: Montgomery, G.G. (Ed.), The Evolution and Ecology of Armadillos, Sloths and Vermilinguas. Smithsonian Institution Press, Washington, pp. 265–268.

Storrs, E.E., Walsh, G.P., Burchfield, H.P., Binford, C.H., 1974. Leprosy in the armadillo: new model for biomedical research. Science 183, 851–852.

Strimmer, K., von Haeseler, A., 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. 13, 964–969.

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.

Sullivan, J., Holsinger, K.E., Simon, C., 1995. Among-site rate variation and phylogenetic analysis of 12S rRNA sigmodontine rodents. Mol. Biol. Evol., 988–1001.

Suzuki, Y., Glazko, G.V., Nei, M., 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. Proc. Natl. Acad. Sci. USA 99, 16143–16168.

Swofford, D.L., 2002. PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4.0b10. Sinauer Associates, Sunderland, MA.

Swofford, D.L., Olsen, G.J., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), Molecular Systematics. Sinauer Associates, Sunderland, MA, pp. 407–514.

Swofford, D.L., Waddell, P.J., Huelsenbeck, J.P., Foster, P.G., Lewis, P.O., Rogers, J.S., 2001. Bias in phylogenetic estimation and its relevance to the choice between parsimony and likelihood methods. Syst. Biol. 50, 525–539.

Taulman, J.F., Robbins, L.W., 1996. Recent range expansion and distributional limits of the nine-banded armadillo (*Dasypus novemcinctus*) in the United States. J. Biogeogr. 23, 635–648.

van Dijk, M.A.M., Paradis, E., Catzeflis, F., de Jong, W.W., 1999. The virtues of gaps: xenarthran (Edentate) monophyly supported by a unique deletion in αa-crystallin. Syst. Biol. 48, 94–106.

Vizcaíno, S.F., 1995. Identificación específica de las "mulitas", género *Dasypus* (Mammalia, Dasypodidae), del noroeste argentino. Descripción de una nueva espécie. Mastozoología Neotropical 2, 5–13.

Waddell, P.J., Kishino, H., Ota, R., 2002. Very fast algorithms for evaluating the stability of ML and Bayesian phylogenetic trees from sequences data. Gen. Informat. 13, 82–92.

Webb, S.D., 1985. The interrelationships of tree sloths and ground sloths. In: Montgomery, G.G. (Ed.), The Evolution and Ecology of Armadillos, Sloths and Vermilinguas. Smithsonian Institution Press, Washington, pp. 105–112.

Wetzel, R.M., 1981. Systematics, distribution, ecology, and conservation of South American Edentates. In: Mares, M.A., Genoways, H.H. (Eds.), Mammalian Biology in South America. Pymtuning laboratory of ecology, University of Pittsburgh, Pittsburgh, pp. 345–375.

Wetzel, R.M., 1985. Taxonomy and distribution of armadillos, Dasypodidae. In: Montgomery, G.G. (Ed.), The Evolution and Ecology of Armadillos, Sloths and Vermilinguas. Smithsonian Institution Press, Washington, pp. 23–46.

Wetzel, R.M., Mondolfi, E., 1979. The subgenera and species of long-nosed armadillos, genus *Dasypus* L. In: Eisenberg, J.F. (Ed.), Vertebrate Ecology in the Northern Neotropics. Smithsonian Institution Press, Washington, pp. 43–63.

Whelan, S., Lio, P., Goldman, N., 2001. Molecular phylogenetics: state-of-the-art methods for looking into the past. Trends Genet. 17, 262–272.

Whittingham, L.A., Slikas, B., Winkler, D.W., Sheldon, F.H., 2002. Phylogeny of the tree swallow genus, *Tachycineta* (Aves: Hirundinidae), by Bayesian analysis of mitochondrial DNA sequences. Mol. Phylogenet. Evol. 22, 430–441.

Wilcox, T.P., Zwickl, D.J., Heath, T.A., Hillis, D.M., 2002. Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. Mol. Phylogenet. Evol. 25, 361–371.

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

Yang, Z., 1996a. Among-site rate variation and its impact on phylogenetic analyses. Trends Ecol. Evol. 11, 367–372.

Yang, Z., 1996b. Maximum-likelihood models for combined analyses of multiple sequence data. J. Mol. Evol. 42, 587–596.

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

Yang, Z., Rannala, B., 1997. Bayesian phylogenetic inference using DNA sequences: a Markov chain Monte Carlo method. Mol. Biol. Evol. 14, 717–724.