Available online at www.sciencedirect.com
SCIENCE DIRECT®
Molecular Phylogenetics and Evolution 33 (2004) 922–935
MOLECULAR PHYLOGENETICS AND EVOLUTION
www.elsevier.com/locate/ympev

Rabbits, if anything, are likely Glires

Emmanuel J.P. Douzery[a,*], Dorothée Huchon[b]

[a] *Laboratoire de Paléontologie, Phylogénie et Paléobiologie, CC064, Institut des Sciences de l'Evolution UMR 5554/CNRS, Université Montpellier II; Place E. Bataillon, 34 095 Montpellier Cedex 05, France*
[b] *Department of Zoology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel*

Received 1 July 2004
Available online 16 September 2004

Abstract

Rodentia (e.g., mice, rats, dormice, squirrels, and guinea pigs) and Lagomorpha (e.g., rabbits, hares, and pikas) are usually grouped into the Glires. Status of this controversial superorder has been evaluated using morphology, paleontology, and mitochondrial plus nuclear DNA sequences. This growing corpus of data has been favoring the monophyly of Glires. Recently, Misawa and Janke [Mol. Phylogenet. Evol. 28 (2003) 320] analyzed the 6441 amino acids of 20 nuclear proteins for six placental mammals (rat, mouse, rabbit, human, cattle, and dog) and two outgroups (chicken and xenopus), and observed a basal position of the two murine rodents among the former. They concluded that "the Glires hypothesis was rejected." We here reanalyzed [loc. cit.] data set under maximum likelihood and Bayesian tree-building approaches, using phylogenetic models that take into account among-site variation in evolutionary rates and branch-length variation among proteins. Our observations support both the association of rodents and lagomorphs and the monophyly of Euarchontoglires (= Supraprimates) as the most likely explanation of the protein alignments. We conducted simulation studies to evaluate the appropriateness of lissamphibian and avian outgroups to root the placental tree. When the outgroup-to-ingroup evolutionary distance increases, maximum parsimony roots the topology along the long *Mus-Rattus* branch. Maximum likelihood, in contrast, roots the topology along different branches as a function of their length. Maximum likelihood appears less sensitive to the "long-branch attraction artifact" than is parsimony. Our phylogenetic conclusions were confirmed by the analysis of a different protein data set using a similar sample of species but different outgroups. We also tested the effect of the addition of afrotherian and xenarthran taxa. Using the linearized tree method, [loc. cit.] estimated that mice and rats diverged about 35 million years ago. Molecular dating based on the Bayesian relaxed molecular clock method suggests that the 95% credibility interval for the split between mice and rats is 7–17 Mya. We here emphasize the need for appropriate models of sequence evolution (matrices of amino acid replacement, taking into account among-site rate variation, and independent parameters across independent protein partitions) and for a taxonomically broad sample, and conclude on the likelihood that rodents and lagomorphs together constitute a monophyletic group (Glires).

Keywords: Rodentia; Lagomorpha; Glires; Maximum likeliood; Bayesian inference

What, if anything, is a rabbit?
(Wood, 1957)

1. Introduction

Two orders of placental mammals are typically small gnawing animals: the rodents (e.g., mice, rats, dormice, squirrels, and guinea pigs) and the lagomorphs (e.g., rabbits, hares, and pikas). Rodentia and Lagomorpha are usually clustered to form the Glires clade based on the shared traits of enlarged, ever-growing incisors adapted for gnawing. Rodents possess one pair of incisors on the upper jaw while Lagomorphs possess two pairs. For this reason, Rodentia have also been termed Simplicidentata and Lagomorpha Duplicidentata.

The taxon Glires, first proposed by Linnaeus in 1758, remains controversial after literally centuries of debate.

* Corresponding author. Fax: +33 4 67 14 36 10.
E-mail address: douzery@isem.univ-montp2.fr (E.J.P. Douzery).

1055-7903/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2004.07.014

In 1957, Albert E. Wood reviewed the many conflicting theories in an article titled, "What, if anything, is a rabbit?" Based on morphological characters, Lagomorpha has been variously considered as related to Triconodonta (a group of fossil Mesozoic mammals), Marsupialia, Artiodactyla, Condylarthra (fossil "ungulates"), Primates, Insectivora, Macroscelidea, or Scandentia (reviewed by Wood, 1957; Li et al., 1987). However, the Glires concept has never been rejected, since none of the alternative hypotheses appears satisfactory. The notion of Glires has been revived by paleontological discoveries, notably the Eurymilidae and Mimotonidae (fossil mammals considered to be close to the ancestral stock of Rodentia and Lagomorpha, e.g., Li et al., 1987). Almost at the same time, this notion was also challenged by the first extensive multigene analyses and studies of the mitochondrial genome (e.g., D'Erchia et al., 1996; Graur et al., 1991; Graur et al., 1996). These papers suggested that rodents might be paraphyletic and unrelated to lagomorphs. This conclusion led morphologists to reconsider the position of lagomorphs. Recent studies support a close relationship between rodents and lagomorphs (e.g., Landry, 1999; Luckett and Hartenberger, 1993; Meng et al., 1994; Wallau et al., 2000). In parallel, it has been argued that conclusions from molecular data about the paraphyly of Glires reflect small species samples, the use of oversimplified models of evolution, tree rooting problems, or long-branch attraction artifacts in tree construction (e.g., Adachi and Hasegawa, 1996; Halanych, 1998; Lin et al., 2002a; Philippe, 1997; Philippe and Douzery, 1994; Reyes et al., 2004; Sullivan and Swofford, 1997). Many recent molecular studies indicate that Glires are monophyletic. Some individual nuclear genes such as pepsinogen C (Narita et al., 2001), IRBP (DeBry and Sagel, 2001; Stanhope et al., 1992), ε-globin, and GHR (Waddell and Shelley, 2003) favor the monophyly of Glires, as well as various combined nuclear and mitochondrial data sets. Lagomorphs join rodents in analyses of 5.7 kb for 28 mammals (Madsen et al., 2001), 2.4 kb for 42 taxa (Huchon et al., 2002), 5.1 kb for 50 taxa (Delsuc et al., 2002), 9.8 kb for 64 taxa (Murphy et al., 2001a), 16.4 kb for 44 taxa (Murphy et al., 2001b), 17 kb for 44 taxa (Amrine-Madsen et al., 2003), and proteins for 38–47 taxa (Waddell et al., 2001). The monophyly of Glires is also supported by the analysis of indels in nuclear genes (de Jong et al., 2003).

Following the approach of Graur et al. (1991, 1996); Misawa and Janke (2003) concatenated proteins, maximized sequence length with no missing data at the expense of broad taxon sampling, and analyzed 6441 amino-acid sites for six placental mammals (rat, mouse, rabbit, human, cattle, and dog) and two outgroups (chicken and xenopus). The authors surprisingly observed a basal position within placental mammals of the two murine rodents. They indicated that "all branches of [their] tree are well supported except by the Bayesian method with rate variation," but concluded that "the Glires hypothesis was rejected."

We respond to the study by Misawa and Janke (2003) by reanalyzing their protein alignments under probabilistic tree-building approaches. Using phylogenetic models that take into account among-site rate variation, mouse, rat, and rabbit sequences appear monophyletic. However, the data do not contain placental taxa that are likely to disrupt the monophyly of Glires (e.g., Scandentia: *Tupaia*) (Waddell and Shelley, 2003). We also investigate whether the use of distant—lissamphibian and avian—outgroups is appropriate to root the placental tree. By simulation, we show that maximum parsimony artefactually roots the topology along the long *Mus-Rattus* branch, whereas maximum likelihood—a more robust tree-building approach (Gaut and Lewis, 1995)—is less sensitive to this undesirable property, even for distant outgroups. Our reanalysis shows that Misawa and Janke's misleading conclusions are likely to be the result of phylogenetic analyses based on models of sequence evolution that fit the data poorly, and on the use of outgroups that are too divergent. Correcting for these problems suggests the monophyly of Glires.

2. Materials and methods

2.1. Data sets

Four data sets were analyzed, the nuclear and mitochondrial ones of Misawa and Janke (2003) as well as two data sets based on Murphy et al. (2001b). The nuclear data set of Misawa and Janke (2003) includes 20 proteins (6441 amino-acid sites) and the mitochondrial data set the 12 H-strand encoded proteins (3559 sites). The Murphy et al. (2001b) data sets were chosen in order to study the impact of taxon sampling on the monophyly of Glires, because they include the sequences of mouse, rat, rabbit, human, dog, and bovine (boreoeutherian placentals), and marsupial outgroups, with afrotherian and xenarthran (non-boreoeutherian) placentals included or excluded. Because Misawa and Janke (2003) worked on protein sequences, the data set of Murphy et al. (2001b) was translated into protein. Some minor corrections were made to the DNA alignments of Murphy et al. (2001b), to better take into account the reading frame. The concatenated protein alignment based on Murphy et al. (2001b) included 15 protein-coding genes for a total of 4333 amino acids: A2AB, ADORA3, ADRB2, ATP7A, BDNF, BRCA1, CNR1, EDG1, IRBP, PNOC, RAG1, RAG2, TYR, vWF, and ZFX. Two data sets were considered: an 8-taxon matrix and a 10-taxon matrix. The 8-taxon matrix is

composed of the former 6 boreoeutherians with two marsupial outgroups: *Didelphis* and a diprotodontian. The 10-taxon matrix is the same as the 8-taxon matrix, with the addition of two taxa: the elephant representing the Proboscidea and sloth representing the Xenarthra. These sequences were added in order to include representative taxa from the four major placental clades. Sequence alignments are available upon request.

2.2. Bayesian inference of phylogeny

The nuclear data set of Misawa and Janke (2003), and the reduced 8- and 10-taxon versions of the Murphy et al. (2001b) data set were analyzed using the Bayesian approach implemented in MrBayes version 3.0b4 (Ronquist and Huelsenbeck, 2003), under a JTT model of protein evolution (Jones et al., 1992). Among-site rate variation was described by a discrete Gamma distribution with four categories (Γ_4) (Yang, 1996). Four Metropolis-coupled Markov chains Monte Carlo (MCMCMC) were run for 100,000 generations with the program default prior distributions, and trees were sampled every 20 generations. The choice of the total number of generations represents a compromise between computing time and rapidity of MCMCMC convergence due to the limited number of taxa (7 ingroups plus one outgroup). Among the 5000 saved trees, the number to be discarded as a burn-in was determined empirically after checking for log-likelihood ($\ln L$) stationarity. The reliability of nodes was measured by their posterior probabilities (PP). Topologies were also compared according to their tree posterior probability (TPP).

To circumvent the problem of overcredibility of Bayesian posterior probabilities for measuring the reliability of tree nodes (Suzuki et al., 2002; Waddell et al., 2001, 2002), bootstrapped posterior probabilities (BootPP) were computed, i.e., PP estimated by MCMCMC after bootstrap resampling of characters (Douady et al., 2003; Waddell et al., 2002). Here, we computed BootPP after 50 bootstrap replicates on amino acids, conducted exactly under the same conditions as those of the original MCMCMC run. Half of the sampled trees were discarded as a burn-in to ensure that MCMCMC reached stationarity. Following Douady et al. (2003) and Waddell et al. (2002), BootPP were calculated by the majority rule consensus of the 50×2500 remaining trees.

2.3. ML inference of phylogeny

The nuclear data set of Misawa and Janke (2003) was also analyzed under maximum likelihood (ML) by PAML (Yang, 1997), version 3.1, and using the $\mathrm{JTT} + \mathrm{F} + \Gamma_4$ model. The option "+F" allows incorporation of the amino acid frequencies of the protein considered. The statistical significance of the differences in log-likelihood between the 105 possible rootings of rodents, lagomorphs, primates, carnivores, and cetartiodactyls by *Xenopus* and *Gallus* was evaluated by the non-parametric test of Kishino and Hasegawa (1989), conducted with the conservative Shimodaira (2002) correction for multiple tree comparisons (SH test). The difference in log-likelihoods of the best root [R] and the alternative 104 rootings [A] ($\delta = \ln L[R] - \ln L[A]$), the standard error ($\sigma$) of this difference, and the confidence P_{SH} values were computed by PAML 3.1.

Because of the conservative behavior of the SH test (Shimodaira, 2002), we evaluated the statistical significance of $\ln L$ decrease for alternative rooting by using the SOWH test (Swofford et al., 1996), based on Monte Carlo simulation, and also known as parametric bootstrap. We used PSeq-Gen (Grassly et al., 1997), version 1.1, to generate 100 matrices of 8 taxa and 6441 amino acids, simulated under a $\mathrm{JTT} + \mathrm{F} + \Gamma_4$ model of protein evolution. The reference topology was ((human: 0.059608, ((rabbit: 0.080175, (mouse: 0.035348, rat: 0.030112) :0.110667): 0.019303, (dog: 0.062351, bovine: 0.086717): 0.020118): 0.012799): 0.305140, chicken: 0.226552, and xenopus: 0.783356). This tree, where primates are basal, corresponded to the best alternative under the constraint that the root is not on the branch leading to the dog plus bovine clade. For each of the 100 matrices obtained, PAML was used to calculate the $\ln L$ of the "basal Primates" and "Glires" trees, with a full optimization of the α parameter of the Gamma distribution. Because of the a priori definition of the topologies compared, this version of the SOWH test corresponded to the mnemonic code "*priPfud*" defined by Goldman et al. (2000).

The same analysis was repeated with 100 simulations under the reference topology corresponding to the second best root (murine rodents are basal): ((((human: 0.070593, (dog: 0.062105, bovine: 0.086517): 0.020670): 0.017181, rabbit: 0.081891): 0.018747, (mouse: 0.035158, rat: 0.030281): 0.094743): 0.297491, chicken: 0.223176, xenopus: 0.777447). PAML was again used to calculate the $\ln L$ of the "basal murines" and "Glires" trees.

2.4. Separate versus combined analysis

The program COMBINE (Pupko et al., 2002) was used to study the effect of the ML model used to combine the different genes on the highest likelihood phylogeny obtained. The nuclear and the mitochondrial amino acid data sets of Misawa and Janke (2003) were analyzed under four different models of sequence evolution. Three methods of branch length estimation were considered: the concatenate model assumed the same set of branch lengths for all proteins; the proportional model assumed that branch lengths are proportional among proteins; and the separate model assumed a different

set of branch lengths for each protein. For the concatenate model, two models of among-site rate variation were used: the homogeneous model that assumes that all sites have the same rate of evolution; and the heterogeneous Gamma model (one α parameter was used for the concatenate data set, i.e., the "one-alpha" model). For the proportional and the separate models, we assumed different Gamma distributions for each protein, to describe among-site rate variation (different α parameters were considered for each gene, i.e., the "*N*-alpha" model). The concatenate model without among-site rate variation and the concatenate model with a single α parameter, are the models used by Misawa and Janke (2003). The proportional and the separate models with *N* alpha parameters are the best models identified by Pupko et al. (2002) to combine sequences.

The nuclear genes were analyzed under the JTT model and the mitochondrial genes under the mtREV model (Adachi and Hasegawa, 1996). Three topologies were compared under the four models (Fig. 1): the best nuclear topology of Misawa and Janke (2003) ("basal-rodents tree"), the best mitochondrial topology of Misawa and Janke (2003) ("mitochondrial tree"), and the topology that supports the monophyly of Glires ("Glires tree," e.g., Madsen et al., 2001; Murphy et al., 2001a; Murphy et al., 2001b; Huchon et al., 2002; Delsuc et al., 2002; Amrine-Madsen et al., 2003 ; Waddell et al., 2001; Waddell and Shelley, 2003). For each topology, the α parameter of the Gamma distribution was estimated with the program COMBINE. This program uses a discrete Gamma distribution with four categories. The different trees were compared under the same model using one-tailed Kishino–Hasegawa test (KH test; Kishino and Hasegawa, 1989). The correction of Shimodaira (2002) for multiple tree comparisons is not implemented in the program COMBINE. To compare the different models, the Akaike Information Criterion (AIC) was used (Sakamoto et al., 1986). The statistical difference between two AIC values was evaluated using the test of Linhart (1988) which is similar to the KH test (Kishino and Hasegawa, 1989).

2.5. Simulation study of the impact of outgroup distance to the ingroup

An analysis of the impact of the evolutionary distance between the ingroup and the outgroup was conducted on the nuclear data set of Misawa and Janke (2003). Based on the $\mathrm{JTT}+\mathrm{F}+\Gamma_4$ model of protein evolution, we used PSeq-Gen 1.1 to generate five sets of 100 matrices – 8 taxa $\times$ 6441 amino acids, with empirical amino acid frequencies deduced from the original data set, Gamma parameter $\alpha = 0.63$, and branch lengths imported from the best tree identified by PAML ("Glires tree"): (((human: 0.079, (rabbit: 0.080, (mouse: 0.035, rat: 0.030): 0.111): 0.019): 0.010, (dog: 0.064, bovine: 0.087): 0.011): 0.308, chicken: 0.229, xenopus: 0.782) – each under the same topology, but with five varying lengths for the branches leading to the ingroup, to *Gallus*, and to *Xenopus*. We changed the distance of the outgroup to the ingroup by rescaling these three branches by factors 0.01, 0.10, 1.00 (hence corresponding to the original data), 10, and 100. Then, PAML 3.1 was used to compare the seven possible rootings along the branches of the model tree, for each of the 5×100 simulated matrices, and under two optimality criteria. These comparisons were first conducted under maximum parsimony (MP) by measuring the number of extra steps required to observe the alternative topologies relative to the most parsimonious one; and then under ML by measuring the δ/σ ratio of the $\ln L$ difference between two competing topologies (δ) and its standard error (σ). For both approaches, the number of times the true (simulated) topology was recovered was recorded.

2.6. Relaxed molecular clock dating

To estimate divergence ages among placentals using the nuclear protein data set of Misawa and Janke (2003), we applied the Bayesian relaxed molecular clock approach of Kishino et al. (2001) and Thorne et al.

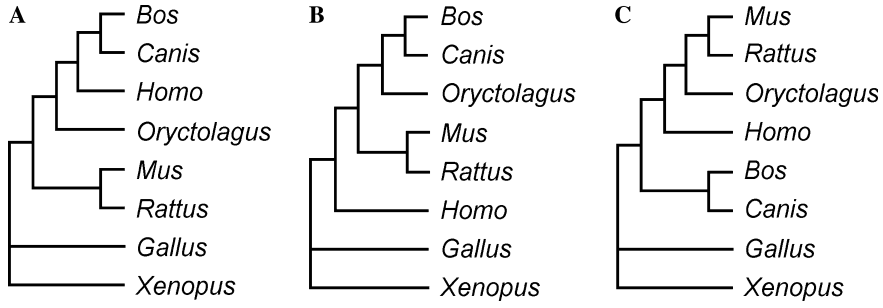


Fig. 1. Topologies compared under four different models for combining different genes. (A) "Basal-rodents tree" following Misawa and Janke (2003). (B) "Mitochondrial tree" following Misawa and Janke (2003). (C) "Glires tree."

(1998), implemented in the DIVTIME 5b package. Following Misawa and Janke (2003) and others (e.g., Kumar and Hedges, 1998), we used 310 Mya for the split between the chicken and the placentals, by constraining the root age of the ingroup to be bounded between 309 and 311 Mya (but see Reisz and Müller, 2004). The program ESTBRANCHES (Thorne et al., 1998) was used to estimate the variance–covariance matrix of branch lengths under a JTT model of protein evolution. Then, the program DIVTIME (Thorne et al., 1998) was used to estimate the posterior divergence times after 10,000 sampling of the MCMC – each sample separated by 100 cycles – and preceded by 100,000 cycles of burn-in. The following priors for Gamma distributions were used: 310 ± 155 Mya for root age in the absence of constraints on node times, 0.10 ± 0.05 amino acid replacements/Mya/100 sites at root node, and 0.003 ± 0.003 for the rate autocorrelation per Mya. The maximum number of time units between root and tip was set to 540 Mya. To evaluate whether different methods of clock relaxation gave similar results, the same analysis with the same paleontological constraint was performed under the non-parametric rate smoothing approach (Sanderson, 1997) implemented in the software r8s, with optimization via Powell's method. The optimal exponential penalty was 1.05 for smoothing the local estimates of substitution rates from a branch to the neighboring branch.

3. Results and discussion

3.1. Bayesian inference of phylogeny based on nuclear sequences

The MCMCMC analysis of the 20 concatenated nuclear proteins under a model incorporating among-site rate variation reached stationarity very quickly (Fig. 2). Independent runs with varying number of generations (100,000, 300,000, or 1,000,000) and tree sampling (every 20 or 100 generations) yielded exactly the same topology and posterior probabilities. The maximum posterior probability (MAP) tree is depicted in Fig. 3. Clade posterior probabilities support the association of *Oryctolagus* with *Mus* + *Rattus* (Glires: PP = 1.00), *Homo* + Glires (PP = 0.80), and *Bos* + *Canis* (PP = 0.99). In Fig. 2, we see that the Markov chains massively visited two topologies, the MAP one (i.e., the "Glires topology"), and the same topology but with a rooting on the branch leading to the primates (i.e., the "mitochondrial topology"). This can be quantified by the tree posterior probabilities, which range from 0.794 for the MAP tree, to 0.198 (*Homo* rooting), and then to 0.004 (*Bos* rooting), 0.003 (*Mus* + *Rattus* rooting), and 0.001 (*Canis* rooting) (Fig. 3).

The exact relationship between Bayesian posterior probabilities and bootstrap percentages is still poorly understood (Alfaro et al., 2003; Cummings et al., 2003; Douady et al., 2003; Waddell et al., 2002).

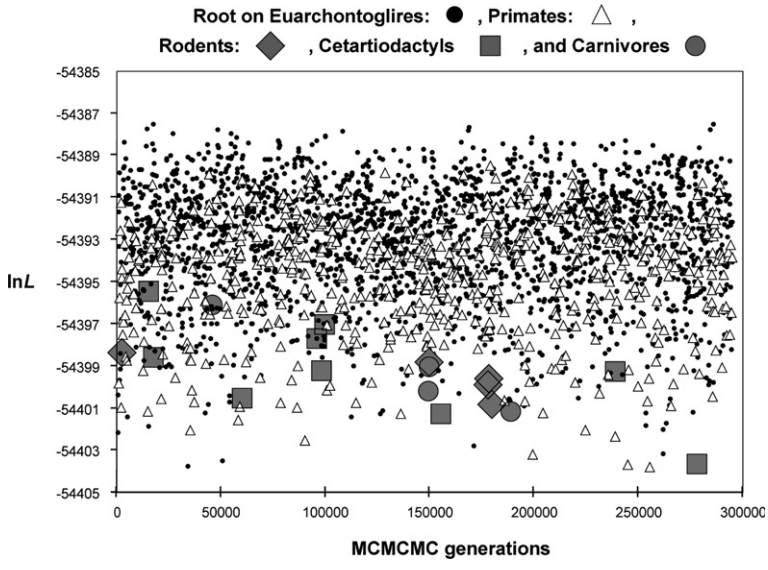


Fig. 2. Log-likelihood ($\ln L$) of the trees as a function of the number of MCMCMC generations for the nuclear data set of Misawa and Janke (2003), comprising 20 concatenated proteins. The topologies visited by the chains actually corresponded to five different rootings of the placental tree ((*Bos*, *Canis*), (*Homo*, (*Oryctolagus*, (*Mus*, *Rattus*)))), and are indicated by black circles (root on the branch leading to euarchontoglires, i.e., the "Glires tree"), white triangles (root on the branch leading to human, i.e., the "mitochondrial tree"), gray diamonds (root on the branch leading to murine rodents, i.e., the "basal-rodents tree" found by Misawa and Janke (2003) for the same data set), gray squares (root on the branch leading to bovine), and gray circles (root on the branch leading to dog). The former two topologies are massively favored by the Bayesian analysis.

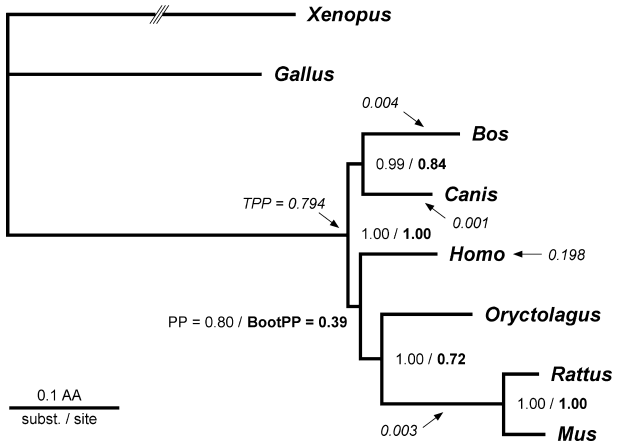


Fig. 3. Maximum posterior probability tree obtained after 100,000 MCMCMC generations under the JTT + Γ_4 model for the 20 concatenated nuclear proteins of Misawa and Janke (2003). For each node, the posterior probability (PP) is given, together with its bootstrapped PP (BootPP in bold) computed after 50 replicates. Arrows indicate the five best rootings of the placental tree, with italicized values corresponding to the tree PP (TPP). The marginal log-likelihood of the tree is $\ln L = -54391.0$, with Gamma parameter $\alpha = 0.63$ (95% credibility interval: 0.58–0.67), and total tree length = 1.89 (1.81–1.97). Branch lengths are proportional to the expected number of replacements per each of the 6441 amino acid (AA) sites. The length of the branch leading to *Xenopus* was reduced by a factor of three.

Credibility values of the Bayesian phylogenetic inference seem to constitute an overestimated index of the reliability of tree nodes under some circumstances (Cummings et al., 2003; Douady et al., 2003; Suzuki et al., 2002; Waddell et al., 2001) whereas bootstrap percentages might be too conservative (Hillis and Bull, 1993; Wilcox et al., 2002). To circumvent this problem, some authors have suggested computing bootstrapped posterior probabilities (Douady et al., 2003; Waddell et al., 2002). While time consuming, this approach helps to discriminate between moderately and strongly supported nodes for which initial PP stands above 0.95. Here, the bootstrapped Bayesian analysis is consistent with the monophyly of Glires (BootPP = 0.72, Fig. 3) and *Bos* + *Canis* (BootPP = 0.84), whereas Primates stands in trifurcation relative to the former two clades (BootPP = 0.39). Though its bootstrapped PP is not maximal, the Glires clade is retrieved, which is in sharp contrast with the conclusions of Misawa and Janke (2003) obtained under NJ, MP, ML, quartet puzzling, and Bayesian analysis without rate variation on the same data set.

3.2. Separate versus combined ML analysis

It has been shown that the model used to combine different sequences under the ML criterion can have an impact on the best phylogeny obtained (Pupko et al., 2002). For this reason, we tried to identified the most appropriate model for the nuclear and the mitochondrial data sets of Misawa and Janke (2003).

For the nuclear data set, the best model is the separate model with 20 α parameters (Table 1). The differences in AIC between the best model and the others is highly significant ($P_{AIC} < 0.001$; data not shown). Under this model, the nuclear data set favors the "Glires tree" relative to the "mitochondrial tree" and the "basal-rodents tree" (Fig. 1). In order to verify that the "Glires tree" was indeed the best tree we computed, under the separate model, the likelihood of the 15 trees that describe all the possible relationships between the five following taxa: Rodentia (*Mus* + *Rattus*), *Oryctolagus*, *Homo*, Laurasiatheria (*Canis* + *Bos*), and the outgroups (*Xenopus*, *Gallus*). The "Glires tree" remains the best tree among the 15 others, although the Kishino–Hasegawa test indicates that the differences in likelihood are not significant. The "Glires tree" is not significantly better than the "basal-rodents tree" ($P_{KH} = 0.17$), but is significantly better than the "mitochondrial tree" ($P_{KH} = 0.02$). Similar results were obtained with the proportional model with twenty α parameters, and with the concatenate model with one α parameter. The "Glires tree" was not significantly better than the "basal-rodents tree" ($P_{KH} = 0.33$ under the proportional model, and $P_{KH} = 0.34$ under the concatenate model),

Table 1
Akaike Information Criterion (AIC) values for the 20 nuclear proteins data set of Misawa and Janke (2003)

Branch length model Among-site rate variation model		Concatenate Homogeneous	Concatenate One alpha	Proportional 20 alpha	Separate 20 alpha
	df	13	14	52	280
Mitochondrial tree	ln *L*	−55670.02	−54435.76	−53561.77	−52795.58
	AIC	111366.03	108899.52	107227.54	106151.15
Basal-rodents tree	ln *L*	*−55571.40*	−54393.13	−53528.98	−52761.46
	AIC	*111168.80*	108814.27	107161.95	106082.91
Glires tree	ln *L*	−55608.09	*−54386.81*	*−53522.41*	*−52745.53*
	AIC	111242.19	*108801.61*	*107148.81*	***106051.06***

df is the number of degrees of freedom, ln *L* is the Log-likelihood value, and AIC is $-2 \times \ln L + 2 \times df$. "One alpha" indicates that a single α parameter of the Gamma distribution was used under the concatenate model; "20 alpha" indicates that different α parameters were used for each gene under the proportional and the separate models. The best AIC value is indicated in bold. The best ln *L* and AIC values are italicized for each model.

but it was significantly better than the "mitochondrial tree" ($P_{KH} = 0.03$ under the proportional model, and $P_{KH} = 0.009$ under the concatenate model). When the homogeneous model was used with the concatenate model the best tree was the "basal-rodents tree," in agreement with Misawa and Janke (2003). However, KH tests indicate that this tree is not significantly better than the "Glires tree" ($P_{KH} = 0.07$). This result is at variance with the high bootstrap values that support the "basal-rodents tree" in Misawa and Janke (2003). The mitochondrial tree was also rejected under the homogeneous model ($P_{KH} = 0.001$).

For the mitochondrial data set, there was no statistical difference between the separate model with 12 α parameters and the proportional model with 12 α parameters ($P_{AIC} = 0.15$). However, these two models were significantly better than the concatenate model with one α parameter or with the homogeneous model ($P_{AIC} < 0.001$). Because the predictions based on a complex model are more prone to error than those based on a simpler model (Nei and Kumar, 2000, p. 154), the best model for the mitochondrial data set was the proportional model with 12 α parameters. In opposition to Misawa and Janke (2003), our analysis of the mitochondrial data set always indicated that the "basal-rodents tree" was the best tree relative to "the mitochondrial tree" and the "Glires tree" (Table 2). This discrepancy between the observation of Misawa and Janke (2003) and our results might result from the fact we used the mtREV model. Under the JTT model of amino acid substitution and the concatenate homogeneous model, the "mitochondrial tree" ($\ln L = -25951.76$) was the best tree compared with the "basal-rodents tree" ($\ln L = -25956.96$) and the "Glires tree" ($\ln L = -25989.03$). However, KH tests indicated that the "basal-rodents tree" was not significantly better than the two others when among-site rate variation is taken into account ($P_{KH} > 0.05$), whatever the model considered. Only under the concatenate homogeneous model did the "Glires tree" appear to be significantly less likely than the "basal-rodents tree" ($P_{KH} = 0.04$).

The mitochondrial genome based-analysis did not support the monophyly of Glires (Table 2). There is currently no consensus concerning the phylogenetic positions of lagomorph and rodent taxa based on mitogenomic analysis. Few mitochondrial studies support the monophyly of Glires once the tree is rooted (e.g., Cao et al., 2000; Pupko et al., 2002; Reyes et al., 2004; Waddell et al., 2001), many do not support it (e.g., Arnason et al., 2002; Schmitz et al., 2000), and the likelihood of alternative topologies is usually not significantly different (Arnason et al., 2002; Cao et al., 2000; Pupko et al., 2002). This discrepancy might come from the fact that mitochondrial sequences seem to contain less phylogenetic signal than the nuclear genes at deep taxonomic levels (Springer et al., 2001; but see Reyes et al., 2004). A complementary explanation might come from the observation that "the placement of the outgroup sequence can have a confounding effect on the ingroup tree, whereby the ingroup is correct when using the ingroup sequences alone, but with the inclusion of the outgroup the ingroup tree becomes incorrect" (Holland et al., 2003). In their mitochondrial analysis, Lin et al. (2002a,b) found monophyly of Glires in their unrooted tree, but rooting the tree by marsupials and a monotreme then induced a shift in the root location, that led to the paraphyly of Glires. However, it has recently been shown that a better agreement between mitochondrial and nuclear topologies is found when comprehensive taxon sampling is used, compositional bias is taken into account, and robust methods are used (Reyes et al., 2004).

About the effect of different ML models of branch length estimation on the inferred phylogeny, our results on the nuclear and mitochondrial proteins confirm the observations of Pupko et al. (2002). A concatenate model is less appropriate for phylogenetic analysis than are separate or proportional models. Unfortunately, these models are time consuming and not available in most phylogenetic analysis packages. Additionally, models that account for among-site rate variation here appear significantly better than the homogeneous model ($P_{AIC} < 0.001$).

Table 2
Akaike Information Criterion (AIC) values for the 12 mitochondrial protein data set of Misawa and Janke (2003)

Branch length model Among-site rate variation model		Concatenate Homogeneous	Concatenate One alpha	Proportional 12 alpha	Separate 12 alpha
	df	13	14	36	168
Mitochondrial tree	$\ln L$	−24618.68	−23931.57	−23767.79	−23614.30
	AIC	49263.37	47891.14	47607.58	47564.61
Basal-rodents tree	$\ln L$	*−24610.63*	*−23920.05*	*−23758.09*	*−23607.91*
	AIC	*49247.27*	*47868.10*	*47588.19*	***47551.82***
Glires tree	$\ln L$	−24646.17	−23929.87	−23767.57	−23616.30
	AIC	49318.34	47887.74	47607.14	47568.59

df is the number of degrees of freedom, $\ln L$ is the Log-likelihood value, and AIC is $-2 \times \ln L + 2 \times df$. "One alpha" indicates that a single α parameter of the Gamma distribution was used under the concatenate model; "12 alpha" indicates that different α parameters were used for each gene under the proportional and the separate models. The best AIC value is indicated in bold. The AIC value under the separate model is not statistically better than the AIC value under the proportional model. The best $\ln L$ and AIC values are italicized for each model.

3.3. Analysis of alternative rooting

The SH test for multiple tree comparisons was used to evaluate the significant differences among alternative rooting. The null hypothesis is that all 105 rootings compared—including the ML one—are equally good explanations of the data. We computed the log-likelihood of the 105 possibilities of connecting rodents, lagomorphs, primates, carnivores, and cetartiodactyls into a bifurcating subtree, rooted by *Xenopus* and *Gallus*. These $15 \times 7 = 105$ rootings arise because there are 15 possible unrooted trees with five placental orders, each with $2 \times 5 - 3 = 7$ branches, i.e., seven potential locations for the root. The best rooting—i.e., the one yielding the highest likelihood—was the "Glires tree." The root was along the branch separating Carnivora + Cetartiodactyla—members of Laurasiatheria—from Glires + Primates—members of Euarchontoglires (= Supraprimates) (Fig. 4). The 104 other rootings were ranked by increasing level of statistical rejection by the SH test (P_{SH} values ranging from 0.97 to less than 0.001). Among them, 85 were statistically rejected ($P_{SH} < 0.05$). The 19 other rootings occur along the branches of three unrooted trees among 15 possible ones. In these three unrooted trees, Rodentia, Lagomorpha, and Primates are directly connected (Fig. 4). The third best likelihood score corresponded to the rooting observed by Misawa and Janke (2003).

The first and second best alternative rootings (Fig. 4) differed from the best rooting by, respectively, 1.5 and 5.8 units of $\ln L$. Two SOWH tests were conducted to evaluate the statistical significance of the difference in $\ln L$ between the best tree and the two first alternative rootings. The tests rejected the two alternative rootings at $P < 0.01$, as the $\ln L$ of the most likely rooting was never greater than the $\ln L$ of the alternative rootings used to simulate the 100 parametric bootstrap replicates.

3.4. Simulations: effect of distance between outgroup and ingroup

To test whether the lissamphibian (*Xenopus*) and avian (*Gallus*) outgroup was too distant from the ingroup to provide a reliable root, a Monte Carlo simulation study was conducted. We evaluated the ability of MP and ML to recover the simulated topology after 100 replicates by varying the outgroup-to-ingroup distance. The lengths of the three branches leading to xenopus, chicken, and placentals were thus rescaled from 0.01 to 100 times (Table 3). While the original tree corresponds to a scaling factor of 1.00, the situation for which the outgroup was very close to the direct placental ancestor corresponded to a scaling factor of 0.01. Conversely, a random rooting behavior of the outgroup was simulated by a scaling factor of 100, and we should expect a random choice of ingroup branches, proportionally to their length (Wheeler, 1990). Scaling factors of 0.1 and 10 represented two intermediates between the two extreme situations.

For scaling factors of 0.01 and 0.1, both MP and ML recovered the true ingroup root (Table 3). For very low outgroup-to-ingroup distances (scaling factor of 0.01), alternative roots were all rejected, by $\Delta = 47$ to 136 extra

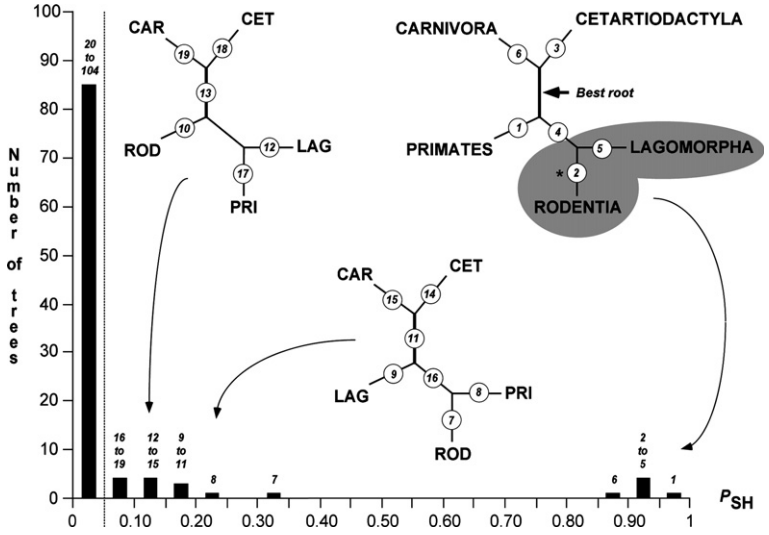
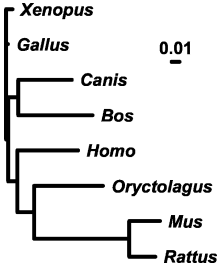
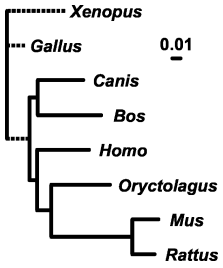
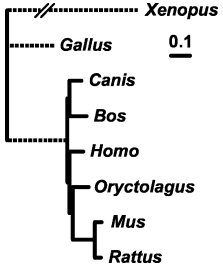
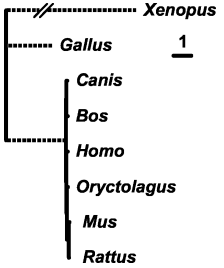
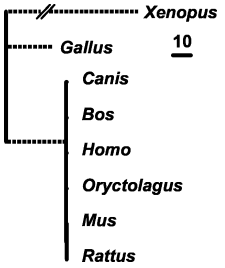


Fig. 4. Confidence levels of the Shimodaira–Hasegawa test for the location of the root of the placental tree based on the nuclear data set of Misawa and Janke (2003). The 105 possible rootings of the placental tree by *Xenopus* and *Gallus* were evaluated and ranked according to their increasing level of rejection. Log-likelihoods were computed under the JTT + F + Γ_4 model. 85 rooting possibilities were statistically rejected. The 19 remaining positions of the root, which are not statistically rejected, are located on three unrooted trees. Those unrooted trees always group rodents, lagomorphs and primates. The circled numbers on the unrooted trees indicate the various alternatives positions of the root. The numbers inside the circle, and the italicized numbers on histogram bars, reflect the ranking of the rooted trees according to their [P_{SH}]. The star indicates the best tree according to Misawa and Janke (2003). The vertical dashed line corresponds to the $P_{SH} = 0.05$ rejection level. CAR, CET, LAG, PRI, and ROD stand for Carnivora, Cetartiodactyla, Lagomorpha, Primates, and Rodentia, respectively.

Table 3
Impact of the evolutionary distance between the lissamphibian and avian outgroup and the ingroup for recovering the simulated topology under maximum parsimony (MP) and maximum likelihood (ML) based on the nuclear protein data set of Misawa and Janke (2003)

	 Scaling = 0.01		 Scaling = 0.1		 Scaling = 1 (real)		 Scaling = 10		 Scaling = 100	
	MP	ML	MP	ML	MP	ML	MP	ML	MP	ML
True root [4%]	100% $\Delta = 0$	100% $\delta/\sigma = 0$	100% 0	100% 0	67% 2 ± 3	100% 0	0% 40 ± 10	41% 0.2 ± 0.3	0% 43 ± 10	0% 0.8 ± 0.6
Root #1 [15%]	0% 47 ± 7	0% 4.8 ± 0.5	0% 43 ± 8	0% 4.5 ± 0.5	10% 12 ± 8	0% 2.9 ± 0.5	1% 29 ± 11	16 % 0.6 ± 0.5	0% 31 ± 11	10% 0.8 ± 0.6
Root #2(*)[21%]	0% 136 ± 12	0% 7.7 ± 0.5	0% 119 ± 12	0% 7.4 ± 0.4	4% 24 ± 12	0% 5.7 ± 0.5	88 % 0 ± 1	3% 1.3 ± 0.7	90% 0 ± 2	18% 0.7 ± 0.6
Root #3 [17%]	0% 51 ± 7	0% 4.5 ± 0.4	0% 44 ± 7	0% 4.3 ± 0.4	15% 9 ± 7	0% 3.1 ± 0.5	7% 17 ± 12	6% 0.7 ± 0.5	8% 21 ± 13	32% 0.6 ± 0.6
Root #4 [4%]	0% 47 ± 7	0% 4.8 ± 0.4	0% 43 ± 8	0% 4.5 ± 0.5	2% 17 ± 8	0% 2.9 ± 0.5	0% 36 ± 9	14% 0.6 ± 0.5	0% 39 ± 8	1% 1.0 ± 0.9
Root #5 [15%]	0% 135 ± 12	0% 7.7 ± 0.5	0% 120 ± 11	0% 7.4 ± 0.5	0% 38 ± 12	0% 5.7 ± 0.5	3% 21 ± 9	3% 1.3 ± 0.7	2% 20 ± 10	21% 0.7 ± 0.6
Root #6 [12%]	0% 49 ± 7	0% 4.5 ± 0.4	0% 43 ± 7	0% 4.3 ± 0.4	2% 16 ± 7	0% 3.1 ± 0.5	1% 30 ± 13	17% 0.7 ± 0.6	0% 33 ± 12	18% 0.7 ± 0.6

Outgroup-to-ingroup distances are expressed as a scaling factor (ranging from 0.01 to 100) for the three (dashed) branches leading to *Xenopus* and *Gallus*. A scaling factor of 1.00 corresponds to the real data. Seven locations of the root are evaluated: the true root is the one used for simulations and corresponds to the highest-likelihood rooting (cf. Fig. 4). The six alternative rootings (1–6) are numbered according to Fig. 4 and correspond to the outgroup branching on *Homo* (#1), murine rodents (#2), *Bos* (#3), Glires (#4), *Oryctolagus* (#5), and *Canis* (#6). Random rooting of the ingroup by the outgroup would occur at a percentage given between brackets, assuming that each of the nine ingroup branches randomly attracts the root according to its length. One hundred Monte Carlo simulations of 6441 amino acids for eight taxa were conducted for each of the five outgroup-to-ingroup distances. Each entry of the table contains two lines. The first line reports the percentage of times the simulated topology was recovered among the 100 replicates. The second line gives for the MP reconstruction the number of extra steps (Δ) relative to the best tree found with the simulated data set, or for the ML reconstructions the ratio between the difference in likelihood relative to the best tree and the standard error of this difference (δ/σ). All values are expressed as mean ± standard error over the 100 replicates. For the scaling factor of 1.00, outgroup-to-ingroup branches (dashed lines) are eight times longer than ingroup branches on average (full lines). (*) Root found by Misawa and Janke (2003).

steps, and $\delta/\sigma = 4.5$ to 7.7 standardized log-likelihoods. These values represent the average differences (over the 100 simulations) between the score (number of steps under MP, or $\ln L$ units under ML) of the simulated topology (root between euarchontoglires and laurasiatherians) and the score of the six alternative rootings. These δ/σ ratios corresponded to $P_{SH} < 0.01$. For moderate outgroup-to-ingroup distances (scaling factor of 0.10), alternative rootings were all rejected by $\Delta = 43$ to 120 extra steps and $\delta/\sigma =$ 4.3 to 7.4 (corresponding to $P_{SH} < 0.05$). Under the conditions of Misawa and Janke's (2003) data set (scaling factor of 1.00), ML again recovered the true root in all cases (mean δ/σ ranged from 2.9 to 5.7, corresponding to $P_{SH} < 0.12$). However, the performance of MP began to decrease, with only 67% of true root recovery. The mean number of extra steps to discriminate among alternative rootings also diminished (Δ ranged from 9 to 38). In the case of excessive outgroup-to-ingroup distance, MP never recovered the true root (Table 3), but massively favored rooting along the *Mus* + *Rattus* branch. For a scaling factor of 10, the performance of ML began to diminish. The true root was recovered in 41% of the cases, though with little discriminating power (δ/σ for alternative roots ranged from 0.6 to 1.3, corresponding to P_{SH} of 0.27 to 0.61). For a scaling factor of 100, ML randomly rooted the ingroup, with recovery percentages being approximately those expected under a random rooting proportionally to ingroup branch lengths. The two preferred rootings were along the bovine and the murine branches, though all seven possibilities behaved similarly, with δ/σ of 0.6–0.8, corresponding to P_{SH} of 0.46–0.57.

As expected, the ingroup was more correctly rooted when the length of the branches leading to lissamphibian and avian outgroup taxa were shorter (Graham et al., 2002; Huelsenbeck et al., 2002; Wheeler, 1990). A clear difference of behavior in recovering the true root was observed between MP and ML for increasing outgroup-to-ingroup distances. MP artificially favored a rooting along the longest ingroup branch—the one leading to murine rodents (cf. Fig. 3)—which might explain the topology observed by Misawa and Janke (2003). For the 6441 amino acid data set, the critical interval of outgroup-to-ingroup distance scaling where the performance of MP decreased was between 0.10 and 1.00, whereas the performance of ML decreased for outgroup-to-ingroup distance scaling between 1.00 and 10.00. More importantly, MP converged towards an artifactual rooting with increasing outgroup-to-ingroup distance by systematically picking-up the root along the long *Mus*–*Rattus* branch. This is likely to be diagnostic of a long-branch attraction phenomenon (Felsenstein, 1978) of the murine branch by the distant outgroup. This feature was reinforced as the branch lengths of the lissamphibian and avian taxa were increased by the scaling factor. In contrast, and under our simulation conditions, ML did not strongly support one given alternative topology when amino acid information was insufficient for reaching proper conclusions, as was the case here when the outgroup randomly rooted the ingroup. The root was placed randomly according to the branch lengths of the ingroup. Such differences of behavior between MP and ML have been previously observed by Swofford et al. (2001). Because Misawa and Janke's (2003) data fall in the area where ML still correctly localizes the ingroup root, the conclusion of this simulation study is that the ML method should confidently root the placental ingroup, despite the apparent very long phylogenetic distance between mammals and the lissamphibian and avian taxa.

3.5. Other nuclear proteins: the Murphy et al. (2001b) data set

Bayesian analysis of the 15 concatenated proteins from Murphy et al. (2001b) for the same ingroup sampling as Misawa and Janke (2003) indicated the monophyly of Glires, but *Homo* falls into an external position relative to the five other placentals. When two additional placentals were added, one afrotherian and one xenarthran, a tree in which Glires, euarchontoglires, and laurasiatherians are monophyletic was recovered (Fig. 5; Waddell et al., 2001). This point emphasizes the need for denser taxon sampling to increase the reli-

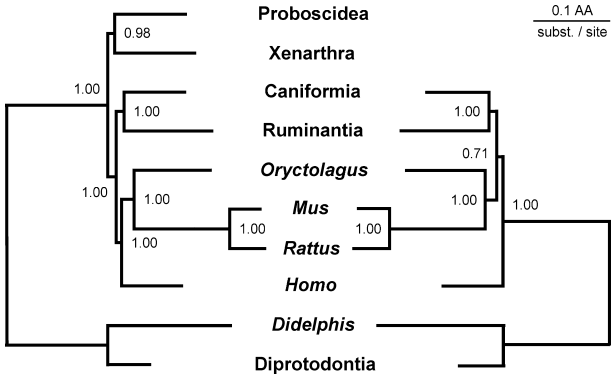


Fig. 5. Bayesian analysis of 15 concatenated nuclear proteins from the data set of Murphy et al. (2001b) with two different taxon samplings among placentals. The search of the maximum posterior probability tree was conducted with and without the afrotherian (elephant) and xenarthran (sloth) species. For the 8-taxon data set, the same placentals as in Misawa and Janke (2003) were considered, but a closer (marsupial) outgroup was used. For each node, the posterior probability (PP) is given. The 10-taxon tree (on the left) has a marginal log-likelihood of $\ln L = -35402.8$, with Gamma parameter $\alpha = 0.54$ (95% credibility interval: 0.49–0.59), and total tree length = 1.32 (1.26–1.38). The 8-taxon tree (on the right) has a marginal log-likelihood of $\ln L = -31106.1$, with Gamma parameter $\alpha = 0.55$ (95% credibility interval: 0.50–0.61), and total tree length = 1.11 (1.06–1.17). Branch lengths are proportional to the expected number of replacements per each of the 4333 amino acid (AA) sites.

ability of phylogenetic inferences (Adachi and Hasegawa, 1995; Lecointre et al., 1993; Philippe and Douzery, 1994; Zwickl and Hillis, 2002). Among the four major placental clades, the elephant and the sloth break the long isolated branch ancestral to placentals, and facilitate the rooting by more distant taxa like marsupials (Delsuc et al., 2002) or even avians or lissamphibians.

3.6. Molecular dating with relaxed clocks

Our Bayesian divergence ages based on Misawa and Janke's (2003) data set show that mice and rats separated ~11 Mya with a 95% credibility interval of 7–17 Mya (Fig. 6). Using the non-parametric rate smoothing approach for clock relaxation yields a *Mus–Rattus* split at 13 Mya. These results contrast with the 38–44 Mya estimate found by Misawa and Janke (2003) on the same data set but with the linearized tree method of molecular dating, and with a topology with murines basalmost among placentals. Springer et al. (2003) observed, for a larger data set, that the age of the mouse/rat split depends on the tree rooting. Divergence of the two murid taxa is estimated at 16 Mya when the Afrotheria are basal in the placental tree (Murphy et al., 2001b) or 35 Mya when the murid rodents are the most basal placentals (Arnason et al., 2002). The difference in topology between our best tree (Fig. 3), the topology of which is in agreement with Madsen et al. (2001) and Murphy et al. (2001a,b), but which contrasts with that of the tree published by Misawa and Janke (2003), is likely to explain the three times more recent estimate obtained here for the split between mice and rats.

Other Bayesian divergence times (Fig. 6) are ~55 Mya for the first split among Glires (95% credibility interval = 40–74 Mya), and ~77 Mya (60–98 Mya) for the placentals represented here (all boreoeutherians). These estimates are 1.2–1.5 times younger than those obtained by Springer et al. (2003) on nuclear genes only. This might be explained by the sparse taxon sampling considered here, combined with the lack of representatives of the two other major placental clades—Afrotheria and Xenarthra—and the very distant fossil reference. However, even if our estimate of the mouse/rat split is 1.5 times deeper—i.e., 16 Mya—the present molecular estimate conforms more with the 12–14 Mya of the paleontological estimates (Catzeflis et al., 1992) and to the 16 Mya of some recent molecular studies based on local (Douzery et al., 2003) or relaxed (Adkins et al., 2003; Springer et al., 2003) molecular clocks than to other molecular estimates of 23 Mya (Adkins et al., 2001), 33 Mya (Nei et al., 2001), 41 Mya (Kumar and Hedges, 1998), or 42 Mya (Huchon et al., 2000) based on lineage-specific or global clock methods. It should be noted that all these molecular dating estimates ignore sources of uncertainty due to errors in sequence determination, sequence alignment, model assumptions, topology, edge lengths, ancestral polymorphism, and calibration (Waddell et al., 1999).

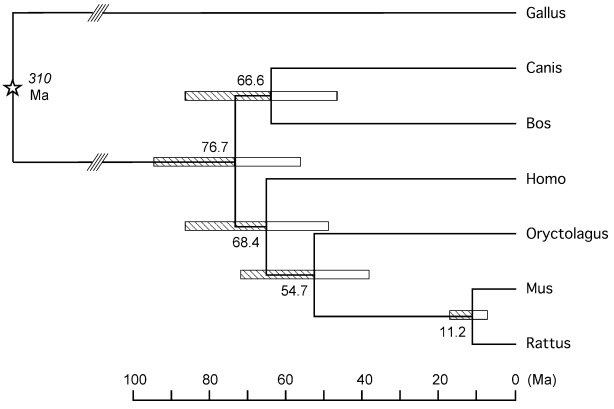


Fig. 6. Chronogram of 6 placentals and one avian outgroup, with Bayesian divergence times estimated on the nuclear data set of Misawa and Janke (2003). For each node, the posterior divergence time is given in millions years (My). The star indicates that the root age was a priori bounded between 309 and 311 Ma, based on the synapsid/diapsid paleontological calibration at 310 Ma. 95% credibility intervals are indicated by horizontal rectangles, in white and hatched for lower (recent) and upper (ancient) intervals respectively. Branch lengths are proportional to time in million years. Branches leading to the chicken and to the placentals were reduced by a factor of four.

4. Conclusion

Misawa and Janke's (2003) phylogenetic conclusions are drawn from congruent results from independent tree-building methods. However, their analyses fail to take among-site rate variation into account. The single situation where Misawa and Janke's (2003) results did not support a basal position for rodents is when the tree was inferred using the Bayesian approach with among-site rate variation. Model comparisons using AIC indicate that accounting for among-site rate variation significantly improves the likelihood of the data. When α parameters are included in the model, the best phylogenetic hypothesis supports the monophyly of Glires, except for the mitochondrial data set. Our ML studies also support the monophyly of Glires. According to statistical tests of alternative topologies (KH and SH tests), we cannot exclude that Glires might be paraphyletic, but this result is hampered by the fact that SOWH tests rejected the two best alternative rootings. Contrary to the claim of Misawa and Janke (2003), the monophyly of Glires is a valid and likely explanation of the alignment based on 20 nuclear proteins, whereas a basal position of murine rodents relative to rabbit, human, bovine and dog is less likely. Additionally, the existence of the Glires clade is supported by other molecular data sets,

as indicated by our protein analysis of a subset of the sequences published by Murphy et al. (2001b). It is also supported by morphological (Landry, 1999; Luckett and Hartenberger, 1993; Meng et al., 1994), paleontological (Archibald et al., 2001; but see Fostowicz-Frelik and Kielan-Jaworowska (2002) for a different opinion), and chromosomal (Stanyon et al., 2003) studies. Our results emphasize once more the importance of accounting for among-site rate variation (e.g., Sullivan and Swofford, 1997; Yang, 1996) in phylogenetic tree reconstructions based on probabilistic approaches. Based on a topology with monophyly of Glires, a Bayesian relaxed molecular clock dating of the age of the *Mus–Rattus* split yields an estimate of 7–17 Mya, compatible with the fossil record.

Acknowledgments

Axel Janke and Kazuharu Misawa kindly communicated their protein alignments. Helpful comments on the manuscript were provided by Frédéric Delsuc, Doyle McKey, and two anonymous reviewers. This work has been supported by the "Genopole Montpellier Languedoc-Roussillon," the "Action Bioinformatique inter-EPST" of the CNRS, and by IFR119 "Biodiversité Continentale Méditerranéenne et Tropicale" (Montpellier) and INFOBIOGEN (Evry, France) computing facilities. Additional access to computers was kindly provided by Hervé Philippe (Montréal, Canada), and Marcel Méchali and Denis Pugnère (Institut de Génétique Humaine de Montpellier [IGH], CNRS UPR 1142). This publication is the contribution No. EPML-003 of the Equipe-Projet multi-laboratoires CNRS-STIC "Méthodes informatiques pour la biologie moléculaire" and No. 2004-041 of the Institut des Sciences de l'Evolution de Montpellier (UMR 5554—CNRS).

References

Adachi, J., Hasegawa, M., 1995. Phylogeny of whales: dependence of the inference on species sampling. Mol. Biol. Evol. 12, 177–179.

Adachi, J., Hasegawa, M., 1996. Instability of quartet analyses of molecular sequence data by the maximum likelihood method: the Cetacea/Artiodactyla relationships. Mol. Phylogenet. Evol. 6, 72–76.

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.

Adkins, R.M., Walton, A.H., Honeycutt, R.L., 2003. Higher-level systematics of rodents and divergence time estimates based on two congruent nuclear genes. Mol. Phylogenet. Evol. 26, 409–420.

Alfaro, M.E., Zoller, S., Lutzoni, F., 2003. Bayes or bootstrap. A simulation study comparing the performance of Bayesian Markov Chain Monte Carlo sampling and bootstrapping in assessing phylogenetic confidence. Mol. Biol. Evol. 20, 255–266.

Amrine-Madsen, H., Koepfli, K., Wayne, R.K., Springer, M.S., 2003. A new phylogenetic marker, apolipoprotein B, provides compelling evidence for eutherian relationships. Mol. Phylogenet. Evol. 28, 225–240.

Archibald, J.D., Averianov, A.O., Ekdale, E.G., 2001. Late Cretaceous relatives of rabbits, rodents, and other extant eutherian mammals. Nature 414, 62–65.

Arnason, U., Adegoke, J.A., Bodin, K., Born, E.W., Esa, Y.B., Gullberg, A., Nilsson, M., Short, R.V., Xu, X., Janke, A., 2002. Mammalian mitogenomic relationships and the root of the eutherian tree. Proc. Natl. Acad. Sci. USA 99, 8151–8156.

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

Catzeflis, F.M., Aguilar, J.-P., Jaeger, J.-J., 1992. Muroid rodents: phylogeny and evolution. Trends Ecol. Evol. 7, 122–126.

Cummings, M.P., Handley, S.A., Myers, D.S., Reed, D.L., Rokas, A., Winka, K., 2003. Comparing bootstrap and posterior probability values in the four-taxon case. Syst. Biol. 52, 477–487.

de Jong, W.W., van Dijk, M.A., Poux, C., Kappe, G., van Rheede, T., Madsen, O., 2003. Indels in protein-coding sequences of Euarchontoglires constrain the rooting of the eutherian tree. Mol. Phylogenet. Evol. 28, 328–340.

DeBry, R.W., Sagel, R.M., 2001. Phylogeny of Rodentia (Mammalia) inferred from the nuclear-encoded gene IRBP. Mol. Phylogenet. Evol. 19, 290–301.

Delsuc, F., Scally, M., Madsen, O., Stanhope, M.J., de Jong, W.W., Catzeflis, F., 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.

D'Erchia, A.M., Gissi, C., Pesole, G., Saccone, C., Arnason, U., 1996. The guinea-pig is not a rodent. Nature 381, 597–600.

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

Douzery, E.J.P., Delsuc, F., Stanhope, M.J., Huchon, D., 2003. Local molecular clocks in three nuclear genes: divergence times for rodents and other mammals, and incompatibility among fossil calibrations. J. Mol. Evol. 57, S201–S213.

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

Fostowicz-Frelik, L., Kielan-Jaworowska, Z., 2002. Lower incisor in zalambdalestid mammals (Eutheria) and its phylogenetic implications. Acta Pal. Pol. 47, 177–180.

Gaut, B.S., Lewis, P.O., 1995. Success of maximum likelihood phylogeny inference in the four-taxon case. Mol. Biol. Evol. 12, 152–162.

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

Graham, S.W., Olmstead, R.G., Barrett, S.C.H., 2002. Rooting phylogenetic trees with distant outgroups: a case study from the commelinoid monocots. Mol. Biol. Evol. 19, 1769–1781.

Grassly, N.C., Adachi, J., Rambaut, A., 1997. PSeq-Gen: an application for the Monte Carlo simulation of protein sequence evolution along phylogenetic trees. Comput. Appl. Biosci. 13, 559–560.

Graur, D., Duret, L., Gouy, M., 1996. Phylogenetic position of the order Lagomorpha (rabbits, hares and allies). Nature 379, 333–335.

Graur, D., Hide, W.A., Li, W.-H., 1991. Is the guinea-pig a rodent?. Nature 351, 649–652.

Halanych, K.M., 1998. Lagomorphs misplaced by more characters and fewer taxa. Syst. Biol. 47, 138–146.

Hillis, D.M., Bull, J.J., 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. Syst. Biol. 42, 182–192.

Holland, B.R., Penny, D., Hendy, M., 2003. Outgroup misplacement and phylogenetic inaccuracy under a molecular clock - A simulation study. Syst. Biol. 52, 229–238.

Huchon, D., Catzeflis, F., Douzery, E.J.P., 2000. Variance of molecular datings, evolution of rodents, and the phylogenetic affinities between Ctenodactylidae and Hystricognathi. Proc. R. Soc. Lond. B 267, 393–402.

Huchon, D., Madsen, O., Sibbald, M.J.J.B., Ament, K., Stanhope, M., 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., Bollback, J.P., Levine, A.M., 2002. Inferring the root of a phylogenetic tree. Syst. Biol. 51, 32–43.

Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. Comput. Applic. Biosci. 8, 275–282.

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.

Kishino, H., Thorne, J.L., Bruno, W.J., 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. Mol. Biol. Evol. 18, 352–361.

Kumar, S., Hedges, S.B., 1998. A molecular timescale for vertebrate evolution. Nature 392, 917–920.

Landry, S.O.J., 1999. A proposal for a new classification and nomenclature for the Glires (Lagomorpha and Rodentia). Mitt. Mus. Nat.kd. Berl. Zool. Reihe 75, 283–316.

Lecointre, G., Philippe, H., Lè, H.L.V., Le Guyader, H., 1993. Species sampling has a major impact on phylogenetic inference. Mol. Phylogenet. Evol. 2, 205–224.

Li, C.-K., Wilson, R.W., Dawson, M.R., Krishtalka, L., 1987. The origin of rodents and lagomorphs. In: Genoways, H.H. (Ed.), Current Mammalogy. Plenum Press, New York, pp. 97–108.

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

Lin, Y.-H., Waddell, P.J., Penny, D., 2002a. Pika and vole mitochondrial genomes increase support for both rodent monophyly and glires. Gene 294, 119–129.

Linhart, H., 1988. A test whether two AIC's differ significantly. South African Statist. J. 22, 153–161.

Luckett, W.P., Hartenberger, J.-L., 1993. Monophyly or polyphyly of the order Rodentia: possible conflict between morphological and molecular interpretations. J. Mammal. Evol. 1, 127–147.

Madsen, O., Scally, M., Douady, C.J., Kao, D.J., DeBry, R.W., Adkins, R., Amrine, H., 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.

Meng, J., Wyss, A.R., Dawson, M.R., Zhai, R., 1994. Primitive fossil rodent from inner Mongolia and its implications for mammalian phylogeny. Nature 370, 134–136.

Misawa, K., Janke, A., 2003. Revisiting the Glires concept - phylogenetic analysis of nuclear sequences. Mol. Phylogenet. Evol. 28, 320–327.

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

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

Narita, Y., Oda, S., Takenaka, O., Kageyama, T., 2001. Phylogenetic position of Eulipotyphla inferred from the cDNA sequences of pepsinogens A and C. Mol. Phylogenet. Evol. 21, 32–42.

Nei, M., Kumar, S., 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Nei, M., Xu, P., Glazko, G., 2001. Estimation of divergence times from multiprotein sequences for a few mammalian species and several distantly related organisms. Proc. Natl. Acad. Sci. USA 98, 2497–2502.

Philippe, H., 1997. Rodent monophyly: pitfalls of molecular phylogenies. J. Mol. Evol. 45, 712–715.

Philippe, H., Douzery, E., 1994. The pitfalls of molecular phylogeny based on four species as illustrated by the Cetacea/Artiodactyla relationships. J. Mammal. Evol. 2, 133–152.

Pupko, T., Huchon, D., Cao, Y., Okada, N., Hasegawa, M., 2002. Combining multiple data sets in a likelihood analysis: which models are the best?. Mol. Biol. Evol. 19, 2294–2307.

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

Reyes, A., Gissi, C., Catzeflis, F., Nevo, E., Pesole, G., Saccone, C., 2004. Congruent mammalian trees from mitochondrial and nuclear genes using Bayesian methods. Mol. Biol. Evol. 21, 397–403.

Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.

Sakamoto, Y., Ishiguro, M., Kitagawa, G., 1986. Akaike Information Criterion Statistics. D. Reidel Pub., Dordrecht.

Sanderson, M.J., 1997. A nonparametric approach to estimating divergence times in the absence of rate constancy. Mol. Biol. Evol. 14, 1218–1231.

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

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

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

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

Stanhope, M.J., Czelusniak, J., Si, J.-S., Nickerson, J., Goodman, M., 1992. A molecular perspective on mammalian evolution from the gene encoding interphotoreceptor retinoid binding protein, with convincing evidence for bat monophyly. Mol. Phylogenet. Evol. 1, 148–160.

Stanyon, R., Stone, G., Garcia, M., Froenicke, L., 2003. Reciprocal chromosome painting shows that squirrels, unlike murid rodents, have a highly conserved genome organization. Genomics 82, 245–249.

Sullivan, J., Swofford, D.L., 1997. Are guinea pigs rodents? The importance of adequate models in molecular phylogenetics. J. Mammal. Evol. 4, 77–86.

Suzuki, Y., Glazko, G.V., Nei, M., 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. Proc. Natl. Acad. Sci. USA 99, 16138–16143.

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, 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.

Thorne, J.L., Kishino, H., Painter, I.S., 1998. Estimating the rate of evolution of the rate of molecular evolution. Mol. Biol. Evol. 15, 1647–1657.

Waddell, P.J., Cao, Y., Hasegawa, M., Mindell, D.P., 1999. Assessing the cretaceous superordinal divergence times within birds and placental mammals by using whole mitochondrial protein sequences and an extended statistical framework. Syst. Biol. 48, 119–137.

Waddell, P.J., Shelley, S., 2003. Evaluating placental inter-ordinal phylogenies with novel sequences including RAG1, ?-fibrinogen, ND6, and mt-tRNA, plus MCMC-driven nucleotide, amino acid, and codon models. Mol. Phylogenet. Evol. 28, 197–224.

Waddell, P.J., Kishino, H., Ota, R., 2001. A phylogenetic foundation for comparative mammalian genomics. Gen. Inform. 12, 141–154.

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

Wallau, B.R., Schmitz, A., Perry, S.F., 2000. Lung morphology in rodents (Mammalia, Rodentia) and its implications for systematics. J. Morphol. 246, 228–248.

Wheeler, W.C., 1990. Nucleic acid sequence phylogeny and random outgroups. Cladistics 6, 363–367.

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.

Wood, A.E., 1957. What, if anything, is a rabbit. Evolution 11, 417–425.

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

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

Zwickl, D.J., Hillis, D.M., 2002. Increased taxon sampling greatly reduces phylogenetic error. Syst. Biol. 51, 588–598.