

Reanalysis of Murphy et al.'s Data Gives Various Mammalian Phylogenies and Suggests Overcredibility of Bayesian Trees

Kazuharu Misawa, Masatoshi Nei

Institute of Molecular Evolutionary Genetics and Department of Biology, The Pennsylvania State University, 328 Mueller Laboratory, University Park, PA 16802, USA

Received: 19 December 2002 / Accepted: 5 January 2003

Abstract. Murphy and colleagues reported that the mammalian phylogeny was resolved by Bayesian phylogenetics. However, the DNA sequences they used had many alignment gaps and undetermined nucleotide sites. We therefore reanalyzed their data by minimizing unshared nucleotide sites and retaining as many species as possible (13 species). In constructing phylogenetic trees, we used the Bayesian, maximum likelihood (ML), maximum parsimony (MP), and neighbor-joining (NJ) methods with different substitution models. These trees were constructed by using both protein and DNA sequences. The results showed that the posterior probabilities for Bayesian trees were generally much higher than the bootstrap values for ML, MP, and NJ trees. Two different Bayesian topologies for the same set of species were sometimes supported by high posterior probabilities, implying that two different topologies can be judged to be correct by Bayesian phylogenetics. This suggests that the posterior probability in Bayesian analysis can be excessively high as an indication of statistical confidence and therefore Murphy et al.'s tree, which largely depends on Bayesian posterior probability, may not be correct.

Key words: Mammalian phylogeny — Bayesian method — Credibility value

Introduction

Murphy et al. (2001) reported that the phylogeny of 18 orders of placental mammals including 42 different species was resolved by Bayesian phylogenetics. In their study, they used 19 nuclear and 3 mitochondrial (mt) genes for constructing a Bayesian phylogenetic tree and obtained high posterior probabilities for almost all clades. However, this tree is different from Misawa and Janke's (2003) tree obtained by using 20 nuclear proteins for five placental mammalian orders. The later tree was supported with high statistical confidence in all analyses by Bayesian, maximum likelihood (ML), maximum parsimony (MP), and neighbor-joining (NJ) methods. The Murphy tree is also different from Arnason et al.'s (2002) tree based on 12 mt proteins for 60 different mammalian species.

Part of the differences in these inferred trees is apparently due to the fact that the Bayesian method on which Murphy et al. (2001) depended often gives overconfidence of a tree inferred (Suzuki et al. 2002). However, it is possible that the topological differences are caused by differences in the type of data used and the way the data were analyzed. For example, Murphy et al. (2001) used coding and noncoding nuclear DNA as well as mtDNA sequences, whereas Misawa and Janke (2003) and Arnason et al. (2002) used primarily protein sequences. Furthermore, Murphy et al.'s sequences had many alignment gaps or undetermined nucleotide sites. Therefore, although the total number of nucleotides in the entire concatenated sequences was 16,367 base pairs (bp), the

number of nucleotide sites shared by all the sequences was only 3170 bp.

For these reasons, we decided to reanalyze the Murphy data in several different ways. First, we chose 13 placental mammalian species of major interest such that they retain as many nucleotide sites and cover as many orders as possible. In this study we used only nuclear protein-coding genes. Second, we analyzed both DNA and protein sequences by Bayesian, ML, NJ, and MP methods. In these analyses we used both simple and complex substitution models except in MP analysis. Third, we used both sequences with and sequences without gaps. These phylogenetic analyses produced different trees for the same set of mammalian species. The results obtained are presented here.

Materials and Methods

Sequence Data

The entire set of DNA sequences used in Murphy et al.'s paper was provided by Dr. Mark Springer. We decided to use only 14 nuclear protein-coding genes in this study. We excluded mt genes, because they appear to evolve differently from nuclear genes (Nei and Glazko 2002). Nuclear noncoding genes were also eliminated, because they included many deletions/insertions and we wanted to use protein sequences as well as DNA sequences. The genes used in the present analysis were β -2 adrenergic receptor (ADRB2), adenosine A3 receptor (ADORA3), α -2B-adrenergic receptor (A2AB), ATPase, Cu^{2+} transporting, α polypeptide (ATP7A), endothelial differentiation, sphingolipid G-protein-coupled receptor 1 (EDG1), breast and ovarian cancer protein 1 (BRCA1), cannabinoid receptor 1 (CNR1), brain-derived neurotrophic factor (BDNF), zinc finger protein (ZFX), recombination activating protein 1 (RAG1), recombination activating protein 2 (RAG2), tyrosinase (TYR), von Willebrand factor (VWF), and retinoid-binding protein (IRBP). The aligned sequences of these genes are available at the Web site <http://mep.bio.psu.edu/databases/>.

Species

As mentioned above, we chose 13 species that nearly maximized the number of shared nucleotide sites. The species used were cat (*Felis catus*), dog (mainly *Canis familiaris* genes), cow (mainly *Bos taurus* genes), pig (*Sus scrofa*), horse (mainly *Equus caballus* genes), human (*Homo sapiens*), elephant (mainly *Loxodonta africana* genes), sea cow (*Trichechus manatus*), elephant shrew (*Macroscelides proboscideus*), armadillo (*Chaetophractus villosus*), rabbit (mainly *Oryctolagus cuniculus* genes), mouse (*Mus musculus*), and rat (*Rattus norvegicus*). Opossum (mainly *Didelphis virginiana* genes) was used as the outgroup.

Tree Reconstruction

Bayesian phylogenetic analysis was performed by using the program MrBayes 2.01 (Huelsenbeck and Ronquist 2001). In this analysis, the initial frequency was set to be 0.25 for each of the four nucleotides of DNA sequences and 0.05 for each of the 20 amino acids of protein sequences. We used the Jukes–Cantor (JC) model (default option) as the simple model and the general time reversible + invariable sites + gamma-rate variation (GTR + I + Γ) model

as the complex model for DNA sequences. The simple and complex models used for protein sequences were the Poisson and the Poisson + I + Γ models, respectively. We obtained Bayesian trees both by including gap and missing nucleotide sites and by excluding them. When gaps were included, they were treated as missing data.

In all of our analyses we employed one cold chain and three incrementally heated chains. The temperature parameter value used was 0.2. A random tree was used for starting each chain of trees. A uniform prior distribution was employed for different topologies. The four chains were run for 600,000 generations, and trees were sampled every 20 generations from the last 300,000 generations. A total of 15,000 sampled trees was used for inferring a Bayesian tree.

Maximum likelihood (ML) trees for DNA sequences were obtained by using the tree bisection and reconnection (TBR) search in PAUP* version 4.0 (Swofford 1998). We used the JC and the GTR + I + Γ models as the simple and the complex models for DNA sequences. The confidence levels of interior branches were evaluated by the bootstrap test with 100 replications. The protein likelihood tree was constructed by using the PROTML program (Adachi and Hasegawa 1996) with the Poisson model only. The among-site variation was not used, because this option was not available in PROTML. The confidence levels of interior branches were evaluated by a pseudo-bootstrap-resampling method with the RELL model, but they are called the bootstrap values here.

Maximum parsimony (MP) trees were also obtained using the TBR search in PAUP* version 4.0 (Swofford 1998). The search procedure for MP trees was essentially the same as that for ML trees. The number of bootstrap replications was again 100. We used only the standard unweighted parsimony method.

Neighbor-joining (NJ) trees (Saitou and Nei 1987) were obtained by using MEGA2 (Kumar et al. 2001). We used the JC and the Tamura–Nei + gamma (TN + Γ) models as the simple and the complex models of nucleotide substitution and the Poisson and the Poisson + gamma models for amino acid substitution (Nei and Kumar 2000, Chap 2). Here we did not use the GTR + I + Γ model for DNA sequences, because there is no analytical formula for obtaining distance estimates for this model. Instead, we used the TN + Γ model, which is the most parameter-rich model available for estimating sequence distances analytically.

Although it is possible to compute pairwise distances even between sequences with gaps using the pairwise deletion option (MEGA2), we did not use this option because the distance relationships are easily disturbed when the number of nucleotides varies extensively from sequence to sequence (Nei and Kumar 2000, p 49). We therefore constructed NJ trees using only sequences without gaps.

Results

Bayesian Trees

The Bayesian trees obtained are presented in Fig. 1 along with the ML trees. Trees A–D in this figure were obtained from DNA sequences, whereas trees E–H were obtained from protein sequences. Trees A, C, E, and G were obtained by using sequences without gaps, and the remaining trees were obtained by using sequences with gaps. The topology of the Bayesian tree for a given set of data and a given type of substitution model was the same as that of the ML tree except in tree E. However, the posterior probability for a given interior branch or a given clade of a Bayesian tree was almost always higher than or equal

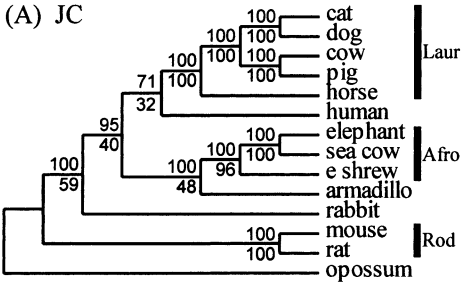
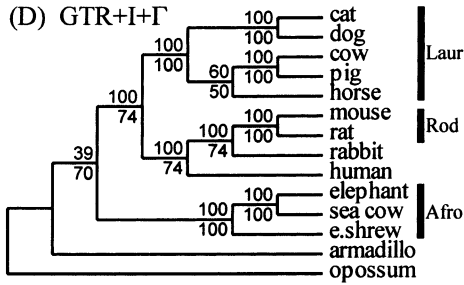
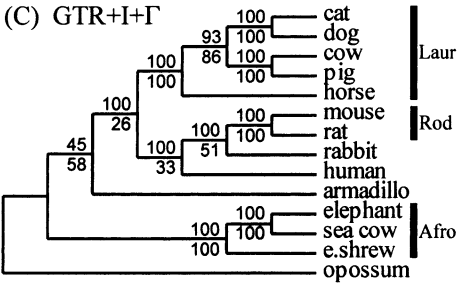
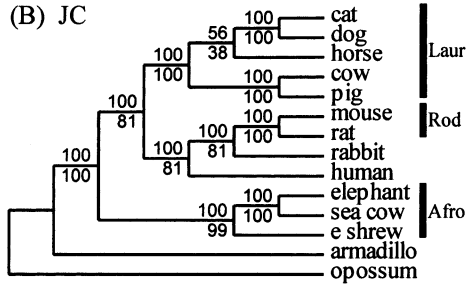
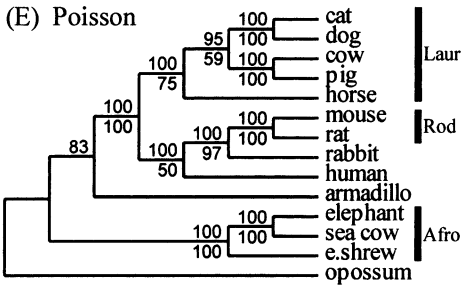
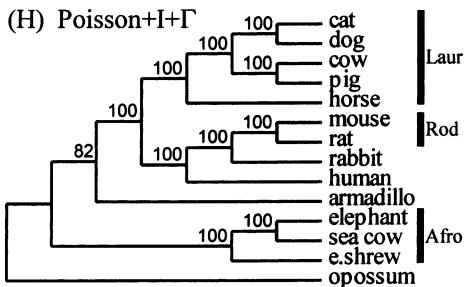
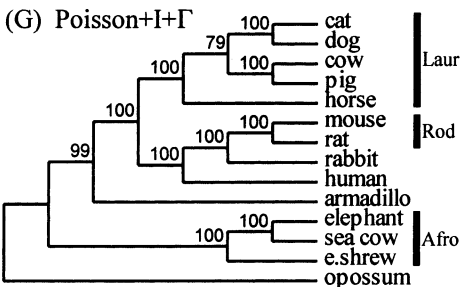
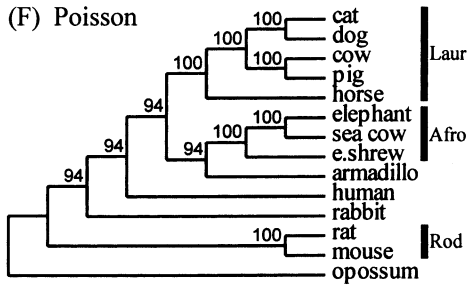
DNA without gaps (6387 bp)**DNA with gaps (13079 bp)****Protein without gaps (2090 aa)****Protein with gaps (4359 aa)**

Fig. 1. Bayesian and ML trees of 13 placental mammalian species. “DNA [or protein] with gaps” implies that DNA (or protein) sequences with alignment gaps and undetermined sites were used, whereas in the case of “DNA [or protein] without gaps” all these sites were eliminated from the analysis. The JC model was used as a simple model for DNA sequences evolution, and the GTR + I + Γ model was used as a complex model. Similarly, the Poisson model was used as a simple model for protein sequence evolution, and the Poisson + I + Γ model was used as a

complex model. The Bayesian and ML trees showed the same topology except in tree E, where armadillo clustered with Afrotheria. In trees F, G, and H, the ML tree was not constructed, because the computer program (PROTML) we used could not produce them. The percentage posterior probability for each interior branch is given above the branch, whereas the percentage bootstrap value is given below the branch. Bootstrap values are based on 100 replications. Laur, Laurasiatheria; Afro, Afrotheria, Rod, Rodentia.

to the bootstrap value for the ML tree. In Fig. 1, the posterior probability for an interior branch is given above the branch, whereas the bootstrap value is given below the branch. The bootstrap values for protein trees are not presented except for tree E, because none of the computer programs available provides the values for these protein trees.

The trees presented in Fig. 1 have some common features. All the trees include three clusters, which may be called Laurasiatheria (Laur), Afrotheria (Afro), and Rodentia (Rod) following Murphy et al. (2001). Here Laurasiatheria includes cat, dog, cow, pig, and horse, and Afrotheria includes elephant, sea cow, and elephant shrew. Rodentia is composed of only mouse and rat here. The phylogenetic relationships of Laurasiatheria, Afrotheria, and Rodentia in trees B, C, D, E, G, and H are the same but are different from those in trees A and F. Note that in tree A rodents are basal within placental mammals and this is supported by a posterior probability of 100%. The same branching pattern is supported in tree F as well, though the posterior probability for the division between rodents and other placental mammals is 94%. In other trees, rodents, rabbit, and human form a monophyletic cluster (Euarchontoglires [Murphy et al. 2001]), and this cluster is related to Laurasiatheria as a sister group. This splitting pattern is supported with a posterior probability of 100%. It is interesting to note that the topology of trees A and F is contradictory with that of other trees with respect to the position of rodents, yet both topologies are supported by high posterior probabilities. This indicates that the statistical confidence as judged by Bayesian posterior probabilities can be unduly high, because two topologies cannot be correct at the same time.

In Fig. 1, the phylogenetic position of armadillo varies considerably with tree. However, the position of this species is generally supported by low posterior probabilities, so that this is not a serious problem in phylogenetic analysis.

ML Trees

ML trees are presented in the case of DNA sequences (A, B, C, and D) and in tree E for protein sequences. As mentioned earlier, the topology of the ML tree is identical to that of the Bayesian tree when the same DNA data set and the same substitution model are used (trees A, B, C, and D). This is because essentially the same ML model of DNA evolution is used. However, the bootstrap value is often much lower than the Bayesian posterior probability. Therefore, according to ML bootstrap values, the branching pattern of rodents, rabbit, human, and armadillo is not well established in any of the four DNA trees. In protein tree E the ML tree was slightly different from

the Bayesian tree. That is, armadillo clustered with Afrotheria in this tree, though the bootstrap support for this cluster was not high (the tree not shown).

MP Trees

The four MP trees constructed with DNA and protein sequences are presented in Fig. 2. Trees A and C in this figure correspond to trees A and E in Fig. 1, because parsimony methods do not assume any complicated substitution models. Interestingly, the phylogenetic relationships of Laurasiatheria, Afrotheria, and Rodentia in tree A in Fig. 2 are the same as those in tree A in Fig. 1, but they are not the same as those of Murphy et al.'s tree. The phylogenetic positions of the remaining species are not the same for the two trees. However, the bootstrap values for the MP tree are generally as low as those for the ML tree in tree A of Fig. 1. Therefore, the two trees are not statistically different from each other. The relationships of the three major clusters of placental mammals in tree C in Fig. 2 are the same as those of tree E in Fig. 1 as well as those of the Murphy tree.

In trees B and D in Fig. 2, rodents are basal to all other placental mammals and therefore contradictory with the topology of the Murphy tree. However, the bootstrap values for critical interior branches are quite low.

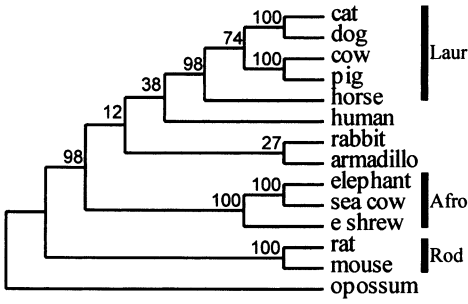
NJ Trees

There are four NJ trees based on DNA and protein sequences with the simple and the complex substitution models (Fig. 3). Interestingly, the phylogenetic relationships of the three major clusters of mammalian species are the same in all trees, and rodents are essentially basal to all other placental mammals. These trees are different from the Murphy tree, but the bootstrap values for critical interior branches are again quite low. Therefore, one cannot say that these topologies are contradictory with the Murphy tree.

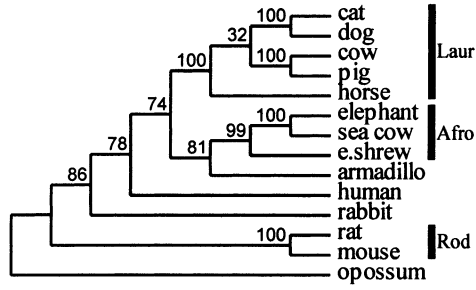
Discussion

We have seen that the posterior probability in Bayesian analysis is generally much higher than the bootstrap value in ML, MP, and NJ analyses and that two different topologies can be strongly supported by Bayesian probabilities. This is clearly illogical, and this conflict suggests that Bayesian posterior probabilities can give overconfidence of statistical support. This is consistent with the conclusion obtained by Suzuki et al.'s (2002) computer simulation. Therefore, we should be cautious about the use of Bayesian probabilities in phylogenetic analysis. The fact that we obtained various topologies

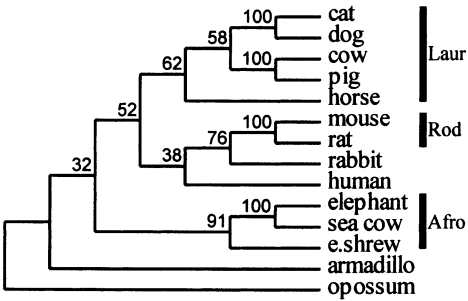
(A) DNA data without gaps (6387 bp)



(B) DNA data with gaps (13079 bp)



(C) Protein data without gaps (2090 aa)



(D) Protein data with gaps (4359 aa)

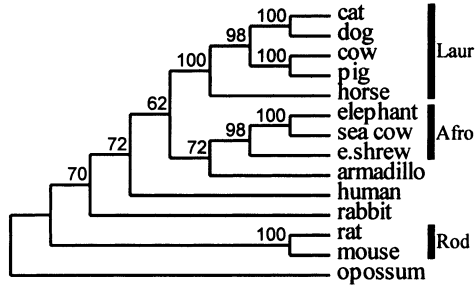
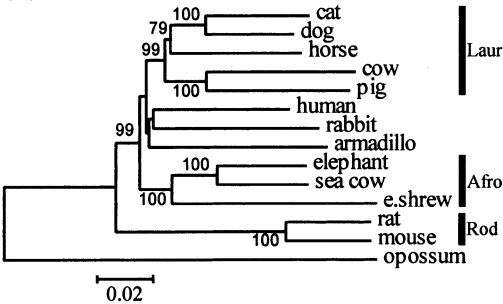


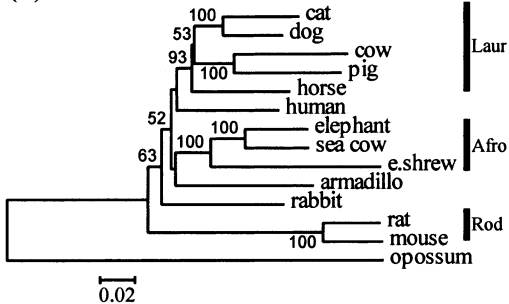
Fig. 2. MP trees. Only unweighted parsimony analysis was conducted. Bootstrap values are based on 100 replications.

DNA data without gaps (6387 bp)

(A) JC

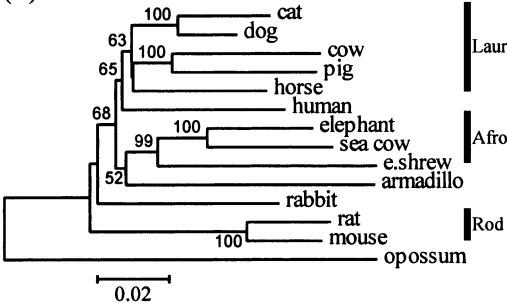


(B) TN+Γ



Protein data without gaps (2090 aa)

(C) Poisson



(D) Poisson+Γ

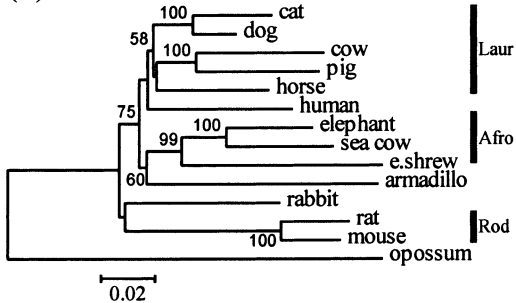


Fig. 3. NJ trees. The JC model was used as a simple model, whereas the Tamura–Nei + Γ model (Tamura and Nei 1993) was used as a complex model. The number of bootstrap replications was 1000. The bootstrap values below 50% are not presented.

may be caused partly by the small number of mammalian orders used here, but if this is the only reason, the posterior probabilities should not be so high in different trees.

In our data analysis DNA sequences and protein sequences often gave different tree topologies. Different substitution models for DNA and protein sequences also produced different trees. Furthermore, different tree-building methods generated different topologies. These results suggest that the sequence data used by Murphy et al. (2001) are not sufficient to resolve the mammalian phylogeny.

In Bayesian and ML methods it is customary to use very sophisticated models of nucleotide substitution with the understanding that the actual pattern of nucleotide substitution must be very complicated. Theoretically, this approach is justified if most nucleotides evolve following a given sophisticated substitution model and the number of nucleotides used is very large. In reality, this condition rarely holds, and in this case simpler models tend to give correct topologies more often than sophisticated ones (e.g., Tatenos et al. 1994; Nei 1996; Yang 1997; Takahashi and Nei 2000; Sullivan and Swofford 2001). The reason for this appears to be that the topology of a tree is not a usual parameter that is often used in standard statistics (Nei 1996; Nei and Kumar 2000, p. 83) and many parameters used in sophisticated models often introduce extra noise in the estimation process. In the present study we used both simple and complicated models, but the results obtained were nearly the same if we ignore interior branches with low ML bootstrap values. This suggests that one need not use too complicated models for constructing ML trees as shown by Takahashi and Nei (2000). A simple model also saves computational time tremendously compared with a complicated model.

As mentioned earlier, Murphy et al.'s (2001) tree is different from Misawa and Janke's (2003) tree, which was for five placental mammalian orders using chicken and *Xenopus* as the outgroups. The latter tree can be written ((((((Artiodactyla, Carnivora) Primata) Lagomorpha) Rodentia) Chicken) *Xenopus*), whereas the Murphy tree for the five mammalian orders with Marsupialia as the outgroup can be written (((Artiodactyla, Carnivora) ((Rodentia, Lagomorpha) Primata)) Marsupialia). Therefore, the topologies of the two trees are considerably different. Yet the Misawa–Janke tree was highly supported by the Bayesian posterior probabilities as well as by the bootstrap values for the ML, MP, and NJ trees. Furthermore, the mt tree for 60 mammalian species obtained by Arnason et al. (2002) is consistent with the Misawa–Janke tree rather than with the Murphy tree.

Nevertheless, we cannot say that the Misawa–Janke tree is more reliable than the Murphy tree at this stage, because the former tree included only 5 placental

Whale	ATCCAGAA-----AAGGAATG
Dolphin	ATCCAGAA-----AAGGAATG
Hippo	ATCCAGGAAATCCAGGAAAGGAACG
Pig	ATCCAGCAAAATCCAGAAAAGGAATG
Cow	ATCCAGGAAATCCAGAAAAGGAATG
Llama	ATCCAGGAGATTCAGAAAAGGAATG
Rhino	ATTCAGGAAATCCAGAAAAGGATTG
Tapir	ATTCAGGAAATCCAGAAAGAG---TG
Horse	ATCCAGGAAATCCAGAAAAGGAATG
Cat	ATCCAGGAAATCCAGAAAAGGAATG
Dog	ATCCAAGAGATCCGGAAAGAGGACTG
Pangolin	ATATGGGAAATCCAGAAAAGGAATG
Flying fox	ATCTAGAA-----AAGGAATG
Fruit bat	ATCCAGGAAATCCAGAAAAGGAATG
Leaf nosed bat	ATCCAGGAAATCCAGAAAAGGAATG
Round-ear bat	ATCCAGG--AATCCAGGAAAGGAATG
Free tailed bat	ATCCAGAA-----AAGGAATA
Hedgehog	GTCCAGAGAATTCAGAAAAGGCATG
Shrew	ATCCAGGAAATCCAGAAAAGGAATG
Mole	ATCCGGGAAATTCAGAAAAGGAATA
Mouse	AACCTAGAAGTCCCCAAAAGGACTG
Rat	AACTTAGAAGTCCCCAAAAGGACTG
Porcupine	ATCCAGGAAATCCAGAAAAGGAATG
Guinea pig	ATCCAGGAAATCCAGAAAAGGAATG
Squirrel	ATCCAGGAAATTCAGAAAAGGAATG
Rabbit	AACCAGGAGATTCAGAAAAGGAATG
Flying lemur	ATCCAGGAAATCCAGAAAAGGAGTG
Tree shrew	ATCCAGGAAATCCAGAAAAGGAA--
Lemur	ACCTAGGA-----AAGGAATG
Human	ATCCAGGAAATGCAGAAGAGGAATG
Sloth	ATCCAGAA-----AAGGAATG
Anteater	ATTCAGAA-----AAGGAATG
Armadillo	ATCCAGAA-----AATGAATG
Tenrec	ATCCAGAA-----AAGGAGTG
Golden mole	ATCCAGAA-----AAGGAATG
E. shrew	ATCCAGAA-----AGGGACTA
Aardvark	ATCCAGAA-----AAGGAATG
Sea cow	ATCCAGAA-----AAGGAATG
Elephant	ATCCAGAA-----AAGGAATG
Hyrax	ATCCGGAA-----AAGGAGTG
Wombat	GCCCAAGG-----CAGGAATG

Fig. 4. Nine-base nucleotide deletion (or insertion) identified in the *BRCA1* gene used in Murphy et al.'s study. This deletion occurred in nucleotide positions 1805 to 1824. This gene was not sequenced in one of the two elephant shrew species (*Macroscelides proboscideus*).

mammalian orders, whereas Murphy et al. considered 18 orders. Interestingly, Poux et al. (2002) discovered two sets of amino acid deletions that are shared by primates, lagomorphs, and rodents. If these deletions represent irreversible shared derived characters (Nei and Kumar 2000, p. 140), they support the Murphy tree rather than the Misawa–Janke tree. However, it is still possible that these shared deletions have occurred independently, because a deletion of the same set of nucleotides is known to occur independently in different evolutionary lineages. For example, Soodyall et al. (1996) showed that the Co2/tRNA (Lys) 9-bp deletion in human mitochondrial DNA occurred independently in South Pacific Asian populations and sub-Saharan African populations. Furthermore, the two deletions shared by primates, lagomorphs, and rodents may not necessarily indicate a true evolutionary clade, if the ancestral populations of these

orders were large and polymorphic with respect to the deletion proteins (alleles) and allelic sorting did not coincide with speciation history (Nei and Kumar 2000, p. 143). Note also that even if they did, they support a monophyly of only three orders. They do not give any information about the evolutionary relationships of other orders.

To find any other insertions/deletions (indels) that might support the Misawa–Janke or the Murphy tree, we examined the DNA sequences used in Murphy et al.'s study. We found two indels that are shared by many species in the gene *BRCA1*. One was the 9-base deletion (positions 905–913) shared by species belonging to Cetartiodactyla, and the other was the 9-base indel (positions 1806–1824) presented in Fig. 4. The distribution of this indel among different species supports neither the Murphy tree nor the Misawa–Janke tree. Actually, it does not support any other currently plausible phylogenetic relationships of mammalian orders. Therefore, if there were no sequencing errors, this suggests that the indel occurred in several lineages independently. This warns us about uncritical use of indel data in phylogenetic analysis.

In general, however, indels have strong phylogenetic signals (Venkatesh et al. 2002), and if we find many other indels defining other clades, we may be able to clarify the phylogeny of mammalian orders more precisely. Particularly, if more complete genomic sequences become available for many different orders of mammals and large DNA block duplications which are likely to be unique (e.g., a 2-megabase block duplication of the human immunoglobulin κ gene region [Schable and Zachau 1993]) or unique genes confined to a group of related species (e.g., dimeric immunoglobulin genes confined to Camelidae [Su et al. 2002]) are identified, we will be able to study phylogenetic trees more efficiently.

Acknowledgments. We thank Axel Janke and Yoshiyuki Suzuki's comments on the first draft of this paper. This study was supported by a research grant from the National Institutes of Health to M.N. (GM20293).

References

- Adachi J, Hasegawa M (1996) MOLPHY: Programs for molecular phylogenetics, version 23. Institute of Statistical Mathematics, Tokyo
- Arnason U, Adegoke JA, Bodin K, Born EW, Esa YB, Gullberg A, Nilsson M, Short RV, Xu X, Janke A (2002) Mammalian mitogenomic relationships and the root of the eutherian tree. *Proc Natl Acad Sci USA* 99:8151–8156
- Huelsenbeck JP, Ronquist F (2001) MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* 17:754–755
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: Molecular evolutionary genetics analysis software. *Bioinformatics* 17:1244–1245
- Misawa K, Janke A (2003) Revisiting the Glires concept—phylogenetic analysis of nuclear sequences. *Mol Phylogenet Evol* 28(2):320–327
- Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady CJ, Teeling E, Ryder OA, Stanhope MJ, de Jong WW, Springer MS (2001) Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science* 294:2348–2351
- Nei M (1996) Phylogenetic analysis in molecular evolutionary genetics. *Annu Rev Genet* 30:371–403
- Nei M, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, New York
- Nei M, Glazko GV (2002) Estimation of divergence times for a few mammalian and several primate species. *J Hered* 93:157–164
- Poux C, van Rheede T, Madsen O, de Jong WW (2002) Sequence gaps join mice and men: Phylogenetic evidence from deletions in two proteins. *Mol Biol Evol* 19:2035–2037
- Saitou N, Nei M (1987) The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Schable KF, Zachau HG (1993) The variable genes of the human immunoglobulin kappa locus. *Biol Chem Hoppe Seyler* 374:1001–1022
- Soodyall H, Vigilant L, Hill AV, Stoneking M, Jenkins T (1996) mtDNA control-region sequence variation suggests multiple independent origins of an "Asian-specific" 9-bp deletion in sub-Saharan Africans. *Am J Hum Genet* 58:595–608
- Su C, Nguyen VK, Nei M (2002) Adaptive evolution of variable region genes encoding an unusual type of immunoglobulin in camelids. *Mol Biol Evol* 19:205–215
- Sullivan J, Swofford DL (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
- Suzuki Y, Glazko GV, Nei M (2002) Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. *Proc Natl Acad Sci USA* 99:16138–16143
- Swofford DL (1998) PAUP*, phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer, Sunderland, MA
- Takahashi K, Nei M (2000) Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. *Mol Biol Evol* 17:1251–1258
- Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* 10:512–526
- Tateno Y, Takezaki N, Nei M (1994) Relative efficiencies of the maximum-likelihood, neighbor-joining, and maximum-parsimony methods when substitution rate varies with site. *Mol Biol Evol* 11:261–277
- Yang Z (1997) How often do wrong models produce better phylogenies? *Mol Biol Evol* 14:105–108