Rewriting Evolution—"Been There, Done That"

David Penny*

Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand

*Corresponding author: E-mail: d.penny@massey.ac.nz.

Accepted: March 25, 2013

Abstract

A recent paper by a science journalist in *Nature* shows major errors in understanding phylogenies, in this case of placental mammals. The underlying unrooted tree is probably correct, but the placement of the root just reflects a well-known error from the acceleration in the rate of evolution among some myomorph rodents.

Key words: placental mammals, phylogeny errors, rates of evolution.

There is a rather famous local T-shirt from the 1970s that was a take-off of those who had bought theirs at exotic locations around the world—it was simply "Been there, done that." However, the message applies very well to claims by a science writer in a three-page *Nature* article last year that untested methods were supposedly "rewriting" the tree of placental mammal relationships (Dolgin 2012). The supposedly "new" phylogeny placed the root of the placental tree within the rodents—contradicting both established molecular and fossil phylogenies! However, the article as a whole is a major attack on the whole field of molecular phylogenetics. Several problems that have been solved in the past are made afresh in that report, and it is instructive/helpful to analyze these difficulties to help avoid similar problems in the future. Unfortunately, the author makes at least four classical mistakes about evolutionary trees: not separating out the problem of rooting the tree (Penny 1976); ignoring the mathematical studies about problems when there are differences in the mutational process, such as in mutation rates (e.g., Hendy and Penny 1989); not specifying the model/mechanism of evolution; and therefore not being able to determine an appropriate optimality criterion (Steel and Penny 2000), and thus cannot evaluate the time scale over which their characters are relevant for phylogeny (Mossel and Steel 2004).

Overall, the author is certainly in no position to challenge well-established basic phylogenies of placental mammals (e.g., Meredith et al. 2011; dos Reis et al. 2012)—even though there are still some well-known uncertainties about a few aspects of the placental tree (such as the precise point of rooting the tree) (Meredith et al. 2011; McCormack et al. 2012), see later. We tried to take the issue up with *Nature* itself, but because of the length of the response, they did not wish any discussion of these fundamental issues. However, it is almost certainly an issue that all journals should have a procedure for handling, at least in a general way, questions that the journals themselves have raised, even if they are by science reporters.

First, there is no outgroup for the claimed rooted tree (Figure 1 of Dolgin [2012] is shown unrooted as fig. 1*A*). Their unrooted tree is standard, but their rooted tree is not (their tree is shown in fig. 1*C*). The current accepted tree (fig. 1*B*) is also consistent with the unrooted tree (fig. 1*A*), so the apparent controversy is not about the unrooted tree; it is about the position of the root of the tree—point 1. This is a very important point, and it means that there is really no problem with the type of data used, it is simply that the position of the root is critical to interpreting the tree. It is unclear how Dolgin (2012) roots the tree, but it appears to be midpoint rooting (see Penny et al. [1995], which is known to be sensitive to rate differences, see the next point.

Second, we have seen the tree in figure 1*C* before (D'Erchia et al. 1996; see also Lin, Waddell, et al. 2002), it was long ago falsified because it is quite well known that there have been significant acceleration in the mutational processes in at least some myomorph rodents. Compared with humans, they have more than a 3-fold increase in mutation rate per cell division (Drake 1999). This rate difference initially led to some difficulties reconciling nuclear and mitochondrial trees (e.g., Lin et al. 2002). Improving the number and distribution of eutherian samples meant, there was the expected basic consilience between nuclear and mitochondrial sequences and that has been maintained as more and more sequences became

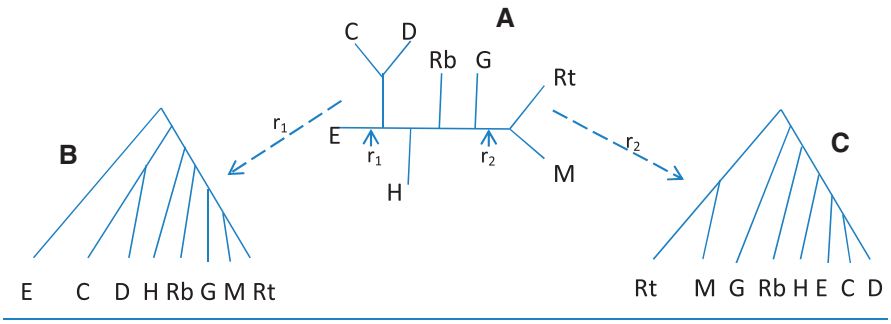

Fig. 1.—The mammalian tree in question. The same unrooted tree (*A*) can be rooted at the arrow "r_1" to give the standard tree for placental mammals (*B*) or at position "r_2" to give the first divergence within rodents (as in *C*). It is therefore essential to specify the position of the root, and this is usually done with an "outgroup"—for example, the marsupials for the placental mammals. C, cat; D, dog; E, elephant; G, guinea pig; H, human; M, mouse; Rb, rabbit; Rt, rat.

available. A similar problem has been found, and solved, in bird evolution—Some early trees gave the root of the avian tree in the faster evolving passerine song birds (see Harrison et al. 2004). Break up the long branches in either the placental or the avian case, and the traditional tree (for placentals) or the paleognath/neognath tree (for birds) reappears. So point 2 is that there are some well-known problems with establishing the root of the tree when there are differences in either mutation rates or nucleotide composition (Lockhart et al. 1992). In a statistical sense, these are "systematic" errors, rather than "sampling" errors.

Although the two issues given above are sufficient to account for the unusual tree in Dolgin (2012), there are other problems. The third point is that it is essential to understand as much as possible about the processes of evolution for the data used. It is certainly not sufficient to claim that a particular analysis is "simpler" (as in Dolgin 2012) because it is known, for example, that some "simple" optimality criteria (such as parsimony) might be the maximum-likelihood estimator for a "no common mechanism" model. However, in principle, this allows a very large increase in the number of parameters and is subject to the long-branch attraction effect mentioned earlier (Hendy and Penny 1989). So under many circumstances, it cannot guarantee to produce the correct tree—even with an unlimited number of characters. So in this sense, it is not "statistically consistent" (see Felsenstein 2003). Thus, an apparently "simple" model does not guarantee consistency of the method. In practice, it is necessary to go well beyond whether a character is "present/absent" and formally evaluate the time scale over which a character will be informative for a given mutation rate (Mossel and Steel 2004). These latter authors show an important limit for Markov models, and their mathematical results certainly help explain the limitation at deeper times. These are the third and fourth points.

We must always be ready to revise our scientific knowledge, and we should welcome new tests of any aspect of science. There is a very interesting aspect of evolutionary theory that it might be difficult to finalize the last details of the placental tree (see Song et al. 2012). Although there are about 3×10^{20} rooted trees for 19 orders of mammals $((2n-3)!!$, where n is the number of taxa, and the !! notation is multiplying by every second number—$1 \times 3 \times 5 \ldots$). We are about down to no more than 10^1 or 10^2 trees left to consider. It has been known for about 30 years that although individual gene trees are highly similar, they are not usually identical. This happens because of coalescence/lineage sorting (Pamilo and Nei 1988), hybridization, differential selection on sequences (at either the primary sequence or tertiary structural level), ambiguous homologies (including from polyploidy and/or copy number variation), gene conversion, lateral gene transfer, etc. So we still have more work to do!

Perhaps the bad news is that unfortunately evolutionary relationships cannot be done seriously by amateurs? However, the good news is that there is a well-studied mathematical basis for tree building (e.g., Felsenstein 2003; Semple and Steel 2003). Certainly, the small nonprotein coding RNAs (ncRNAs) are an important addition to our knowledge of gene regulation and help inform us about important dynamic aspects of evolution (Hoeppner et al. 2012). However, there is no evidence, yet, that simple "presence/absence" of ncRNAs evolves at a time scale that makes them useful for deeper phylogeny—that has to be established separately. There has been some careful work that showed, for example, the occurrence of a vault RNA throughout vertebrates, but that study used gene location data (synteny) to help show homology (Stadler et al. 2009). We certainly require full testing/evaluation of new and/or untried methods of analysis. However, the present attack on evolutionary study is just full of classical mistakes. We must always determine the time scale over which characters are relevant, it is essential to solve the mathematical problems, and it is also necessary to determine the correct placement of the root of the tree. Only then can we start to move beyond what we know already. As far as we can

tell, the current difficulty is not any property of microRNAs, but the mutational processes in myomorph rodents. The unrooted tree of Dolgin (2012) is standard; it is just the position of the root of the tree that is the problem. Although we should all welcome the testing of our current knowledge by new data and new approaches, there are also known standards that must be met.

Literature Cited

D'Erchia AM, Gissi C, Pesole G, Saccone C, Arnason U. 1996. The guinea-pig is not a rodent. Nature 381:597–600.
Dolgin E. 2012. Rewriting evolution. Nature 486:460–462.
dos Reis M, et al. 2012. Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. Proc Biol Sci. 279:3491–3500.
Drake JW. 1999. The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes. Ann N Y Acad Sci. 870:100–107.
Felsenstein J. 2003. Inferring phylogenies. Sunderland (MA): Sinauer.
Harrison GL, et al. 2004. Four new avian mitochondrial genomes help get to basic evolutionary questions in the late cretaceous. Mol Biol Evol. 21:974–983.
Hendy MD, Penny D. 1989. A framework for the quantitative study of evolutionary trees. Syst Zool. 38:297–309.
Hoeppner MP, Gardner PP, Poole AM. 2012. Comparative analysis of RNA families reveals distinct repertoires for each domain of life. PLoS Comput Biol. 8(11):e1002752.
Lin Y-H, et al. 2002. 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 PJ, Penny D. 2002. Pika and vole mitochondrial genomes add support to both rodent monophyly and glires. Gene 294:119–129.
Lockhart PJ, et al. 1992. Controversy on chloroplast origins. FEBS Lett. 301: 127–131.
McCormack JE, et al. 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. Genome Res. 22:746–754.
Meredith RW, et al. 2011. Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. Science 334:521–524.
Mossel E, Steel M. 2004. A phase transition for a random cluster model on phylogenetic trees. Math Biosci. 187:189–203.
Pamilo P, Nei M. 1988. Relationships between gene trees and species trees. Mol Biol Evol. 5:568–583.
Penny D. 1976. Criteria for optimising phylogenetic trees and the problem of finding the root of a tree. J Mol Evol. 8:95–116.
Penny D, et al. 1995. Improved analyses of human mtDNA sequences support a recent African origin for *Homo sapiens*. Mol Biol Evol. 12: 863–882.
Semple C, Steel M. 2003. Phylogenetics. Oxford: Oxford University Press.
Song S, Liu L, Edwards SV, Wu S. 2012. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. Proc Natl Acad Sci U S A. 109: 14942–14947.
Stadler PF, et al. 2009. Evolution of vault RNAs. Mol Biol Evol. 26: 1975–1991.
Steel MA, Penny D. 2000. Parsimony, likelihood and the role of models in molecular phylogenetics. Mol Biol Evol. 17:839–850.

Associate editor: Bill Martin