Heredity (2005) 94, 461–462

www.nature.com/hdy

NEWS AND COMMENTARY

Molecular clocks

Closing the gap between rocks and clocks

K Cranston and B Rannala

Heredity (2005) **94,** 461–462. doi:10.1038/sj.hdy.6800644
Published online 16 February 2005

A new study provides an advance in evolutionary research through reconciling data from the fossil record and the molecular clock. Estimating species divergence times from molecular sequence data via phylogenetic trees is possible with the molecular clock, which allows the separation of rate and time by assuming a constant rate of molecular evolution. Unfortunately, species divergence times estimated using the molecular clock typically appear much more ancient than dates based on the fossil record. The reason for this discordance has been widely debated (Benton and Ayala, 2003; Brochu *et al*, 2004), and one explanation is the statistical bias in molecular-based estimates of ages. For example, ignoring among-lineage rate variation can cause an upward bias in age estimates (Aris-Brosou and Yang, 2003). Recent advances in phylogenetic theory have allowed for rate variation, but age estimates obtained using these new methods continue to disagree with paleontological estimates. A new study by Douzery *et al*, 2004 applies a Bayesian relaxed clock method to a large eukaryotic data set and obtains much better agreement between molecular dates and the fossil record.

Predicting the timing of evolutionary events from fossil, morphological and molecular data is a challenging estimation problem. It starts with a phylogenetic tree, where branch lengths measure the amount of evolution between species. This measure confounds rate and time, meaning that we can interpret a long branch as either a long period of time or a high rate of evolution. Assuming a constant rate of evolution (the molecular clock hypothesis) allows separation of rate from time, and estimation of the elapsed time between divergences. Most data sets, however, do not demonstrate rate constancy, over time or among lineages. In these cases, relaxed molecular clock methods allow different branches on the tree to have different evolutionary rates. One method models the rate of evolution as autocorrelated between branches such that the rate after a speciation event depends on the rate in the common ancestor (Thorne *et al*, 1998). By combining such a model of rate evolution with calibration points from the fossil record, we can estimate divergence times without assuming a molecular clock.

Even with relaxed clock methods, divergence times estimated from molecular data are often far more ancient than those predicted from the paleontological record (Hedges and Kumar, 2003). This discrepancy can be uncomfortably large, sometimes hundreds of millions of years. The Douzery *et al* study reduces the gap between molecular and fossil dates using three strategies: (i) increasing the size of the data set, both in width (number of genes) and in depth (number of taxa); (ii) using a large number of fossil calibration points; and (iii) incorporating uncertainty in both evolutionary rates and fossil calibration points.

Divergence time estimates depend greatly on the accuracy of the phylogeny, which can be best inferred with large amounts of data. Previous analyses used a small number of taxa and estimated times based on a limited number of genes. Phylogenetic trees and species divergence times inferred for different genes may often be incongruent due to factors such as lineage sorting and errors of phylogenetic inference. The Douzery *et al* data set is large, including 129 proteins in 39 eukaryotes with over 30 000 positions. Divergence times are defined at the species level, and inclusion of a large number of genes increases the likelihood that the inferred tree approaches the true species tree.

Combining genes in an analysis can create new difficulties. It is well known, for example, that substitution rates vary greatly across genes and that accounting for this variation is important for the accuracy of phylogenetic analyses. A weakness of the Douzery *et al* methodology is that they endeavor to accommodate among-gene rate variation by applying a single common gamma distribution to model rate variation across all sites in a composite sequence of concatenated genes. However, adjacent sites within the same gene will have rates that are more similar than predicted under this model. An alternative approach would be to estimate an average rate for each gene and assume that rates are drawn from a common distribution with a mean rate that is shared across sites within each specific gene.

Increasing the number of species potentially improves our ability to accurately reconstruct the phylogeny and also allows for a greater number of fossil calibration points. The accuracy of divergence times tends to be greater for speciations closer to calibration points, so a denser distribution of these points will improve the overall accuracy of the analysis. In addition, the use of multiple calibrations highlights inconsistencies between molecular and fossil dates, and between different fossil dates.

Quality of data is as important as quantity. Measuring divergence times requires accurate estimates of evolutionary rates and fossil calibration points. Neither of these quantities is known without error, and such uncertainty must be incorporated into the analysis. The Bayesian relaxed clock method (Thorne *et al*, 1998) used by Douzery *et al* allows fossil calibrations to be defined as age ranges, rather than single dates. Then, when estimating the rates, this method incorporates variable rates among branches by integrating over a range of possible rates rather than inferring a fixed value for each branch. This will tend to produce age estimates that are more conservative (eg, bracketed by larger confidence intervals).

Limitations of the divergence time method cause the authors to ignore phylogenetic uncertainty. In this study, a single phylogenetic tree is input as the true tree for the analysis. For a large data set, a single tree cannot adequately represent the true evolutionary history of the species under investigation. Future studies will most likely average over the posterior distribution of compatible topologies. An ideal method would simultaneously infer the topology, evolutionary rates and divergence times, given the molecular data and fossil calibration ranges as input.

Estimates of divergence times should continue to improve as more sequences, especially whole genomes, are collected. Without accurate estimates of biological timepoints, it is impossible to

conduct evolutionary studies that make inferences about the effects of environmental change on biodiversity, or study the differences in microevolutionary processes between species. The Douzery *et al* study is an important first step toward refining the accuracy of divergence times based on composite fossil and molecular data sets. The study also provides an important stimulus for theoretical evolutionary biologists, highlighting the limitations of existing models and suggesting directions for future research. The divergence time estimation problem is an interesting combination of fossil, morphological and molecular data. To proceed, paleontologists, morphologists and molecular evolutionists must individually and collectively refine their methods in the hope that they will ultimately reach a consensus about the timing of the origins of major species groups.

K Cranston is at the Department of Medical Genetics, 539 Medical Sciences Building, University of Alberta, Edmonton, Alberta, Canada T6G 2H7. B Rannala is at the Genome Center and Section of Evolution and Ecology, University of California Davis, One Shields Avenue, Davis, CA 95616, USA.

e-mail: bhrannala@ucdavis.edu

Aris-Brosou S, Yang Z (2003). *Mol Biol Evol* **20**: 1947.
Benton MJ, Ayala FJ (2003). *Science* **300**: 1698.
Brochu CA, Sumrall CD, Theodor JM (2004). *J Paleontol* **78**: 1.
Douzery EJ, Snell EA, Bapteste E, Delsuc F, Philippe H (2004). *Proc Natl Acad Sci USA* **101**: 15386.
Hedges SB, Kumar S (2003). *Trends Genet* **19**: 200.
Thorne JL, Kishino H, Painter IS (1998). *Mol Biol Evol* **15**: 1647.