EXTENDED PDF FORMAT
SPONSORED BY

BIO-RAD

http://discover.bio-rad.com



Dating the Tree of Life
Michael J. Benton, *et al.*
Science **300**, 1698 (2003);
DOI: 10.1126/science.1077795

The following resources related to this article are available online at www.sciencemag.org (this information is current as of December 18, 2006):

Updated information and services, including high-resolution figures, can be found in the online version of this article at:
http://www.sciencemag.org/cgi/content/full/300/5626/1698

Supporting Online Material can be found at:
http://www.sciencemag.org/cgi/content/full/300/5626/1698/DC1

A list of selected additional articles on the Science Web sites **related to this article** can be found at:
http://www.sciencemag.org/cgi/content/full/300/5626/1698#related-content

This article **cites 37 articles**, 20 of which can be accessed for free:
http://www.sciencemag.org/cgi/content/full/300/5626/1698#otherarticles

This article has been **cited by** 60 article(s) on the ISI Web of Science.

This article has been **cited by** 18 articles hosted by HighWire Press; see:
http://www.sciencemag.org/cgi/content/full/300/5626/1698#otherarticles

This article appears in the following **subject collections**:
Evolution
http://www.sciencemag.org/cgi/collection/evolution

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:
http://www.sciencemag.org/help/about/permissions.dtl

Downloaded from www.sciencemag.org on December 18, 2006

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published weekly, except the last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. Copyright c 2005 by the American Association for the Advancement of Science; all rights reserved. The title SCIENCE is a registered trademark of AAAS.

Dating the Tree of Life

Michael J. Benton[1]* and Francisco J. Ayala[2]

The relative merits of molecular and paleontological dates of major branching points in the tree of life are currently debated. In some cases, molecular date estimates are up to twice as old as paleontological dates. However, although it is true that paleontological dates are often too young (missing fossils), molecular dates are often too old (statistical bias). Intense study of the dating of major splits in the tree of mammals has shown rapprochement as fossil dates become older and molecular dates become younger.

The reconstruction of segments of the "tree of life" has long been a driving force for systematists. Since the mid-1980s, there has been an exponential growth in the number of phylogenetic papers published each year (*1*). The tree of life project, whose end point is the construction of the single phylogenetic tree linking all species living and extinct, promises to be a substantial, international research program involving thousands of biologists. The scientific aim is the same as that set out by Darwin (*2*): to understand where life came from, the shape of evolution, and the place of humans in nature and to determine the extent of modern biodiversity and where it is threatened (*3*, *4*).

A key concern in this project is the calibration of phylogenies against time. This surfaced in the 1960s with the first attempts to estimate divergence dates in a phylogenetic tree from molecular evidence. Since then, the value of the molecular and morphological or paleontological approaches (*5*–*7*) has been recognized. However, some commentators indicate that, in cases of dispute, molecular dates should generally (*8*–*10*) or always (*11*, *12*) be preferred. We suggest that there are no such simple solutions. First, more morphological or paleontological and molecular trees and dates agree than disagree. Second, although paleontological dates by definition are always underestimates (providing specimens are correctly identified), it may be that molecular dates are always overestimates. Third, close study and care of calibration points can lead to rapprochement, where apparently disputed ages eventually converge on an agreed date.

Early Origins of Major Clades

In some noted cases, the molecular age estimates for origins of groups are about twice as old as the oldest fossils. The range of molecular estimates for the origin of metazoans is 600 to 1500 million years ago (Ma) (*9*, *13*–*16*), with many recent estimates narrowing it down to 700 to 1000 Ma (*15*, *17*–*19*). There is fossil evidence of Precambrian metazoans but nothing before about 600 Ma. The new molecular consensus, however, is that basal splits among major animal clades happened about 1000 Ma and that the modern phyla, such as molluscs, arthropods, brachiopods, and echinoderms, diverged about 600 to 800 Ma (Table 1). There are three reasonable explanations for these discrepancies (*20*): (i) the molecular and paleontological dates may mark different events (*16*, *21*), for example, the genetic divergence of lineages (molecular date) and the acquisition of hard skeletons (paleontological date); (ii) the fossil dates could be too young (*8*, *9*, *13*) as a result of an absence of fossils from much of the Precambrian, because either they lacked skeletons, they were microscopic, they did not become incorporated into the rocks, or they have been missed by paleontologists; or (iii) the molecular dates could be too old as a result of unaccounted-for variations in the rates of molecular evolution, incorrect calibration points, or inadequate correction for other biases (*14*, *22*, *23*).

The first vascular land plants are found as fossils in the Silurian, and earlier evidence from possible vascular plant spores may extend the range back to the Ordovician, 475 Ma (*24*), considerably younger than a molecular estimate of 700 Ma (*25*). A similar gap exists for angiosperms, with the oldest generally accepted fossils being from the Early Cretaceous (120 to 130 Ma) (*24*). DNA sequence evidence places the divergence of angiosperms in the Mid Jurassic, 140 to 190 Ma (*26*), but the date could be much older, Carboniferous (290 Ma) (*27*), if it turns out that the sister group of angiosperms is the gymnosperms.

For modern birds, molecular estimates place the split of basal clades and modern orders at 70 to 120 Ma (*12*, *28*, *29*). Although many supposed Cretaceous representatives of modern bird orders have been cited (*28*, *29*), most have been disputed, generally because the fossils are isolated elements (*30*). The oldest uncontroversial fossils of modern bird orders date from the Paleocene (60 Ma), much younger than most molecular estimates of origins.

The dating of the radiation of modern placental (eutherian) mammals also seemed to be an example of unusually early molecular dates. The paleontological view (*31*) is that placentals split from marsupials some time in the Early Cretaceous (144 Ma). The first molecular dates (*8*, *12*, *32*) seemed much older: origin of eutherians in the Late Jurassic

Table 1. Estimated ages (Ma) of branching points in lower parts of the tree of Metazoa (animals). Metazoa are all animals, including sponges (Porifera). Bilateria are essentially all metazoans except sponges and coelenterates (for example, corals, sea anemones, and hydroids). Deuterostomia are echinoderms, chordates (backboned animals and relatives), and some smaller groups. The data sources [G, "gene" (DNA or RNA); and P, protein or amino acid] are noted, and numbers of nuclear genes, or use of MtDNA and rRNA (ribosomal RNA) or enzymes (E) are in parentheses. Ref., reference number; est., estimated; and min., minimum. Blank entries indicate unavailable data.

Data source	Metazoa (Chordata-Porifera)	Bilateria (Chordata-Arthropoda)	Deuterostomia (Chordata-Echinodermata)	Ref.
G (8 G)		1200 ± 100	1001 ± 100	(*13*)
P (64 E)	930 ± 115	790 ± 60	590	(*17*)
G (4 G)	940 ± 80	700 ± 80		(*15*)
G (18 G)		670 ± 60	600 ± 60	(*14*)
G (22 G)		830 ± 55		(*18*)
G (50 G)	1350 ± 150 (est.)	993 ± 46		(*57*)
G (22 G)		659 ± 131		(*22*)
P (10 E)		627 ± 51		(*19*)
G (MtDNA; 18S rRNA)		588 min.	586/589 min.	(*20*)

[1]Department of Earth Sciences, University of Bristol, Bristol BS8 1RJ, UK. [2]Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697–2525, USA.

*To whom correspondence should be addressed. E-mail: Mike.Benton@bris.ac.uk

Downloaded from www.sciencemag.org on December 18, 2006

(150 to 170 Ma), split of major placental groups in the Early Cretaceous (100 to 130 Ma), and split of modern placental orders in the mid- to Late Cretaceous (80 to 100 Ma). The oldest fossil representatives of modern mammalian orders dated from the Paleocene and Eocene (50 to 65 Ma).

A survey of recent literature suggests that such examples are not typical and that most paleontological and molecular dates agree. This is true for intraphylum splits in many animal groups (*19*, *33*), the origin and divergences of major insect clades (*34*), early (Paleozoic) splits among basal vertebrates (*35*) and tetrapods (*12*, *32*), and most intraordinal splits among birds (*28*, *29*) and mammals (*32*, *36*, *37*). Furthermore, in a comparison of 206 trees of mammals founded on molecular and morphological data (*38*), congruence was commoner than noncongruence. Morphological trees were nearly twice as good as molecular trees in terms of matching between the rank orders of branching points (nodes) and oldest fossils, whereas morphological trees were 10% better than molecular trees in terms of stratigraphic consistency of the nodes. Among the molecular trees, those developed on the basis of DNA or RNA data were better than those developed on the basis of protein sequences, at least in rank order of nodes and stratigraphic consistency of nodes. Protein trees, however, were best in terms of minimizing the proportion of ghost range (the postulated minimum missing fossil record implied by a tree). Fossil and molecular data are not always at odds, but both approaches have drawbacks.

Under- and Overestimating Dates

Fossils can only underestimate actual dates. Paleontologists will never find the first member of a clade, so by definition the oldest fossil must be younger than the origin of its group. Diagenesis, metamorphism, and erosion remove rocks (and included fossils) from the record and paleontologists cannot sample the earth's surface exhaustively, so much is missed (*39*). Fossil occurrence may be closely correlated with the vicissitudes of rock preservation (*40*, *41*).

The importance of these factors has long been debated (*1*). According to a pessimistic view, the fossil record is so tied to the rock record that posited mass extinctions, even the Cretaceous-Tertiary boundary (K-T) event, could be artifacts of the rock record (*41*). A more optimistic view is that the K-T event and other mass extinctions are real and that the statistical manipulations used to throw doubt on them must be so crude as to be themselves doubtful (*4*). Indeed, the order of fossils in the rocks is more often in agreement with the implied order of branching events in cladograms than not (*42*–*44*). These assessments have been made with the use of new age and clade metrics (*42*–*44*) that allow assessment of the reliability of fossil records and trees. The time difference between lineage divergence and the acquisition of a recognizable synapomorphy may be important biologically but unimportant geologically; disputes are measured in millions and tens of millions of years, not thousands.

It is often proposed that molecular dates are correct (with error bars) and that methods exist to correct for error (*8*–*10*, *12*). However, critics have pointed out several pervasive biases that make molecular dates too old. First, if calibration dates are too old, then all other dates estimated from them will also be too old (*22*). The commonly used date for the initial divergence of the bird and mammal lines based on fossils (310 Ma) may be accurate (*31*) or marginally too old (*22*), but other divergence dates (such as the primate-rodent at 110 Ma, arthropod-chordate at 993 Ma, fungal-metazoan at 1100 Ma, nematode-chordate at 1177 Ma, and plant-fungal-metazoan at 1576 Ma) that are commonly used (*15*, *18*, *25*, *32*) are all on the basis of previous molecular studies. Some of these dates are incompatible: The nematode-chordate date (1177 Ma) cannot be older than the fungal-metazoan date (1100 Ma), because the first branching point is higher in the tree than the second. The choice of maximal dates such as these merely promulgates maximal estimates, all of which are probably too old. To use any of these dates injects circularity into the procedure, and to use several does not help because they are not independent of each other (*14*, *20*).

A second biasing factor is that undetected fast-evolving genes could bias estimates of timing. Empirical and statistical studies of vertebrate sequences suggest that such non-clock-like genes may be detected and that they do not affect estimates of dating (*32*). Others, however, have found that the statistical tests commonly used to exclude such sequences have unacceptably low power and could produce consistent overestimations of dates of divergence (*14*, *16*, *20*, *24*). This is because they cannot reliably reject short molecular sequences that show higher-than-normal rates of evolution, and hence the calculated time since divergence is higher than it should be. This problem may be avoided by using longer concatenated sequences and appropriate correction factors (*45*).

A third source of bias relates to polymorphism. Two species often become fixed for alternative alleles that existed as a polymorphism in their ancestral species. If so, the divergence time estimated from the DNA sequences corresponds to the origin of the polymorphism, which predates the divergence of the species (*46*). It is hard to judge the impact of this, but in cases of balanced polymorphisms estimated dates could be millions of years too old. Extreme cases of this are the human lymphocyte antigen and major histocompatability complex genes (*47*).

A fourth biasing factor is that molecular time estimates show asymmetric distributions, with a constrained younger end but an unconstrained older end. A typical plot of age estimates from different genes is right-skewed, with a large number of values at the left-hand (younger) end and a long tail of ever-older values to the right (Fig. 1). This is because rates of evolution are constrained to be nonnegative (so the lower boundary is nonelastic), but the rates are unbounded above zero (so the upper boundary is elastic) (*48*). Simply taking an arithmetic mean of the estimated divergence times on the basis of all possible rates of evolution consistently overestimates the true date. This overestimation becomes more marked as the rate of molecular evolution decreases and/or the sequences become shorter. The overestimates also grow as target times become increasingly remote, so this could be a particular problem for estimates of dates in the Precambrian, for example, for the diversification of life, the

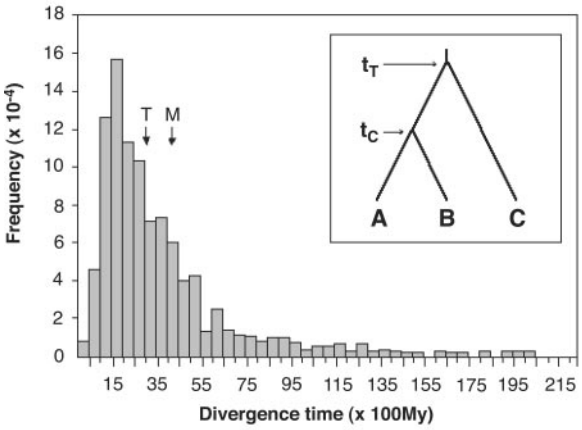


Fig. 1. Skewing and age bias in estimating molecular dates. (**Inset**) Tree topology for lineages A, B, and C. t_C and t_T represent, respectively, calibration and target times. The main panels shows a frequency distribution of 1000 estimates of the divergence time between lineages C and AB in the inset, set to have occurred 3000 Ma ago and obtained with the use of a short (75 residues), slow-evolving (one replacement per site per 10^{10} years) protein and with the use of the split between A and B, set to 300 Ma ago, as a calibration point. T and M represent target (3000 Ma) and estimated mean (4084 Ma) times. [Modified from (*48*)]

Downloaded from www.sciencemag.org on December 18, 2006

plant-fungi-animals splits, and the radiation of animal phyla (*45*, *48*).

The common assumption that molecular dates will improve as molecular data sets become larger (*8*, *13*, *45*) may not be born out (*49*, *50*). Estimated dates may indeed converge, but they may converge on consistent overestimates (*48*). Careful choice of genes may be a more appropriate strategy, with a focus on long and fast-evolving (yet alignable) sequences. The discrepancy between fossil and molecular dates for ancient parts of the tree of life may, however, always remain because of a combination of nonpreservation of critical early fossils and overestimation biases that cannot readily be corrected in the molecular dates.

Rapprochement and Prospects

In attempting to reconstruct the single tree of life, systematists have access to three essentially independent data sets (*42*–*44*): fossils, morphological cladograms, and molecular trees. Some parts of the tree of life are beginning to show a rapprochement as older fossils and younger molecular dates converge on a single conclusion.

A good case is the timing of the basal splits in the tree of modern mammals. The debate was polarized by rather loose statements that contrast the fossil record, where modern orders of mammals appear in the fossil record only after 65 Ma, in the Tertiary, with molecular dates that posit entirely Cretaceous (before 65 Ma) origins (*8*, *10*, *12*). However, further analysis of the nodes in the tree has revealed that fossil and molecular evidence are in accord for 14 of the 18 mammalian orders differentiated after the end of the Cretaceous [Supporting Online Material (SOM) Text, table S1, fig. S1].

Rapprochement is to be expected; only one tree and one set of dates can be correct. But how does it happen? In the case of the ape tree, some early molecular dates were too young, and the fossil dates were too old. The paleontological error was partly a result of misclassified and missing fossils. New finds have filled the gap back to 6 to 7 Ma on the human line, but there are no fossils yet on the chimp line. In the case of the splitting of modern mammal orders, some early discussions were misinformed: Taxonomic grades were confused, and certain Cretaceous fossils were ignored. New finds have filled some gaps (*51*), and other gaps are highlighted for further fossil hunting, especially in the Late Cretaceous of Africa and South America.

Are the congruent results better? Paleontological tree-making has improved methodologically since the 1960s by the widespread use now of cladistic methods. Some of the earlier disagreements followed from confused claims about identifications of fossils on the basis of sloppy character definition. Among molecular practitioners, there is a debate about whether one should use the maximum number of genes (*12*, *32*) or select only those that may retain a strong phylogenetic signal (*16*, *28*). In the case of metazoan origins (Table 1), molecular dates that approach the fossil dates have been achieved more by adjusted calibration dates and different statistical filtering procedures [compare with (*13*, *14*)] than by the use of different kinds of protein or DNA-RNA data. In the case of mammals (table S1), analyses published after 2000 seem to give more dates in agreement with fossil dates than earlier analyses, but there is no clear trend. Earlier analyses with discrepant human-rodent dates were mainly on the basis of mitochondrial DNA (MtDNA) sequencing (*52*–*54*), but recent analyses including MtDNA genes (*37*, *55*, *56*) offer dates more in line with paleontological estimates. The changes could have as much to do with filtering and statistical processing of the data as with the choice of genes. There is no regular matching of age estimates and numbers or types of genes, but this will be a fruitful area for further consideration.

In the quest for the tree of life, it is arid to claim that either fossils or molecules are the sole arbiter of dating or of tree shape. It is more reasonable to accept that both data sets have their strengths and weaknesses and that each can then be used to assess the other.

References and Notes

1. M. Pagel, *Nature* **401**, 877 (1999).
2. C. Darwin, *On the Origin of Species by Means of Natural Selection* (John Murray, London, 1859).
3. A. Purvis, A. Hector, *Nature* **405**, 212 (2000).
4. M. J. Benton, in *Telling the Evolutionary Time: Molecular Clocks and the Fossil Record*, P. J. Donoghue, M. P. Smith, Eds. (Taylor and Francis, London, in press).
5. C. Patterson, Ed., *Molecules and Morphology in Evolution: Conflict or Compromise?* (Cambridge Univ. Press, Cambridge, 1987).
6. A. Adoutte, S. Tillier, R. DeSalle, *Mol. Phylogenet. Evol.* **9**, 331 (1998).
7. P. J. Donoghue, M. P. Smith, Eds., *Telling the Evolutionary Time: Molecular Clocks and the Fossil Record* (Taylor and Francis, London, in press).
8. S. Easteal, *Bioessays* **21**, 1052 (1999).
9. G. A. Wray, *Genome Biol.* **3**, 1 (2001).
10. A. B. Smith, K. J. Peterson, *Ann. Rev. Ecol. Syst.* **30**, 65 (2002).
11. D. Graur, *Trends Ecol. Evol.* **8**, 141 (1993).
12. S. B. Hedges, P. H. Parker, C. G. Sibley, S. Kumar, *Nature* **381**, 226 (1996).
13. G. A. Wray, J. S. Levinton, L. H. Shapiro, *Science* **274**, 568 (1996).
14. F. J. Ayala, A. Rzhetsky, F. J. Ayala, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 606 (1998).
15. N. Nikoh *et al.*, *J. Mol. Evol.* **45**, 97 (1997).
16. L. D. Bromham, A. Rambaut, R. Fortey, A. Cooper, D. Penny, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 12386 (1998).
17. D.-F. Feng, G. Cho, R. F. Doolittle, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13028 (1997).
18. X. Gu, *J. Mol. Evol.* **47**, 369 (1998).
19. M. Lynch, *Evolution* **53**, 319 (1999).
20. L. D. Bromham, M. D. Hendy, *Proc. R. Soc. Lond. Ser. B* **267**, 1041 (2000).
21. A. Cooper, R. Fortey, *Trends Ecol. Evol.* **13**, 151 (1998).
22. M. S. Y. Lee, *J. Mol. Evol.* **49**, 385 (1999).
23. F. Rodriguez-Trelles, R. Tarrio, F. J. Ayala, in (7).
24. P. S. Soltis, D. E. Soltis, V. Savolainen, P. R. Crane, T. G. Barraclough, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 4430 (2002).
25. D. S. Heckman *et al.*, *Science* **293**, 1129 (2001).
26. M. J. Sanderson, J. A. Doyle, *Am. J. Bot.* **88**, 1499 (2001).
27. P. Kenrick, *Nature* **402**, 358 (1999).
28. A. Cooper, D. Penny, *Science* **275**, 1109 (1997).
29. M. Van Tuinen, S. B. Hedges, *Mol. Biol. Evol.* **18**, 206 (2001).
30. G. J. Dyke, *Geol. J.* **36**, 305 (2001).
31. M. J. Benton *J. Mol. Evol.* **30**, 409 (1990).
32. S. Kumar, S. B. Hedges, *Nature* **392**, 917 (1998).
33. J. Schütze *et al.*, *Proc. R. Soc. Lond. B* **266**, 63 (1999).
34. M. W. Gaunt, M. A. Miles, *Mol. Biol. Evol.* **19**, 748 (2002).
35. A. Graybeal, *Syst. Biol.* **43**, 174 (1994).
36. E. Eizirik, W. J. Murphy, S. J. O'Brien, *J. Hered.* **92**, 212 (2001).
37. U. Arnason *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8151 (2002).
38. M. J. Benton, *Mol. Phylog. Evol.* **9**, 398 (1998).
39. D. M. Raup, *Science* **177**, 1065 (1972).
40. A. B. Smith, *Philos. Trans. R. Soc. London Ser. B* **356**, 1 (2001).
41. S. E. Peters, M. Foote, *Nature* **416**, 420 (2002).
42. M. A. Norell, M. J. Novacek, *Science* **255**, 1690 (1992).
43. M. J. Benton, R. Hitchin, *Hist. Biol.* **12**, 111 (1996).
44. M. J. Benton, M. Wills, R. Hitchin, *Nature* **403**, 534 (2000).
45. M. Nei, P. Xu, G. Glazko, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 2497 (2001).
46. R. R. Hudson, J. A. Coyne, *Evolution* **56**, 1557 (2002).
47. F. J. Ayala, *Science* **270**, 1930 (1995).
48. F. Rodriguez-Trelles, R. Tarrio, F. J. Ayala, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8112 (2002).
49. G. J. P. Naylor, W. M. Brown, *Syst. Biol.* **47**, 61 (1998).
50. M. J. Sanderson, *Mol. Biol. Evol.* **19**, 101 (2002).
51. Q. Ji. *et al. Nature* **416**, 816 (2002).
52. S. Easteal, C. Collet, D. Betty, *The Mammalian Molecular Clock* (Landes, Austin, TX, 1995).
53. U. Arnason, A. Gullberg, A. Janke, X. Xu, *J. Mol. Evol.* **43**, 650 (1996).
54. A. Janke, X. Fu, U. Arnason, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 1276 (1997).
55. S. Horai, K. Hayasaka, R. Kondo, K. Tsugane, N. Takahata, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 532 (1995).
56. P. J. Waddell, Y. Cao, M. Hasegawa, D. P. Mindell, *Syst. Biol.* **48**, 119 (1999).
57. D. Y.-C. Wang, S. Kumar, S. B. Hedges, *Proc. R. Soc. Lond. Ser. B* **266**, 163 (1999).
58. We thank S. B. Hedges, Z.-X. Luo, K. Padian, and anonymous reviewers for comments. Supported by the Leverhulme Trust (M.J.B.).

Supporting Online Material

www.sciencemag.org/cgi/content/full/300/5626/1698/DC1
SOM Text
Fig. S1
Table S1
References and Notes

Downloaded from www.sciencemag.org on December 18, 2006