



I N T O F O C U S

MOLECULAR DATES FOR THE CAMBRIAN EXPLOSION: IS THE LIGHT AT THE END OF THE TUNNEL AN ONCOMING TRAIN?

Lindell Bromham

The much-debated discrepancy between molecular and palaeontological estimates for the metazoan radiation has become a lot fuzzier lately. There has always been variation between molecular date estimates, but up until a few years ago most molecular dates were clearly Precambrian. But new “relaxed clock” methods, that allow the rate of molecular evolution to vary between lineages, have changed this picture: there are now a number of molecular dating studies that provide date estimates compatible with a radiation of major metazoan lineages in the early Cambrian. Has the discrepancy between molecules and fossils dissolved?

The history of molecular clock analyses is one of increasing statistical sophistication. The earliest molecular clock studies of metazoan history assumed a strict clock (a consistent rate of molecular evolution across all lineages), such as the pioneering studies of Runnegar (1982) and Wray et al. (1996). However, comparative studies of molecular evolution demonstrated that rates of change are not always uniform across lineages (e.g., Bromham et al. 1996). To avoid this problem, many researchers relied upon clock tests (such as the relative rates test) to exclude rate-variable sequences and select apparently well-behaved, clock-like genes (e.g., Hedges et al. 2004; Wang et al. 1999). But all of these molecular dating studies produced surprisingly old dates for metazoan radiation, suggesting that the major splits in the animal kingdom occurred deep in the Precambrian. Even

studies that allowed some phyla to have different rates of molecular evolution gave dates of origin long before the first putative metazoan fossils (Bromham et al. 1998). But now a new kind of molecular date estimation, which allows every branch of the phylogeny to have a different rate of molecular evolution, is producing strikingly different results.

The most reliable way to estimate a rate of molecular evolution for every branch of a phylogeny is to use some kind of external constraints, such as fossil calibration dates. This would work best for a group that had such a complete fossil record that all nodes in the phylogeny could be confidently dated. In practice, this is rare (and, if all dates are already known with certainty, then there is little need for molecular dates for that group). Ideally, we would like to have molecular dating methods that will work on any data, even in the presence of substantial rate variation, without requiring many fixed calibration points. If you can't estimate different rates directly from calibration dates, then the only way of allowing rates to vary is to employ a model that allows you to predict the most likely pattern of rate change. There are now a number of methods that do this (for a review, see Welch and Bromham 2005). Much to the delight of many researchers, these rate-variable methods have produced date estimates that are much more compatible with fossil evidence, placing the major metazoan divergences in or just before the early Cambrian (Aris-Brosou and Yang 2002; Aris-Brosou and Yang 2003; Peterson et al. 2004).

“Relaxed clocks” were developed in response to observations that variation in the rate of molecular evolution is widespread (e.g., Drummond et al. 2006; Kishino et al. 2001; Sanderson 2002). So, at first glance, these new methods seem to be a more realistic approach to molecular dating, because they allow rate variation and they get the “right answer” for the metazoan radiation. But, like traditional molecular clock analyses, relaxed clocks must make very strong assumptions about the way that rates change over the tree (Welch and Bromham 2005). Typically this involves specifying not only models of rate change (usually favouring many small rate changes over fewer large ones), but also the distribution of speciation and extinction events in time. Given that these important assumptions concern aspects of a lineage’s evolutionary history about which we often know little, the parameter estimates used are usually chosen for reasons of statistical expedience, not because they are accurate reflections of the processes of molecular evolution (Welch et al. 2005).

For example, some methods use a model of rate change borrowed from descriptions of physical processes. The Ornstein-Uhlenbeck process (OUP) was formulated to describe the motion of a particle retarded by friction, and as such it is biased towards decreases in rate. What is reasonable for a moving particle is not necessarily reasonable for molecular rates, for which there is no a priori reason to believe that decreases in rate are more likely than increases. Unlike the motion of a particle, rates of molecular change are not expected to ever decrease to zero. An alternative model assumes an exponential distribution of rate changes. Exponential models are widely used in biology, but their use in relaxed clocks once again favours decreases in rates over rate increases (see Welch and Bromham 2005; Welch et al. 2005).

While the overall approach of relaxed clocks may seem more biologically appealing, the core assumptions of many of these methods are not necessarily a more realistic description of the data. This is not necessarily bad, but two things need to be ascertained: (1) the robustness of the methods to violations of these assumptions; (2) an empirical appreciation of how often we expect real data to violate these assumptions. The effect of these assumptions on date estimates can be seen in published data studies that present estimates obtained under several different models. For example, Peterson et al. (2004) demonstrated the sensitivity of date estimates to the assumption of variation in rates across sites: when all sites were assumed to evolve at the same rate, they obtained

date estimates for the metazoan diversification broadly compatible with the appearance of the Ediacaran fauna (573 Myr ago), but allowing rates to vary between sites gave a much older date estimate (656 Myr). Several studies have demonstrated that choosing between alternative models of rate variation or branch length estimation can dramatically change the date estimates obtained (Ho et al. 2005; Welch et al. 2005).

So, as molecular dating methods have become more sophisticated, it has become possible to arrive at a very wide range of possible date estimates by varying the assumptions made in the estimation process. This means that any reasonable position, and quite a few unreasonable ones, can now be supported by molecular date estimates. The most obvious conclusion to be drawn from this array of estimates is that molecular dates cannot be taken at face value: the data and the methods employed can dramatically change the date estimates gained. So how do we know which molecular dates are right, and which are wrong? The most attractive solution is to select the molecular dates most compatible with other lines of evidence. In effect, this makes molecular date estimates redundant – if we, quite sensibly, opt to believe only those molecular date estimates that match our prior expectations, then this is equivalent to saying that we gain no extra information from using molecular data. So it can be argued that we should ignore molecular dates altogether: if we know that we cannot take molecular dates at face value, and if the only way of determining which dates we can trust is to compare them to the fossil record, then why not just forget about molecular dates and use the fossils on their own?

Of course, it is also possible to make this kind of argument about many sources of historical data. For example, at a recent Royal Society conference (“Major Steps in Cell Evolution”, September 2005), many eminent scientists presented reasonable interpretations of the same data that enabled them to reach entirely contradictory opinions. The same rock sections were labelled as containing cells or abiotic inclusions. Phylogenetic trees were interpreted as evidence both for and against a late origin of methanogenesis. Geochemical analysis provided evidence of early eukaryotes, or for their late arrival. Snowball Earth episodes were short or long, thick or thin. There were many vigorous debates. But nobody claimed that, because two knowledgeable palaeontologists can look at the same rock and draw different conclusions, we should therefore scrap the fossil record altogether. Instead, these sometimes passionate exchanges generally ended with an agreement that more data

is needed, and more work needs to be done on its interpretation.

I would argue that we should take the same attitude to the development of molecular clocks. Molecular dates can be misleading. There are examples where they are clearly wrong, and there are cases where they are evidently imprecise. But to give up on molecular data altogether would be equivalent to rejecting all fossil evidence on the basis that the existence of Lazarus taxa demonstrates the unreliability of the geological record. The surprising history of the coelacanth, "extinct" in the fossil record since the Cretaceous, yet happily swimming the oceans today, does not mean the vertebrate fossil record is worth nothing.

There is historical information in DNA. Genomes change as organisms evolve. The longer two lineages have been separated in evolutionary time, the greater the difference between their genomes will be. But the relationship between genetic distance and time is complex, and its predictive power may be limited in many cases. The molecular clock is unlikely ever to replace the fossil record as the primary source of information on evolution in deep time. But it has a critical role to play as an alternative historical narrative, potentially complementing the biases and gaps of the palaeontological record.

The amount of genetic divergence between modern animals is far greater than we would expect from half a billion years of evolution, given what we know about rates of molecular evolution in the Phanerozoic. This observation suggests either that the animal kingdom has a deep hidden history, or that the rates of molecular evolution in the early part of the metazoan radiation were somehow must faster than they have been for most of the Phanerozoic. Just as the astounding burst of animal diversity in the early Cambrian needs to be explained (whether due to rapid evolution or palaeontological artefact), the molecular data for the metazoan radiation also requires explanation (whether deep divergence, fast early rates, or measurement artefact). I find that the more I learn about molecular clocks, the more I despair that they can ever be trusted. But the reason I keep working in the field of molecular clocks, rather than giving up, is this: wouldn't it be great if we could make them work?

REFERENCES

Aris-Brosou, S. and Yang, Z. 2002. Effects of models of rate evolution on estimation of divergence dates with special reference to the metazoan 18S ribosomal RNA phylogeny. *Systematic Biology*, 51:703–714.

Aris-Brosou, S. and Yang, Z. 2003. Bayesian models of episodic evolution support a late precambrian explosive diversification of the metazoa. *Molecular Biology and Evolution*, 20:1947–1954.

Bromham, L., Rambaut, A., and Harvey, P.H. 1996. Determinants of rate variation in mammalian DNA sequence evolution. *Journal of Molecular Evolution*, 43(6): 610–621.

Bromham, L., Rambaut, A., Fortey, R., Cooper, A., and Penny, D. 1998. Testing the Cambrian explosion hypothesis by using a molecular dating technique. *Proceedings of the National Academy of Sciences, USA*, 95:12386–12389.

Drummond, A., Ho, S.Y.W., Phillips, M.J., and Rambaut, A. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology*, in press.

Hedges, S.B., Blair, J.E., Venturi, M.L., and Shoe J.L. 2004. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. *BMC Evolutionary Biology*, 4:2.

Ho, S.Y.W., Phillips, M.J., Drummond, A.J., and Cooper, A. 2005. Accuracy of rate estimation using relaxed-clock models with a critical focus on the early metazoan radiation. *Molecular Biology and Evolution*, 22:1355–1363.

Kishino, H., Thorne, J.L., and Bruno, W.J. 2001. Performance of a divergence time estimation method under a probabilistic model of rate evolution. *Molecular Biology and Evolution*, 18:352–361.

Peterson, K.J., Lyons, J.B., Nowak, K.S., Takacs, C.M., Wargo, M.J., and McPeek, M.A. 2004. Estimating metazoan divergence times with a molecular clock. *Proceedings of the National Academy of Sciences, USA*, 101:6536–6541.

Runnegar, B. 1982. A molecular-clock date for the origin of the animal phyla. *Lethaia*, 15:199–205.

Sanderson, M.J. 2002. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Molecular Biology and Evolution*, 19:101–109.

Wang, D.Y.-C., Kumar, S., and Hedges, S.B. 1999. Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. *Proceedings of the Royal Society of London B*, 266:163–171.

Welch, J.J. and Bromham, L. 2005. Molecular dating when rates vary. *Trends in Ecology and Evolution*, 20:320–327.

Welch, J.J., Fontanillas, E., Bromham, L. 2005. Molecular dates for the "Cambrian explosion": the influence of prior assumptions. *Systematic Biology*, 54:672–678.

Wray, G.A., Levinton, J.S., and Shapiro, L.H. 1996. Molecular evidence for deep Precambrian divergences among metazoan phyla. *Science*, 274:568–573.