

Stratigraphic data have a role in phylogenetic analysis

Whatever the quality of the fossil record it must contain information about time of a species' genesis and extinction. *Professor Charles Marshall* holds that we should not ignore this information but rather find ways to incorporate it in a 'total evidence' approach.

CHARLES MARSHALL

One of the most interesting (and contentious) issues in phylogenetic analysis is whether stratigraphic data should be used in phylogenetic reconstruction. Here I argue that, at least in principle, stratigraphic data provide important phylogenetic information, and suggest how that information might be used in a meaningful way.

Stratigraphic data are phylogenetically informative

Cladograms may be used to predict the order in which species should appear in the fossil record. Thus, with a perfect fossil record, a mis-match between the order of appearance in the fossil record and that predicted by the cladogram would indicate an error in the cladogram. Thus, perfect stratigraphic data would provide an effective test of phylogenetic hypotheses.

But the fossil record is incomplete, and the order in which species appear in the fossil record may not reflect the order in which they have originated¹. In a slightly less than perfect fossil record, one would still expect the true order of species originations to be largely preserved. With a more degraded fossil record, however, we lose the ability to discern the true order of species originations. Without some way of evaluating the degree of incompleteness in the fossil record it is difficult to see how stratigraphic data can be used to test phylogenetic hypotheses.

Quantifying the incompleteness of the fossil record

Two methods have been developed for assessing the statistical significance of the differing degrees of incompleteness implied by different proposed phylogenies. Both approaches use the richness of the fossil record of each species as a guide to the confidence^{2,3} or likelihood⁴ that the true time of origin lies within some stratigraphic interval of the first appearance in the fossil record.

The maximum likelihood method enables one to immediately assess the relative degree of support for phylogenetic tree (topology) over another given an observed stratigraphic record⁴. A critical issue is immediately raised: if the best topology implied by character data differs from the best topology implied by stratigraphic data, how does one find the topology that represents the best compromise between the two? Two basic approaches have been proposed, one probabilistic, the other heuristic.

Probabilistic approaches to combining data

Disparate data types can be combined effectively if likelihoods (probabilities) can be assigned to each topology for each type of data. For each topology the likelihoods under each type of data are simply multiplied, and the tree with the highest score selected as the best topology. In addition, the range of topologies that account for 95% of the likelihood can also be found, making it possible to assess the range of hypotheses that are consistent with the data.

For DNA data maximum likelihood methods are well developed⁵ and

stratigraphic likelihoods can now be assessed. Thus, the machinery is now in place for formally combining DNA data and stratigraphic data in phylogenetic analysis. However, likelihood values are often particularly sensitive to the model of DNA evolution (for the character cladogram) or fossil preservation (for the stratigraphic analysis) used. Thus, it is important to check that the data fit the models used before relying on the likelihood scores derived from those models.

It has also been suggested that likelihoods can be assigned to cladograms based on morphological data⁶. If the method proves robust, then the basic framework for combining morphological, molecular and stratigraphic data is in place.

The probabilistic approach to combining character with stratigraphic data in phylogenetic analyses requires explicit models for evolution and preservation. This is a most exciting area at the interface between phylogeny and stratigraphy. A great deal of work needs to be done before it can be brought to fruition, especially for stratigraphic data. For example, it is becoming clear that simple models of random fossilization⁴ will be inappropriate in many geological settings⁷.

An heuristic approach

In the absence of these methodological developments, a powerful alternative for combining stratigraphic and morphological data is stratocladistics⁸. At a chosen level of stratigraphic resolution, the number of gaps in the stratigraphic record implied by a particular cladogram is counted against that topology (a parsimony count). Thus, character and stratigraphic data are combined by finding the topology with the smallest combined character and stratigraphic parsimony debt. This is in effect a total evidence approach using both character and stratigraphic data.

A balance of views

There is a strong desire to find the single best phylogenetic hypothesis. However, at least in the near future, the most significant progress will be made if the range of topologies most consistent with one or more types of data are reported. This will help focus attention on different properties and issues associated with each type of data, attention that should enable us to use the richness of information in the fossil record to help elucidate the history of life. The challenge is not whether stratigraphic data have a role in phylogenetics, the question is how best to use them.

Charles Marshall

Institute for Geophysics and Planetary Physics, UCLA, Los Angeles, USA

References

1. Marshall, C.R. Determining stratigraphic ranges. in *The adequacy of the fossil record* (Donovan, S.K. & Paul, C.R.C. eds) 23-53 (John Wiley and Sons, London, 1998).
2. Marshall, C.R. The fossil record and estimating divergence times between lineages: Maximum divergence times and the importance of reliable phylogenies. *J. Mol. Evol.* **30**, 400-408 (1990).
3. Springer, M.S. Molecular clocks and the incompleteness of the fossil record. *J. Mol. Evol.* **41**, 531-538 (1995).
4. Huelsenbeck, J.P. & Rannala, B. Phylogenetic methods come of age: testing hypotheses in an evolutionary context. *Science* **276**, 227-232 (1997).
5. Huelsenbeck, J.P. & Crandall, K.A. Phylogeny estimation and hypothesis testing using maximum likelihood. *Ann. Rev. Ecol. Syst.* **28**, 437-466 (1997).
6. Wagner, P.J. A likelihood approach for evaluating estimates of phylogenetic relationships among fossil taxa. *Paleobiol. in the press* (1998).

7. Holland, S.M. The stratigraphic distribution of fossils. *Paleobiol.* **21**, 92-109 (1995).
8. Clyde, W.C. & Fisher, D.C. Comparing the fit of stratigraphic and morphologic data in phylogenetic analysis. *Paleobiol.* **23**, 1-19 (1997).