

The value of fossil data

Is the incompleteness of the fossil record any reason to exclude the information that it contains? Professor Chris Paul argues that stratigraphic data is being treated inconsistently compared to other forms of data.

CHRIS PAUL

All contributors to this debate agree that the fossil record is incomplete. Disagreements concern whether, or at what level of completeness, data from the fossil record can contribute to phylogenetic reconstruction. The fossil record is indeed incomplete, preserved fossils represent a biased sample, the discovery of fossils is not random and those found reveal only partial information about the original organisms. Nucleic acids, for example, have a very limited survival time.

These facts are perfectly true, but if they are applied to living organisms a surprisingly similar situation exists¹. Data from the fossil record are frequently ignored because they are known to be incomplete. The implication is that complete data are essential, as [Marshall](#) explicitly states. Why then are incomplete data on morphology or nucleic acid sequences from living species no barrier at all to phylogenetic reconstruction?

Completeness of the biological record

Modern estimates of biodiversity suggest that about 10% of living species have been described^{2,3}. Even if we had described all living species, they represent at most 3-5 % of all species that lived in the Phanerozoic, let alone since life evolved^{4,5}. Where organisms live affects the likelihood of their being studied. So the biological and palaeontological records are equally incomplete, biased and poorly known.

If incomplete data are a serious concern, surely information dependent upon recently developed techniques, such as nucleic acid sequencing, is the most incomplete and therefore the most suspect? The arguments must be consistent. Partial information cannot be used to dismiss fossil data, whilst at the same time arguing that extremely incomplete molecular data are entirely reliable.

Strengths of the fossil record

The two strengths of the fossil record are that it provides information about extinct forms of life we would never imagine existed from studying only living organisms, and that it is the only reliable source of information about the history of life on earth. These involve two kinds of information: intrinsic characters (usually skeletal morphology) and an extrinsic sequence of events (stratigraphic occurrences of fossils).

Some fossils are close to the latest common ancestor of large, diverse modern clades. Archaeopteryx reveals more about what really defines birds than all modern bird species combined because it reveals characters which first set birds apart, undistorted by 150 million years of subsequent evolution.

Reliability of the preserved sequence of fossils

To my mind the most important contribution of the fossil record is the sequence of events it preserves, which is highly reliable⁴. First (or last) occurrences of fossil species can only be preserved in the wrong order with respect to the order in which they evolved if the species coexisted. Estimates suggest that during the last 550 million years, perhaps 3-5 % of species co-existed at any one time⁴. In 95-97 % of random comparisons between pairs

of fossil species there is no possibility that they could be preserved in the wrong order.

This argument does not depend on the completeness of the fossil record. If the record only contained two different fossils, they would (with 95-97% confidence) occur in the right stratigraphic order with respect to the order in which they evolved. The probabilities are also that the order of occurrence of fossil species reflects the order of branching in a phylogeny of the clades to which they belong. Fortey & Jefferies⁶ randomly reduced a known phylogeny of 100 species to just 3, simulating a 3% fossil record. The three species occurred in the correct order with respect to their branching pattern.

The fossil record provides a valid, independent test of phylogenetic hypotheses (cladograms) derived from character analysis. Stratigraphy also provides evidence for the types of relationship. Sister taxa with overlapping ranges arose by branching (cladogenesis) and are genuine sisters; those whose ranges do not overlap represent sequential taxa in an unbranched (anagenetic) lineage and are ancestor and descendent. Confidence intervals on known ranges then help to decide whether the relationship is direct (parent child) or not (grandparent).

The best interpretation of the sequential species of the chalk sea urchin *Hagenowia*⁷ is a lineage of ancestor-descendant species. Cladistics assumes that sister groups always arose at the same time. Anagenesis and cladogenesis appear to be equally common in the fossil record, which implies that cladistics generates 'ghost ranges'⁸ for roughly half the taxa considered. Thus cladistics always implies that the fossil record is more incomplete than it really is and cannot be used to test the completeness of the fossil record⁹⁻¹¹. Cladistics cannot recognize ancestors, whereas ancestors undoubtedly existed^{12,13}. Pearson points out that many lineages of microfossils can be traced continuously. To argue that none of these preserves recognizable ancestors is folly, especially when morphological intermediates occur at exactly the stratigraphic level predicted by inferred ancestor-descendant relationships as Simms has shown for the Jurassic crinoid *Isocrinus*¹⁴.

Conclusions

No science is based on complete information, and the fossil record is comparable to any other scientific data set. Ancestors undoubtedly existed and to argue that we should not even try to recognize them, as [Siddall](#) does, is untenable. Fossils which meet all the morphological criteria and occur in the correct stratigraphic position are good candidates. When tracing a family tree, discovering uncles is better than refusing to admit that anyone had a parent.

Science advances by testing hypotheses. The fossil record may be incomplete but it provides a valid independent test of phylogenetic hypotheses derived from character analysis. Surely even a poor test is better than no test at all?

Chris Paul

Department of Earth Sciences, University of Liverpool, UK

References

1. Paul, C.R.C. How complete does the fossil record have to be? *Rev. Esp. Palaeont.* **7**, 127-133 (1992).
2. Nicol, D. The number of living animal species likely to be fossilized. *Florida Scientist*, **40**, 135-139 (1977).
3. Paul, C.R.C. Adequacy, completeness and the fossil record. in *The adequacy of the fossil record* (Donovan, S. K. & Paul, C. R. C. eds) 1-22 (Wileys, Chichester, 1998).
4. Paul, C.R.C. The adequacy of the fossil record reconsidered. *Spec. Pap. Palaeont.* **33**, 7-16 (1985).
5. Sepkoski, J. *et al.* Phanerozoic marine diversity and the fossil record. *Nature*, **293**, 435-437 (1981).
6. Fortey, R.A. & Jefferies, R.P.S. and phylogeny - a compromise

approach. *Syst. Ass. Spec. Vol.* **21**, 198-234 (1985).

7. Norrell, M. A. Taxic origin and temporal diversity: the effect of phylogeny. in *Extinction and phylogeny* (Novacek, M.J. & Wheeler, Q.D. eds) 89-118 (Columbia University Press, NY, 1992).
8. Gale, A.S. & Smith, A.B. The palaeobiology of the Cretaceous irregular echinoids *Infulaster* and *Hagenowia*. *Palaeontol.* **25**, 11-42 (1982).
9. Clyde, W.C. & Fisher, D.C. Comparing the fit of stratigraphic and morphologic data in phylogenetic analysis. *Paleobiol.* **23**, 1-19 (1997).
10. Wagner, P.J. Stratigraphic tests of cladistic hypotheses. *Paleobiol.* **21**, 153-178 (1995).
11. Huelsenbeck, J.P. & Rannala, B. Maximum likelihood estimation of phylogeny using stratigraphic data. *Paleobiol.* **23**, 174-180 (1997).
12. Paul, C.R.C. The recognition of ancestors. *Historical Biology* **6**, 239-250 (1992).
13. Foote, M. On the probability of ancestors in the fossil record. *Paleobiol.* **22**, 141-151 (1996).
14. Simms, M. British Lower Jurassic crinoids. *Palaeontogr. Soc. Monogr.* 581 (1989).