

Why phylogenetic hypotheses need testing by stratigraphic data

The fossil record contains many gaps and the quality of sampling can vary immensely over time and among taxa. But, argues *Dr Peter Wagner*, when constructing phylogenies no source of information should be ignored. The challenge is knowing how to weight such data.

PETER WAGNER

Whenever and however relationships between different species (phylogenies) are inferred, they implicitly either support or contradict evolutionary hypotheses¹. Important research programs have developed recently around this idea^{1,2}. However the conclusions of such studies are only valid if accurate phylogenetic reconstructions are assumed.

Testing the fundamental assumptions behind macroevolutionary conclusions is essential. Unfortunately, popular methods for phylogenetic reconstruction, such as parsimony, can fail even under simple models of independent character evolution³⁻⁵. Thus, additional tests for phylogenetic inferences must be found.

Here I shall briefly outline:

- 1) the need for additional tests of phylogenetic hypotheses,
- 2) how stratigraphic data can yield such tests, and
- 3) how we can incorporate stratigraphic data into phylogenetic analyses.

Homoplasy masks hierarchy

Hypothesized phylogenies predict hierarchical distributions of homologies⁶. However, characters usually show local 'hierarchy' instead of true global hierarchy, indicating a failure to identify homologies correctly.

Recent analyses⁷ of various fossil groups indicate that frequencies of independently derived similarities (homoplasies) increase as species groups (clades) age. In addition hierarchical signals deteriorate as clades age. Thus, homoplasies among geologically younger taxa obscure the global hierarchy among geologically older taxa.

Phylogenetic predictions concerning non-hierarchical congruence are inexact. Rampant homoplasy also can produce high levels of congruence^{8,9}. Even randomly produced data show some general congruence^{10,11}. Unsurprisingly, simulations confirm that parsimony analysis is less accurate as homoplasy increases and global hierarchy decreases^{4,12}.

Homoplasy vs. homology

Differentiating homology and homoplasy hypotheses often requires multiple independent data sets. If the examples of congruence highlighted by one set of characters, for example molecular characters, produce a similar phylogenetic tree to that from say morphological data then one can feel confident about this predicted phylogeny^{13,14}. Such techniques have been applied to taxa with both fossil (extinct) and living (extant) members¹⁵, but cannot be used on taxa with single character sets (that is most extinct clades).

Computer simulated evolution with finite character states shows that parsimony often underestimates actual evolution^{16,17}. This is true even when the shape of the tree produced by parsimony analyses matches the known course of simulated evolution. Such differences become more prominent when

the simulations differ from the initial assumptions used in the subsequent analyses^{16,18}. If the most likely length of a tree is not the minimum length perhaps the abductive assumption that "least is best"¹⁹ in phylogenetic analyses deserves re-evaluation.

Testing phylogeny with stratigraphy

Phylogenies also make necessary predictions about the time of appearance and duration of extinct lineages²⁰. These in turn make probabilistic predictions about the range of strata in which specific fossils can be expected^{21,22}. If hypotheses make predictions about data, then they can be tested by those data.

There are two ways to exploit this information: traditional tests of the probability of failing to sample taxa over an amount of strata²³⁻²⁵, and tests of the likelihood of alternative phylogenetic hypotheses occurring given the available stratigraphic²⁶ (and even morphologic¹⁶) data.

Probabilistic tests²² have been used extensively to test hypotheses about extinctions²⁷. Stratigraphic data can also test hypotheses on the origins of groups implied from phylogenies. Such data can identify examples of congruence that do not reflect homology. The shortest phylogenetic tree with no significant gaps can then be found²⁸. In this way stratigraphy is used semi-deductively to reject hypotheses about relationships, not semi-inductively to estimate them.

Stratocladistics and Likelihood

Stratocladistics²⁸ represents a variation of a simple likelihood test²⁹. Summing steps is equivalent to summing the negative log probabilities of character change if²⁹:

- 1) the probabilities of each character are the same for all clade members,
- 2) characters never change twice on the same branch, and
- 3) characters always evolve independently.

The most likely tree then is the one invoking the fewest steps.

Negative log probabilities of implied stratigraphic gaps (regions of the tree where no examples exist in the fossil record) can be summed along with the negative log probabilities of character change if the stratigraphic "characters" are weighted as:

$$-\ln(1-R) = W * -\ln(P[c])$$

where 1-R is the probability of not sampling a taxon over a particular interval³⁰, W is the relative weight of the stratigraphic character, and P[c] is the probability of character change along a branch. The most likely tree then is the one with the lowest sum of character changes and weighted gaps.

(Type I and II errors can become common with numerous probability tests³¹. However, nodes that produce numerous unique branches can overcome significant stratigraphic gaps²⁸. This offers a buffer against Type I errors making such methods preferable to probabilistic approaches I have previously advocated.)

Unsound assumptions?

Stratocladistics favours trees with the fewest examples of homoplasy and shortest stratigraphic gaps. Indeed many likelihood tests of stratigraphic ranges favour no gaps in stratigraphy²⁶.

Neither assumption is beyond criticism. I have already explained why the minimum homoplasy premise likely is unsound. In addition evaluation of

preservation rates³², implicit to different phylogenetic hypotheses¹⁶, often shows that the lack of gaps in the fossil record is unlikely even when the record is fairly continuous¹⁷.

Phylogenetically implied gaps imply sampling intensities, the likelihoods of which can be tested statistically³². Testing an implicit prediction of a phylogenetic hypothesis in this way is preferable to making an assumption about sampling, even if that assumption is derived from observed data.

Monte Carlo simulations can estimate tree likelihoods given observed congruence, without assuming "morphologic clocks"¹⁶. This allows the likelihood of alternative trees to be calculated as the product of the likelihood in regard to stratigraphic data and the likelihood based on morphology,

$$(L[\text{tree}|\text{stratigraphy}] * L[\text{tree}|\text{morphology}]) \text{ (ref. 33)}$$

Such tests will reject most trees as implausible, while retaining the most likely and those within an acceptable range³³. This is the approach that I currently favour.

Uneven sampling

Stratigraphic data are not panaceas for phylogenetic reconstruction. Sampling of some taxa is sufficiently poor that long gaps are anticipated. Stratigraphic correlations are sometimes so poor that one is uncertain whether trees even indicate gaps. More importantly, extremely heterogeneous sampling can reduce the power of such statistical tests^{30,32} and render all hypotheses nearly equally likely. Sampling heterogeneity can also exaggerate the improbability of gaps³⁴, although this affects some tests less than others³⁵.

Inconsistent sampling does not preclude using stratigraphic data in statistical tests. However, methods (statistical or otherwise) should not be used without first testing their assumptions. Further development of statistical tests of both stratigraphic and morphologic data will permit better understanding of both geological history and character evolution^{36,37}. Improved phylogenetic inferences will be natural by-products of these advances.

Concluding thoughts

A reconstructed phylogeny is more than a series of nodes to be named - it offers a summary of evolutionary patterns. Most importantly, it gives information about the rules governing these patterns and where/when those rules differ or change.

If we are to use phylogenies as tests of macroevolutionary hypotheses, then the phylogenies themselves must be rigorously tested to help ensure that our conclusions about evolution are not artefacts of our assumptions about evolution.

Peter Wagner

Field Museum of Natural History, Chicago, USA

References

1. Smith, A.B. *Systematics and the fossil record - Documenting evolutionary patterns*. (Blackwell, Oxford, 1994).
2. Harvey, P.H. & Pagel, M.D. *The comparative method in evolutionary biology*. (Oxford, Oxford, 1991).
3. Rohlf, F.J., Chang, W.S., Sokal, R.R. & Kim, J. Accuracy of estimated phylogenies: effects of tree topology and evolutionary model. *Evolution* **44**, 1671-1684 (1990).
4. Huelsenbeck, J.P. When are fossils better than extant taxa in phylogenetic analysis? *Syst. Zool.* **40**, 458-469 (1991).
5. Huelsenbeck, J.P. & Hillis, D.M. Success of phylogenetic methods in the four-taxon case. *Syst. Biol.* **42**, 247-264 (1993).

6. Estabrook, G.F. Cladistic methodology: A discussion of the theoretical basis for the induction of evolutionary history. *Ann. Revs Ecol. Evol.* **3**, 427-456 (1972).
7. Wagner, P.J. Saturation of cladistic character space - debunking the myth of the infinite. *Geol. Soc. Am. Abstracts with Program* **30**, A-326 (1998).
8. Klöser, M., Farris, J.S., Kluge, A.G. & Bult, C. Skewness and permutation. *Cladistics* **8**, 275-287 (1992).
9. Alroy, J. Four permutation tests for the presence of phylogenetic structure. *Syst. Biol.* **43**, 430-437 (1994).
10. Archie, J.W. A randomization test for phylogenetic information in systematic data. *Syst. Zool.* **38**, 239 - 252 (1989).
11. Faith, D.P. & Cranston, P.S. Could a cladogram this short have arisen by chance alone? On permutation tests for cladistic structure. *Cladistics* **7**, 1-28 (1991).
12. Archie, J.W. in *Homoplasy - the recurrence of similarity in evolution* (Sanderson, M.J. & Hufford, L. eds) 153-187 (Academic, San Diego, 1996).
13. Templeton, A. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and apes. *Evol.* **37**, 221-244 (1983).
14. Larson, A. in *Molecular approaches to ecology and evolution* (Scheirwater, B., Streit, B., Wagner, G.P. & DeSalle, R. eds.) 371-390 (Birkhäuser Verlag, Basel, 1994).
15. Littlewood, D.T.J. & Smith, A.B. A combined morphological and molecular phylogeny for sea urchins (*Echinoidea: Echinodermata*). *Phil Trans. R. Soc. London B* **347**, 213-234 (1995).
16. Wagner, P.J. A likelihood approach for estimating phylogenetic relationships among fossil taxa. *Paleobiology* **24**, 430-449 (1998).
17. Wagner, P.J. Phylogenetics of Ordovician - Silurian Lophospiridae (*Gastropoda: Murchisoniina*): The importance of stratigraphic data. *Am. Malacolog. Bull.* **in the press** (1999).
18. Kuhner, M.K. & Felsenstein, J. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Molec. Biol. Evol.* **11**, 459-468 (1994).
19. Sober, E. *Reconstructing the past* (The MIT Press, Cambridge, Mass., 1988).
20. Smith, A.B. Patterns of diversification and extinction in early Palaeozoic echinoderms. *Palaeontol.* **31**, 799-828 (1988).
21. Paul, C.R.C. in *Problems of phylogenetic reconstruction* (Joysey, K.A. & Friday, A.E. eds) 75-117 (Academic, London, 1982).
22. Strauss, D. & Sadler, P.M. Classical confidence intervals and Bayesian probability estimates for ends of local taxon ranges. *Math. Geol.* **21**, 411-427 (1989).
23. Marshall, C.R. Confidence intervals on stratigraphic ranges. *Paleobiol.* **16**, 1-10 (1990).
24. Jackson, J.B.C. & Cheetham, A.H. Phylogeny reconstruction and the tempo of speciation in cheilostome Bryozoa. *Paleobiol.* **20**, 407-423 (1994).
25. Wagner, P.J. Stratigraphic tests of cladistic hypotheses. *Paleobiol.* **21**, 153-178 (1995).
26. Huelsenbeck, J.P. & Rannala, B. Maximum likelihood estimation of topology and node times using stratigraphic data. *Paleobiol.* **23**, 174-180 (1997).
27. Marshall, C.R. & Ward, P.D. Sudden and gradual molluscan extinctions in the latest Cretaceous of the western European Tethys. *Science* **274**, 1360-1363 (1996).
28. Fisher, D.C. in *Interpreting the hierarchy of nature - from systematic patterns to evolutionary process theories* (Grande, L. & Rieppel, O. eds) 133-171 (Academic, Orlando, FL, 1994).
29. Felsenstein, J. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. *Biol. J. Linn. Soc.* **16**, 183-196 (1981).
30. Foote, M. & Raup, D.M. Fossil preservation and the stratigraphic ranges of taxa. *Paleobiol.* **22**, 121-140 (1996).

31. Sokal, R.R. & Rohlf, F.J. *Biometry*, 2nd ed. (W. H. Freeman, New York, 1981).
32. Foote, M. Estimating taxonomic durations and preservation probability. *Paleobiol.* **23**, 278-300 (1997).
33. Edwards, A.W.F. *Likelihood - expanded edition* (Johns Hopkins University, Baltimore, 1992).
34. Marshall, C.R. in *New approaches for studying speciation in the fossil record* (Erwin, D.H. & Anstey, R.L. eds) 208-236 (Columbia University, New York, 1995).
35. Marshall, C.R. Confidence intervals on stratigraphic ranges with nonrandom distributions of fossil horizons. *Paleobiol.* **23**, 165-173 (1997).
36. Meacham, C.A. in *Cladistics: perspectives on the reconstruction of evolutionary history* (Duncan, T. & Stuessy, T.F. eds) 152-165 (Columbia University, New York, 1984).
37. Sharkey, M.J. A hypothesis-independent method of character weighting for cladistic analysis. *Cladistics* **5**, 63-86 (1989).

I thank M. Foote, H. Larsson, S. Lidgard, R. Lupia, C. Sidor, A. Smith and J. Alroy for comments and discussion.