coefficients listed in Table 2; resulting melt compositions are shown in Fig. 1 as circles joined by solid curves (the white circle is the model melt of the assumed depleted source; the black circles are for melts of sources to which 1–7 wt% of the crustal component has been added in 1% increments). Sediments and altered crustal rocks vary considerably in their composition (particularly in their Ba/Th ratios; Fig. 1h) and therefore our model curves are best regarded as averages over which the range of melts of natural mantle–crust mixtures might scatter. The calculated solid curves in Fig. 1 assume that partition coefficients for trace elements and oxygen isotope fractionation between melt and residue during melting are independent of enrichment. This approximation is likely to be incorrect for the La/Sr and Sm/La ratios if enriched material is present as mafic domains because Sm and Ti are relatively compatible in garnet (a major constituent of high-pressure mafic rocks). We have therefore shown an alternative model in Fig. 1e and f in which MORBs are mixtures of 10% batch partial melts of depleted peridotite and 10% batch partial melts of eclogite having a composition equal to our recycled-crust model component; eclogite is assumed to be 50% garnet and 50% clinopyroxene. Eclogite probably melts to a different extent from peridotite at a given set of conditions. The direction and extent of these differences are expected to vary with temperature, pressure and water fugacity; rather than specifying these variables we have adopted the simplifying assumption that eclogite and peridotite components melt to comparable extents. The slopes of curves calculated by this model are not significantly changed by adopting other reasonable assumptions about the melting behaviour of eclogite, although for some models the amount of recycled material in the source may be significantly different from the fraction of melt derived from that material. This alternative differs little from the first model for indices other than La/Sr and Sm/La as a result of the subtle differences between melt–peridotite and melt–eclogite distribution coefficients for other elements plotted in Fig. 1. We therefore omit the results of this alternative calculation from panels other than Fig. 1e and f. We have assumed that oxygen isotope compositions during melting are constant (that is, variations in the $^{18}$O of melts are directly proportional to variations in their sources). Data sources for our model calculations are referenced in the Supplementary Information.

Received 12 July; accepted 9 December 1999.


Supplementary information is available on Nature’s World-Wide Web site (http://www.nature.com) or as paper copy from the London editorial offices of Nature.

Acknowledgements

We thank P. Michael; we also thank D. Anderson, P. Asimow, A. Halliday and C. Langmuir for comments on the manuscript. We thank the Chevron Corporation for donation of the mass spectrometer used for isotopic measurements.

Correspondence and requests for materials should be addressed to J.M.E. (e-mail: elier@gps.caltech.edu).

Quality of the fossil record through time

M. J. Benton*, M. A. Wills† & R. Hitchin*†

* Department of Earth Sciences, University of Bristol, Bristol BS8 1RJ, UK
† Oxford University Museum of Natural History, Parks Road, Oxford OX1 3PW, UK

Does the fossil record present a true picture of the history of life1–3, or should it be viewed with caution4–6? Raup7 argued that plots of the diversification of life2 were an illustration of bias: the older the rocks, the less we know. The debate was partially resolved by the observation8 that different data sets gave similar patterns of rising diversity through time. Here we show that new assessment methods, in which the order of fossils in the rocks (stratigraphy) is compared with the order inherent in evolutionary trees (phylogeny), provide a more convincing analytical tool: stratigraphy and phylogeny offer independent data on history. Assessments of congruence between stratigraphy and phylogeny for a sample of 1,000 published phylogenies show no evidence of diminution of quality backwards in time. Ancient rocks clearly preserve less information, on average, than more recent rocks. However, if scaled to the stratigraphic level of the stage and the taxonomic level of the family, the past 540 million years of the fossil record provide uniformly good documentation of the life of the past.

The reduction in quality of the fossil record backwards in time seems self-evident. Fossils in ancient rocks are more likely to have been eroded, crushed, melted, subducted, not collected or misunderstood than younger fossils5. The demonstration7 that similar patterns of diversification were found from different data sets seemed to indicate that, at least when viewed on a broad scale, the fossil record did correctly document the history of life. Indeed, palaeobiologists subsequently used the fossil record as a literal source of data on the history of the diversification of life6,9 and of mass extinctions10,11 without applying any correction factors for possible time-related bias. However, the different data sets compared7 were all subject to the same geological biasing factors, and that study did not use independent data to demonstrate that older parts of the fossil record could be trusted. This lingering doubt about the long-term quality of the fossil record has resurfaced in debates about the timing of origin of major groups of organisms: molecular studies suggest that the Metazoa (animals)12 and modern bird and mammal orders13,14 apparently originated much earlier.
Numbers of trees in each partition, the mean tree size and the standard error of the range of tree sizes are shown. In these samples, 95% of all trees fall in the range of the mean tree size.

Table 1 Time-related partitions of the dataset of 1,000 cladograms

<table>
<thead>
<tr>
<th>Dataset partition</th>
<th>Number of trees</th>
<th>Mean tree size</th>
<th>s.e.</th>
<th>Mean range of origins (Myr)</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. All trees</td>
<td>1,000</td>
<td>10.359</td>
<td>0.196</td>
<td>126.80</td>
<td>4.19</td>
</tr>
<tr>
<td>2. Origins entirely Palaeozoic</td>
<td>156</td>
<td>8.468</td>
<td>0.345</td>
<td>73.17</td>
<td>5.99</td>
</tr>
<tr>
<td>3. Origins a mixture of Palaeozoic and post-Palaeozoic dates</td>
<td>261</td>
<td>11.388</td>
<td>0.417</td>
<td>266.36</td>
<td>10.25</td>
</tr>
<tr>
<td>4. Origins entirely Mesozoic</td>
<td>106</td>
<td>9.075</td>
<td>0.425</td>
<td>82.70</td>
<td>5.73</td>
</tr>
<tr>
<td>5. Origins a mixture of Mesozoic and Cenozoic dates</td>
<td>243</td>
<td>12.086</td>
<td>0.471</td>
<td>123.85</td>
<td>4.12</td>
</tr>
<tr>
<td>6. Origins entirely Cenozoic</td>
<td>234</td>
<td>9.282</td>
<td>0.352</td>
<td>29.92</td>
<td>2.29</td>
</tr>
</tbody>
</table>

Numbers of trees in each partition, the mean tree size and the standard error of the range of tree sizes are shown. In these samples, 95% of all trees fall in the range of the mean tree size ±2 standard errors (s.e.) thus 10.359 ± 0.392, for all trees.

Table 2 Results for comparisons of cladograms, with stratigraphic data

<table>
<thead>
<tr>
<th>Dataset partition</th>
<th>SCI Mean</th>
<th>s.e.</th>
<th>RCI Mean</th>
<th>s.e.</th>
<th>GER Mean</th>
<th>s.e.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.551</td>
<td>0.008</td>
<td>31.130</td>
<td>4.587</td>
<td>0.562</td>
<td>0.010</td>
</tr>
<tr>
<td>2</td>
<td>0.618</td>
<td>0.021</td>
<td>62.064</td>
<td>3.446</td>
<td>0.529</td>
<td>0.031</td>
</tr>
<tr>
<td>3</td>
<td>0.493</td>
<td>0.014</td>
<td>16.247</td>
<td>14.442</td>
<td>0.515</td>
<td>0.018</td>
</tr>
<tr>
<td>4</td>
<td>0.601</td>
<td>0.028</td>
<td>44.566</td>
<td>4.368</td>
<td>0.636</td>
<td>0.033</td>
</tr>
<tr>
<td>5</td>
<td>0.496</td>
<td>0.013</td>
<td>19.033</td>
<td>6.098</td>
<td>0.627</td>
<td>0.018</td>
</tr>
<tr>
<td>6</td>
<td>0.604</td>
<td>0.018</td>
<td>33.604</td>
<td>8.459</td>
<td>0.532</td>
<td>0.024</td>
</tr>
</tbody>
</table>

Data shown according to the stratigraphic consistency index, SCI, the relative completeness index, RCI and the gap excess ratio, GER. Mean values for each metric are given, together with the standard error (s.e.). Dataset partitions are defined in Table 1.

than was expected from the fossil evidence. These new results could be mistaken for, or they could indicate long episodes of missing fossil record. However, there is no doubt that older parts of the fossil record cannot ever be as well known as more recent parts.

One solution is to compare independent sources of data. New methods for inferring phylogenies, such as cladistics applied to the morphology of fossil and living taxa, and molecular sequencing techniques, provide representations of evolutionary trees that are essentially independent of stratigraphic data. Hence, it is possible to compare age data from fossils in the rocks with clade data from molecular and morphological trees using a range of congruence metrics, the stratigraphic consistency index (SCI), the relative completeness index (RCI) and the gap excess ratio (GER) (see Methods).

In our study, we compiled a database of 1,000 morphological and molecular cladograms from the literature (see Methods). The cladograms were sorted into those in which all lineages arose in the Palaeozoic, Mesozoic or Cenozoic, as well as mixed cases with Palaeozoic and post-Palaeozoic origins and mixed Mesozoic/Cenozoic origins. Sample sizes and tree sizes were relatively uniform across all five temporal divisions (Table 1). Remarkably, the relative congruence of age and clade data for all five time divisions showed no clear increase from most ancient to most recent (Table 2; Fig. 1).

In more detail, SCI values for all time partitions of the data set are very similar, with mean values in the range 0.493–0.618, and with most lying close to the mean value for the whole dataset of 0.551. Indeed, the highest mean SCI value overall, 0.618, is for the Palaeozoic-only values, and the lowest mean values (0.493, 0.496) are for the mixed Palaeozoic/Mesozoic and Mesozoic/Cenozoic partitions (Table 2, Fig. 1). The RCI measures showed greater variation, but no clear relationship to time (Table 2; Fig. 1). Mean values range widely, from 11.362 to 62.064, around the mean for the whole dataset of 31.130. The highest value is for the Palaeozoic partitions, which is not surprising as the RCI depends on a measure of total known range (see Methods), and for clades arising in the Palaeozoic, known ranges may potentially extend to the present day. The GER measures show little evidence of time bias (Table 2; Fig. 1). Mean values range from 0.491 to 0.636, but most lie close to the whole-sample mean of 0.562. The mean value for Palaeozoic cladograms (0.529) is not significantly different from that for the Cenozoic (0.532).

The expected result, that the fossil record of stratigraphic first appearance should become worse with increasing age is not confirmed. These counter-intuitive results might be the result of temporal biases in the data set such as (1) variable cladogram quality, (2) differing major taxonomic groups, (3) differing approaches to cladogram construction, (4) the quality of age dates and stratigraphy, (5) the total age span of origins, (6) the taxonomic level of terminal taxa, (7) tree size and (8) tree balance.

The first four of these possible sources of error imply highly unlikely assertions, such as that most Palaeozoic cladograms are markedly better than post-Palaeozoic (1), that the Palaeozoic groups are more amenable to cladistic analysis than the post-Palaeozoic (2), that systematists working on groups with Palaeozoic origins use better techniques than those working on post-Palaeozoic groups (3) and that Palaeozoic age dates and stratigraphy are better than post-Palaeozoic (4).

It could be argued that the first of these problems, variable cladogram quality, makes our results meaningless. The assertion would simply be that cladograms are so full of error that there is no reason to expect congruence with any stratigraphic signal in a large-scale analysis such as ours. There is no absolute measure of cladogram quality. However, cladograms tend to match stratigraphy better than predicted by null models, and there is evidence that published morphological and molecular phylogenetic trees are probably generally close to the truth.

The remaining four possible biasing factors all turn out to be differently distributed in the current data set. The total span of time occupied by the origins of the groups in an assessed cladogram (5)
GER is standard range length, the known fossil record. SCI is the ratio of consistent to inconsistent nodes in a cladogram. RCI is observed tree with SCI calculated according to the distribution of ranges in b. b. the observed tree and observed distribution of stratigraphic range data, yielding an RCI of 66%. GER is derived from G_min and G_max values calculated in c and d. The stratigraphic ranges from b rearranged on a pectinate tree to yield the smallest possible MIG or G_min. d. The stratigraphic ranges from b rearranged on a pectinate tree to yield the largest possible MIG or G_max.

Figure 2 Calculation of the three congruence metrics for age versus clade comparisons. SCI is the ratio of consistent to inconsistent nodes in a cladogram. RCI is observed tree calculated from an observed tree with SCI calculated according to the distribution of ranges in b. b. the observed tree and observed distribution of stratigraphic range data, yielding an RCI of 66%. GER is derived from G_min and G_max values calculated in c and d. The stratigraphic ranges from b rearranged on a pectinate tree to yield the smallest possible MIG or G_min. d. The stratigraphic ranges from b rearranged on a pectinate tree to yield the largest possible MIG or G_max.

Methods

Data base

The data set consisted of 1,000 cladograms, including one cladogram of ‘all life’, 33 cladograms of plants, nine cladograms of coelenterates, one cladogram of molluscs, 179 cladograms of arthropods, 14 cladograms of brachiopods, one cladogram each of bryzoans and graptolites, 60 cladograms of echinoderms, 34 cladograms of basal deuterostomes including calcichordates, 157 cladograms of fishes and 510 cladograms of tetrapods, including 26 of amphibians, 203 of reptiles, 8 of birds and 269 of mammals, extracted piecemeal from many sources. The Fossil Record 2 (ref. 28) was the major source of stratigraphic data for dates of origin of families and suprafamilial taxa. Some cladograms included genera and species, and their dates of origin were generally determined from data provided in the paper that presented the cladogram. Origins and extinctions of clades were assessed to the level of the stratigraphic stage (mean duration of the 79 time units used for the Phanerozoic is 6.8 Myr). Geological dates for these stages (in Myr) were taken from one source.26 The data set was divided into stratigraphic divisions listed in Table 1, and the entire data set is available as Supplementary Information.

Letters to nature

An alternative assertion could be that the age/clade metrics cannot in fact detect major gaps in the fossil record. Perhaps the fossil record does deteriorate seriously backwards in time, but the preserved parts of the fossil record are equally complete through time. For example, a sequence of Cretaceous fossils with ammonites might be no more complete than a Silurian sequence with graptolites, some 300 million years older. The cladogram metrics could not be expected to give equivalent values. However, perhaps the good Silurian record represents only a tiny fraction of the total record, a much smaller proportion than is represented by the Cretaceous ammonites. In other words, the older record is substantially worse, and our metrics might fail to detect it. This would mean that the fossil record is itself a biased sample of past life. However, soft-bodied organisms are equally poorly known from rocks of all ages, for example. The age/clade metrics circumvent the problem to a certain extent, as the phylogenies include preservable and non-preservable taxa alike. Further tests of the fossil record/true record issue are required. Our results are confirmed by recent studies that consider the preservation potential of different fossil groups.

Our finding that the fossil record does not diminish in quality with time seems at first impossible, because it is a demonstrable fact that geological activity destroys ancient rocks and ancient fossils. We should consider the scaling of observations in terms of both stratigraphy and taxonomy. Experience shows that major changes in the dating of fossils do not occur at the level of geological systems or stages, but at the finer divisions of substages and zones. Likewise, orders and families are often relatively stable, while new discoveries constantly alter the definitions of genera and species of fossils. The stability of longer time intervals and larger taxonomic categories perhaps reflects an adequate (if incomplete) fossil record. However, global studies of diversification at species and zonal levels would generally be meaningless because the incompleteness of more ancient parts of the fossil record renders it inadequate in most cases for such studies. It is important to distinguish between ‘completeness’ and ‘adequacy’. Early parts of the fossil record are clearly incomplete, but they can be regarded as adequate to illustrate the broad patterns of the history of life.
Age versus clade congruence

Three measures were used to assess age versus clade congruence (Fig. 2): the stratigraphic consistency index, SCI (ref. 24), the relative completeness index, RC1 (ref. 25) and the gap excess ratio, GER (ref. 21). The SCI is the ratio of consistent to inconsistent nodes in a cladogram, and it can range from 0 to 1.0 in a fully pectinate (unbalanced) tree, but the minimum value lies between 0 and 0.5 in balanced trees12,13. The RC1 and GER depend on numerical age estimates of the branching points on a cladogram, and the calculation of ‘ghost ranges’. The ghost range14 is the difference in age, or number of stratigraphic intervals, between the oldest known fossils of two sister taxa. The RC1 is assessed as the ratio between the sum of ghost ranges to the sum of recorded fossil ranges in any cladogram. The GER focuses solely on the estimated dates of origin of groups, and compares the sum of actual ghost ranges in a cladogram with the theoretical minimum number of ghost ranges if the various branches in the cladogram are rearranged. The metrics were calculated using the software ‘Ghosts 2.4’, developed by M.A.W., which assesses all three metrics (SCI, RC1, and GER) for individual cladograms, or for large batches of cladograms (available from http://palaeo.gly.bris.ac.uk/cladestrat/cladesstrat.html).

Received 7 June; accepted 15 November 1999.


Supplementary information is available on Nature’s World-Wide Web site (http://www.nature.com) or as paper copy from the London editorial office of Nature, and may also be viewed at http://palaeo.gly.bris.ac.uk/cladestrat/cladesstrat.html.

Acknowledgements

We thank the Leverhulme Trust and NERC for continued funding of our work, and E. Farra, M. Foote and P. N. Pearson for helpful comments on the manuscript.

Correspondence and requests for materials should be addressed to M.J.B. (e-mail: mike.benton@bristol.ac.uk).

© 2000 Macmillan Magazines Ltd