

## Multiple origins of life

(extinction/bioclade/Precambrian/evolution/stochastic processes)

DAVID M. RAUP\* AND JAMES W. VALENTINE†

\*Department of the Geophysical Sciences, University of Chicago, Chicago, Illinois 60637; and †Department of Geological Sciences, University of California, Santa Barbara, California 93106

Contributed by David M. Raup, February 22, 1983

**ABSTRACT** There is some indication that life may have originated readily under primitive earth conditions. If there were multiple origins of life, the result could have been a polyphyletic biota today. Using simple stochastic models for diversification and extinction, we conclude: (i) the probability of survival of life is low unless there are multiple origins, and (ii) given survival of life and given as many as 10 independent origins of life, the odds are that all but one would have gone extinct, yielding the monophyletic biota we have now. The fact of the survival of our particular form of life does not imply that it was unique or superior.

The formation of life *de novo* is generally viewed as unlikely or impossible under present earth conditions. However, conditions on the primitive earth seem to have been more appropriate for life origins. The oldest known rocks in which fossils could appear are about 3,500 Myr old, and they have yielded structures that are generally interpreted as the remains of prokaryotes (1). This puts the origin of life in an interval with no rock record and for which we have no direct evidence of earth conditions.

The early age of the first known fossils suggests that life originated readily when suitable conditions became available. Thus, the idea that the origin of life was exceedingly improbable is brought into question; perhaps life can start easily, given a reasonably propitious environment. This possibility has strengthened speculation that life may be widespread through the universe. It also strengthens speculation that earthly life may have developed more than once.

Life forms are made possible by the remarkable properties of polypeptides. It has been argued that there must be many potential but unrealized polypeptides that could be used in living systems (2). The number of possible primary polypeptide structures with lengths comparable to those found in living systems is almost infinite. This suggests that the particular subset of polypeptides of which organisms are now composed is only one of a great many that could be associated in viable biochemistries.

There is no taxonomic category available to contain all life forms descended from a single event of life origin. Here, we term such a group, earthly or otherwise, a bioclade. If more than one bioclade survives, life is polyphyletic. If only one survives, it is monophyletic.

The presence of an oxidizing atmosphere and of complex ecosystems today effectively precludes the origination of new bioclades (3), and the same argument may be extended to the entire Phanerozoic time. However, immediately after the first origin of life, potential niches for primitive organisms would not have been filled, and the possibility of additional life origins can be seriously entertained. Given a primitive earth contain-

ing a full assortment of prebiotically synthesized organic building blocks for life and conditions appropriate for life origins, it is possible that rates of bioclade origins might compare well enough with rates of diversification. Indeed, there has been speculation that life may have been polyphyletic (1, 4). However, there is strong evidence that all living forms are descended from a single ancestor. Biochemical and organizational similarities and the “universality” of the genetic code indicate this. Therefore, the question for this paper is: If there were multiple bioclades early in life history, what is the probability that only one would have survived? In other words, if life originated more than once, would we know it?

### CLADE EXTINCTION AND SURVIVAL

Our approach to the problem will be to use simple stochastic branching models. Any evolutionary system where hybridization between lineages is negligible can be viewed as a branching system. Lineages originate as branches off preexisting lineages, and they are terminated by extinction, with the result being describable as a classic birth–death process. Within a given branching system, or tree, there inevitably are clusters of lineages descended from single founders, or ancestors. Such clusters are called clades. The bioclade, defined above, encompasses all clades (living and extinct) descended from one origination of life.

Birth–death models have been applied widely to problems of the evolution and extinction of clades within our bioclade. Rates of branching and extinction for much of the Phanerozoic record (600 Myr B.P. and younger) are reasonably well known and have been used for analyses of survivorship at various levels in the taxonomic hierarchy (5, 6).

A simple but straightforward approach to clade survivorship is expressed in the following equation for the probability,  $P_{s,t}$ , that a clade that starts with a single founder at some time  $t = 0$  will survive at least until time  $t = t$  (7):

$$P_{s,t} = 1 - \frac{q(e^{(p-q)t} - 1)}{pe^{(p-q)t} - q}, \quad [1]$$

where  $p$  is the probability (per lineage per Myr) of lineage branching and  $q$  is the comparable probability of lineage extinction. Time ( $t$ ) is expressed in Myr.

The time-homogeneous model implied by Eq. 1 requires the assumption that  $p$  and  $q$  are stochastically constant through time. This assumption is valid as a first approximation in some paleontological situations but not in others. If, for example, a clade goes through a mass extinction, the extinction rate,  $q$ , goes up sharply for a brief time, causing the clade to pass through a bottleneck of low diversity. Under such circumstances, the probability of survival of the clade is lower than would be predicted by Eq. 1. Conversely, a short-term increase in origination rate (or decrease in extinction rate) will increase the probability of

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

clade survival because standing diversity will go up. In view of this, the applicability of the time-homogeneous model depends on the nature of the departures from the assumption of temporal constancy of turnover rates. As will be shown below, the time-homogeneous model is probably conservative in the present context because it overestimates the probability of survival of bioclades.

Given the diversity of life that we know to have existed at the start of Phanerozoic time, what is the probability that all species were the descendants of a single founder? That is, what is the probability that all but one bioclade went extinct? Given  $D$ , the number of independent origins of life, the probability,  $M_t$ , of monophyly (one surviving bioclade) is:

$$M_t = \frac{D P_{s,t}(1 - P_{s,t})^{D-1}}{1 - (1 - P_{s,t})^D} \quad [2]$$

where  $P_{s,t}$  is defined in Eq. 1.

By solving Eq. 2 for an array of combinations of the constants, we can estimate the probability of monophyly for the group of lineages existing at some later time. If the later time is 600 Myr B.P. (latest Precambrian),  $t$  in Eq. 2 may be taken to be 3,000 Myr. As an example of the computation, let us assume that five bioclades started at 3,600 Myr B.P. and that subsequent lineages had an average duration of 100 Myr, approximately equivalent to an extinction rate,  $q$ , of 0.01. Let us assume further that  $p$  is enough larger than  $q$  so that the expected species diversity in the latest Precambrian period was 1,000; this calls for a  $p$  value of 0.01177 because

$$S_t = S_0 e^{(p-q)t}, \quad [3]$$

where  $S_0$  is the number of lineages at time = 0, and  $S_t$  is the number at time =  $t$ , or

$$(p - q) = \frac{\ln S_t - \ln S_0}{t} \quad [4]$$

and

$$p = (p - q) + q. \quad [5]$$

Solving Eq. 2 with these constants yields  $M_t = 0.70$ . This means that the probability is 0.70 that the 1,000 species alive at 600 Myr B.P. descended from one of the five founders of life and, thus, all belong to a single bioclade. But it also indicates a statistical expectation that 30% of the latest Precambrian species belonged to other bioclades.

### DIVERSIFICATION AND SURVIVAL AMONG PROKARYOTES

In truly primitive clades, we are dealing with preprokaryotic grades of organization wherein propagation was entirely asexual and species lineages did not exist in the same sense that they do in most eukaryotes. Thus, we cannot bring any direct evidence to bear on speciation and extinction rates of the earliest organisms. Nevertheless, we know which parameters are important, and this can form some general basis for further consideration of the problem.

Speciation as understood today involves isolation of gene pools and genetic changes that usually create niches that are in some way different from those of the parent species. What limits the number of species in existence is a matter of dispute, but few would argue with the generality that the structure of the environment provides for a finite number of distinctive niches and that diversity is further scaled by the resource requirements of the organisms that happen to be present. These two factors are linked to a third, the extent of ecological access of species to

the unoccupied regions of biospace (8). Both the character of the environment and the adaptive potential of the biota are subject to change; thus, greater or less biospace may come to be occupied over time. The Phanerozoic fossil record indicates an overall expansion of occupied biospace (9, 10). The vastly more spotty Precambrian record of prokaryotes is at least consistent with such a pattern (1).

Prokaryotes are also asexual and lack gene pools, so we must use the notion of distinctive niches to characterize prokaryotic diversity. Prokaryotic niche diversity is based chiefly on biochemical distinctiveness rather than on morphology (11). Therefore, dimensions of ecospace that define prokaryotic diversity must be those that define biochemical activities. Extinction in these forms implies that a distinctive region of biospace is vacated, at least temporarily.

It is commonly assumed that many of the earliest life forms, including preprokaryotes, were heterotrophs, feeding chiefly upon organic molecules that had collected through prebiotic synthesis (2, 3, 12, 13). Although a surprising number of compounds can be created in prebiotic syntheses (14, 15), it is unlikely that they ever reached the rich variety required to support a biota that was diverse by modern standards. Some primitive forms may have used the metabolites of other forms as energy sources, adding to diversity, but this would fall far short of the contribution to diversity of the multilevel trophic webs of the Phanerozoic. Chemautotrophs also may have been present early, and a modest variety of chemautotrophic prokaryotes exists today, metabolizing substances such as methane and sulfides (11). The number of inorganic chemical energy sources is very limited, however. It follows that trophic resource variety can hardly have been the basis for high levels of diversity among primitive life forms.

Another possible source of niche diversity is habitat partitioning. The primitive earth was presumably about as physically varied as the modern earth. However, modern anaerobic prokaryotes are notoriously hardy and generalized in their physical requirements. Furthermore, there were no complex organisms to provide specialized host environments. Therefore, the most likely conclusion is that neither the physical habitat nor the trophic resources provided a source of niche partitioning to support high diversities of primitive life forms.

Extinction rate should be correlated with the rate of environmental change and the long ranges of some prokaryote morphologies suggest low rates. However, as noted above, morphology in these organisms is probably a poor guide to biochemistry. One can argue that the physical environment on the primitive earth was probably more volatile than today: tectonically associated processes should have been far more rapid than during Phanerozoic time, and large body impact rates were higher. Whether such an unstable situation would translate into truly high extinction rates is conjectural, however, because the primitive organisms may have been hardy with respect to physical conditions. Competition between kinds of organisms may have played a role, but there is no evidence either way.

It is clear from the foregoing discussion that existing knowledge of Precambrian diversities and turnover rates is poor. Our best guess is that diversities were low and lineages were long-lived, but hard data are lacking. Therefore, we must explore the mathematical consequences of a wide range of branching and extinction rates in order to place limits on the probability of bioclade survival.

### ANALYSIS

We have performed a number of calculations using Eq. 2. The following ranges for the constants were used.

Extinction rate,  $q$ , was varied between 0.001 and 1.0. This corresponds to a range of lineage duration from 1 to 1,000 Myr. Mean durations of species lineages in the Phanerozoic vary from 1 to 10 Myr.

Origination rate,  $p$ , was varied over a range sufficient to produce latest Precambrian diversities from 100 to 10,000 lineages. A species diversity for the latest Precambrian (600 Myr B.P.) of from 1,000 to 5,000 is most reasonable (16).

Time,  $t$ , was taken as 3,000 Myr, as in the sample calculation presented earlier. The number of bioclades,  $D$ , founded independently by single lineages at 3,600 Myr B.P. was varied from 1 to 100. That is, we considered up to 100 separate origins of life.

$M_t$ , the probability of monophyly, was computed for a matrix of values of  $p$ ,  $q$ , and  $D$  chosen from the ranges given above. Initial computations indicated, however, that any increase in  $p$ , other things being equal, lowers  $M_t$ . Because we are interested in a conservative estimate of  $M_t$ , we need be concerned only with the highest reasonable estimate of  $p$ . This value is the one that predicts a late Precambrian diversity of 10,000 species. With this held constant, the results of varying the extinction rate  $q$  and the number of independent origins of life  $D$  are given in Table 1.

The highest numbers of Table 1 indicate conditions under which multiple origins of life would not be recognized. As one moves down and to the right in the table, the probability of recognizing polyphyly increases, so that the probability is overwhelming in the lower right corner:  $M_t \approx 0$ , where lineage duration is 1,000 Myr and there were 100 bioclades at the start.

A diagonal across the table separates a region where latest Precambrian life would appear monophyletic from a region where polyphyly, if present, would likely be recognized. We can conclude that if Precambrian lineage durations were <50 Myr, there could have been several to many separate origins of life without our knowing it!

We consider this conclusion to be conservative for several reasons.

(i) The value used for latest Precambrian diversity is conservative in that it is higher than any existing estimates from the fossil record and, as noted above, the higher the latest Precambrian diversity figure, the lower the computed value of  $M_t$ .

(ii) A mean lineage duration for Precambrian organisms of 50 Myr is about as high as can be conceived in the present state of knowledge.

(iii) The mostly likely departure from the time-homogeneous model is that caused by mass extinction. Mass extinction increases the likelihood of apparent monophyly by eliminating

Table 1. The probability,  $M_t$ , that all latest Precambrian species belong to a single bioclade, assuming survival of life\*

Bioclades, starting no.	Lineage duration (1/q), Myr							
	1	10	20	30	40	50	100	1,000
2	1	0.99	0.97	0.96	0.95	0.93	0.88	0.41
3	1	0.97	0.95	0.92	0.90	0.88	0.77	0.16
4	1	0.96	0.93	0.89	0.86	0.82	0.68	0.06
5	0.99	0.95	0.90	0.86	0.82	0.78	0.61	0.02
6	0.99	0.94	0.88	0.83	0.78	0.73	0.54	0.01
7	0.99	0.93	0.86	0.80	0.74	0.69	0.48	0
8	0.99	0.92	0.85	0.78	0.71	0.65	0.42	0
9	0.99	0.91	0.83	0.75	0.68	0.62	0.37	0
10	0.99	0.90	0.81	0.73	0.65	0.58	0.33	0
20	0.98	0.82	0.66	0.53	0.42	0.33	0.10	0
50	0.96	0.63	0.38	0.22	0.12	0.06	0	0
100	0.93	0.42	0.15	0.05	0.02	0	0	0

\* Latest Precambrian diversity is held constant at 10,000.

Table 2. Probability that life will survive to the latest Precambrian eon\*

Bioclades, starting no.	Lineage duration (1/q), Myr							
	1	10	20	30	40	50	100	1,000
1	0.003	0.03	0.06	0.08	0.11	0.13	0.23	0.75
2	0.01	0.06	0.11	0.16	0.21	0.25	0.41	0.94
3	0.01	0.09	0.16	0.23	0.29	0.35	0.55	0.99
4	0.01	0.11	0.21	0.30	0.37	0.44	0.66	1
5	0.02	0.14	0.26	0.36	0.44	0.51	0.74	1
6	0.02	0.17	0.30	0.41	0.50	0.58	0.80	1
7	0.02	0.19	0.34	0.46	0.56	0.63	0.85	1
8	0.02	0.21	0.38	0.51	0.60	0.68	0.88	1
9	0.03	0.24	0.42	0.55	0.65	0.72	0.91	1
10	0.03	0.26	0.45	0.59	0.69	0.76	0.93	1
20	0.06	0.45	0.70	0.83	0.90	0.94	1	1
50	0.14	0.78	0.95	0.99	1	1	1	1
100	0.26	0.95	1	1	1	1	1	1

\* It is assumed that  $p$  is that value which would predict a final diversity of 10,000 species.

bioclades. Because the model used here ignores mass extinction, the values of  $M_t$  in Table 1 are probably conservative—that is, erring on the low side.

(iv) The model does not allow for competitive exclusion of one bioclade by another. Any such result of competition would decrease the chances of additional bioclades surviving to the latest Precambrian. Whereas we do not know whether competition was a significant element, our use of a model which excludes competition is the conservative position in that it lowers the computed value of  $M_t$ .

A cognate question can be asked: Under the conditions of Table 1, what is the probability that life itself (at least one bioclade) would have survived? We can approach this from Eq. 1 for  $P_{s,t}$ , the probability of survival of a clade that is founded by a single lineage. Consider the case where lineage duration is 1 Myr, with  $t = 3,000$  Myr and  $p$  adjusted to predict a latest Precambrian diversity of 10,000. This is the case corresponding to the first column in Table 1.  $P_{s,t}$  for these conditions is 0.003, which suggests that a bioclade would have had a very small chance of surviving for 3,000 Myr. As the number of separate origins of life goes up, however, so does the probability that at least one will survive. The binomial probabilities for this survival are given in Table 2.

Table 2 is somewhat difficult to interpret. Life did survive in the only biologic system known. This event may have had a very low probability (0.003, for example) or a high one. We have no way of investigating the question. Nevertheless, it seems reasonable to postulate that survival of life on earth was not a fluke. With this in mind, moving down and to the right in Table 2 makes life's survival more credible. But to do so is to move into the part of Table 1 where polyphyly should be present and recognized. It is a premise of this analysis that all known life is monophyletic. Therefore, we conclude that reality may be somewhere near the middle of Tables 1 and 2, with mean lineage durations between 20 and 50 Myr and with multiple origins of life. It is in this region of the tables where survival of life has a substantial probability (Table 2) but where multiple origins of life are unlikely to be recognized (Table 1).

## DISCUSSION

Although our analysis is less than rigorous, owing to the paucity of data on evolutionary turnover among primitive organisms, we conclude that multiple origins of life in the early Precambrian is a reasonable possibility. The fact that all present-day

life appears to have descended from a single ancestor does not void the possibility of multiple origins because most such origins would have aborted as a consequence of the birth–death process at the level of lineages. With a time-homogeneous model, at least 10 extinct bioclades could be “hidden” in the Precambrian if mean lineage duration was less than about 50 Myr. The possible number of extinct bioclades would be increased by most departures from the homogeneous model or by competition between bioclades, or by both.

From the foregoing, one can speculate that bioclades with far more potential than our own may have been extinguished in the Precambrian by ill chance.

It is also possible that our own bioclade was actually superior to any contemporary bioclade(s) and survived for that reason. What is most unlikely is that our bioclade is the best of all possible bioclades.

We thank Stephen M. Stigler for mathematical discussions. Research on which this article is based was funded in part by National Aeronautics and Space Administration Grant NAS 2-73 to J. W. V., who also owes thanks to the Department of Geology, Field Museum of Natural History, (Chicago, IL) where this was prosecuted under the Visiting Scientist Program.

1. Awramik, S. M. (1981) in *Ecological Crises in Ecologic and Evolutionary Time*, ed. Nitecki, M. H. (Academic, New York), pp. 83–102.
2. Cairns-Smith, A. G. (1971) *The Life Puzzle* (Oliver & Boyd, Edinburgh, Scotland).
3. Oparin, A. I. (1938) *The Origin of Life* (Dover, New York), 2nd Ed. (1st Ed. in Russian, 1924).
4. Pirie, N. W. (1959) in *The Origin of Life on Earth*, eds. Clark, F. & Synge, R. C. M. (Pergamon, London), pp. 76–83.
5. Raup, D. M. (1978) *Paleobiology* 4, 1–15.
6. Raup, D. M. (1981) *Acta Geol. Hisp.* 16, 25–33.
7. Bailey, N. J. T. (1964) *The Elements of Stochastic Processes, With Applications to the Natural Sciences* (Wiley, New York), Eq. 8.46.
8. Simpson, G. G. (1953) *The Major Features of Evolution* (Columbia Univ. Press, New York).
9. Bambach, R. K. in *Biotic Interactions in Recent and Fossil Benthic Communities*, eds. Tevesz, M. J. & McCall, P. L. (Plenum, New York), in press.
10. Bambach, R. K. in *Phanerozoic Diversity Patterns: Profiles in Macroevolution*, ed. Valentine, J. W. (Princeton Univ. Press, Princeton, NJ), in press.
11. Kluiver, A. J. & van Niel, C. B. (1956) *The Microbe's Contribution to Biology* (Harvard Univ. Press, Cambridge, MA).
12. Haldane, J. B. S. (1929) *Rationalist Annual*, Vol. 3.
13. Orgel, L. E. (1973) *The Origin of Life: Molecules and Natural Selection* (Wiley, New York).
14. Fox, S. W. & Dose, K. (1972) *Molecular Evolution and the Origin of Life* (Freeman, San Francisco).
15. Miller, S. J. (1982) in *Mineral Deposits and the Evolution of the Biosphere*, eds. Holland, H. D. & Schidlowski, M. (Springer, Berlin).
16. Valentine, J. W., Foin, T. C. & Peart, D. (1978) *Paleobiology* 4, 55–66.