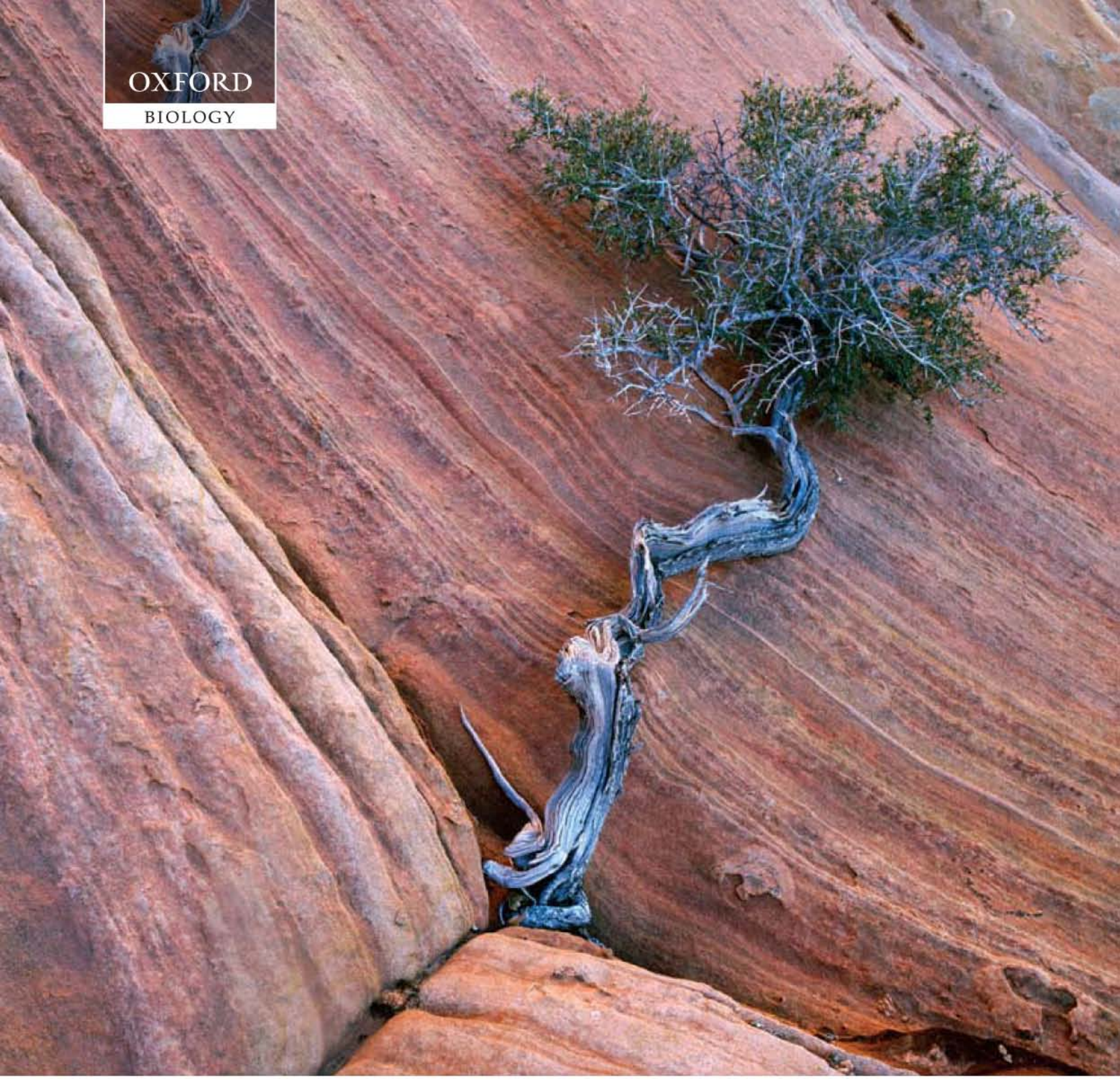

OXFORD
BIOLOGY

the TIMETREE *of* LIFE

edited by **S. BLAIR HEDGES** *and* **SUDHIR KUMAR**

foreword by James D. Watson

Discovering the Timetree of Life

S. Blair Hedges[a,*] *and Sudhir Kumar*[b]
[a]Department of Biology, 208 Mueller Laboratory, Pennsylvania State University, University Park, PA 16802-5301, USA; [b]Center for Evolutionary Functional Genomics, The Biodesign Institute, Arizona State University, Tempe, AZ 85287-5301, USA
*To whom correspondence should be addressed (sbh1@psu.edu)

Abstract

The two primary components of evolutionary history are the relationships of organisms (phylogeny) and their times of divergence. Together they form a timetree: a phylogenetic tree scaled to time. The fossil record initially provided the timescale, but this has been supplemented in recent years with the application of molecular clocks. The Timetree of Life is now being discovered, largely through phylogenetic and chronological analyses of DNA and protein sequences. The addition of a temporal dimension to the tree of life is driving major advances in evolutionary biology, providing a better understanding of the mechanisms of evolution, and revealing the reciprocal interactions between life and the environment throughout Earth history.

The evolutionary tree of life describes how species are related and organized in the greatest of biological hierarchies. In a letter to Huxley, 2 years before publication of *Origin of Species* (*1*), Charles Darwin predicted:

> The time will come, I believe, though I shall not live to see it, when we shall have fairly true genealogical trees of each great kingdom of Nature.

With great delight, we can say that the time *has* come, and that it is now. Certainly, many important details remain to be worked out, such as deep branching patterns among major taxonomic groups and the interrelationships of many species, but much of the tree of life already has taken shape (*2*). This revolution in evolution has occurred largely through advances in molecular biology over the last half century, building on a foundation laid by paleontology and comparative biology. It would not have been possible without many discoveries, progressively building on previous work, such as the structure of DNA (*3*), methods to sequence proteins and DNA (*4–6*), a technique to rapidly amplify DNA (*7*), and advances in statistical methods of data analysis. Some—perhaps most—of the resulting molecular phylogenies have corroborated trees based on morphology and cell biology, but many findings were unexpected including the discoveries of archaebacteria (*8*) and an African clade of mammals (*9*) to name just two. Our current understanding of the tree of life draws from fossils, morphology, and—especially in the last two decades—many molecular phylogenies.

However, a phylogenetic tree provides only half of the picture. Evolutionary history has two primary components—relationships and timescale—and both are important. Together, they form something that does not have a specific name; hence we use the word "timetree" for any tree scaled to time. It is preferred over the more general term "chronogram" which does not indicate that a tree is involved, or "phylogeny." (Phylogeny is the specific branching order—relationships—without the temporal component.) In past decades, the two words have been used separately ("time tree"), although rarely, with the compound form appearing only a few times in recent years (e.g., *10, 11*). The word "phylochronology" has been applied recently to the study of populations through time using ancient DNA methods (*12*), but it is just as applicable to the study of timetrees in general. The dimension of time provides a direct connection with other fields of science, and the ability to relate biological evolution with climate change and Earth history in general. A timetree of life offers a more complete view of the framework of evolutionary history than the tree of life alone, and the utility of this perspective is broadly recognized (*13*).

Timetrees were not invented with molecular data. In fact, Darwin's only figure in the *Origin of Species* (*1*) is essentially a timetree—a hypothetical one—scaled to generations rather than years. Subsequently, paleontologists (e.g., *14*) were the major producers of timetrees, because fossils initially provided the only information to establish the evolutionary timescale. This changed in the early 1960s. Enough protein sequence information became available to show that molecular change is more predictable and quantifiable than morphological change (*15*), an observation that has come to be known

S. B. Hedges and S. Kumar. Discovering the Timetree of Life. Pp. 3–18 in *The Timetree of Life*, S. B. Hedges and S. Kumar, Eds. (Oxford University Press, 2009).

as the "molecular clock" (*16*). Subsequently, studies of viral DNA evolution, where dates of divergence are known, confirmed that time-dependent change occurs at the molecular level (*17*). The basic principle of molecular clocks is that there is a strong enough correlation between molecular change and geologic time, such that the rate can be used to measure time in parts of the tree lacking a calibration time. Correlations between fossil and molecular times are often strong (*18*), but large differences have been encountered as well (*19–21*), leading to intense scrutiny of both types of data and the generation of new hypotheses to explain the difference.

The theoretical basis of molecular clocks is a separate issue from the application of clocks for the same reason that early civilizations counted days—quite accurately—by watching the sun move across the sky without knowing how that process occurred. From the beginning, it has been suggested that the explanation for the molecular clock—in part or in whole—lies in the selective neutrality of mutations and substitutions (*15, 16, 22*). Because mutations occur randomly, this would mean that molecular clocks are stochastic clocks, in the same class as geologic clocks that rely on random isotopic decays. This is in contrast to regular clocks, such as those based on pendulums or atomic resonance. While it is likely that molecular clocks are stochastic clocks, it is less certain whether they are driven by neutral mutations. Explanations involving selection have been proposed as well (*23–25*) and no consensus has been reached as to the basis for the clock-like change in molecular data (*26*). But lineage-specific variation should not be confused with the wide variation in rates of change among genes and proteins, which is the result of selection on function (*27, 28*). Moreover gene-specific variation based on selection is fully compatible with the hypothesis of neutrality and stochasticity in any given gene or protein. While the theoretical underpinning of molecular clocks and the dynamics of rate variation continue to be studied (*26*), the field has firmly entered an empirical phase. Divergence times are being estimated in many groups of organisms and this effort is growing at a rapid pace, driven by the greater accessibility of DNA sequence data (Fig. 1).

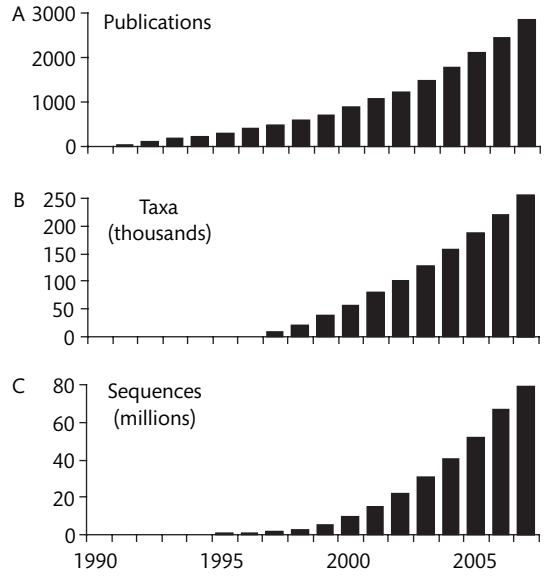

Fig. 1 The rapid expansion of published molecular time estimates. (A) Number of research publications that have discussed molecular clocks or molecular time estimates. Data are from the *Web of Science* online resource, with results checked for appropriateness to molecular evolutionary clocks. The search included only titles, abstracts, and keywords and therefore represented only a fraction of the total number of articles, although not all identified articles present new time data. (B) Thousands of taxa in Genbank (NCBI). (C) Millions of DNA sequences in Genbank during the same period. All counts are cumulative.

Advances in methodology

Methodology for molecular time estimation has developed considerably since the 1960s, in parallel with improved methods of data acquisition. Initially, most researchers estimating divergence times used distance data from immunological, allozyme, and DNA–DNA hybridization methods (e.g., *20, 29*) until DNA sequencing became more accessible (*30*). Relative rate tests were used almost from the beginning to test the significance of clock-like behavior and to adjust time estimates based on lineage-specific rate differences (*31–33*). Methods for accommodating lineage-specific rate variation in a molecular clock analysis continued to be developed (*33, 34*) and were useful in some focused studies (e.g., *35*) but were not mature for routine analyses. Instead, most molecular clock studies before the year 2000 took an approach that used statistical tests to identify and exclude genes and lineages that violated rate constancy before estimating time (*36, 37*). In the last decade, further improvements were made in methods for accommodating rate variation among lineages (*38–42*). These methods have reduced the need to exclude species or genes that show rate variation, facilitating molecular clock analysis. As a consequence, most molecular clock studies now use these "relaxed clock" methods.

Any device for timing requires calibration, and this is perhaps the most critical aspect of molecular clocks and the one most often debated. If timing is done with a gene deemed to be evolving at a constant rate, only a single calibration is needed, by definition—the rate established at any one point in the tree can be propagated to all points in the tree. For this reason, emphasis was often placed on the quality of the calibration (*18*), whether it was from a single fossil or an average rate from multiple fossils. The closer the calibration is to the true evolutionary divergence, the more accurate the resulting time estimates. Relaxed clock methods can work with a single calibration point, but they are expected to do better with multiple calibrations because modeling of rate of evolution across a tree can only be done reliably with multiple reference points. It is better to use as many good calibration points as possible, but only one calibration is required to estimate time. The number of calibrations used in any study is dictated by practical considerations—what is available—and, most importantly, by the quality and distribution of calibrations (*37*).

For groups lacking calibrations and showing rate constancy in a gene (or genes), there is nothing wrong with using a rate established in another group of organisms as long as the assumptions are explicitly stated. This is an approach used frequently in the past for mitochondrial DNA analyses (e.g., *43, 44*), as it provided valuable information on the timing of events otherwise unavailable. This flexibility is necessary because calibration points of any kind are absent from a vast majority of nodes in the Tree of Life, requiring rates of change to be extrapolated from one group or node having calibrations to another lacking calibrations, if divergence times are to be estimated at all. In fact, rate extrapolation—between one part of a tree and another—has to be done in every molecular clock study, regardless of the number of calibration points.

As mentioned earlier, the best calibration point is one that is closest to the true evolutionary divergence. Such calibrations are rare and usually restricted to divergences caused by the separation of land (e.g., separation of two continents) for terrestrial organisms, or water for aquatic organisms, and that have associated geologic dates. Fossil calibrations are always minimum times of divergence and therefore will result in time estimates that are minimums as well, more recent than the true divergence. A major change in the use of calibrations has come with development of relaxed clock methods that permit both maximum and minimum calibrations in estimating divergence times. If both types of calibrations are used in a study, the resulting times can be considered as estimates of the true divergence time. The difficulty comes in identifying valid maximum times of divergence for calibration. The age of the Earth (4600 million years ago, Ma) is a global maximum for all nodes in the Tree of Life, but it is too old for most purposes. The best maximum calibrations are those that involve emergence of land areas (e.g., islands), providing a maximum time for diversification of terrestrial organisms restricted to that area (*45*). Establishing a maximum in the fossil record is more difficult. In 1996, we proposed a method for establishing a maximum calibration using transitional fossils, in that case involving the transition from fishes to tetrapods ~380–360 Ma (*19*). A series of fossils documents the morphological transition from lobe-finned fishes to stem tetrapods (*46, 47*), thus constraining the maximum time of any divergence within tetrapods, such as the split between mammals and birds. However, fossils recording such evolutionary transitions are rarely available.

Another method of establishing a maximum calibration in a relaxed clock analysis is to use the age of the earliest fossil evidence for a lineage (*48, 49*). This approach is problematic. Its use is tantamount to interpreting the fossil record as the true record of evolutionary history, which is *guaranteed* to underestimate the true time of divergence. A related method uses the age of the oldest fossil of the most closely related group to the node in question as the maximum age for the node (*50, 51*). This would often result in a narrow interval between the minimum and maximum time of divergence of two lineages, and its use would force relaxed clock methods to produce estimates that are not older than the maximum calibration. Consequently, the power of sophisticated statistical methods would fail to be realized in molecular clock analysis and their outcomes will effectively not be different from reading the fossil record as a literal interpretation of the Timetree of Life (or, at the least, biased by this method). In general, the oldest fossil of a lineage can only establish a minimum for that lineage, not a maximum for another lineage (*52, 53*).

Another method of determining the maximum constraint of a node from the fossil record again considers the earliest fossils of closest relatives, but it instead places emphasis on the absence of fossils of the clade in question from earlier deposits that otherwise should contain representatives (*53*). Such "soft maximum" constraints (*54*) have been proposed, along with minimum calibration constraints, for a moderate number of divergences among animals (*53, 55, 56*). However, soft maximum

calibration points tied to assumptions regarding the absence of earlier fossils may cause considerable underestimation of divergence time in the same way as mentioned earlier. A case in point concerns the continuing debate over the Cambrian Explosion model, which suggests that most animal phyla originated (diverged from one another) during a relatively short period of time in the latest Precambrian and early Cambrian, ~550 Ma (*57–59*). Instead, most studies using molecular clocks have found that divergences among phyla occurred hundreds of millions of years before their appearance in the fossil record (reviewed in *60–62*); those studies claiming younger time estimates (*63–68*) have been shown to be flawed (*60, 62, 69–71*). If the soft maximum constraint of 581 Ma (*53*) is used in relaxed clock studies, it will cause the inferred time estimates to be close to fossil time estimates. In other words, if an investigator believes that the maximum date for protostome deuterostome divergence is close to 581 Ma, then the test of the Cambrian explosion hypothesis by molecular data is a pointless exercise.

In addition to minimum constraints, we have proposed that probability distributions describing the likelihood of the true divergence time for a calibration node be considered in molecular clock analyses (*18, 19, 37*). For example, we have described how the divergence of birds and mammals is associated with a higher probability closer to the minimum (310 Ma) than to the maximum (370 Ma), which leads to a distribution with a similar probability rather than one with a central tendency (e.g., a normal distribution) or uniform distribution (*19*). Based on the birth–death model of diversification, a long-tailed (logistic) distribution for calibrations has been favored (*53*), although other distributions are possible (*37*). In fact, a distribution can be envisioned in which the highest density (likelihood) is near the oldest time when only the maximum constraint is known, and in which the tail of the distribution is toward the younger time. Some of these probability distributions of calibrations have already been used while estimating the confidence intervals of divergence times (*72*) and in Bayesian relaxed clock analyses (*41*).

While probability distributions for calibrations may be useful, they should also be used with caution. For example, the use of nonuniform distributions (e.g., logistic, exponential) may agree with some patterns of diversification (*53*), but they may not be the best model for all divergences in a data set—and hence may impose a bias—if a nonbiological process affecting fossilization is responsible. For example, oxygen levels in the atmosphere rose sharply in the late Precambrian, 580–542 Ma, from 1% to 10% of the present level to nearly 100% of the present level (*73*). Before 580 Ma, it is unlikely that there would have been sufficient oxygen for large, hard-bodied animals. Because the probability of fossilization apparently changed dramatically at 580 Ma, the logistic distribution would no longer be appropriate. Instead, a uniform calibration probability distribution might be more appropriate. Considering this, and the fact that we have a poor understanding of gaps in the fossil record in general (*74*), a uniform distribution (or, the assignment of a maximum or minimum calibration point without a distribution) may be the least biased approach for most or all nodes unless there is justification for using a particular distribution. Clearly, this is one area of research in great need of attention from both modelers of fossil preservation potential (*74–77*) and developers of relaxed clock methods (*38, 39, 41, 42, 54*).

The current state of knowledge

The Timetree of Life (*78*) summarizes the current state of knowledge on the Timetree of Life, with some caveats. The first is that it includes only living organisms, sampled by molecular methods (the molecular record). A synthesis of living and fossil groups is an important future goal (see later). Secondly, there is unequal coverage among the kingdoms and phyla, directly related to the limited availability of molecular data for certain groups of organisms. Thus, plants, cartilaginous fishes, amphibians, reptiles, and mammals are relatively well covered, whereas protists, fungi, invertebrates, and ray-finned fishes are poorly represented. Coverage of prokaryotes is moderate: roughly two-thirds of the families of archaebacteria and one-third of the families of eubacteria are included in these timetrees. Thirdly, the book covers only the families and groups above that taxonomic level. To include genera and species would have been impossible in a single volume, although such expanded coverage is planned for the future in a similar format, and is available elsewhere in an online database (*11*). Another caveat is that taxonomic ranks above the species level are partly arbitrary with regards to the temporal depth of included divergences, resulting in disproportionate coverage of evolutionary lineages among the groups (*79*). Lastly, while it is an authority-based synthesis, opinions often differ among experts and therefore some biases will necessarily be present in how the data and conclusions are presented. Despite these caveats, the book chapters are windows into

the literature of an area and mostly (if not all) all available divergence times relevant to the scope of the chapter. New divergence time data or analyses are included in 13 chapters: Archaebacteria, Eubacteria, Eukaryota, Monocots, Metazoa, Scaphopoda, Aculeata, Coleoptera, Chondrichthyes, Dipnoi, Serpentes, Crocodylia, and Piciformes.

All individual timetrees in *The Timetree of Life* were assembled here into a single timetree (Fig. 2). It includes all three superkingdoms and 1610 families (or family-level

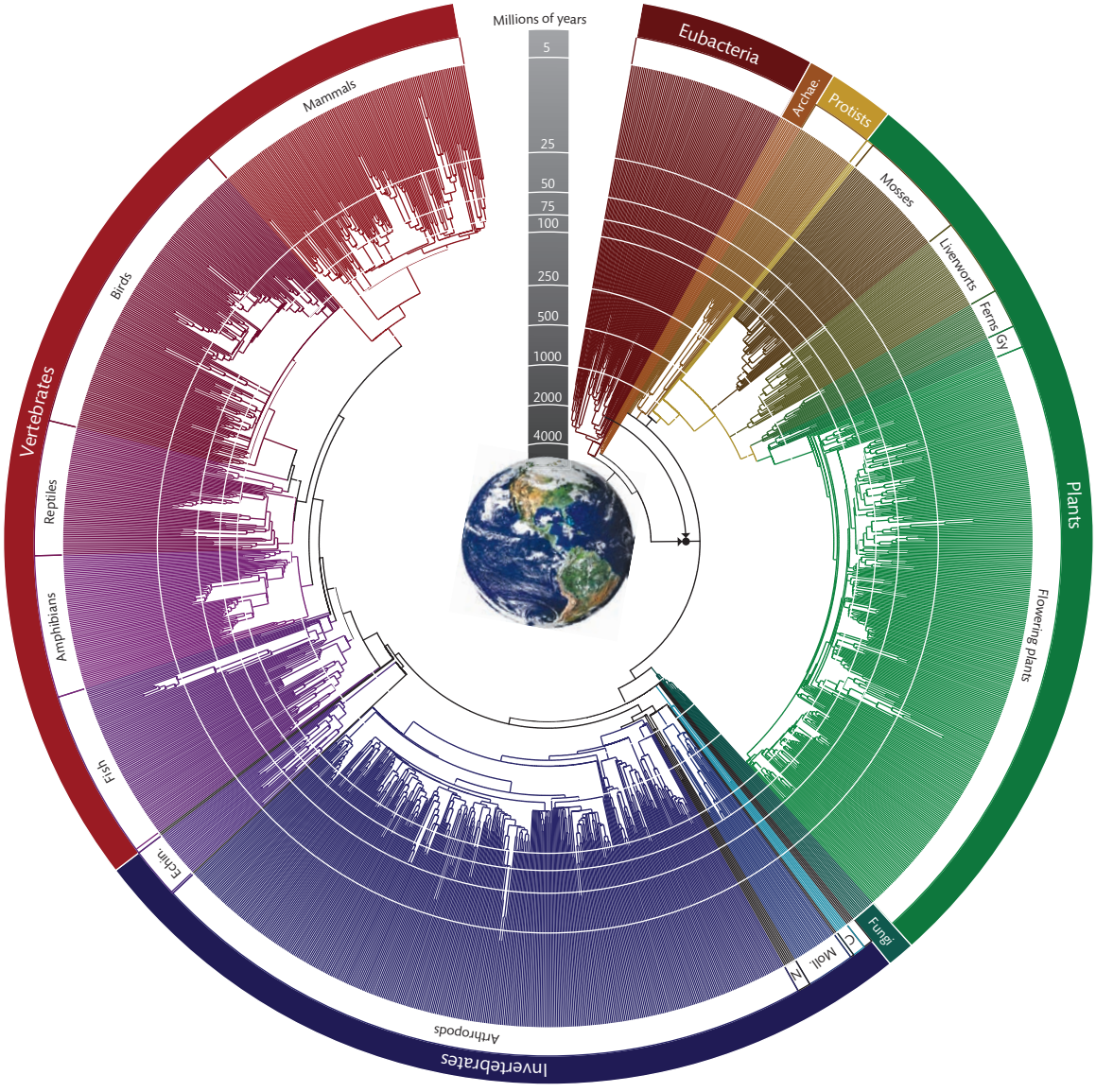

Fig. 2 The Timetree of Life from an assembly of individual timetrees. Each of the 1610 terminal branches represents a family or family-level taxon, although only major clades are labeled. This global timetree was assembled from all of the timetrees in *The Timetree of Life* (*78*), with very little additional manipulation or editing. For three nodes lacking molecular time estimates (Scaphopoda/Cephalopoda, Decapodiformes/Octapodiformes, and Anabantoidei/Notothenoidei), the ages of the earliest pertinent fossils were considered in setting the time. Also, Nematoda was joined with Priapulida and Arthropoda, and set to the same time of divergence. A total of 36 nodes in 13 chapters (0.2% of all nodes) were slightly older than parental nodes in other chapters and were adjusted (average, 6.7%) to eliminate the conflict. *Abbreviations*: C (Cnidaria), Gy (Gymnosperms), Echin. (Echinodermata), and Moll. (Mollusks).

taxa). It is unlikely that a complete timetree of all life will be available in the foreseeable future (or ever) given the uncertainties in some parts of the tree, problems of horizontal gene transfer, incompleteness of the fossil record, ongoing extinctions through human impact, and inability to collect and sample—with molecular methods—all living species for practical reasons. Nonetheless, this first major synthesis could be called the Timetree of Life, while acknowledging that it is a work in progress by many people, and that it is not too far from the true version. The next decade or two will see the Timetree of Life come into much better focus, but it is unlikely there will be any one day where the community can proclaim that the Tree of Life or Timetree of Life has been achieved. In the future, history will probably record the discovery of the Timetree of Life as having happened during a short period—perhaps in the first two or three decades of the twenty-first century.

Families through time

This synthesis and analysis of the timetrees and divergence times in *Timetree of Life* provides the opportunity to make broad comparisons across all of life and reveals some new patterns. One useful comparison is the average age of a family lineage (the elapsed time since the divergence with its closest relative). Taxonomic ranks are often used in comparative studies, even though their biological meaning is unclear (*80*). Therefore, it is useful to know how family lineages differ in age among groups. Sparse coverage of families will bias lineages towards older ages, and therefore some caution must be used in evaluating the results (Table 1), especially in the poorly sampled groups such as invertebrates and ray-finned fishes. There is great disparity among groups, as is already known from fossil evidence (*81*). However, molecular time estimates should provide a better quantification of the difference because they correspond to lineage originations rather than the earliest occurrence as recorded by fossils.

Typical families within the three superkingdoms differ greatly in age (Table 1). Families of Archaebacteria and Eubacteria, on average, are 25 and 14 times, respectively, as old as those of eukaryotes. However, plant and animal families are nearly identical in age, on average (~100 Myr old), despite their long and separate histories of taxonomic practice. Among animals, the average age of an invertebrate family (144 Myr) is about twice that of a vertebrate family (69.7 Myr). Within vertebrates, the

Table 1. Ages of family lineages of organisms as measured by molecular clocks, based on an analysis of data in *The Timetree of Life* (*78*).

Group	Average age (Myr)	No. families	SE
Archaebacteria	2567	13	300
Eubacteria	1388	89	80.6
Eukaryotes	102.4	1378	1.9
Land plants	103.1	463	3.1
Mosses	156.3	59	10.1
Liverworts	167.3	42	12.5
Ferns	209.9	21	18.9
Gymnosperms	229.2	12	19.4
Flowering plants	73.9	329	1.5
Eudicots	69.3	235	1.8
Monocots	85.4	77	2.2
Animals	102.1	915	2.3
Invertebrates	143.7	401	3.2
Mollusks	183.9	26	20.2
Nematodes	249.8	5	34.3
Spiders	158.6	26	10.3
Bees, ants, & stingless wasps	117.1	22	6.2
Beetles	127.6	183	3.4
Lacewings	206.5	17	7.3
True flies	142.9	51	7.5
Crustaceans	185.1	44	11.6
Sea urchins	95.9	27	13.2
Vertebrates	69.7	514	2.5
Jawless fishes	482.3	2	0.0
Cartilaginous fishes	143.0	57	7.8
Ray-finned fishes	53.4	15	13.7
Lungfishes	172.3	3	52.3
Frogs and toads	92.2	59	6.0
Salamanders	146.6	10	8.8
Caecilians	118.9	7	25.0
Turtles	91.3	14	11.3
Lizards, snakes, and amphisbaenians	84.2	54	5.6
Crocodilians	76.7	3	12.9
Birds	42.8	149	1.7
Mammals	37.1	141	1.3

Note: Age was measured as the time of divergence between that family and its closest relative.
Myr = million years and SE = standard error.

fishes show wide variation in the average age of families, ranging from 482 Myr in jawless fishes (two families) to 53 Myr in a small selection of ray-finned fishes. The non-avian reptiles (turtles, lizards, snakes, amphisbaenians, and crocodilians) have similar mean ages, dating to the late Cretaceous (91–77 Myr). The three orders of amphibians also have average ages in the Cretaceous, although a bit earlier (146–92 Myr). Families of birds and mammals typically are younger (Cenozoic) and are similar in average age (43 and 37 Myr, respectively).

Fossils and molecules

A comparison between 46 fossil minimum divergence times (*55*) and the corresponding molecular times (Fig. 3) shows a high correlation ($r = 0.96$), as has been observed previously in broad surveys (*18, 36*). Only two fossil dates were older than molecular dates: Ochotonidae vs. Leporidae and Aves vs. Crocodylia. Both are controversial and have been discussed in the literature (*82, 83*). In the other comparisons, fossil times are 27.5% (4–57%) younger than molecular times on average, with the largest differences involving the earliest animal divergences (the Cambrian Explosion).

A maximum calibration (e.g., from the fossil record) should be older than the (unknown) true divergence time. Likewise, if the molecular divergence time is an unbiased estimate of the true divergence time, and its variance is not excessive, it should be younger than maximum calibrations in a majority of cases. However, while the correlation coefficient of molecular vs. soft maximum times was high (0.96), nearly half (48%) of the soft maximum divergence times (*55*) were younger than the corresponding molecular time estimates (Fig. 2). They ranged from 99% older to 44% younger. This indicates that either (i) the molecular time estimates are overestimates of the true divergence, (ii) these soft maximum times are underestimates of the true divergence, or (iii) the variance in the molecular estimates is so large that a large fraction of estimates—by chance—are older than the fossil maximum estimates. The results of the diversification analyses in the next section give some independent support for the accuracy of molecular time estimates, which would suggest that the soft maximum times (*55*) are underestimates of the true divergence.

The diversification of life

If we were to know the total number of species that existed at all times in the past, we would have a complete view of the rate of evolutionary diversification (speciation) through time. This information would be valuable

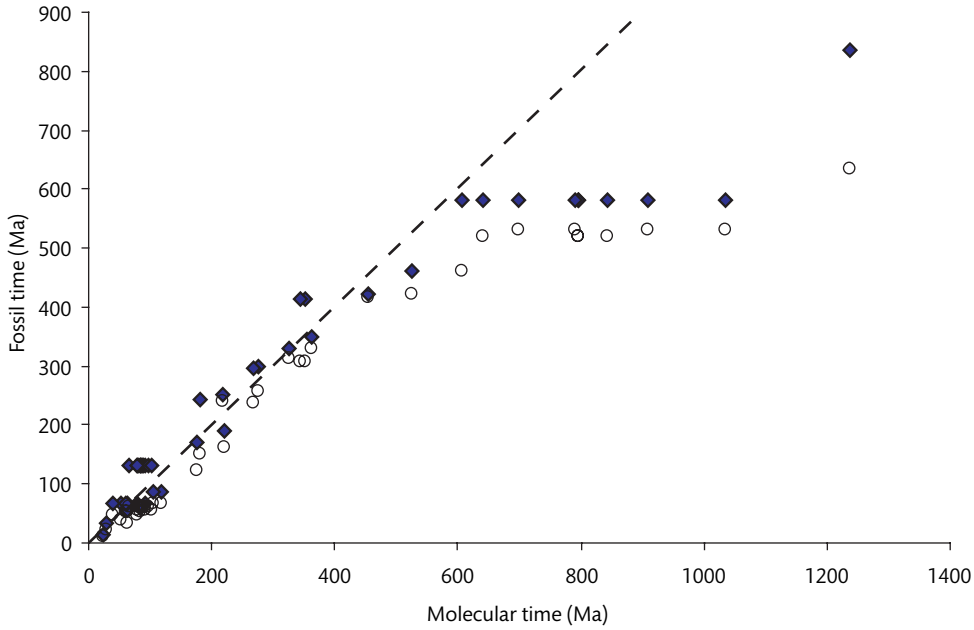

Fig. 3 Relationship between molecular time estimates (*x*-axis), based on data from *The Timetree of Life* (*78*) and fossil time estimates (*y*-axis) (*55*). Open circles = fossil minimum times; closed diamonds = fossil soft maximum times; dashed line = 1:1 relationship; Ma = million years ago.

for understanding not only how diversification in some groups (e.g., predators) was affected by diversification in other groups (e.g., prey), but also the relationship between biological diversification and Earth history. Syntheses of the fossil record, usually focusing on genera or families, already have alluded to some of these patterns (*75, 84*), although evidence from molecular clocks has not always reproduced them (*19, 21*), leading to debates over evolutionary mechanisms. But the two estimates are not expected to agree all the time. Origination events in the fossil record measure the first occurrence of a taxon, recognizable by morphology, which must postdate the phylogenetic divergence of that lineage and its closest relative, measured by molecular clocks. In theory, the time difference between the phylogenetic divergence and the fossil origination could be slight (~1%), but that would require an extraordinarily complete fossil record and the evolution of diagnostic morphological characters over a very short period of time. In practice, however, a substantial difference between the fossil and molecular time estimates is expected, and such a difference (28%, average) was found, as discussed in the previous section. Also, at least some of the pattern of diversification recorded in the fossil record can be attributed to known biases such as peaks in originations ascribed to rare sites of fossilization (*75, 85*).

Despite the significant taxonomic gaps, the data assembled in *The Timetree of Life* provide an opportunity, for the first time, to compare global, fossil-based patterns of diversification (*75, 84, 86, 87*) with those measured by molecular clocks. The two data sets are based on family-level (and above) taxa, with 7186 families in the fossil record database (*81*) and 1610 families in the synthesis here of molecular times. The molecular data set does not include extinct lineages whereas the fossil data set lacks many extant lineages. The fossil record database also contains many more marine and invertebrate taxa than are currently available in the molecular database. On the other hand, the early history of life is much better represented by the molecular data set. There are essentially no fossils of any groups of extant taxa—aside from cyanobacteria and a handful of uncontested eukaryote fossils (*88, 89*)—before about 600 Ma (*81*). In this sense, the molecular data provide the first glimpse of diversification patterns in the Precambrian. Also, divergences can be sampled continuously and evenly throughout the molecular timetree, whereas fossils are necessarily assigned to geologic periods which vary in length. In the analyses detailed here, we follow previous authors in presenting diversification curves as lineage originations, both cumulatively and as originations or rates binned to time intervals (*75, 84*). Because of the relatively small fraction of exclusively marine families in the molecular data set, a distinction was not made between marine and continental taxa.

The global diversification curve (Fig. 4A and B) shows a relatively smooth exponential increase (linear on a log scale) except for the last 300 Myr where the rate trajectory shifts steeply upward. (The slight plateau seen in the most recent intervals, the last ~30–40 Myr, is probably an artifact of the taxonomic scope of the data being restricted to the family level and above). This is a much different and smoother curve than has been observed in past analyses of diversification based on the fossil record. The results of those fossil analyses have been debated as to whether they conform to a dampened exponential curve (*75, 86, 87, 90, 91*) or a logistic model (*84, 92–95*). The explanation for the exponential curve is that diversification proceeds continuously at about the same rate without being limited by competition. In contrast, the second school of thought contends that competition among lineages for ecological niches causes diversification to follow a logistic curve, with an early rapid rate and a late slow rate, eventually reaching a plateau. These analyses (Fig. 4) support the first, expansionist model, and show a good fit to a standard exponential curve for most (93%) of the history of life.

The rate curve (Fig. 4C) also shows the sharp increase in origination rate in the last ~300 Myr. The rate between 4000 and 400 Ma averages 17% per 200 Myr whereas the rate for the last 200 Myr is 64%. A biological explanation for this dramatic rate increase is not immediately obvious. The timing could suggest that it is related to the colonization of land and diversification of terrestrial organisms in the late Paleozoic Era, especially in the Carboniferous and Permian (359–251 Ma). If so, it could also be associated with a major pulse in atmospheric oxygen ~340–250 Ma (*96, 97*). However, there is not a complete concordance between oxygen levels and diversification, because oxygen declined in the early Mesozoic at the same time that the lineage origination rate continued to increase.

Yet another, and perhaps, simpler explanation is that the rate spike of the last ~300 Myr is an artifact of the extinction process and the fact that only living lineages are being examined. This well known "Pull of the Present" bias results when an excess of recent lineages are sampled that will soon become extinct and removed from the surviving curve of continuous lineages. In other words, If we traveled back through time and sampled

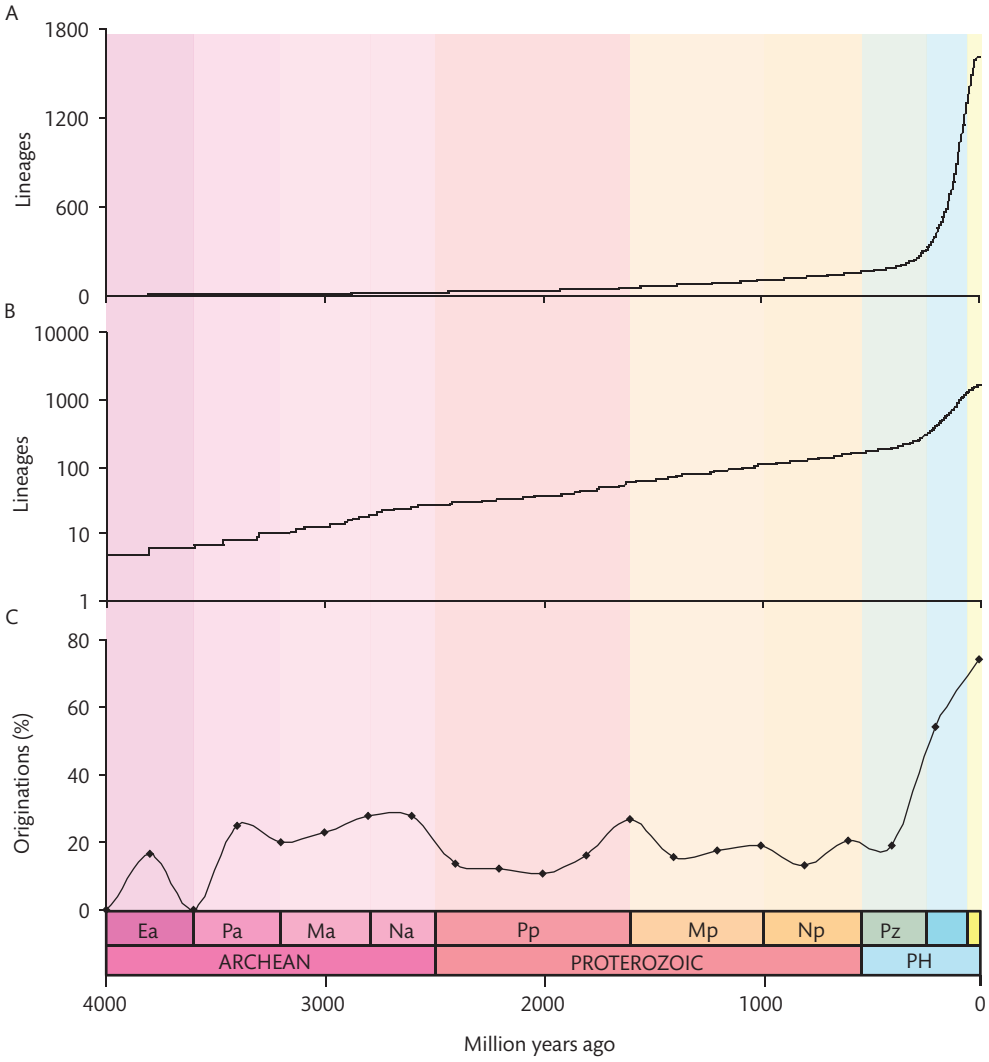

Fig. 4 Diversification curves of life based on analyses presented here. (A) Cumulative curve, showing the total number of lineages at any point in time, based on all timetrees and lineages (1610) in *The Timetree of Life* (*78*). The curve samples data in 1-Myr intervals and is not smoothed. (B) Cumulative curve as in A, in 1-Myr intervals, plotted on a log axis (also not smoothed). (C) Rate curve, showing the number of lineage originations as a percentage of standing diversity (originations/total lineages at that point in time), in 200-Myr intervals. *Abbreviations*: Ea (Eoarchean), Ma (Mesoarchean), Mp (Mesoproterozoic), Na (Neoarchean), Np (Neoproterozoic), Pa (Paleoarchean), PH (Phanerozoic), Pp (Paleoproterozoic), and Pz (Paleozoic).

lineages living at 300 Ma, we might see the same spike in diversification, but occurring in the preceding 300 Myr (600–300 Ma). This is similar—but not identical—to the "Pull of the Recent" bias in the fossil record thought to be caused by the greater number of fossil sites in more recent times but still debated vigorously (*87, 98, 99*).

Turning now to the last billion years, finer sampling intervals reveal more details in the diversification rate curve (Fig. 5A–C). Surprisingly, it shows a noticeable depression at 250 Ma, which is the time of the Permian–Triassic extinctions—the largest in the fossil record. The depression corresponds roughly to a 50% decrease in rate of lineage origination. In the finer sample of 10-Myr intervals, it is a low of 2.7% surrounded by peaks of ~7% (≅20 lineages) (Fig. 5B) while in the 20-Myr-interval plot it is a low of 6.7% bordered by peaks of 11–13% (= 32–47 lineages) (Fig. 5C). This Permian–Triassic rate depression is unlikely the result of a bias from fossil calibrations, because molecular time estimates are often considerably older than fossil estimates

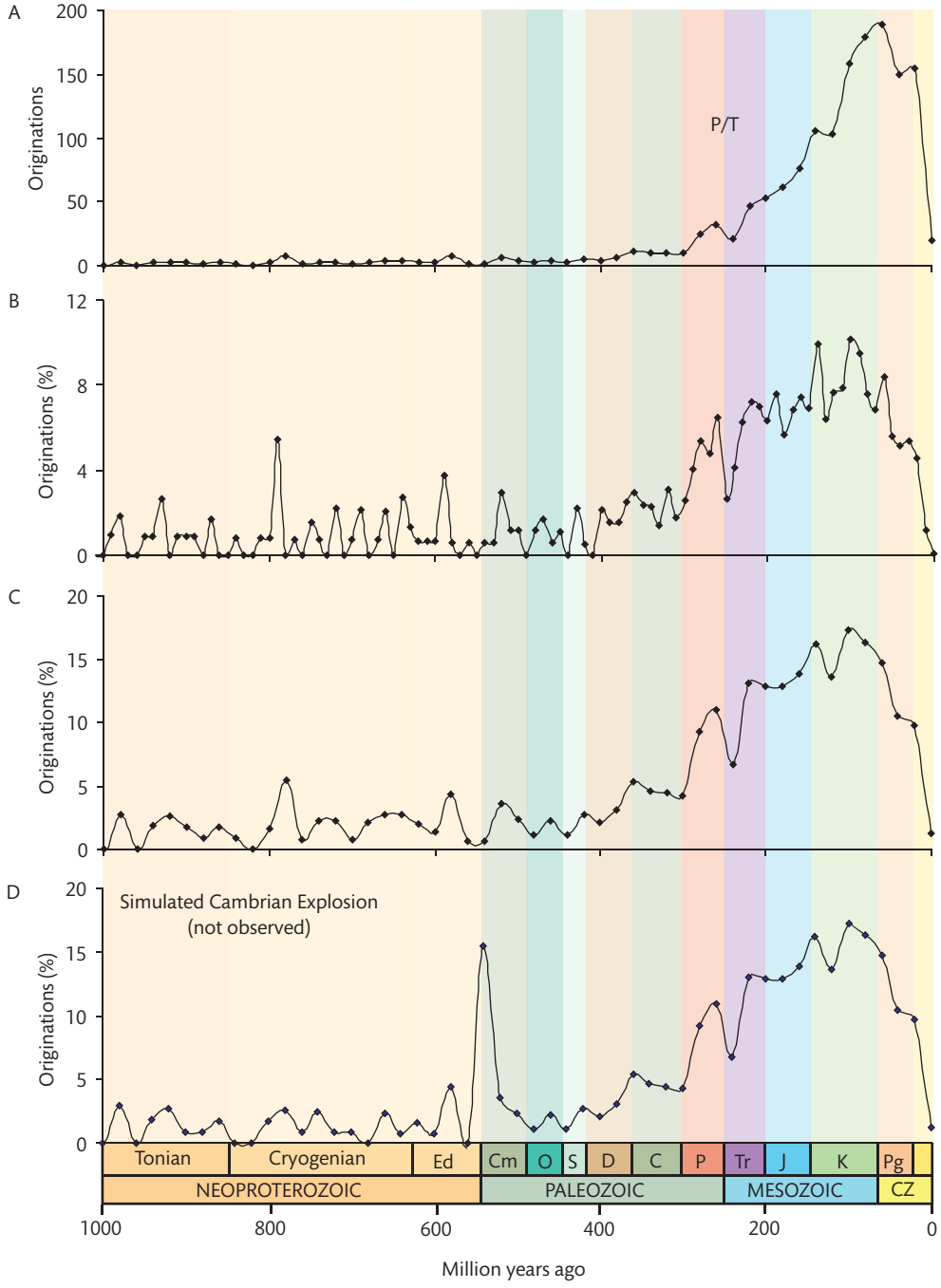

Fig. 5 Diversification curves of life, for the last 1 billion years, based on analyses presented here. (A) Rate curve, showing the total number of lineage originations in 20-Myr intervals. (B) Rate curve, plotted as a percent of standing diversity (originations/total lineages) in 10-million-year intervals. (C) Rate curve, plotted as a percent of standing diversity (originations/total lineages) in 20-Myr intervals. (D) Same rate curve as in C, but showing expected distribution if all animal lineages originated no earlier than the late Precambrian (560 Ma), corresponding to the Cambrian Explosion hypothesis. This distribution was not observed. *Abbreviations*: C (Carboniferous), Cm (Cambrian), CZ (Cenozoic), D (Devonian), Ed (Edicaran), J (Jurassic), K (Cretaceous), O (Ordovician), P (Permian), Pg (Paleogene), P/T (Permian–Triassic extinctions), S (Silurian), and Tr (Triassic).

(or calibrations), as noted earlier. A detailed analysis of diversification in mammals (*100*) did not reveal a rate effect corresponding to the Mesozoic–Cenozoic extinction event (66 Ma). Our analysis also did not find any clear rate effect at 66 Ma, although the taxonomic limitations of this data set (families and above) reduced resolution of events during the Cenozoic. If the rate depression at ~250 Ma is an effect of the Permian–Triassic extinctions, it adds increased confidence in the accuracy of molecular time estimates, and hence The Timetree of Life (Fig. 2).

The Cretaceous peaks in the rate curves, ~140–100 Ma (Fig. 5A–C) correspond to a time when the supercontinents were breaking up, possibly explaining an increased rate of diversification, as has been suggested for bird and mammal ordinal lineages (*19*). Lineages adapting to a great diversity of niches within a continent are more likely to survive the vagaries of the extinction process, in the long term, than those diversifying in more localized settings. Alternatively, the Cretaceous peaks may be the product of two artifacts in combination: the Pull of the Present bias causing an increased rate of recent diversification (since ~300 Ma) combined with the family-level taxonomic bias in the data set, causing a decline in rates in the most recent sampled intervals (<100 Ma).

Concerning the Cambrian Explosion model, the diversification curves (Fig. 5) do not show a pulse in originations at that time (~550 Ma) and therefore do not support this hypothesis. This is not a surprise because nearly all molecular clock analyses of animal origins have found divergence times to be much older than those estimated from the fossil record (*60–62*), as discussed earlier. However, to test whether the Cambrian Explosion signal could be present in the data and obscured by other patterns, we set the maximum possible time estimate for any animal lineage to 560 Ma. The resulting artificial rate curve (Fig. 5D) shows an unusually high level of lineage originations (15.4%/20 Myr) forming a sharp peak at that time (~560 Ma). Such a high level of originations is not reached again until 140 Ma. To see if relaxing the maximum constraint to an earlier time would still result in a noticeable peak, we made it 100 Myr older. The sharp peak remained in the resulting distribution, being diminished only slightly (now 14.2%) and shifted to the left (to 660 Ma). Therefore the signal resulting from the Cambrian Explosion model would be unusual and noticeable in the data, if it were present, either in the Cambrian or late Precambrian (as old as 660 Ma). This reinforces the conclusion, as noted earlier, that the Cambrian Explosion represents an environmental change—probably the result of greatly increased levels of atmospheric oxygen—leading to better preservation, rather than a time when animal phyla originated.

Some patterns are evident within groups, as discussed in individual chapters of *The Timetree of Life*, although the number of lineages is limited in most cases. The rate curve for prokaryotes (Fig. 6) reveals the great age of typical families of Eubacteria and Archaebacteria, with nearly all divergences occurring in the Precambrian. Several peaks in rate are observed when either 100-Myr intervals (Fig. 6A) or 200-Myr intervals (Fig. 5B) are used. The earliest originations (>3000 Ma), before the first peak, represent mostly hyperthermophiles (both superkingdoms) and methanogens (Archaebacteria). All peaks correspond to diversification of Eubacteria. The first is in the late Archean (~2500–2800 Ma), followed by a notable depression (~2500–2000 Ma), and then a cluster of peaks in the late Paleoproterozoic through end of the Precambrian (~1600–600 Ma). The depression corresponds to the Great Oxidation Event (GOE), the first conclusive evidence of a rise in atmospheric oxygen (*101, 102*). Because oxygen was likely toxic to anoxygenic organisms, we may speculate that the GOE caused a mass extinction event, which led to depressed origination rates. The curve shows that, following the depression, a large amount of diversification occurred during the middle to late Proterozoic, corresponding to the time when eukaryotes were beginning their diversification. Diversification curves for other groups (Fig. 7) largely reflect aspects of their evolution already well established, and discussed in individual chapters (*78*). This includes differences in the time of onset of diversification and the average age of family lineages.

All of these distributions have limitations for drawing conclusions about the evolutionary history of life. Determining whether the variation in diversification rate discussed above is significant may require many more lineages and finer taxonomic sampling. Future syntheses that integrate the fossil record, and include groups that are missing here, will provide a more complete view of diversification. However, the addition of taxa below the family level (genera and species) may not change broad patterns in eukaryotes, except for the last ~100 Myr.

A global repository of divergence times

A global consortium maintains public sequence data and alignments but until recently there was no public database for molecular divergence times. We created TimeTree (http://www.timetree.org) (*11*) to fill this gap. Although the hierarchical nature of the data required complexity

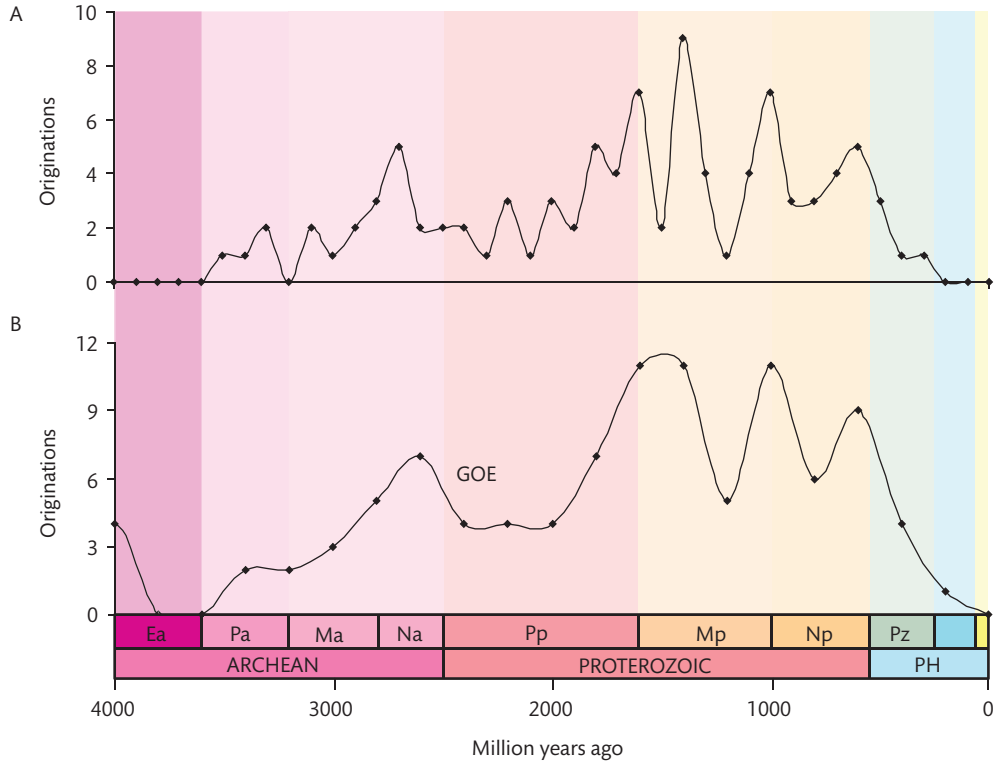

Fig. 6 Diversification rate curve for prokaryotes. (A) Lineage originations (family level and above) sampled in 100-Myr intervals. (B) Lineage originations sampled in 200-Myr intervals. *Abbreviations*: Ea (Eoarchean), GOE (Great Oxidation Event), Ma (Mesoarchean), Mp (Mesoproterozoic), Na (Neoarchean), Np (Neoproterozoic), Pa (Paleoarchean), PH (Phanerozoic), Pp (Paleoproterozoic), and Pz (Paleozoic).

in the database design, it also permitted broader utility in the data presented. TimeTree makes use of a single, large, and conservative guide tree (a version of the Tree of Life) and maps divergence times from a diverse array of published timetrees and divergence time estimates. A query to the database consists of asking for the divergence of species (or taxon) A from species B. The results show all of the pertinent published studies and time estimates bearing on that species divergence, and time estimates are summarized in different ways for the user.

A good illustration of how the system works is the divergence of cat and dog. After the user searches for those two species, TimeTree identifies the two most-inclusive groups containing those taxa, the suborders Feliformia (cat) and Caniformia (dog) of the mammalian order Carnivora. All published times of divergence between Feliformia and Caniformia are then assembled, because every one traces through the same node. For example, the true divergence time of mongoose (Feliformia) and raccoon (Caniformia) is identical to that of cat and dog. Therefore, any single time estimate between two species might be used for hundreds or thousands of other pairwise comparisons and, conversely, many time estimates might pertain to the divergence of two species never actually sequenced. This means that every database query involves tree-based computation and analysis rather than simple retrieval of tabular data.

In the results, additional information is provided such as error estimates, links to abstracts and sequence data, and summary statistics (unweighted, and weighted by number of genes). This gives the user the opportunity to evaluate the published results and determine, visually, if there is a consensus in the field regarding the time estimate for a node in question. That approach differs from the one taken in *The Timetree of Life* (*78*), where experts for each group evaluate the evidence and draw conclusions, sometimes favoring one result over another. Both approaches have their advantages and disadvantages. The number of genes in a study is only one variable of many, some of which are hard to represent in a database. An expert familiar with the literature of a particular group would be the best person to evaluate the different timetrees and time estimates for that group. On the other

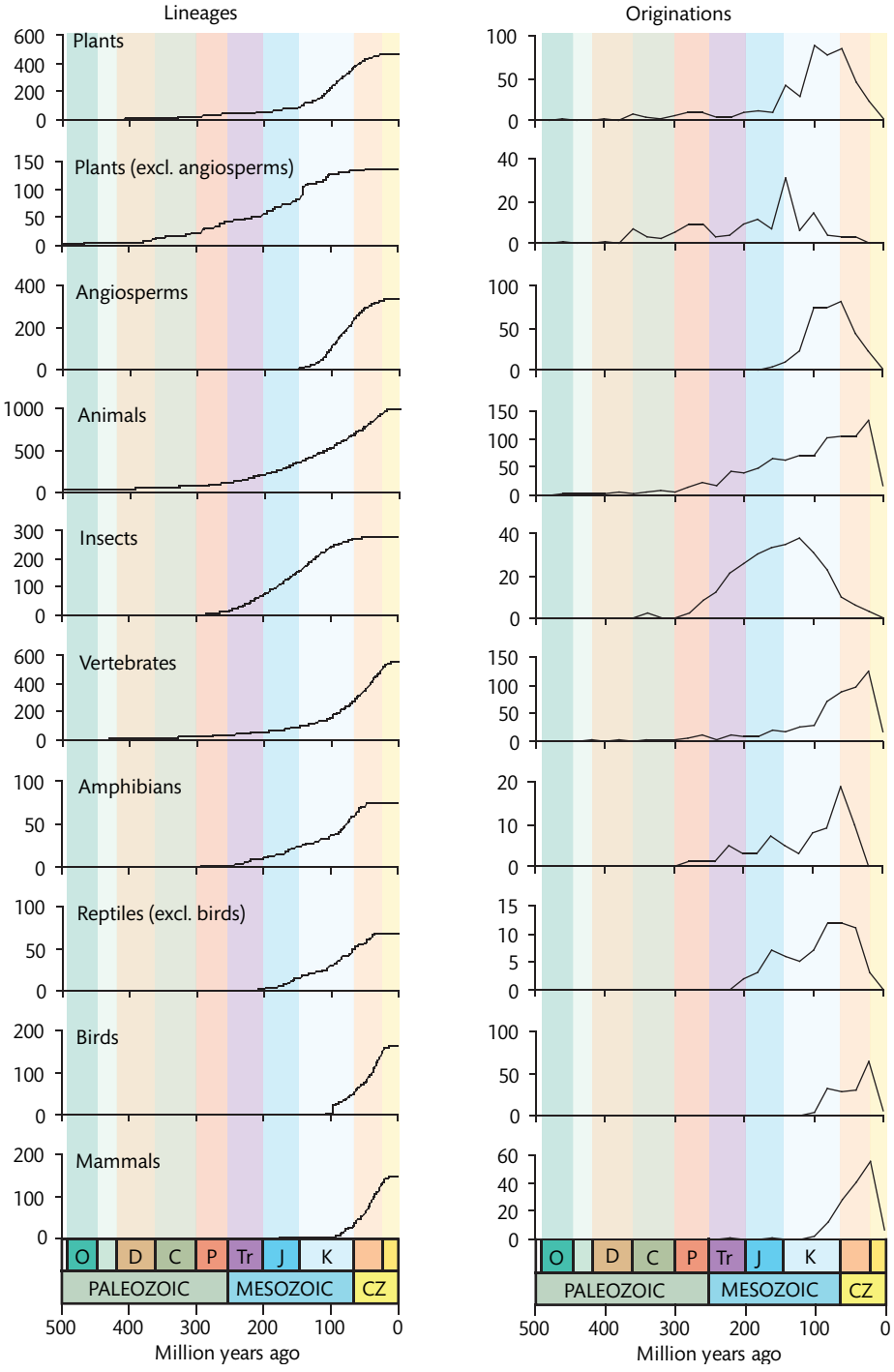

Fig. 7 Diversification curves of selected grou ps. Cumulative curves are on left, showing the total number of lineages (family level and above) sampled in 1-Myr intervals; rate curves are on right, showing lineage originations sampled in 20-Myr intervals. *Abbreviations*: C (Carboniferous), CZ (Cenozoic), D (Devonian), J (Jurassic), K (Cretaceous), Tr (Triassic).

hand, many nodes in the Timetree of Life do not have a relevant expert. For those nodes, and for others where the volume of data has made separate evaluation difficult, a database approach is preferred.

Future prospects

Extinct species and groups of taxa need to be incorporated with living taxa to provide a more complete picture of the Timetree of Life. Also, efforts should be made to fill in the major gaps in taxonomic coverage. Protists, fungi, invertebrates, and ray-finned fishes are all poorly represented, even though many molecular phylogenies exist for these groups and calibrations are available. The most difficult problems to be solved will be the earliest divergences in prokaryotes and eukaryotes and placement of the root of the tree, both phylogenetically and temporally. Resolution of those questions may occur with the increased number of prokaryote and eukaryote genomes that will be available in coming years along with development of new methods and approaches, but they are difficult problems to solve. Methodology for calibration and time estimation is still in its early stage of development and will likely see major improvements during the next decade. Besides new algorithms and approaches, there is also a great need for the design of user-friendly software (*103*) to make methods more accessible to the community.

The immense value of having a robust Timetree of Life—for all fields of science—cannot be overstated. It will provide a means of estimating rates of change for almost anything biological—for example, morphological structures, behaviors, genes, proteins, non-coding regions of genomes—in any group of organisms. In that sense it will catalyze a Renaissance in comparative biology. For paleontologists, geologists, geochemists, and climatologists, it will provide a biological timeline for comparison, prediction, and synchronization with Earth history. In turn this will help formulate better hypotheses for how the biosphere has evolved on Earth and provide insights into evolutionary mechanisms in the Universe.

Acknowledgment

We thank all of the contributors to *The Timetree of Life* for making such a synthesis possible. This work was supported by grants to S.B.H. from the U.S. National Science Foundation (NSF) and National Aeronautics and Space Administration (Astrobiology Institute) and to S.K. from the NSF and National Institutes of Health.

References

1. C. Darwin, *On the Origin of Species by Means of Natural Selection* (John Murray, London, 1859).
2. J. Cracraft, M. J. Donoghue, Eds., *Assembling the Tree of Life* (Oxford University Press, Oxford, 2004).
3. J. D. Watson, F. H. C. Crick, *Nature* **171**, 737 (1953).
4. F. Sanger, *Science* **129**, 1340 (1959).
5. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977).
6. A. M. Maxam, W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 560 (1977).
7. K. Mullis *et al.*, *Cold Spring Harbor Symp. Quant. Biol.* **51**, 263 (1986).
8. C. R. Woese, G. E. Fox, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5088 (1977).
9. M. S. Springer *et al.*, *Nature* **388**, 61 (1997).
10. L. M. Hower, S. B. Hedges, *Caribb. J. Sci.* **39**, 298 (2003).
11. S. B. Hedges, J. Dudley, S. Kumar, *Bioinformatics* **22**, 2971 (2006).
12. E. A. Hadly *et al.*, *PLOS Biol.* **2**, 1600 (2004).
13. R. Dawkins, *The Ancestor's Tale: A Pilgrimage to the Dawn of Evolution* (Houghton Mifflin, New York, 2004).
14. A. S. Romer, *Vertebrate Paleontology* (The University of Chicago Press, Chicago, 1966).
15. E. Zuckerkandl, L. Pauling, in *Horizons in Biochemistry,* M. Marsha, B. Pullman, Eds. (Academic Press, New York, 1962), pp. 189–225.
16. E. Zuckerkandl, L. Pauling, in *Evolving Genes and Proteins*, V. Bryson, H. J. Vogel, Eds. (Academic Press, New York, 1965), pp. 97–165.
17. H. Hayashida, H. Toh, R. Kikuno, T. Miyata, *Mol. Biol. Evol.* **2**, 289 (1985).
18. S. Kumar, S. B. Hedges, *Nature* **392**, 917 (1998).
19. S. B. Hedges, P. H. Parker, C. G. Sibley, S. Kumar, *Nature* **381**, 226 (1996).
20. V. M. Sarich, A. C. Wilson, *Science* **158**, 1200 (1967).
21. G. A. Wray, J. S. Levinton, L. H. Shapiro, *Science* **274**, 568 (1996).
22. M. Kimura, *Nature* **217**, 624 (1968).
23. E. Margoliash, *Proc. Natl. Acad. Sci. U.S.A.* **50**, 672 (1963).
24. D. Hartl, D. Dykhuizen, *Nature* **281**, 230 (1979).
25. M. Goodman, G. Braunitzer, A. Stangl, B. Schrank, *Nature* **303**, 546 (1983).
26. S. Kumar, *Nat. Rev. Genet.* **6**, 654 (2005).
27. W.-H. Li, *Molecular Evolution* (Sinauer Associates, Sunderland, Massachusetts, 1997).
28. M. Nei, *Molecular Evolutionary Genetics* (Columbia University Press, New York, 1987).
29. C. Sibley, J. Ahlquist, *J. Mol. Evol.* **20**, 2 (1984).
30. T. D. Kocher *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 6196 (1989).
31. V. M. Sarich, A. C. Wilson, *Proc. Natl. Acad. Sci. U.S.A.* **58**, 142 (1967).

32. F. Tajima, *Genetics* **135**, 599 (1993).
33. N. Takezaki, A. Rzhetsky, M. Nei, *Mol. Biol. Evol.* **12**, 823 (1995).
34. M. Hasegawa, H. Kishino, T. Yano, *J. Hum. Evol.* **18**, 461 (1989).
35. C. D. Schubart, R. Diesel, S. B. Hedges, *Nature* **393**, 363 (1998).
36. S. B. Hedges, S. Kumar, *Trends Genet.* **19**, 200 (2003).
37. S. B. Hedges, S. Kumar, *Trends Genet.* **20**, 242 (2004).
38. M. J. Sanderson, *Mol. Biol. Evol.* **14**, 1218 (1997).
39. J. L. Thorne, H. Kishino, I. S. Painter, *Mol. Biol. Evol.* **15**, 1647 (1998).
40. J. L. Thorne, H. Kishino, *Syst. Biol.* **51**, 689 (2002).
41. A. J. Drummond, S. Y. W. Ho, M. J. Phillips, A. Rambaut, *PLOS Biol.* **4**, 699 (2006).
42. T. Britton, C. L. Anderson, D. Jacquet, S. Lundqvist, K. Bremer, *Syst. Biol.* **56**, 741 (2007).
43. W. M. Brown, M. George, A. C. Wilson, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 1967 (1979).
44. G. F. Shields, A. C. Wilson, *J. Mol. Evol.* **24**, 212 (1987).
45. M. P. Heinicke, W. E. Duellman, S. B. Hedges, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 10092 (2007).
46. P. E. Ahlberg, J. A. Clack, *Nature* **440**, 747 (2006).
47. E. B. Daeschler, N. H. Shubin, F. A. Jenkins, *Nature* **440**, 757 (2006).
48. S. A. Magallon, M. J. Sanderson, *Evolution* **59**, 1653 (2005).
49. M. J. Sanderson, *Am. J. Bot.* **90**, 954 (2003).
50. J. Müller, R. R. Reisz, *BioEssays* **27**, 1069 (2005).
51. R. R. Reisz, J. Müller, *Trends Genet.* **20**, 237 (2004).
52. S. B. Hedges, S. Kumar, M. van Tuinen, *BioEssays* **28**, 770 (2006).
53. M. J. Benton, P. C. J. Donoghue, *Mol. Biol. Evol.* **24**, 26 (2007).
54. Z. H. Yang, B. Rannala, *Mol. Biol. Evol.* **23**, 212 (2006).
55. M. J. Benton, C. J. Donoghue, R. J. Asher, in *The Timetree of Life*, S. B. Hedges, S. Kumar, Eds. (Oxford University Press, New York, 2009), pp. 35–86.
56. P. C. J. Donoghue, M. J. Benton, *Trends Ecol. Evol.* **22**, 424 (2007).
57. S. Conway Morris, *Nature* **361**, 219 (1993).
58. J. W. Valentine, D. Jablonski, D. H. Erwin, *Development* **126**, 851 (1999).
59. G. E. Budd, N. J. Butterfield, S. Jensen, *Science* **294**, U1 (2001).
60. S. B. Hedges, F. U. Battistuzzi, J. E. Blair, in *Neoproterozoic Geobiology and Paleobiology*, S. Xiao, A. J. Kaufman, Eds. (Springer, New York, 2006), pp. 199–229.
61. G. A. Wray, *Genome Biol.* **3**, 1 (2001).
62. J. E. Blair, in *The Timetree of Life,* S. B. Hedges, S. Kumar, Eds. (Oxford University Press, New York, 2009), pp. 223–230.
63. S. Aris-Brosou, Z. Yang, *Syst. Biol.* **51**, 703 (2002).
64. S. Aris-Brosou, Z. Yang, *Mol. Biol. Evol.* **20**, 1947 (2003).
65. E. J. P. Douzery, E. A. Snell, E. Bapteste, F. Delsuc, H. Philippe, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 15386 (2004).
66. K. J. Peterson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 6536 (2004).
67. K. J. Peterson, N. J. Butterfield, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 9547 (2005).
68. K. J. Peterson, J. A. Cotton, J. G. Gehling, D. Pisani, *Philos. Trans. Roy. Soc. Lond. B* **363**, 1435 (2008).
69. J. E. Blair, S. B. Hedges, *Mol. Biol. Evol.* **22**, 387 (2005).
70. L. A. Hug, A. J. Roger, *Mol. Biol. Evol.* **24**, 1889 (2007).
71. A. J. Roger, L. A. Hug, *Philos. Trans. Roy. Soc. Lond. B* **361**, 1039 (2006).
72. S. Kumar, A. Filipski, V. Swarna, A. Walker, S. B. Hedges, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 18842 (2005).
73. D. E. Canfield, S. W. Poulton, G. M. Narbonne, *Science* **315**, 92 (2007).
74. A. B. Smith, *J Geol Soc* **164**, 731 (2007).
75. M. J. Benton, *Science* **268**, 52 (1995).
76. M. Foote, J. P. Hunter, C. M. Janis, J. J. J. Sepkoski, *Science* **283**, 1310 (1999).
77. S. Tavaré, C. R. Marshall, O. Will, C. Soligo, R. D. Martin, *Nature* **416**, 726 (2002).
78. S. B. Hedges, S. Kumar, Eds., *The Timetree of Life* (Oxford University Press, New York, 2009).
79. J. C. Avise, in *The Timetree of Life*, S. B. Hedges, S. Kumar, Eds. (Oxford University Press, New York, 2009), pp. 19–25.
80. J. C. Avise, G. C. Johns, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7358 (1999).
81. M. J. Benton, *The Fossil Record 2* (Chapman & Hall, London, 1993).
82. A. F. Hugall, R. Foster, M. S. Y. Lee, *Syst. Biol.* **56**, 543 (2007).
83. C. A. Matthee, in *The Timetree of Life*, S. B. Hedges, S. Kumar, Eds. (Oxford University Press, New York, 2009), pp. 487–489.
84. J. J. Sepkoski, *Phil. Trans. Roy. Soc. Lond. B* **353**, 315 (1998).
85. A. B. Smith, *Systematics and the Fossil Record* (Blackwell Scientific Publications, London, 1994).
86. M. J. Benton, *Trends Ecol. Evol.* **12**, 490 (1997).
87. M. J. Benton, B. C. Emerson, *Palaeontology* **50**, 23 (2007).
88. S. M. Porter, *Paleon. Soc. Pap.* **10**, 35 (2004).
89. S. M. Porter, in *Neoproterozoic Geobiology and Paleobiology*, S. Xiao, A. J. Kaufman, Eds. (Springer, Dordrecht, The Netherlands, 2006), pp. 1–21.
90. J. W. Valentine, *Palaeontology (Oxford)* **12**, 684 (1969).
91. P. W. Signor, in *Phanerozoic Diversity Patterns: Profiles in Macroevolution*, J. W. Valentine, Ed. (Princeton University Press, Princeton, 1985).
92. D. M. Raup, *Science* **177**, 1065 (1972).
93. J. J. Sepkoski, *Paleobiology* **4**, 223 (1978).
94. J. J. Sepkoski, *Paleobiology* **5**, 222 (1979).

95. J. J. Sepkoski, *Paleobiology* **10**, 246 (1984).
96. R. A. Berner, D. J. Beerling, R. Dudley, J. M. Robinson, R. A. Wildman, *Ann. Rev. Earth Planet. Sci.* **31**, 105 (2003).
97. R. A. Berner, *Geochim. Cosmochim. Acta* **70**, 5653 (2006).
98. J. Alroy *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **98**, 6261 (2001).
99. M. J. Benton, M. A. Wills, R. Hitchin, *Nature* **403**, 534 (2000).
100. O. R. P. Bininda-Emonds *et al.*, *Nature* **446**, 507 (2007).
101. D. Canfield, *Ann. Rev. Earth Planet. Sci.* **33**, 1 (2005).
102. H. D. Holland, *Geochim. Cosmochim. Acta* **21**, 3811 (2002).
103. S. Kumar, J. Dudley, *Bioinformatics* **23**, 1713 (2007).