The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils?

Emmanuel J.P. Douzery, Elizabeth A Snell, Eric Bapteste, Frédéric Delsuc, Hervé Philippe

▶ **To cite this version:**

Emmanuel J.P. Douzery, Elizabeth A Snell, Eric Bapteste, Frédéric Delsuc, Hervé Philippe. The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils?. Proceedings of the National Academy of Sciences of the United States of America, 2004, 101 (43), pp.15386-91. 10.1073/pnas.0403984101 . halsde-00193035

HAL Id: halsde-00193035
https://hal.science/halsde-00193035

Submitted on 30 Nov 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Manuscript information: 30 text pages (including references and figure legends) ; two Figures ; two Tables ; five Supporting Informations (S1 to S5). Characters count:

All text characters plus spaces =	40,231	
Fig. 1 (2-columns, 12 cm high = 360 x 12) =	4,320	
Fig. 2 (1-column, 5 cm high = 180 x 5) =		900
Table 1 (1-column, 6 lines high = 60 x 6) =		360
Table 2 (1-column, 8 lines high = 60 x 8) =		480
Space Allowance		
1 double-column figure (1 x 240) =		240
1 single-column figure (1 x 120) =		120
2 single-column tables (2 x 120) =		240
Total characters in paper =		**46,891**

Classification — Biological Sciences: Evolution.

The timing of eukaryotic evolution: does a relaxed molecular clock reconcile proteins and fossils?

Emmanuel J. P. Douzery [1, *], Elizabeth A. Snell [2], Eric Bapteste [3],

Frédéric Delsuc [1, 4], & Hervé Philippe [3, 4, *]

[1] Paléontologie, Phylogénie et Paléobiologie , CC064, Institut des Sciences de l'Evolution (UMR CNRS 5554), Université Montpellier II, Place E. Bataillon, 34 095 Montpellier Cedex 5, France.

[2] School of Animal and Microbial Sciences, The University of Reading, Whiteknights PO Box 228, Reading RG6 6AJ, United Kingdom.

[3] Phylogénie, Bioinformatique et Génome, UMR 7622 CNRS, Université Pierre et Marie Curie, 9 quai St Bernard Bât. C, 75005 Paris, France.

[4] Canadian Institute for Advanced Research. Département de Biochimie, Université de Montréal, Succursale Centre-Ville, Montréal, Québec H3C3J7, Canada.

(*) Corresponding authors: douzery@isem.univ-montp2.fr (Tel = 33 4-67-14-48-63 ; Fax = 33 4-67-14-36-10) and herve.philippe@UMontreal.CA (Tel = 1 514-343-6720 ; Fax = 1 514-343-2210)

ABSTRACT

The use of nucleotide and amino acid sequences allows improved understanding of the timing of evolutionary events of life on earth. Molecular estimates of divergence times are however controversial and are generally much more ancient than suggested by the fossil record. The limited number of genes and species explored, and pervasive variations in evolutionary rates are the most likely sources of such discrepancies. Here we compared concatenated amino acid sequences of 129 proteins from 36 eukaryotes to determine the divergence times of several major clades, including animals, fungi, plants, and various protists. Due to significant variations of their evolutionary rates and to handle the uncertainty of the fossil record, we used a Bayesian relaxed molecular clock simultaneously calibrated by 6 paleontological constraints. We show that, according to 95% credibility intervals, the eukaryotic kingdoms diversified 950-1,259 million years ago (Mya), animals diverged from choanoflagellates 761-957 Mya, and the debated age of the split between protostomes and deuterostomes occured 642-761 Mya. The divergence times appeared to be robust with respect to prior assumptions and paleontological calibrations. Interestingly, these relaxed clock time estimates are much more recent than those obtained under the assumption of a global molecular clock, yet bilaterian diversification appears to be about 100 million years more ancient than the Cambrian boundary.

INTRODUCTION

Reconstructing and dating the early evolutionary history of eukaryotes has proven a challenging question due to the scarcity of the fossil record, especially for protists (1). In the Archean eon, biomarker compounds characteristic of possibly extinct stem eukaryotes are found ~2,700 million years ago (Mya) (2). Morphological clues for eukaryotes seem to occur in the Paleoproterozoic, but convincing microfossil evidence appears ~1,500 Mya in Lower Mesoproterozoic successions (3). Fossils that can be interpreted as crown eukaryotes are found in Upper Mesoproterozoic / Lower Neoproterozoic rocks, around 1,200 Mya for red algae (4), and 1,000 Mya for stramenopiles (5). Chlorophytes and testate amoebae are found in Middle to Late Neoproterozoic formations at ~750 Mya (5, 6). Special attention has been attracted to the origin of the animal phyla. Bilaterian metazoans first appeared at the end of the Late Proterozoic (~600 Mya at the Doushantuo Formation) (7) and were already quite diversified ~530 Mya (8), suggesting a rapid diversification, known as the Cambrian explosion.

The advent of molecular data started a new era by providing an independent approach to address the history of life on Earth, and thus creating a closer connection between geology, paleontology, and biology (9). The constancy of molecular evolutionary rate over time, *i.e*. the molecular clock hypothesis, was suggested early on (10). Calibration of this clock based on taxa with a rich fossil record allowed the inference of divergence times for other living organisms (11). Yet, DNA and protein age estimates of land plants, fungi, and animal phyla possibly appeared several hundred million years (Myr) older than indicated by paleontology (12).

A major concern in molecular dating is that independent studies based on multiple genes yielded highly variable estimates. For example, the protostome / deuterostome divergence time ranges from 573 to 1,200 Mya, albeit either close to or drastically more ancient than the Cambrian explosion (13-23). The observed discrepancy between clocks and rocks might be due to an extended period of Precambrian metazoan diversification without

any trace in the fossil record (24), but also to biases in molecular dating methods. Among the limitations of the latter approaches, the non-clocklike behavior of genes and proteins (25), the limited power of the relative rate tests used to detect lineage-specific rate variation (26), and the existence of a systematic bias toward overestimation of evolutionary time scales (27) are the most confounding. For example, incorrectly assuming a clocklike property of the sequence data involves higher (22, 28, 29) or lower (30) divergence times. Consequently, dating methods have been developed to relax the molecular clock assumption by allowing rate variations along branches of a phylogenetic tree (28, 31-33).

To estimate the timing of eukaryotic evolution while reducing the impact of stochastic error due to the finite length of the sequences, we used a large data set of 129 combined nuclear proteins for 36 animals, fungi, plants, and protists. Despite the expectation that potential lineage-specific evolutionary rate variations from one protein to another might be counterbalanced in large-scale analysis (34), extensive among species variations in the rate of amino acid replacements were identified. We thus took advantage of a relaxed molecular clock approach developed in a Bayesian framework, which approximates divergence times by a Markov chain Monte Carlo (MCMC) numerical method (32, 33). Such a relaxed clock method has three major advantages: (i) the biologically unverified hypothesis of a constant evolutionary rate all over the history of eukaryotes is unnecessary, (ii) the incorporation of prior constraints on divergence times is preferred over the use of fixed time points in order to handle the inherent uncertainties of paleontological data (fossils never match exactly on nodes of phylogenetic trees), and (iii) several independent calibrations can be used simultaneously. Since divergence times based on this relaxed protein clock were much more recent than those based on a global (constant) clock, computer simulations were used to explore the reliability of the Bayesian dating method.

MATERIAL & METHODS

The assembly of the data set of 129 proteins from 36 eukaryotes and its partitioned maximum likelihood (ML) analysis are provided elsewhere (35). Due to ambiguous alignment, two third of the sites were conservatively removed to ensure the sitewise homology of the remaining 30,399 positions at this deep evolutionary time scale. Alignments are available from HP upon request. We verified orthology through the analysis of each gene with standard phylogenetic methods, and we discarded several proteins, including some that were frequently used for studying eukaryotic evolution (tubulins, actin, and HSP70). Our reasonable taxon coverage in this large multiprotein framework involves that some genes are lacking for some taxa, leading to ~25% of missing character states. The best topology (Fig. 1) was rooted by *Dictyostelium,* as suggested by a derived gene fusion separating opisthokonts (animals, choanoflagellates, fungi) from bikonts (eukaryotes ancestrally with two cilia) (36, 37). However, because of the potential homoplasy of gene fusion and fission characters (38), a more classical rooting along the kinetoplastid branch was also considered.

Relaxed molecular clock analyses

Divergence times were estimated under a Bayesian relaxed molecular clock (33), constrained by six independent fossil references. To explicitly incorporate the inherent uncertainty of paleontological estimates, each primary calibration was defined as a prior time constraint, with lower and / or upper bounds. In a conservative approach, we used the stratigraphic range of the geological period to which key fossils for the node under focus were attributed: (i) the Eudicots / Liliopsida split during Jurassic (144-206 Mya), just before the first record of angiosperm pollen grains in the Neocomian (see (39)) ; (ii) the Bryophyta / Tracheophyta split during Ordovician (443-490 Mya) as attested by the oldest land plant spores (40) ; (iii) the origin of ascomycotans before Devonian (417 Mya ; (41)); (iv) the Actinopterygii / Mammalia split during Devonian (354-417 Mya) as actinopterygians and sarcopterygians remains are dated from this period (42) ; (v) the split between Diptera and Lepidoptera + Hymenoptera during the Devonian-Carboniferous (290-417 Mya) (42) ; and

(vi) the Chelicerata / Insecta split during Cambrian (490-543 Mya) as suggested by the oldest remains of these groups (42). Relaxing the molecular clock assumption involved two steps. First the program ESTBRANCHES (version 5b: ftp://statgen.ncsu.edu/pub/thorne) (33) estimated the branch lengths and their variance-covariance matrix from the concatenated 30,399 sites, under the Jones-Taylor-Thornton (JTT) model. Second, the program DIVTIME (version 5b) estimated the posterior ages of all nodes within the ingroup. Prior Gamma distributions on three parameters of the relaxed clock model were assumed and specified through the mean and standard deviation (SD) of the root age, root rate, and rate autocorrelation: 1,000 Mya (SD = 500 Mya) for the expected time between tips and root without node time constraints, 0.036 (SD = 0.018) amino acid replacements per 100 sites per Myr at the ingroup root node, 0.001 (SD = 0.001) for the parameter (ν) that controls the degree of rate autocorrelation per Myr along the descending branches of the tree. The highest possible time between tips and root was 4,500 Mya. The MCMC were started from random values. After conservatively discarding 100,000 generations as the burn-in, they were run for 10 million generations. A sample was taken each 100 generations. Posterior divergence times were identical for different MCMC started independently, evidencing convergence of the Bayesian dating procedure. The uncertainty on divergence time estimates was characterized by 95% credibility intervals, calculated after sorting the 100,000 MCMC samples, and then reporting the 2,501st and 97,500th values.

Simulation studies

Simulation studies were conducted to evaluate (i) the ability of the global and relaxed clock approaches to accurately estimate divergence times when rates do correlate or not along tree branches, and (ii) the impact of missing data. We generated 10 matrices of 36 taxa and 30,399 amino acid positions (i.e., 1,094,364 character states) under PSeq-Gen (43), by using empirical amino acid frequencies, and a Gamma distribution parameter (α = 0.70) for rate heterogeneity among sites. Each branch length was set equal to the product of its rate (i.e., the average of the rates at the nodes that begin and end it) by its time duration

(i.e., the age difference between its beginning and ending nodes) as measured by DIVTIME. For each of these 10 simulated data sets, we applied the six calibration constraints, and computed the Bayesian divergence times as described in the previous section. Then, we removed each character corresponding to the missing data of the original data set (281,517 amino acids, i.e., ~25% of the data), and again calculated the divergence times. Age estimates computed with and without missing data were then compared for each node.

Moreover, a second set of 10 complete matrices was generated under a global clock model. The same topology, amino acid frequencies, Gamma parameter, and distribution of node times were used, but a single (constant) rate of 0.025 amino acid replacement per 100 sites was enforced for branch length entries. Autocorrelated-rate and constant-rate matrices were analyzed under ESTBRANCHES. Then, MCMC runs were conducted under DIVTIME to estimate divergence times, assuming either rate autocorrelation along branches (mean and SD for $\nu \neq 0$) or a perfect clock (mean and SD for $\nu = 0$).

RESULTS

A molecular time scale for the main eukaryote clades

The phylogeny resulting from the ML analysis of the 30,399 unambiguously aligned amino acid positions (Fig. 1) is in agreement with current opinion about eukaryotic relationships (44). The log-likelihoods (ln*L*) of the phylogenetic tree were $\ln L_{\text{UNCONSTRAINED}} = -779{,}284$ without a molecular clock constraint, and $\ln L_{\text{CLOCK}} = -783{,}020$ under the global clock and a rooting by *Dictyostelium*. A likelihood ratio test significantly rejected the hypothesis of a clocklike behavior of our data ($P < 0.0001$). Thus, the rate of amino acid replacements extensively changed among the eukaryote lineages here analyzed. For example, trypanosomes and nematodes evolved two times and three times faster than mammals, respectively (data not shown).

The relaxed molecular clock based on the 129 proteins was calibrated by six time constraints spread over the phylogenetic tree and defined within green plants, animals, and fungi. Credibility intervals at 95% indicated that the basal split between the major eukaryotic kingdoms documented here occurred 950-1,259 Mya (mean: 1,085), followed by their diversification in less than 200 Myr (Fig. 1). Plantae originated 892-1,162 Mya (mean: 1,010), and red algae branched off 825-1,061 Mya (mean: 928) whereas stem land plants separated from green algae 662-812 Mya (mean: 729). This suggests that the endosymbiosis of a free-living cyanobacterium that led to primary plastids (surrounded by two membranes) occurred between 825 to 1,162 Mya (Fig. 1). Plastids surrounded by four membranes originated from secondary endosymbiosis, in which a red alga was engulfed and retained by a flagellate protist. This event probably happened along the branch leading to stramenopiles + alveolates, i.e., between 767-1,072 Mya, shortly after the primary endosymbiosis (Fig. 1).

Among opisthokonts, most of the inferred divergence times were more recent than previous molecular multigene estimates. The divergence between animals and fungi was for example estimated 872-1,127 Mya (mean: 983 Mya) whereas others provide estimates up to 1,600 Mya (45). The divergence between choanoflagellates and animals was estimated during the Late Proterozoic, between 761 and 957 Mya (mean: 849). The major transition during animal evolution from unicellular to multicellular state might have thus occurred at least 200 Myr before the Cambrian explosion. Interestingly, the divergence time between protostomes and deuterostomes (Fig. 1) was 642-761 Mya (mean: 695). This value is much lower than most of the previously published ones (12, 17, 46), but is close to recent estimates (22, 23), reducing to about 40 Myr the minimum gap between the protein relaxed clock timing and the first indisputable bilaterian fossil (7). Within fungi, the divergence between ascomycetes and basidiomycetes is estimated around 629-837 Mya (mean: 727), whereas a recent study using a global clock method on a similar number of proteins, but a lower number of taxa and calibration points, yielded 1,200 Mya (45).

Global versus relaxed clock estimates

Our molecular time scale for the major eukaryotic clades is on average 1.5 times more recent than the one based on a global clock (12). To evaluate the accuracy and precision of dates estimated under linear versus autocorrelated rates extrapolations beyond calibrations, we conducted simulations. Table 1 recapitulates the age estimates of two selected nodes—bilaterians and our tree root—calculated under either global or relaxed clock assumptions onto data simulated with or without constant rate of evolution. Relaxed clock timings were accurate, though slightly deeper when estimated on constant-rate data. They were also less precise—as attested by 95% credibility interval widths—relative to global clock estimates, because of the greater number of parameters involved (22). However, constant-rate extrapolations conducted with all other conditions being identical yield less accurate, i.e., 1.2 to 1.4 deeper divergence times when actual rates do vary. The simulations therefore indicate that the autocorrelation model of substitution rate evolution does not underestimate ages on clocklike data, whereas a linear extrapolation on rate-variable data markedly overestimates divergence times (Table 1).

Impact of missing data on divergence times

Despite the expectation that the large size of our dataset would have likely buffered the impact of missing data (47), our dating estimates might have been seriously influenced by the absence of about 25% of amino acids. This point was evaluated on simulated protein sequences, by comparing divergence times inferred from complete and gapped alignments (Supporting Information S1). The presence of missing data leads to slightly under- or over-estimated divergence ages, depending on the node under focus, and induce a mean difference of –0.2 % ± 2.8 over all nodes (extreme values: –9.8 to +8.7%). On average, the relative difference is respectively +1.7 % and –2.5 % for ages which are over- and under-estimated by missing data. This corresponds to variations of ca. +17 Myr to –25 Myr for a

1,000 Myr old node. For the root age, these values are indeed twice less than the actual standard deviation measured by the Bayesian approach. Interestingly, the random introduction of an additional 25% of missing data—yielding a data set with 50% of missing amino acids—again provides very similar dates relative to the original data set (data not shown). The relaxed clock method is therefore efficient, even when the alignment contains partial sequences.

Sensitivity of posterior divergence ages upon prior assumptions

Since Bayesian posteriors might depend on priors, we evaluated the sensitivity of our dating results upon *a priori* hypotheses. Different expected number of time units between tree root and tips were tested by comparing posterior ages obtained using different Gamma-distributed priors: 540 ± 270 Mya (the Cambrian boundary), 1,000 ± 500 Mya (earliest body fossils of stramenopiles (5)), 2,000 ± 1,000 Mya (a recent molecular estimate for eukaryotes origin (12)), and 3,000 ± 1,500 Mya (a deep value for stem eukaryotes (2)). Posterior ages were plotted against the four independent root priors for each of the 34 tree nodes. This resulted in 34 regression lines with a mean slope of 0.016 (range: –0.002 ± 0.0005 to 0.067 ± 0.007). Such an average impact over all nodes indicated that a 1,000 Myr variation of the root prior will involve a posterior divergence age variation of only 16 Myr, which is far less than the order of magnitude of the uncertainty on the divergence ages themselves. Posterior divergence times thus appeared only slightly sensitive to huge changes in prior specification of root age. Furthermore, comparison of *a priori* and *a posteriori* age estimates for calibration nodes did not reveal any systematic trend (Supporting Information S2), suggesting that paleontological uncertainties did not bias our calculations since they were incorporated in the dating procedure (31, 33).

Changes in the reference topology may potentially affect molecular dating (39). Since the root location is uncertain among eukaryotes (44), we recalculated divergence times after moving the root from the branch leading to the cellular slime mold to the one leading to

kinetoplastids (Fig. 1). Divergence times in opisthokonts, especially among animals, were not sensitive to this modification (e.g., we obtained 627-738 Mya for bilaterians). The kinetoplastid rooting actually involved deeper divergences among Plantae (899-1,191 Mya). The main clades of eukaryotes also diverged earlier (1,042-1,412 Mya: Supporting Information S3), likely because such an alternative root location involves a pectinate tree shape. Thus, we confirm that the above-mentioned observations of Middle- to Late-Proterozoic cladogeneses among most eukaryotes were not the result of a rooting artifact.

Molecular clock approaches are also sensitive to the choice of calibrations (48). For example, divergence time estimates under the single vertebrate calibration were on average 1.4 times deeper than those measured with the six calibrations (Supporting Information S4). We systematically studied the effect of calibration selection by analyzing all the possibilities of choosing one or combining 2, 3, 4, 5 or 6 calibrations. Whereas divergence times averaged over all possibilities of using a given number of simultaneous calibrations seems not to vary much with those numbers, the minimum and maximum estimates may dramatically vary depending upon the choice of a given combination of fossil references (Table 2). However, a reduction of the uncertainty on divergence times was clearly associated to the increase of the number of simultaneous calibrations. Importantly, our datings were robust against the random exclusion of one or two calibration points.

DISCUSSION

Robustness of the relaxed clock estimates

Molecular clocks are highly controversial, dates from various sources being often so different that some researchers doubt that molecular dating of evolutionary events is possible (49). Three points are of crucial importance: (i) the size of the molecular sample (i.e. stochastic errors due to a limited number of species and/or genes), (ii) the efficiency of the dating method (i.e. systematic errors in the inference), and (iii) the quality of the fossil

calibrations (i.e. amplification of paleontological uncertainties). We tried to reduce the impact of all these potential issues.

Our divergence time estimates should not be seriously affected by the stochastic error since we used a very large data set (129 genes, 30,399 amino acid positions) corresponding to proteins involved in different functional pathways like translation, transcription, cytoskeleton, and metabolism. Contrary to all previous studies with such a large sample of genes, a reasonable taxonomic diversity (36 species) was sampled. Here, half of the total number of proteins contains at most three missing taxa over 36, and 25% of the amino acids are missing in our final alignment. Nevertheless, our dating estimates are robust upon missing data (Supporting Information S1). Indeed, a potentially reduced phylogenetic (and/or dating) accuracy associated with missing data is primarily caused by the occurrence of too few complete characters rather than too many missing positions (47). Presently, since the shortest concatenated sequence includes 7,500 amino acids, our complete alignment retains a strong phylogenetic and dating signal.

Second, our relaxed molecular clock datings (Fig. 1) appear to be robust to the specification of prior distributions on model parameters. This suggests that much posterior dating information regarding measure of divergence times among eukaryotes is attributable to the 129 concatenated proteins rather than to prior specifications. Moreover, the relaxed clock approach retrieves the true divergence times (Table 1) when substitution rates are distributed according to the autocorrelated-rate model developed by Thorne *et al*. (32).

Third, the concomitant use of six calibration constraints chosen among plants, fungi, and animals (Figure 1) also helped to more accurately measure the absolute amino acid replacement rates in these three distinct clades, and stabilized the Bayesian divergence times on the whole tree. An alternative procedure (22, 28) is to use several independent calibrations, and, for each gene, to average the corresponding divergence estimates. However, such an approach is "inferior to a simultaneous analysis of multiple genes under the constraints of multiple calibrations" (22). Accordingly, we have shown that the use of simultaneous calibrations reduces the uncertainty on age estimates (Table 2).

Variable-rate and constant-rate molecular clocks

The greatest discrepancy between age estimates without and with the clock assumption is observed for the *Saccharomyces* / *Candida* split—here estimated at 168-308 (mean: 235) Mya (Fig. 1) compared to 841 Mya (45). These two fungi appear to be fast evolving (data not shown), and such high rates are probably not accommodated by global clock based methods. When we constrained a perfect molecular clock under the Bayesian approach, this divergence was estimated at 439-468 Mya (mean: 454). On average, the divergence times among eukaryotes were 1.7 times deeper with a global clock than with a relaxed clock (Supporting Information S5). The greatest contrast (ratio > 2.0) between global and relaxed clock ages is observed for alveolates, nematodes, and kinetoplastids, i.e., the fastest-evolving taxa for the 129 proteins here evaluated. A similar result was obtained for animals under various models of clock relaxation (22, 28), and for mitochondrial genomes of mammals (29).

How to evaluate which method, either a global or a relaxed clock, is better describing the data at hands? Estimating a time scale by the mean of age extrapolation beyond animal, fungi, and plant calibrations may possibly lead to biased estimates. Whereas interpolation of date estimates are certainly less biased because evolutionary rates are bounded by zero, extrapolation could conduct to rates increasing on average when moving far back in time. Thus, increased rates—faster earlier and slower since the Neoproterozoic / Cambrian transition—may result into apparently too recent estimates for deeper nodes. Two elements suggest that our analyses have not been affected by such a trend. First, we plotted the posterior evolutionary rate of each node against its mean posterior age estimate as measured on our concatenated data set (Fig. 2). No trend towards a rate increase far back in time is observed, and the rate for the last common bilaterian ancestor is close to the mean rate measured along the eukaryotic phylogeny. In particular, our 129 proteins did not record the Late Precambrian episodic burst of evolutionary rates detected on 22 mitochondrial and

nuclear genes (22). Second, simulation studies (Table 1) indicate that, at least in the present case, the autocorrelated rates model does not underestimate divergence times, neither on clocklike nor on relaxed clock sequences.

Comparison with other relaxed molecular clock datings

Recently, two relaxed clock dating studies—among metazoans (23) and photosynthetic eukaryotes (50)—yielded results that are, at first sight, incongruent with our estimates. The use of seven concatenated proteins and the non parametric rate smoothing r8s approach (31) indicate a more recent occurrence of the last common bilaterian ancestor, between 556-592 Mya (23) instead of 642-761 here. Importantly, these estimates were obtained without taking into account among site rate variation. In the same paper, calculations conducted by modeling such a rate variation with a Gamma distribution yield a confidence interval of 636-678 Mya for the bilaterian diversification, i.e., an estimate overlapping with ours.

On the basis of a tree derived from 5 concatenated plastid proteins, calibrated by 5 intervals under the r8s approach, Yoon *et al*. (50) suggested a red-green algae split between 1,591-1,757 Mya, instead of 825-1,061 or 899-1,191 Mya here depending on the root location. Our reanalysis of this data set under the DIVTIME relaxed molecular clock approach suggests a major incompatibility between “green” and “red” calibrations. Under the ML topology inferred from the 5 plastid proteins, and enforcing the angiosperms, spermatophytes, and land plants nodes between 90-130, 290-320, and 432-476 Mya (50) would involve a red algae diversification between 392-761 Mya (mean 551 Ma), and a *Chondrus* / *Palmaria* split between 165-381 Mya. This is incompatible with the respective 1,174-1,222 and 594-603 Mya constraints used in reference (50). Reciprocally, if we use the latter two "red" calibrations, angiosperms, spermatophytes, and land plants would have diversified 106-336, 344-809, and 626-1,237 Mya. Actually, calibrations among land plants only yield a red algae / chlorobionts split between 925-1,576 Ma, a credibility interval that

partially overlaps with ours. Further analyses, including longer plastid sequences, are required to evaluate the degree of congruence between molecular and paleontological data in red algae (4, 50). More generally, molecular dating still needs to be refined by improved sampling of genes and especially of species, and improved methods with more realistic models of rate variation.

Towards a partial reconciliation of clocks and rocks

Deeper molecular dates are actually expected since the fossil record documents only the first appearance of a morphologically recognizable (crown) group, and not the actual time of genetic divergence (51). Previous molecular datings yielded molecular ages always more ancient than paleontological ones, but often to a large extent (12, 15, 20, 46). Interestingly, our molecular dating (Fig. 1) does not appear to be biased in such a systematic way. Molecular ages are more recent than the fossil record for red algae but more ancient for animals. We dated the divergence between green plants and red algae between 825-1,061 Mya (Fig. 1) whereas a fossil of 1,200 Mya has been interpreted as belonging to a specific subgroup of red algae, i.e. bangiophytes (4). Molecular dating with a kinetoplastid rooting would be in better agreement with this fossil (95% credibility interval: 899-1191 ; Fig. 1), but such a rooting might represent a tree reconstruction artifact (36). Actually, single fossils may also bear their load of uncertainty in taxonomy identification and geological ages (11). Illustrating the latter case, the age of the *Bangiomorpha* assemblage is in fact bounded between 1,267 and 723 Mya (52). For bilaterians, the discrepancy between our molecular estimates (642-761 Mya) and the fossil record (7) is reduced to a minimum of 40 Myr. According to protein data, the divergence of the main stem groups of metazoans (Deuterostomia, Ecdysozoa and Lophotrochozoa) would have occurred in the Late Neoproterozoic, but their diversification would date close to the Precambrian-Cambrian boundary. This suggests that the stem lineages would have been organisms with a poor

fossil record because of their small size, taphonomic problems, or undistinguishable morphology.

The use of a relaxed molecular clock on our large data set estimated the diversification time of most extant eukaryotic kingdoms at ~1,100 Mya in the mid-Proterozoic (Fig. 1). Even if cytoskeletal and ecological prerequisites for eukaryote diversification were already established some 1,500 Mya (3), the postulated anoxic and sulfidic redox status of oceans between ~2,000 to 1,000 Mya might have limited the rise of photosynthetic eukaryotes (53). Around 1,000 Mya, the diversification of plants may have been facilitated by the primary endosymbiotic event that led to plastids. Moreover, it has been proposed that a major increase in atmospheric oxygen level occurred between 1,000 and 640 Mya (54). In association with the Neoproterozoic appearance of more fully oxic oceans (53), this major transition towards a more oxygenic environment could possibly have accelerated the diversification of the eukaryotic organisms, thanks to the aerobic respiration provided by mitochondria (55). Close to the Meso- / Neo-Proterozoic transition, choanoflagellates diverged from their closest animal relatives marking the dawn of cellular cooperation in animals. The major eukaryotic kingdoms therefore originated much more recently than previously thought (12), but did not seem to have been affected by the postulated global glaciations ("snowball Earth" hypothesis) between 750 and 600 Mya (56). The present study illustrates how geological and biological records of organismal evolution might be partially reconciliated by using molecular relaxed clock datings from large genomic comparisons.

ACKNOWLEDGEMENTS

We thank Henner Brinkmann, Wilfried de Jong, Christophe Douady, Eric Fontanillas, Peter Holland, Emmanuelle Javaux, Franz Lang, Nicolas Lartillot, David Moreira, and two anonymous referees for helpful suggestions. Debashish Bhattacharya kindly gave us the alignment of 5 plastid genes. This work has been supported by the "ACI Informatique-

Mathématique-Physique en Biologie Moléculaire [ACI IMP-Bio]", the Canada Research Chair Program, and the IFR119 "Biodiversité Continentale Méditerranéenne et Tropicale" (Montpellier) and INFOBIOGEN (Evry, France) computing facilities. This publication is the contribution N° EPML-004 of the Equipe-Projet multi-laboratoires CNRS-STIC "Méthodes informatiques pour la biologie moléculaire" and N° 2004-YYY of the Institut des Sciences de l'Evolution de Montpellier (UMR 5554 - CNRS).

REFERENCES

1. Knoll, A. H. (2003) *Life on a young planet: the first three billion years of evolution on Earth* (Princeton University Press, Princeton).
2. Brocks, J. J., Logan, G. A., Buick, R. & Summons, R. E. (1999) *Science* **285,** 1033-1036.
3. Javaux, E. J., Knoll, A. H. & Walter, M. R. (2001) *Nature* **412,** 66-69.
4. Butterfield, N. J. (2000) *Paleobiology* **26,** 386-404.
5. Porter, S. M. & Knoll, A. H. (2000) *Paleobiology* **26,** 360-385.
6. Butterfield, N. J., Knoll, A. H. & Swett, K. (1994) *Fossils and Strata* **34,** 1-84.
7. Chen, J.-Y., Bottjer, D. J., Oliveri, P., Dornbos, S. Q., Gao, F., Ruffins, S., Chi, H., Li, C.-W. & Davidson, E. H. (2004) *Science* **305,** 218-222.
8. Conway Morris, S. (2000) *Proc Natl Acad Sci U S A* **97,** 4426-4429.
9. Benner, S. A., Caraco, M. D., Thomson, J. M. & Gaucher, E. A. (2002) *Science* **296,** 864-868.
10. Zuckerkandl, E. & Pauling, L. (1965) in *Evolving Genes and Proteins*, eds. Bryson, V. & Vogel, H. J. (Academic Press, New York), pp. 97-166.
11. Bromham, L. & Penny, D. (2003) *Nature Rev. Genet.* **4,** 216-224.
12. Hedges, S. B. (2002) *Nat. Rev. Genet.* **3,** 838-849.
13. Feng, D. F., Cho, G. & Doolittle, R. F. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 13028-13033.
14. Ayala, F. J., Rzhetsky, A. & Ayala, F. J. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 606-611.
15. Gu, X. (1998) *J Mol Evol* **47,** 369-371.
16. Lynch, M. (1999) *Evolution* **53,** 319-325.
17. Wang, D. Y., Kumar, S. & Hedges, S. B. (1999) *Proc R Soc Lond B Biol Sci* **266,** 163-171.
18. Bromham, L. D. & Hendy, M. D. (2000) *Proc R Soc Lond B Biol Sci* **267,** 1041-1047.
19. Cutler, D. J. (2000) *Mol. Biol. Evol.* **17,** 1647-1660.
20. Nei, M., Xu, P. & Glazko, G. (2001) *Proc Natl Acad Sci U S A* **98,** 2497-502.
21. Wray, G. A. (2002) *Genome Biol.* **3,** 0001.1-0001.7.
22. Aris-Brosou, S. & Yang, Z. (2003) *Mol. Biol. Evol.* **20,** 1947–1954.
23. Peterson, K. J., Lyons, J. B., Nowak, K. S., Takacs, C. M., Wargo, M. J. & McPeek, M. A. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 6536-6541.
24. Cooper, A. & Fortey, R. (1998) *Trends Ecol. Evol.* **13,** 151-156.

25. Pagel, M. (1999) *Nature* **401,** 877-884.
26. Bromham, L., Penny, D., Rambaut, A. & Hendy, M. D. (2000) *J Mol Evol* **50,** 296-301.
27. Rodriguez-Trelles, F., Tarrio, R. & Ayala, F. J. (2002) *Proc Natl Acad Sci U S A* **99,** 8112-8115.
28. Aris-Brosou, S. & Yang, Z. (2002) *Syst Biol* **51,** 703-14.
29. Hasegawa, M., Thorne, J. L. & Kishino, H. (2003) *Genes Genet. Syst.* **78,** 267-283.
30. Sanderson, M. J. (2003) *Am. J. Bot.* **90,** 954-956.
31. Sanderson, M. J. (1997) *Mol. Biol. Evol.* **14,** 1218-1231.
32. Thorne, J. L., Kishino, H. & Painter, I. S. (1998) *Mol Biol Evol* **15,** 1647-57.
33. Kishino, H., Thorne, J. L. & Bruno, W. J. (2001) *Mol Biol Evol* **18,** 352-61.
34. Doolittle, R. F., Feng, D. F., Tsang, S., Cho, G. & Little, E. (1996) *Science* **271,** 470-7.
35. Philippe, H., Snell, E. A., Bapteste, E., Lopez, P., Holland, P. W. H. & Casane, D. (2004) *Mol. Biol. Evol.* **9,** 1740-1752.
36. Philippe, H., Lopez, P., Brinkmann, H., Budin, K., Germot, A., Laurent, J., Moreira, D., Müller, M. & Le Guyader, H. (2000) *Philos Trans R Soc Lond B Biol Sci* **267,** 1213-1221.
37. Stechmann, A. & Cavalier-Smith, T. (2002) *Science* **297,** 89-91.
38. Snel, B., Bork, P. & Huynen, M. (2000) *Trends Genet.* **16,** 9-11.
39. Sanderson, M. J. & Doyle, J. A. (2001) *Am. J. Bot.* **88,** 1499-1516.
40. Kenrick, P. & Crane, P. R. (1997) *The origin and early diversification of land plants: a cladistic study* (Smithsonian Institution, Washington, D.C., USA).
41. Taylor, T. N., Hass, H. & Kerp, H. (1999) *Nature* **399,** 648.
42. Benton, M. J. (1993) *Fossil record 2* (Harper Collins, Academic, London).
43. Grassly, N. C., Adachi, J. & Rambaut, A. (1997) *Comput. Appl. Biosci.* **13,** 559-560.
44. Baldauf, S. L. (2003) *Science* **300,** 1703-1706.
45. Heckman, D. S., Geiser, D. M., Eidell, B. R., Stauffer, R. L., Kardos, N. L. & Hedges, S. B. (2001) *Science* **293,** 1129-1133.
46. Wray, G. A., Levinton, J. S. & Shapiro, L. H. (1996) *Science* **274,** 568-573.
47. Wiens, J. J. (2003) *Syst. Biol.* **52,** 528-538.
48. Douzery, E. J. P., Delsuc, F., Stanhope, M. J. & Huchon, D. (2003) *J. Mol. Evol.* **57,** S201-S213.
49. Shields, R. (2004) *Trends Genet.* **20,** 221-222.
50. Yoon, H. S., Hackett, J. D., Ciniglia, C., Pinto, G. & Bhattacharya, D. (2004) *Mol. Biol. Evol.* **21,** 809-818.
51. Bromham, L., Phillips, M. J. & Penny, D. (1999) *Trends in Ecology and Evolution* **14,** 113-118.
52. Butterfield, N. J. (2001) *Precambrian Res.* **111,** 235-256.

53. Anbar, A. D. & Knoll, A. H. (2002) *Science* **297,** 1137-1142.

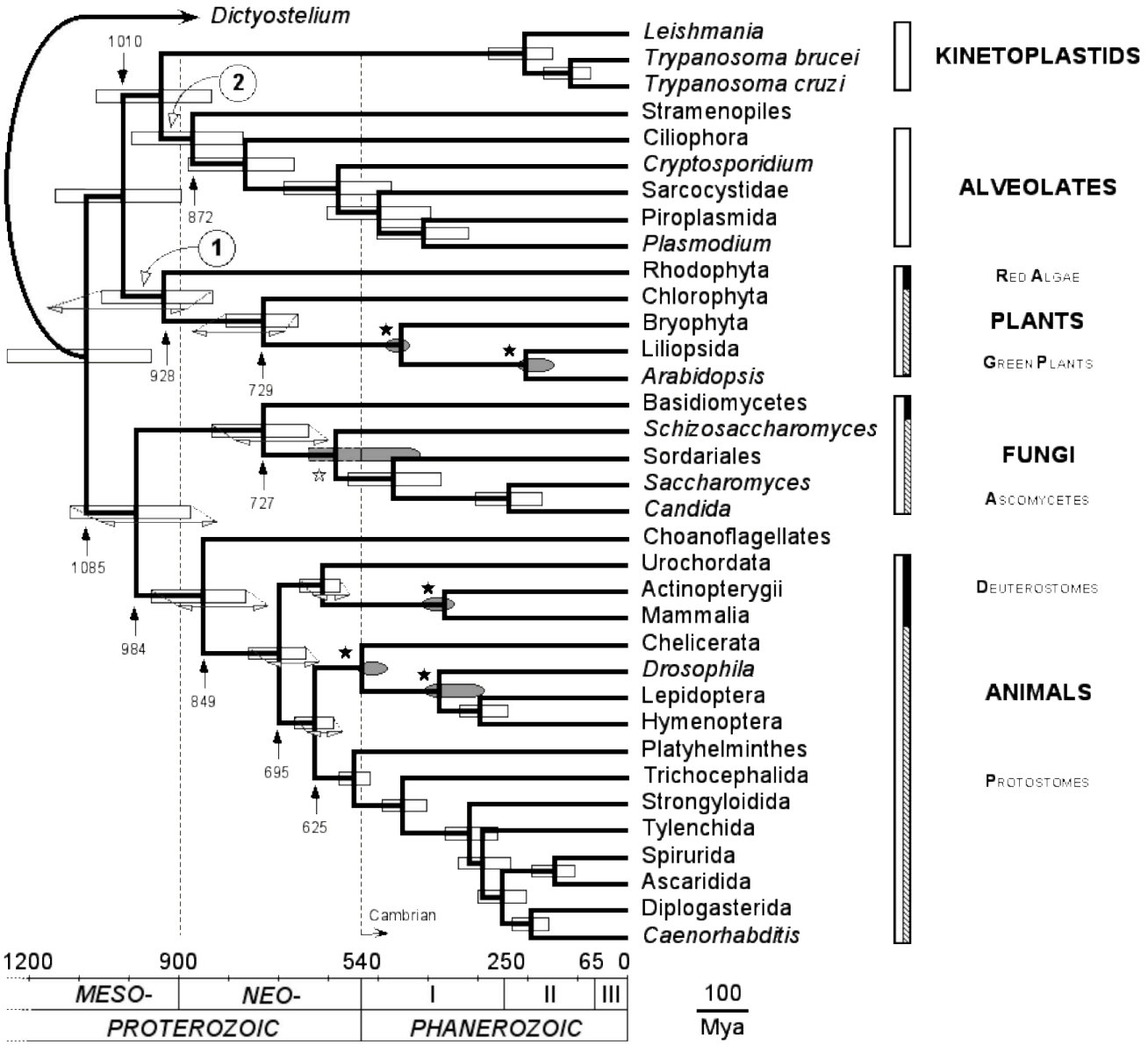
54. Canfield, D. E. & Teske, A. (1996) *Nature* **382,** 127-132.

55. Philippe, H. & Adoutte, A. (1998) in *Evolutionary relationships among Protozoa*, eds. Coombs, G., Vickerman, K., Sleigh, M. & Warren, A. (Kluwer, Dordrecht), pp. 25-56.

56. Lubick, N. (2002) *Nature* **417,** 12-3.

Figure 1. Divergence time estimates (in million years ago [Mya]) among eukaryotes, based on a Bayesian relaxed molecular clock applied to 30,399 amino acid positions. The topology is the highest-likelihood one, with branch lengths proportional to the absolute ages of the subtending nodes. *Dictyostelium* rooted the tree but was pruned from dating analyses. White rectangles delimit 95% credibility intervals on node ages. Stars indicate the six nodes under prior paleontological calibration (lower bound only for the white star ; lower and upper bounds for black stars). Grey areas encompass the bounds between which calibration nodes stand *a posteriori* during the Bayesian search. Primary and secondary plastid endosymbioses are respectively indicated by the circled "1" and "2". Double horizontal arrows and dotted lines indicate the displacement of 95% credibility intervals for 8 selected nodes after re-rooting the tree along the kinetoplastid branch. The Paleozoic, Mesozoic, and Cenozoic are indicated by I, II, and III, respectively. Transitions between Meso- / Neo-Proterozoic and Cambrian are indicated by vertical dashed lines.

Figure 2. Relationship between node rates and node times. For each node of the chronogram, the amino acid replacement rate has been plotted against the estimated divergence times. The Cambrian boundary is indicated by the vertical line. The mean and the ± 2 S.D. of the rates are indicated by the continuous and dashed horizontal lines, respectively.


Dictyostelium
1010
2
872
1
928
729
1085
984
849
695
625
Cambrian
Leishmania
Trypanosoma brucei
Trypanosoma cruzi
Stramenopiles
Ciliophora
Cryptosporidium
Sarcocystidae
Piroplasmida
Plasmodium
Rhodophyta
Chlorophyta
Bryophyta
Liliopsida
Arabidopsis
Basidiomycetes
Schizosaccharomyces
Sordariales
Saccharomyces
Candida
Choanoflagellates
Urochordata
Actinopterygii
Mammalia
Chelicerata
Drosophila
Lepidoptera
Hymenoptera
Platyhelminthes
Trichocephalida
Strongyloidida
Tylenchida
Spirurida
Ascaridida
Diplogasterida
Caenorhabditis
KINETOPLASTIDS
ALVEOLATES
RED ALGAE
PLANTS
GREEN PLANTS
FUNGI
ASCOMYCETES
DEUTEROSTOMES
ANIMALS
PROTOSTOMES
1200
900
540
250
65 0
MESO-
NEO-
I
II
III
PROTEROZOIC
PHANEROZOIC
100
Mya

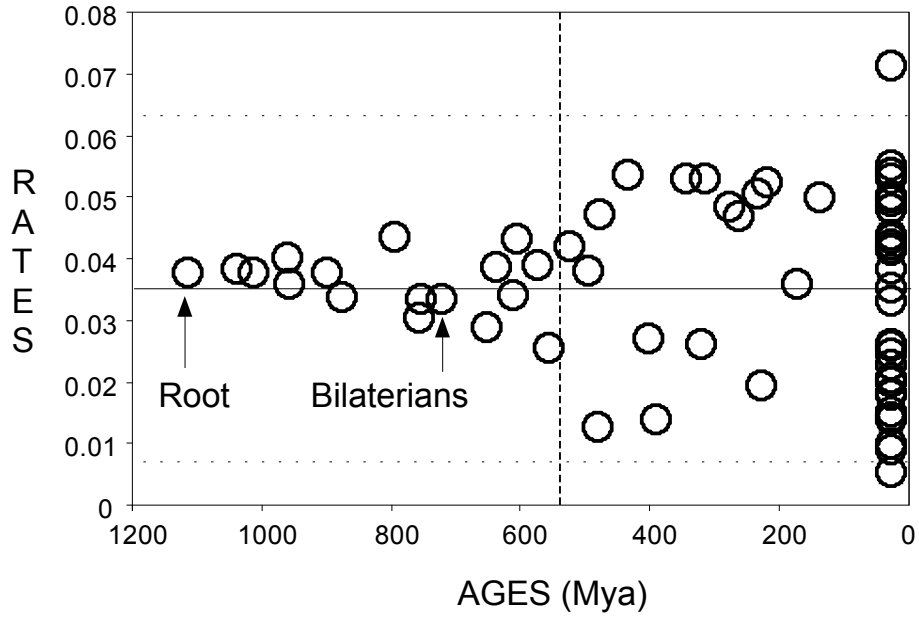

0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
RATES
Root
Bilaterians
1200
1000
800
600
400
200
0
AGES (Mya)

Table 1. Relaxed and global clock average age estimates (in Mya) for bilaterians and the tree root. Divergence times are computed on 30,399 amino acid long data for 36 taxa, simulated under variable (autocorrelated) or constant rates of evolution, by using either a relaxed or a global molecular clock. Expected (simulated) ages are 695 Mya for bilaterians, and 1,085 Mya for the root. The 95% credibility interval widths are given in parentheses. All values are averages given with ± the standard-error on 10 replicates.

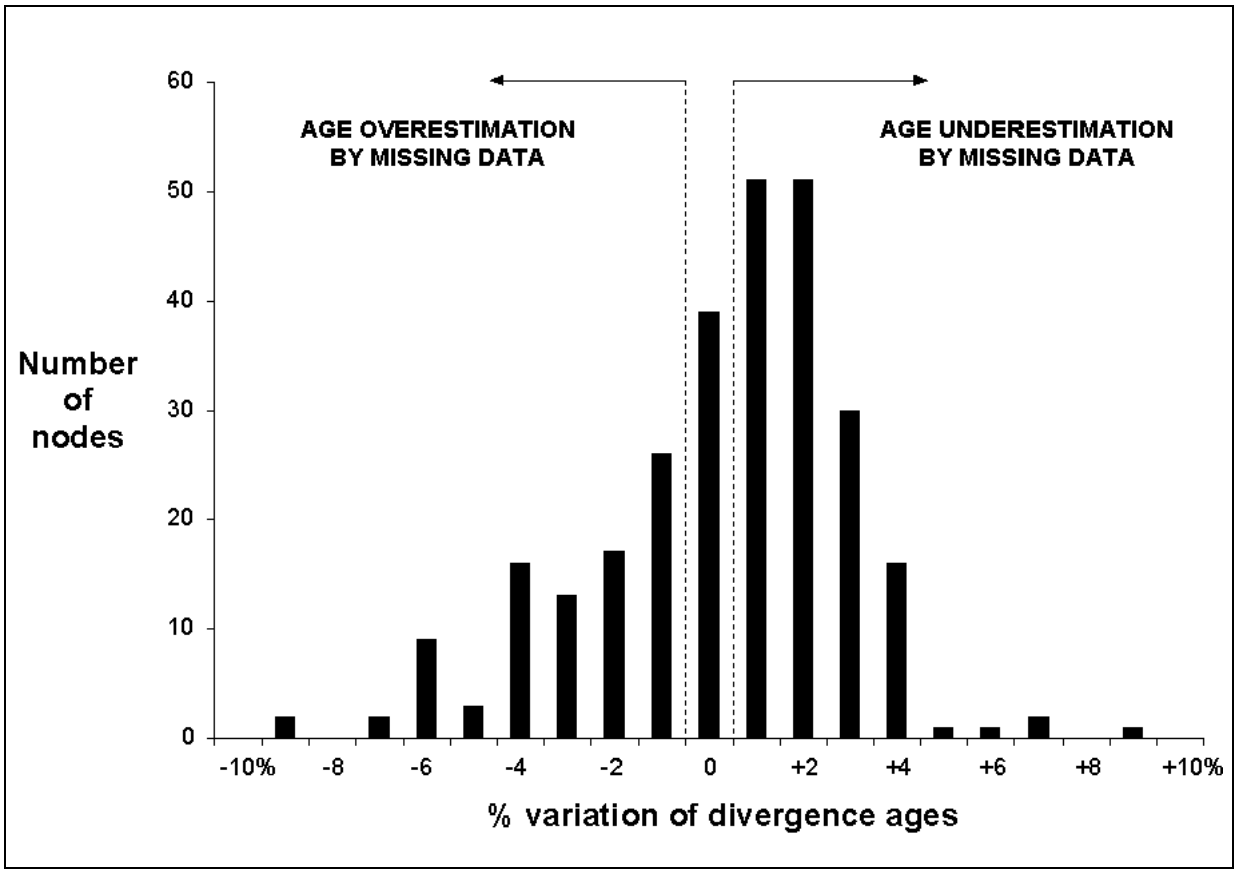
	CLOCK	SIMULATIONS	
		Variable rates	**Constant rate**
BILATERIA	**Relaxed**	692 ± 6 (119 ± 4)	706 ± 7 (52 ± 6)
	Global	864 ± 9 (22 ± 0.4)	700 ± 7 (35 ± 5)
ROOT	**Relaxed**	1,083 ± 15 (308 ± 14)	1,122 ± 11 (89 ± 8)
	Global	1,541 ± 14 (41 ± 1)	1,079 ± 8 (54 ± 7)

Table 2. **Posterior divergence ages for bilaterians and tree root computed on real data under a relaxed molecular clock calibrated by different combinations of one to six time constraints.** Mean ages, range of ages, and mean 95% credibility interval widths (δ CredI) are computed on the 6, 15, 20, 15, 6, or 1 possibilities of incorporation of respectively C = 1, 2, 3, 4, 5, or 6 simultaneous calibrations.

	BILATERIANS		ROOT	
C	**Ages (Range)**	**δ CredI**	**Ages (Range)**	**δ CredI**
1	694 (452-999)	330 ± 201	1,140 (751-1,617)	578 ± 307
2	708 (536-999)	199 ± 97	1,147 (885-1,621)	379 ± 135
3	709 (560-863)	157 ± 73	1,132 (908-1,361)	329 ± 82
4	706 (642-845)	136 ± 56	1,117 (1,029-1,298)	312 ± 68
5	699 (669-757)	123 ± 32	1,099 (1,045-1,174)	301 ± 39
6	695 (—)	119	1,085 (—)	309

SUPPORTING INFORMATION S1.
Sensitivity of Bayesian dating estimates to missing data.

The histogram of the divergence time estimates of the nodes computed with and without missing data is reported. The percentage of variation of divergence ages after introducing 25% random gaps is given, and defined as 100 x [1 - (age of a given node computed with missing data) / (age of the same node computed with complete data)]. It has been computed for 280 nodes = 28 nodes that are not under paleontological constraints x 10 replicates of simulations. Interesting to note, the greatest difference introduced by missing data is always less than 10%, meaning that the potential error on the age of the eukaryotic root due to missing data is inferior to ~100 Myr.



SUPPORTING INFORMATION S2.
Divergence time estimates under a Bayesian relaxed molecular clock for the six calibration nodes.

The age of the six calibration nodes were *a priori* forced to fall into a certain time range, and after the MCMC analysis, posterior estimates of these calibration nodes are reported. Some posterior divergence times lay in the middle of the prior time interval (insects, arthropods), whereas others lay close to one end, either the recent (land plants, deuterostomes) or ancient (flowering plants) bound. Examination of the *a posteriori* calibrations thus did not reveal a systematic trend of posterior estimates to be bound upwards or downwards. The reference topology was rooted by *Dictyostelium*. Geological times containing the posterior estimates are also given.

The posterior Gamma distributions used to relax the molecular clock assumption were characterized by the following mean values, ± 1 standard deviation: ν = 0.00099 ± 0.00030 (prior: 0.001 ± 0.001) for the parameter that controls the degree of rate autocorrelation per million years along the descending branches of the tree, 1,085 ± 79 Mya (prior: 1,000 ± 500 Mya) for the root age, and 0.03775 ± 0.00804 (prior: 0.03641 ± 0.0182) amino acid replacements per 100 sites per million years at the ingroup root.

All absolute ages of the geological periods are taken from the 1999 Geological Time Scale of the Geological Society of America (www.geosociety.org/science/timescale/timescl.pdf).

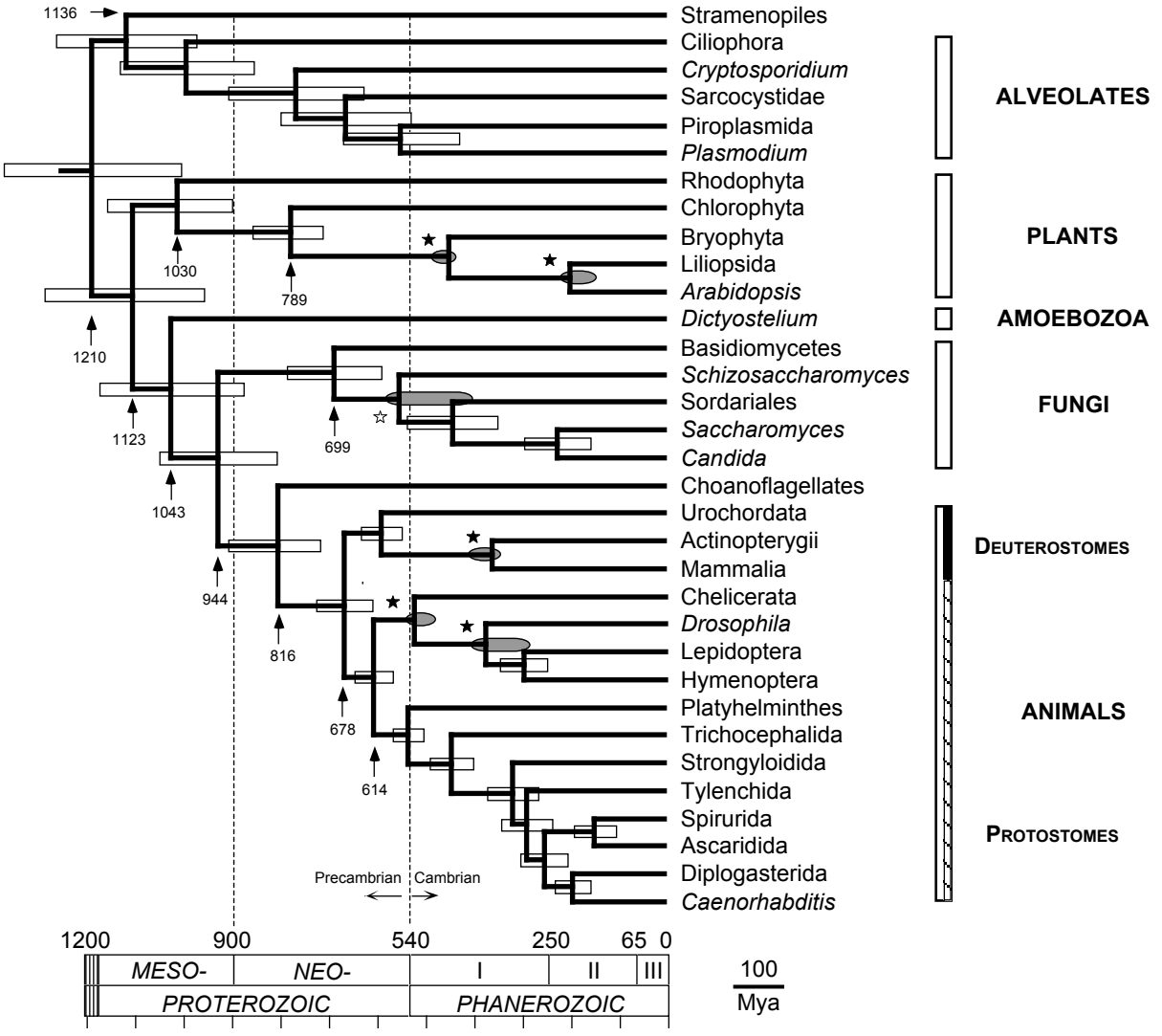
Calibration	**Prior Constraint**	**Posterior 95% Cred. I.**	**Geological Time**
Liliopsida *vs*. eudicots	144-206 Mya	188-206 Mya	Early Jurassic
Bryophyta *vs*. Tracheophyta	443-490 Mya	443-477 Mya	Late Ordovician
Actinopterygii *vs*. Mammalia	354-417 Mya	354-389 Mya	Late Devonian
Diptera *vs*. Lepidoptera + Hymenoptera	290-417 Mya	325-412 Mya	Middle Devonian
Chelicerata *vs*. Insecta	490-543 Mya	502-543 Mya	Cambrian
Schizosaccharomyces vs. Other Ascomycota*	417 Mya	477-687 Mya	Late Proterozoic

* Only a lower bound was constrained.

SUPPORTING INFORMATION S3.

Divergence time estimates (in million years ago [Mya]) for the major groups of eukaryotes, based on a relaxed molecular clock applied to 30,399 amino acid positions, and calibrated by the simultaneous use of six independent paleontological references.

The Kinetoplastidae rooted the tree (leading to a sister-group relationship between Stramenopiles + Alveolates, and a clade containing Plantae associated to *Dictyostelium* + Opisthokonta), but this outgroup was pruned from dating analyses. See the caption of Figure 1 for the conventions used.



Under this *Trypanosoma* + *Leishmania* rooting, the 95% credibility intervals for the divergence age of the following crown groups are: 562-641 Mya for Deuterostomes, 578-654 for Protostomes, 627-738 for Bilaterians, 733-918 for Choanoflagellates + Bilaterians, 833-1,081 for opisthokonts, 162-297 for *Saccharomyces*-*Candida*, 614-797 for Ascomycetes + Basidiomycetes, 705-893 for Chlorophytes + land plants, 899-1,191 for red-green algae, 868-1,178 for alveolates, 983-1,323 for chromalveolates, and 1,042-4,412 for the eukaryotes here considered.

SUPPORTING INFORMATION S4.
Molecular divergence time estimates of some eukaryote groups obtained under a Bayesian relaxed molecular clock calibrated by either a single vertebrate constraint (Actinopterygii *vs*. Mammalia) or the six simultaneous constraints, and comparison with the global clock calibrated by the single vertebrate point.

The reference topology was rooted by the amoeba *Dictyostelium*, and the ingroup topology matched the one depicted on Figure 1. Mean posterior divergence times are given in million years ago, and 95% credibility intervals are provided between parentheses. On the 29 nodes not constrained by prior calibrations, divergence ages based on a single calibration within the slow-evolving vertebrates are 1.4 times deeper than estimates based on the synergetic use of the six calibrations (listed in SUPPORTING INFORMATION S2). This is an indication of the importance of checking dating estimates against multiple independent calibrations. Combining the global clock with the vertebrate calibration involves on average 2.9 deeper age estimates relative to the timings derived from the relaxed clock calibrated by six points.

Number of calibrations →	**One (Vertebrates)**	**Six (Synergy)**	**One (Vertebrates)**
Molecular clock →	**Relaxed**	**Relaxed**	**Global**
Deuterostomes (Urochordates / Vertebrates)	836 Mya (700-997)	610 Mya (572-657)	1171 Mya (1087-1293)
Protostomes (Arthropods / other protostomes)	897 Mya (737-1091)	625 Mya (587-668)	1563 Mya (1450-1728)
Triploblasts (Bilaterians)	996 Mya (813-1217)	695 Mya (642-761)	1597 Mya (1483-1762)
Choanoflagellates / Bilaterians	1237 Mya (991-1535)	849 Mya (761-957)	1988 Mya (1842-2194)
Opisthokonta (Fungi / Animals)	1450 Mya (1150-1815)	984 Mya (872-1127)	2471 Mya (2292-2727)
Saccharomyces* / *Candida	315 Mya (233-417)	235 Mya (168-308)	755 Mya (696-836)
Ascomycetes / Basidiomycetes	1046 Mya (831-1309)	727 Mya (629-837)	1785 Mya (1653-1969)
Chlorophytes / Land plants	999 Mya (779-1268)	729 Mya (662-812)	1490 Mya (1379-1645)
Plants (Chlorophytes / Rhodophytes)	1347 Mya (1066-1690)	928 Mya (825-1061)	2201 Mya (2040-2430)
Alveolates (Ciliophora *vs*. others)	1099 Mya (864-1386)	767 Mya (661-890)	2304 Mya (2134-2544)
Chromalveolates (Stramenopiles *vs*. Alveolates)	1263 Mya (998-1588)	872 Mya (767-1002)	2434 Mya (2255-2686)
Diversification of the eukaryotes here considered	1610 Mya (1269-2025)	1085 Mya (950-1259)	2819 Mya (2615-3112)

SUPPORTING INFORMATION S5.
omparison of the molecular divergence time estimates of some eukaryote groups obtained under the Bayesian relaxed molecular clock and the global clock.

The reference topology was rooted by the amoeba *Dictyostelium*, and the ingroup topology matched the one depicted on Figure 1. Mean posterior divergence times are given in million years ago, and 95% credibility intervals are provided between parentheses. On the 28 nodes not constrained by prior calibrations, global clock estimates are 1.7 times deeper than relaxed clock estimates.

	Relaxed clock	**Global clock**
Deuterostomes (Urochordates / Vertebrates)	610 Mya (572-657)	740 Mya (725-755)
Protostomes (Arthropods / other protostomes)	625 Mya (587-668)	888 Mya (874-902)
Triploblasts (Bilaterians)	695 Mya (642-761)	949 Mya (934-964)
Choanoflagellates / Bilaterians	849 Mya (761-957)	1185 Mya (1162-1208)
Opisthokonta (Fungi / Animals)	984 Mya (872-1127)	1477 Mya (1452-1502)
Saccharomyces* / *Candida	235 Mya (168-308)	454 Mya (439-468)
Ascomycetes / Basidiomycetes	727 Mya (629-837)	1071 Mya (1049-1092)
Chlorophytes / Land plants	729 Mya (662-812)	851 Mya (833-869)
Plants (Chlorophytes / Rhodophytes)	928 Mya (825-1061)	1297 Mya (1272-1322)
Alveolates (Ciliophora *vs*. others)	767 Mya (661-890)	1378 Mya (1348-1409)
Chromalveolates (Stramenopiles *vs*. Alveolates)	872 Mya (767-1002)	1454 Mya (1426-1482)
Diversification of the eukaryotes here considered	1085 Mya (950-1259)	1681 Mya (1654-1709)