Supporting Information

Parfrey et al. 10.1073/pnas.1110633108

SI Text

Calibration Constraints. Calibration constraints (CCs) were assigned from the fossil record and were set to take into account the multiple sources of uncertainty that arise from using paleontological information to calibrate molecular clocks (1–3). The level of informativeness of the priors varied among calibration constraints to reflect to the level of confidence in the timing of the split. For many lineages (including all Proterozoic CCs) the fossil record provides only a minimum divergence time, which is reflected as a very long tail in the prior probability that extends back to ~3500 Ma. In most cases, the CC was placed at the node where the clade with an available fossil split from its sister group; for example the first recorded angiosperm pollen (4) is used to constrain the split of angiosperms from their gymnosperm ancestors (Table 1 and Fig. 2). In cases where the fossil falls within the crown clade, the CC was placed at the base of the clade, as in the Endopterygota where the first Mecoptera fossils constrain the split between *Apis* and *Drosophila* (Table 1). Minimum dates (offsets in BEAST) were assigned conservatively. We used radiometric dates when available, and set the minimum constraint to the youngest edge of the reported confidence interval. Thus, the minimum age of the CC for Arcellinida is 736 Ma, because arcellinid fossils are found in rocks older than 742 ± 6 Ma (Table 1) (5). For fossils assigned to geological stages, we used the upper boundary of the stage according to the 2009 International Stratigraphy Chart published by the International Commission on Stratigraphy (http://www.stratigraphy.org/). For example, angiosperm pollen is first found in Valanginian rocks (4) and so was constrained to a minimum date of 133.9 Ma.

Prior distributions were set in one of two ways depending on the level of uncertainty. For clades with robust fossil records where the maximum age of the clade is unlikely to be substantially earlier than its first occurrence (e.g., angiosperms), the prior distribution was set to include 95% of the probable age of the clade. In contrast, Proterozoic records and fossils of groups with a poor fossilization potential provide only minimum dates for lineage origin and, commonly, no information on maximum clade age (e.g., Arcellinida). In these cases the prior distribution was specified with a very long tail, as assessed in BEAUTi, that extended back to ~3500 Ma.

Selected CCs are discussed here (see Table 1 for details of the remaining CCs). Fossils of the earliest red alga, *Bangiomorpha*, occur in the lower section of the Hunting Formation, Canada, which is bracketed by U-Pb radiometric dates on volcanic rocks of 1267 ± 2 Ma and 723 ± 3 Ma. Direct Pb-Pb dates on carbonates correlative with those containing the fossils yield a much narrow constraint of 1198 ± 24 Ma (6), but this date remains unpublished, and radiometric dating of carbonates can be problematic. The true age of the Hunting *Bangiomorpha* fossils may therefore lie closer to the lower U-Pb age constraint than the upper, because of the sequence stratigraphic position of fossiliferous strata relative to constraining volcanic rocks and chemo- and biostratigraphic data consistent with a later Mesoproterozoic (>1250 Ma) age (6). In most *All 720*, and *Phan* analyses, the minimum date for the *Bangiomorpha* constraint was set at 1174 Ma. Given the importance of the *Bangiomorpha* calibration as potentially the oldest phylogenetically constraining fossil by roughly 450 myr, we also ran the *All 720*, and *Phan* analysis with the constraint for *Bangiomorpha* set at 720 Ma (the minimum age for the Hunting Formation) for comparison (analysis *d*). Because of the controversy surrounding *Bangiomorpha*, we have placed the calibration on the base of the red algae rather than on a particular node within the clade to be conservative.

The single CC in the Excavata is placed within the euglenids. Although the Excavata generally have a poor fossil record, the euglenid *Moyeria* is widely distributed in the Ordovician and Silurian, with an earliest occurrence in the Caradocian (7), dated at 450 Ma. *Moyeria* is thought to have been photosynthetic based on the patterning of its pellicle, indicating an early acquisition of the secondary green alga endosymbiont (8), thus the CC is placed at the split between photosynthetic (*Euglena*) and heterotrophic (*Entosiphon*) euglenids in the tree (Fig. 1).

The calibration constraint for diatoms is based on the earliest diatom fossils from the Valanginian to Hauterivian Myogok Formation in Korea (9); a date of 133.9 Ma is used to represent the upper Valanginian boundary. The CC of this node is younger than in other clock analyses (10) because we do not rely on *Pyxidicula*, a putative Toarcian diatom (11) for which the material has been lost.

We include a CC for ciliates in the *All 720*, and *Phan* analyses that is based on the presence of gammacerane in Neoproterozoic sedimentary rocks (12). Tetrahymenol, the precursor of gammacerane, is commonly found in some ciliates, although it has also been found in bacteria (13). Tetrahymenol production is documented from the Oligohymenophorea and the Plagiopylea (*Trimyena*), which are not included in this analysis, so the CC was placed at the stem of the Oligohymenophorea (*Tetrahymena* and *Paramecium*). This CC was included despite the possibility of bacterial origin, because the 736 Ma constraint is much younger than the date estimated for ciliates (~1150 Ma) without this constraint in the *Phan* analyses.

Root of the Eukaryotic Tree of Life. Although our goal was to elucidate timing of major events in eukaryotic evolution, we also explored the impact of changing the position of the root, because rooted phylogenies are crucial for interpreting the evolutionary events in the history of a lineage. A root must be either provided or estimated for molecular clock analyses (14, 15). However, the root of the eukaryotic tree of life is difficult to determine because the common methods for rooting phylogenies are vulnerable to artifacts caused by rate heterogeneity among lineages of eukaryotes and the vast distance between eukaryotes and archaea or bacteria (16–18). Although numerous hypotheses have been proposed (19–23), the position of the root remains an open debate (16, 17, 24, 25). The most popular hypothesis of recent years places the root of eukaryotes between the Opisthokonta + Amoebozoa (unikonts) and the remaining eukaryotes (bikonts) (19, 26), and previous molecular clock analyses of eukaryotes rooted trees in this manner (10, 27, 28). However, several lines of evidence contradict the unikont/bikont split (23, 24), and alternative roots have been suggested, including at the base of Opisthokonta (23, 29), within Archaeplastida (21, 22), or along the lineage leading to Euglenozoa (20). Rooting the tree of extant eukaryotes along the branch leading to Opisthokonta is supported by ongoing gene-tree species-tree reconciliation work by Gordon Burleigh (University of Florida).

Here, we assess the impact different positions of the root have on estimates of the age of eukaryotes. The root is (*i*) estimated in BEAST using the molecular clock criterion (30); (*ii*) placed between Opisthokonta and the rest of eukaryotes (23, 29); or (*iii*) placed between "Unikonta" and the rest of eukaryotes (19). PhyloBayes requires a fixed topology for molecular dating analyses, hence those analyses were run rooted either on Opistho-

konta or "Unikonta". No strongly supported alternative root emerged from the analyses in which BEAST determined the root based on the molecular clock. Rooting by the molecular clock criterion uses information from branch lengths and has been shown in simulation studies to work reasonably well in situations where there is no outgroup available (31). The position of the root varied across runs, falling between Excavata or Excavata + unikonts and the rest of eukaryotes (analyses *e* and *f*). *All* trees are available in *SI Text* (Figs. S1–S7 and Dataset S1).

Topology of the Eukaryotic Tree. The topology of the trees produced by BEAST through coestimation of phylogeny and dates are consistent with the broad outlines of eukaryotic topologies recovered in other analyses (32, 33). Of note, three relationships emerge in the BEAST topologies that were not present in the RAxML analyses (Dataset S1). RAxML analyses divide Excavata into two clades when all major lineages of Excavata are included (109-taxon set; Dataset S1); however, Excavata is monophyletic in all BEAST analyses (unless the root of eukaryotes falls within Excavata). Second, BEAST consistently places haptophytes as the sister clade of SAR (Fig. 2 and Figs. S1–S7). Finally, cryptomonads consistently branch as sister to kathablepharids and within the clade of primary photosynthetic eukaryotes: red algae, green algae (including plants), and glaucophytes (the Archaeplastida or Plantae hypothesis; Fig. 2 and Figs. S1–S7).

Prior-Only Analyses. The differences in estimated node age between the calibration sets (*Phan* vs. *All*) are driven by the sequence data, rather than the prior distribution of calibration constraints. *All 720* analysis conditions in Table S1 were assessed without the data to determine the impact of priors on estimated divergence dates (prior-only analyses). In BEAST analyses of priors alone, without sequence data, yield dates for major nodes that are 200–800 myr younger than analyses with data. The disparity was much greater for *Phan* analyses: prior-only analyses yielded dates 500–800 myr younger here. For example, the root age is 1478 Ma in the 109-taxon *Phan* analysis *b*, but only 717 Ma when this analysis is run without sequence data. In contrast, all PhyloBayes prior-only analyses produce dates much older than analyses run with the data with the root falling between 3817 Ma and an unreasonable 5047 Ma. These results demonstrate that the CCs chosen did not determine our dates.

SI Materials and Methods

Phylogenetic Analysis. BEAST requires a reasonable starting tree to analyze complex datasets so the initial topology was obtained in RAxML. Two hundred bootstrap replicates followed by an exhaustive maximum likelihood search were done using the MPI version of RaxML 7.0.4 with rapid bootstrapping and the WAG + gamma model (34). The best-fitting amino acid substitution matrix available in BEAST was WAG for all partitions as estimated in ProtTest (35). This resulted in a highly supported topology consistent with that found by RAxML analyses in Parfrey et al. (33) (Dataset S1).

BEAST Model Conditions and Analyses. We ran preliminary analyses in BEAST to assess the impact of several options, including type of molecular clock and partitioning of genes. Analyses were run at Smith College and on the freely available Oslo Bioportal (www.bioportal.uio.no/). Parameters were deemed a poor fit for the data if likelihood values did not converge across four runs of 10 million generations. Based on this criterion, we selected the UCL relaxed clock model combined with unpartitioned genes for subsequent BEAST analyses (analyses *a*–*h*).

A two-pronged approach was used to increase chain mixing, as measured by estimated sample size in Tracer. First, four initial chains of 10 million generations each were run with the best RAxML tree as the starting tree and the remaining priors at default settings (excluding CCs). In the first phase, initial runs were done with a RAxML starting tree that had branch heights (ages) set to 360 so that all nodes were older than the CCs assigned to them. Priors for all parameters (excluding CCs) were left at default settings. Five million generations were removed from each of these chains as burnin (as determined by convergence of likelihood values in Tracer v1.5.4), and chains were combined in LogCombiner v1.5.4 (distributed with BEAST) (36). The final 1 million generations of the preliminary runs were used to generate a starting tree for subsequent analyses that had a robust tree topology with realistic branch heights. Trees were annotated using TreeAnnotator v1.5.4 (36) using the mean node heights and maximum clade credibility tree settings. In the second phase of the analysis, eight runs of 10 million generations each were conducted for each analysis (Table S1). Operator values and prior distributions on substitution rates were informed from the results of initial runs. One million generations were removed from each chain as burnin, and the remaining generations were combined from both log and tree files in LogCombiner v1.5.4. Trees were annotated in TreeAnnotator v1.5.4 and assessed in FigTree v1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/).

Model conditions for BEAST were determined in preliminary analyses of four chains run for 10 million generations each. If likelihood values did not converge across four runs of 10 million generations, as assessed in Tracer v1.5 (distributed with BEAST v1.5.4) (36), the model was deemed a poor fit for the data. The strict clock and uncorrelated exponential molecular clock models were both rejected based on this criterion, as was analyzing the 15 genes as a single partition. If competing models converged in likelihood scores, the likelihoods were compared using Bayes factors (37) as assessed in Tracer, although we did not rely on this metric as the harmonic mean calculation of Bayes factors has been demonstrated to be unreliable (38). In these cases, the estimated divergence dates were compared between competing models. Fixing tree topology to the most likely RAxML tree (Dataset S1) resulted in lower likelihood scores. Further, allowing BEAST to coestimate phylogeny and divergence dates might yield better results for both, so for all subsequent analyses the topology was estimated. *All 720*, and Phanowing BEAST to modify tree topology resulted in a highly supported topology that was broadly consistent with other analyses (32, 33, 39).

Each gene was analyzed as a separate partition for both site models and molecular clock models because the analyses did not converge when the 15 genes were analyzed as a single partition. The WAG amino acid substitution matrix was used for all genes, as it was the best-fitting model available in BEAST as determined by PROTTEST (35). A model of amino acid substitution that included gamma-distributed rate classes was found to be a better fit for the data; however, this resulted in a 10-fold computational cost, thus the gamma correction was used in only a few cases for comparison and yielded similar dates and topologies compared with analyses without a gamma correction. For example, adding gamma correction to BEAST analyses of 109 taxa rooted on opisthokonts with *All* and *Phan* CCs (analyses *a* and *b*) yielded the same topology as Fig. 2 and Fig. S1, respectively, with the age of the root shifted from 1774 Ma to 1668 Ma in analysis *a* and 1478 Ma to 1433 Ma in analysis *b*, where Proterozoic CCs were excluded. These analyses are not included in Table S1 because they were run only 5 million generations (rather than 10 million) due to constraints on compute resources.

The UCL relaxed clock model was found to be the best clock model available in BEAST for these data, as analyses using either a strict molecular clock or an uncorrelated exponential relaxed clock did not converge Only uncorrelated models are implemented in BEAST (36). The UCL relaxed clock is expected to perform better on datasets with deep divergences and rate heterogeneity across the tree, because the SD parameter captures

the variation in rates across the tree (30). The coefficient of variation of the UCL clock ranges from 0.3 to 2.5 for different genes, which indicates that rates vary between 30% and 250% across the tree depending on the gene. Thus, these data are not clock-like.

PhyloBayes Analyses. Analyses were run in PhyloBayes version 3.2f (40). *All* chains were run for at least 1,000 cycles. Calibrations for PhyloBayes were based on the calibrations in BEAST but specified as a date range as required by PhyloBayes with soft bounds at the default setting (Table S3). The model of sequence evolution was the same across all PhyloBayes analyses with a generalized time-reversible (GTR) amino acid substitution matrix (-gtr), a Dirichlet mixture profile (-cat), and a Dirichlet process modeling rates across sites (-ratecat). For each condition, two replicate chains were run. Analyses were run with the tree topology of Fig. 2, which was fixed and rooted either on the Opisthokonta or "Unikonta". Analyses were run under either the CIR molecular clock model, in which rates across branches are autocorrelated, or the UGAM model, which is nonautocorrelated and thus similar to BEAST (40, 41). There is much debate as to whether substitution rates are best modeled as autocorrelated across the tree or uncorrelated (30, 41–43). Autocorrelated models of the molecular clock assume that evolutionary rates along a branch are dependent on the rate of the parent branch (16, 41), whereas uncorrelated models draw rates of evolution for each branch from a distribution (30, 42). These clock models were chosen because CIR was shown to be a good fit for many different datasets (41), and UGAM is similar to the uncorrelated model run in BEAST. CIR (logBF 61) was preferred to UGAM (logBF 32) in Bayes factor analyses comparing clock models to deconstrained models in PhyloBayes. Dates were assessed by running readdiv with 250 generations removed as burn-in for each analysis. The mean dates were averaged, and the error bars were derived from the overall minimum and maximum of the 95% confidence interval for the two chains.

1. Donoghue PCJ, Benton MJ (2007) Rocks and clocks: Calibrating the Tree of Life using fossils and molecules. *Trends Ecol Evol* 22:424–431.
2. Ho SYW, Phillips MJ (2009) Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. *Syst Biol* 58:367–380.
3. Rutschmann F, Eriksson T, Salim KA, Conti E (2007) Assessing calibration uncertainty in molecular dating: The assignment of fossils to alternative calibration points. *Syst Biol* 56:591–608.
4. Crane PR, Friis EM, Pedersen KR (1995) The origin and early diversification of angiosperms. *Nature* 374:27–33.
5. Porter SM, Meisterfeld R, Knoll AH (2003) Vase-shaped microfossils from the Neoproterozoic Chuar Group, Grand Canyon: A classification guided by modern testate amoebae. *J Paleontol* 77:409–429.
6. Butterfield NJ (2000) *Bangiomorpha pubescens* n. gen., n. sp.: Implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology* 26:386–404.
7. Gray J, Boucot AJ (1989) Is *Moyeria* a euglenoid? *Lethaia* 22:447–456.
8. Leander BS, Witek RP, Farmer MA (2001) Trends in the evolution of the euglenid pellicle. *Evolution* 55:2215–2235.
9. Harwood DM, Nikolaev VA, Winter DM (2007) Cretaceous records of diatom evolution, radiation, and expansion. *Paleontol Soc Papers* 13:33–59.
10. Berney C, Pawlowski J (2006) A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. *Proc Biol Sci* 273:1867–1872.
11. Rothpletz A (1896) On the flysch fucoids and a few other fossil algae, as well as diatoms from Liassic sponge reefs (Translated from German). *Z Dtsch Geol Ges* 52: 154–160.
12. Summons RE, Walter MR (1990) Molecular fossils and microfossils of prokaryotes and protists from Proterozoic sediments. *Am J Sci* 290-A:212–244.
13. Kleemann G, et al. (1990) Tetrahymenanol from the phototrophic bacterium *Rhodopseudomonas palustris*: First report of a gammacerane triterpene from a prokaryote. *J Gen Microbiol* 136:2551–2553.
14. Renner SS, Grimm GW, Schneeweiss GM, Stuessy TF, Ricklefs RE (2008) Rooting and dating maples (*Acer*) with an uncorrelated-rates molecular clock: Implications for north American/Asian disjunctions. *Syst Biol* 57:795–808.
15. Sanderson MJ, Doyle JA (2001) Sources of error and confidence intervals in estimating the age of angiosperms from rbcL and 18S rDNA data. *Am J Bot* 88:1499–1516.
16. Roger AJ, Hug LA (2006) The origin and diversification of eukaryotes: Problems with molecular phylogenetics and molecular clock estimation. *Philos Trans R Soc Lond B Biol Sci* 361:1039–1054.
17. Embley TM, Martin W (2006) Eukaryotic evolution, changes and challenges. *Nature* 440:623–630.
18. Tekle YI, Parfrey LW, Katz LA (2009) Molecular data are transforming hypotheses on the origin and diversification of eukaryotes. *Bioscience* 59:471–481.
19. Stechmann A, Cavalier-Smith T (2003) The root of the eukaryote tree pinpointed. *Curr Biol* 13:R665–R666.
20. Cavalier-Smith T (2010) Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. *Biol Lett* 6:342–345.
21. Nozaki H (2005) A new scenario of plastid evolution: Plastid primary endosymbiosis before the divergence of the "Plantae," emended. *J Plant Res* 118:247–255.
22. Rogozin IB, Basu MK, Csürös M, Koonin EV (2009) Analysis of rare genomic changes does not support the unikont-bikont phylogeny and suggests cyanobacterial symbiosis as the point of primary radiation of eukaryotes. *Genome Biol Evol* 1:99–113.
23. Arisue N, Hasegawa M, Hashimoto T (2005) Root of the Eukaryota tree as inferred from combined maximum likelihood analyses of multiple molecular sequence data. *Mol Biol Evol* 22:409–420.
24. Roger AJ, Simpson AGB (2009) Evolution: Revisiting the root of the eukaryote tree. *Curr Biol* 19:R165–R167.
25. Koonin EV (2010) The origin and early evolution of eukaryotes in the light of phylogenomics. *Genome Biol* 11:209.
26. Keeling PJ, et al. (2005) The tree of eukaryotes. *Trends Ecol Evol* 20:670–676.
27. Douzery EJP, Snell EA, Bapteste E, Delsuc F, Philippe H (2004) The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? *Proc Natl Acad Sci USA* 101:15386–15391.
28. Hug LA, Roger AJ (2007) The impact of fossils and taxon sampling on ancient molecular dating analyses. *Mol Biol Evol* 24:1889–1897.
29. Stechmann A, Cavalier-Smith T (2002) Rooting the eukaryote tree by using a derived gene fusion. *Science* 297:89–91.
30. Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *PLoS Biol* 4:e88.
31. Huelsenbeck JP, Bollback JP, Levine AM (2002) Inferring the root of a phylogenetic tree. *Syst Biol* 51:32–43.
32. Hampl V, et al. (2009) Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic "supergroups". *Proc Natl Acad Sci USA* 106: 3859–3864.
33. Parfrey LW, et al. (2010) Broadly sampled multigene analyses yield a well-resolved eukaryotic tree of life. *Syst Biol* 59:518–533.
34. Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML Web servers. *Syst Biol* 57:758–771.
35. Abascal F, Zardoya R, Posada D (2005) ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* 21:2104–2105.
36. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7:214.
37. Suchard MA, Weiss RE, Sinsheimer JS (2001) Bayesian selection of continuous-time Markov chain evolutionary models. *Mol Biol Evol* 18:1001–1013.
38. Xie W, Lewis PO, Fan Y, Kuo L, Chen MH (2011) Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Syst Biol* 60:150–160.
39. Burki F, et al. (2009) Large-scale phylogenomic analyses reveal that two enigmatic protist lineages, telonemia and centroheliozoa, are related to photosynthetic chromalveolates. *Genome Biol Evol* 1:231–238.
40. Lartillot N, Lepage T, Blanquart S (2009) PhyloBayes 3: A Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25:2286–2288.
41. Lepage T, Bryant D, Philippe H, Lartillot N (2007) A general comparison of relaxed molecular clock models. *Mol Biol Evol* 24:2669–2680.
42. Linder M, Britton T, Sennblad B (2011) Evaluation of Bayesian models of substitution rate evolution—parental guidance versus mutual independence. *Syst Biol* 60:329–342.
43. Ho SYW (2009) An examination of phylogenetic models of substitution rate variation among lineages. *Biol Lett* 5:421–424.

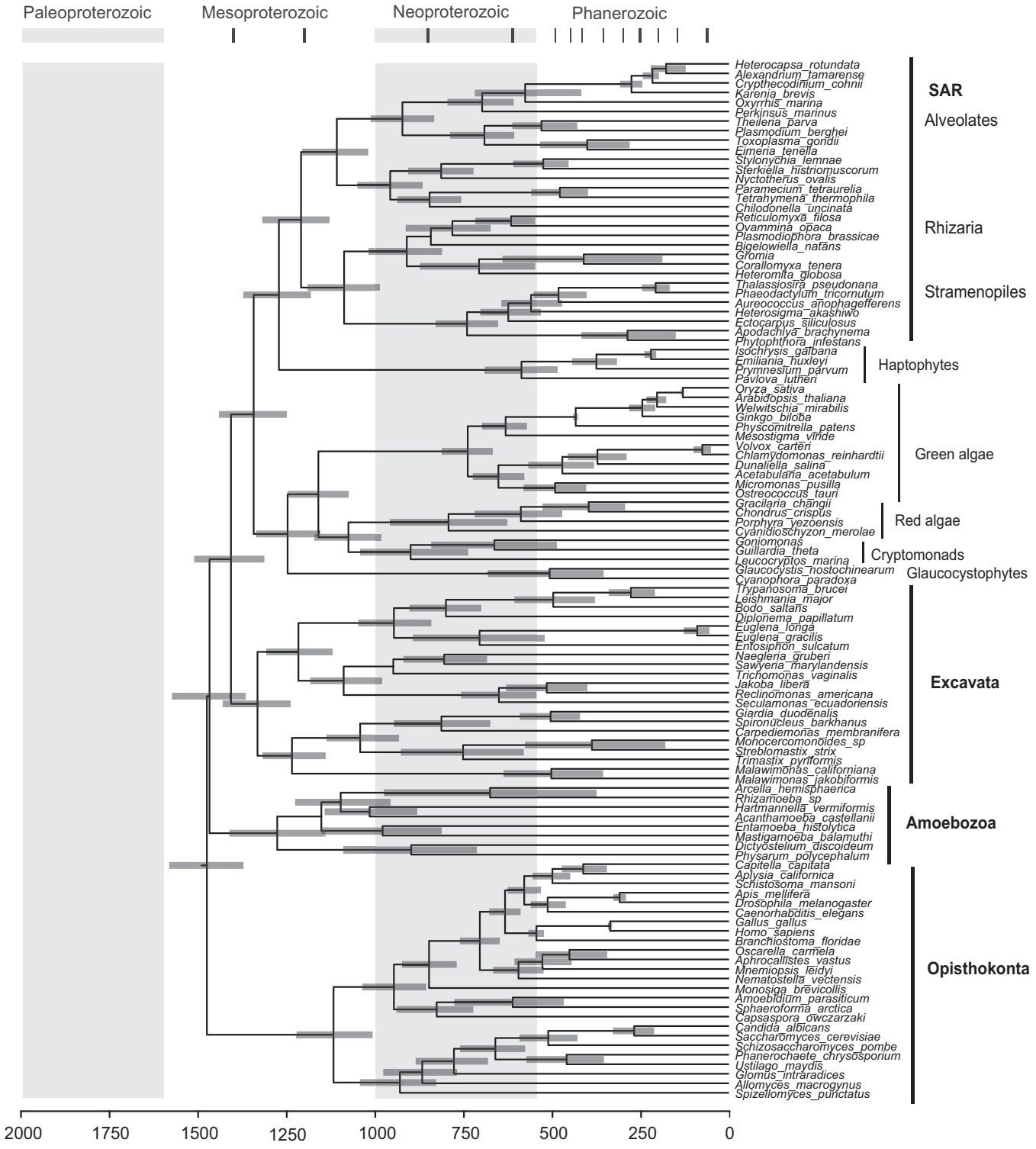

Fig. S1. Time-calibrated tree of eukaryotes using Phanerozoic calibration points, 109 taxa, rooted on Opisthokonta, and constructed in BEAST (analysis *b*). Nodes are at mean divergence times, and gray bars represent 95% HPD of node age. (*Upper*) Geological time scale. (*Lower*) Absolute time scale (in Ma). Thick vertical bars demarcate eras, and thin vertical lines denote periods, with dates derived from the 2009 International Stratigraphic Chart.

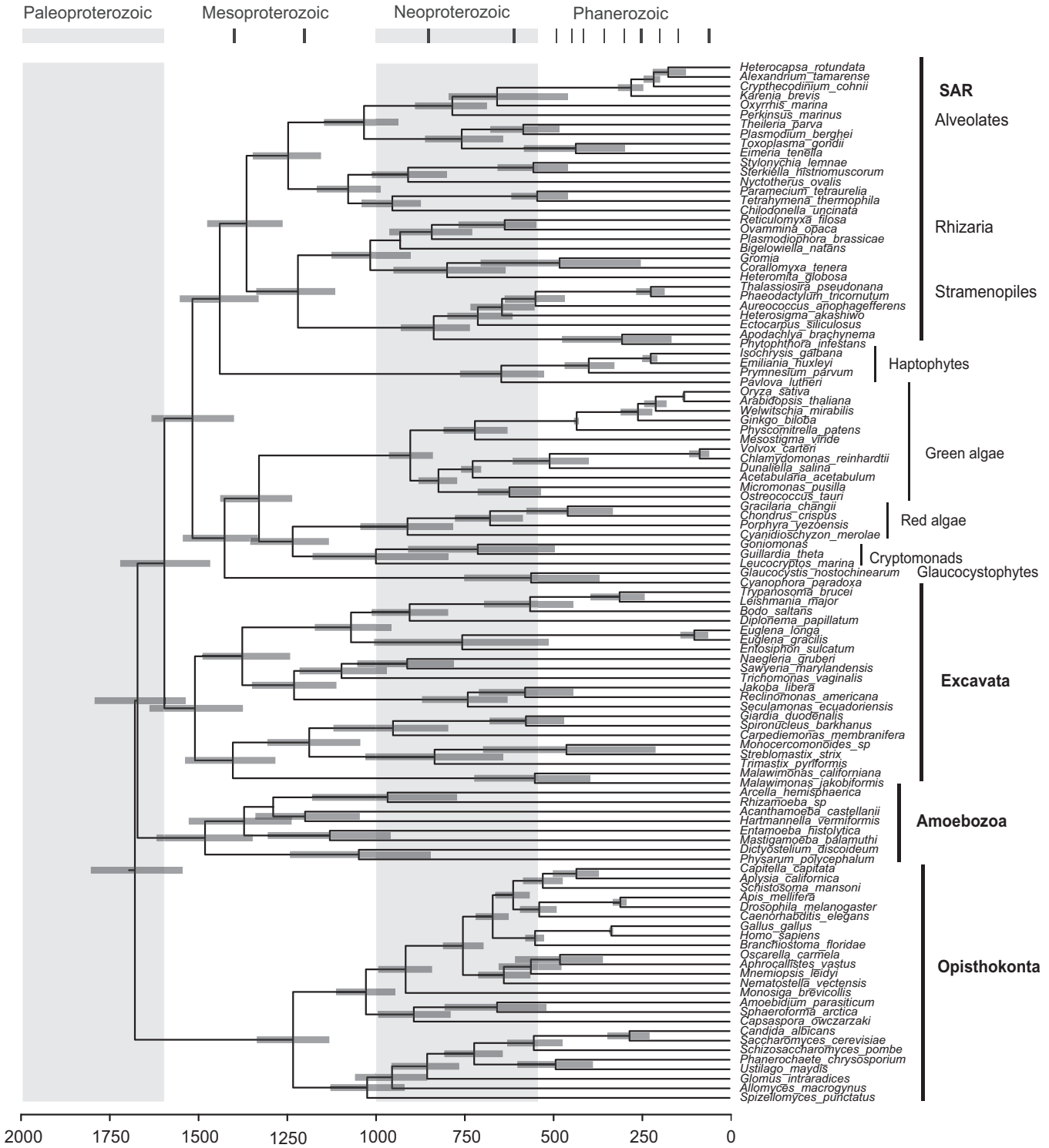

Fig. S2. Time-calibrated tree of eukaryotes using *All* (Proterozoic and Phanerozoic) calibration points with the *Bangiomorpha* CC set at 720 Ma, 109 taxa, rooted on Opisthokonta, and constructed in BEAST (analysis *c*). Other notes as in Fig. S1.

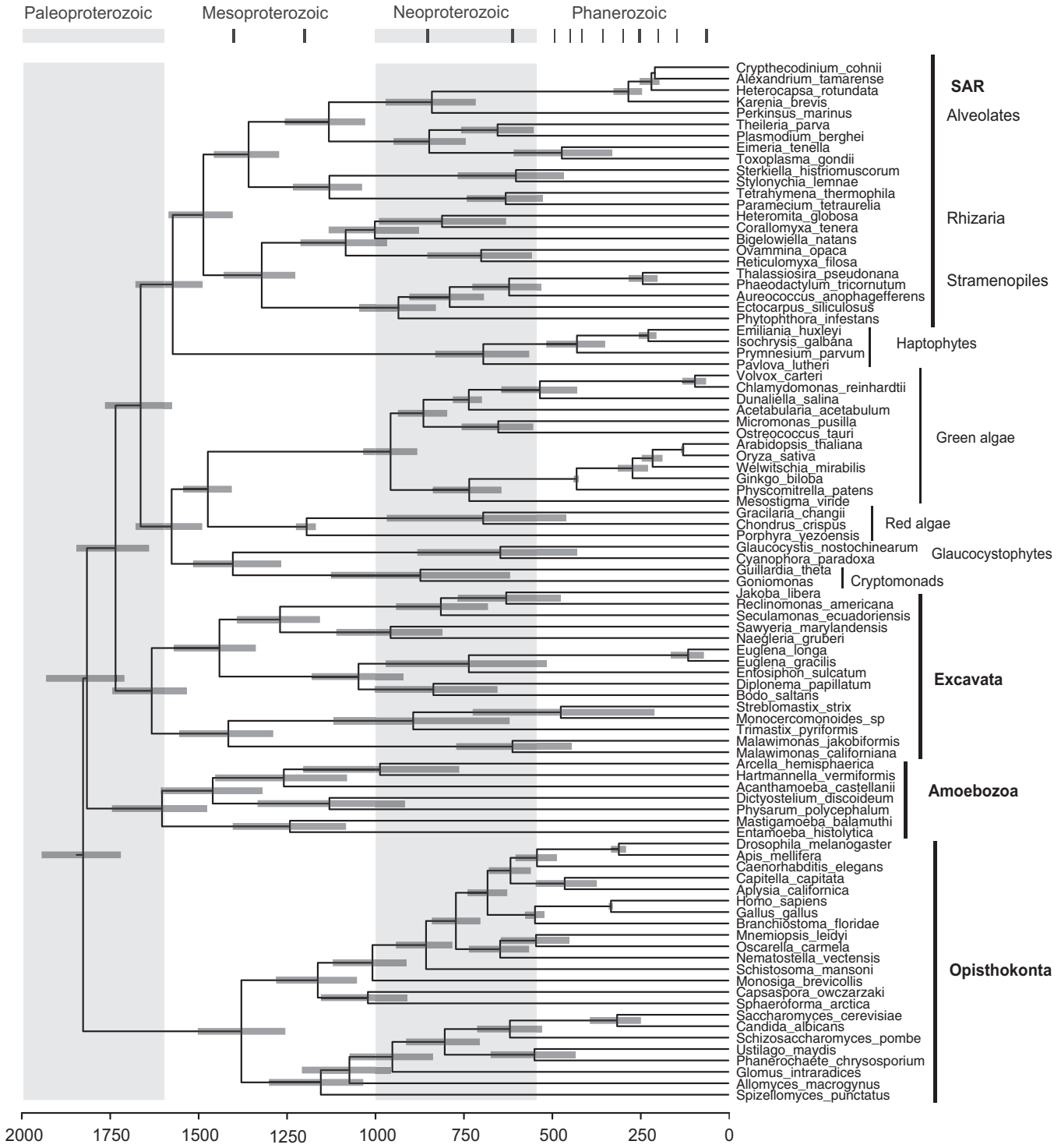

Fig. S3. Time-calibrated tree of eukaryotes using *All* calibration points, 91 taxa, rooted on Opisthokonta, and constructed in BEAST (analysis *d*). Other notes as in Fig. S1.

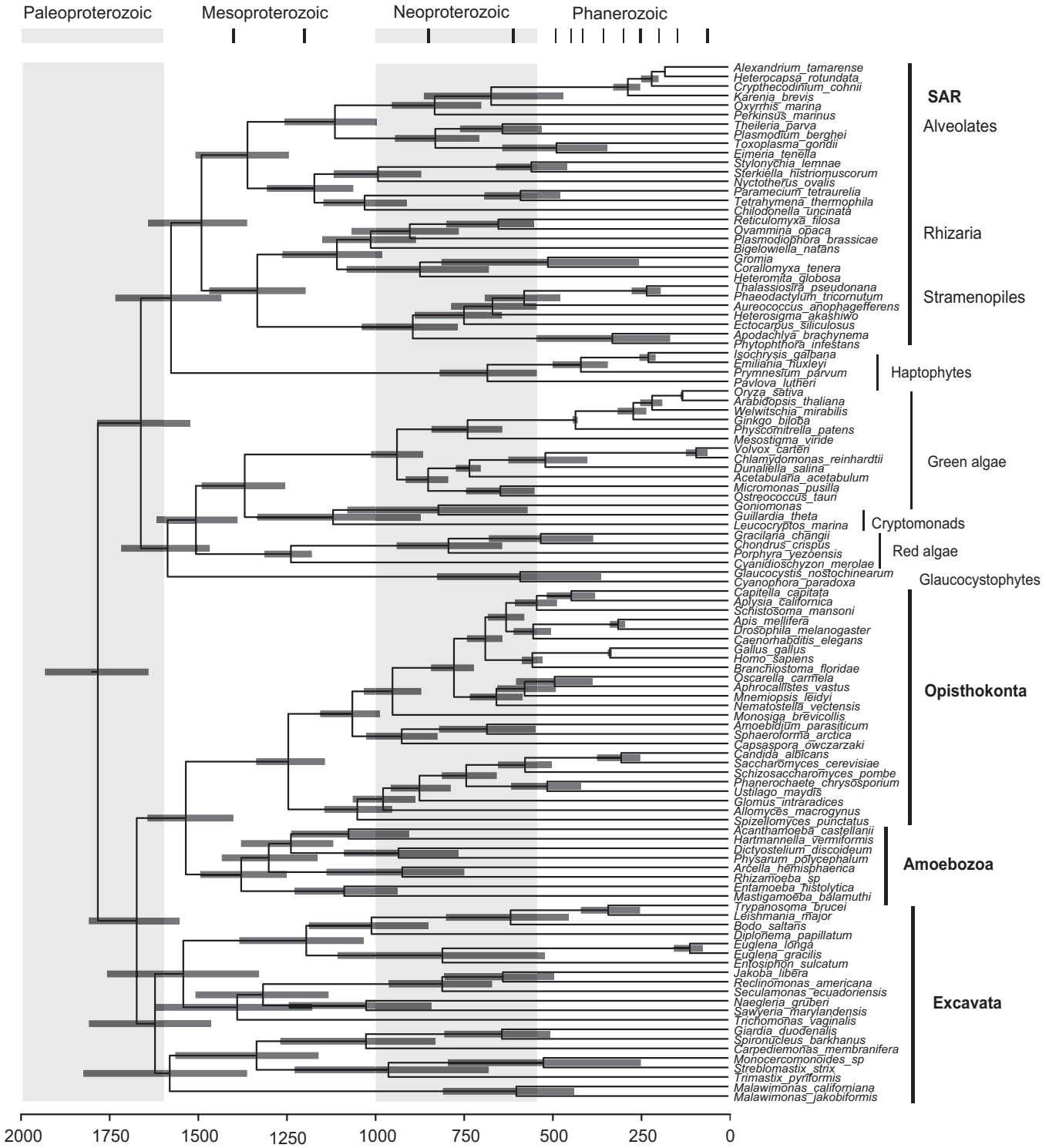

Fig. S4. Time-calibrated tree of eukaryotes using *All* calibration points, 109 taxa, root estimated by BEAST, and constructed in BEAST (analysis *e*). Other notes as in Fig. S1.

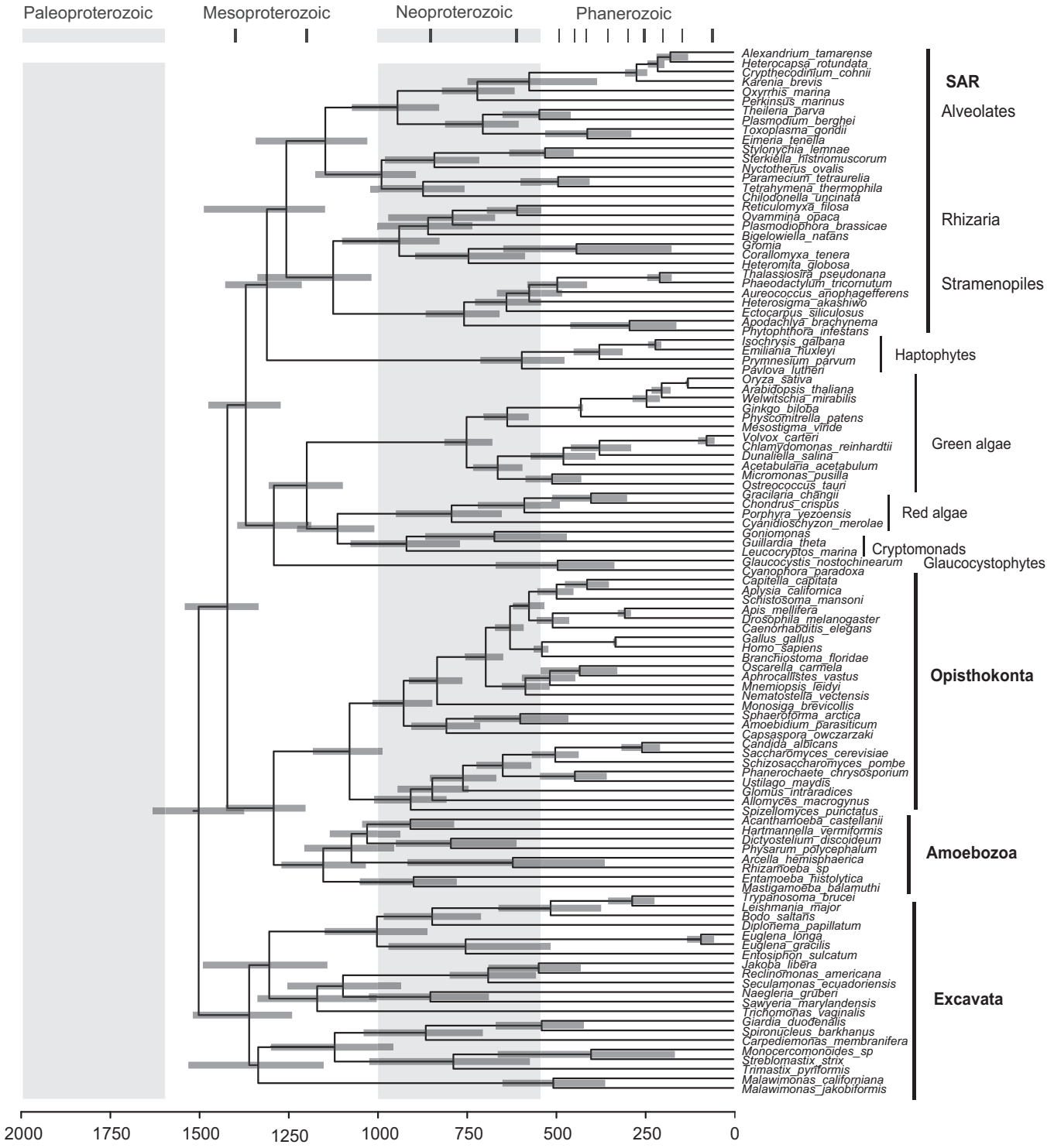

Fig. S5. Time-calibrated tree of eukaryotes using Phanerozoic calibration points, 109 taxa, root estimated by BEAST, and constructed in BEAST (analysis *f*). Other notes as in Fig. S1.

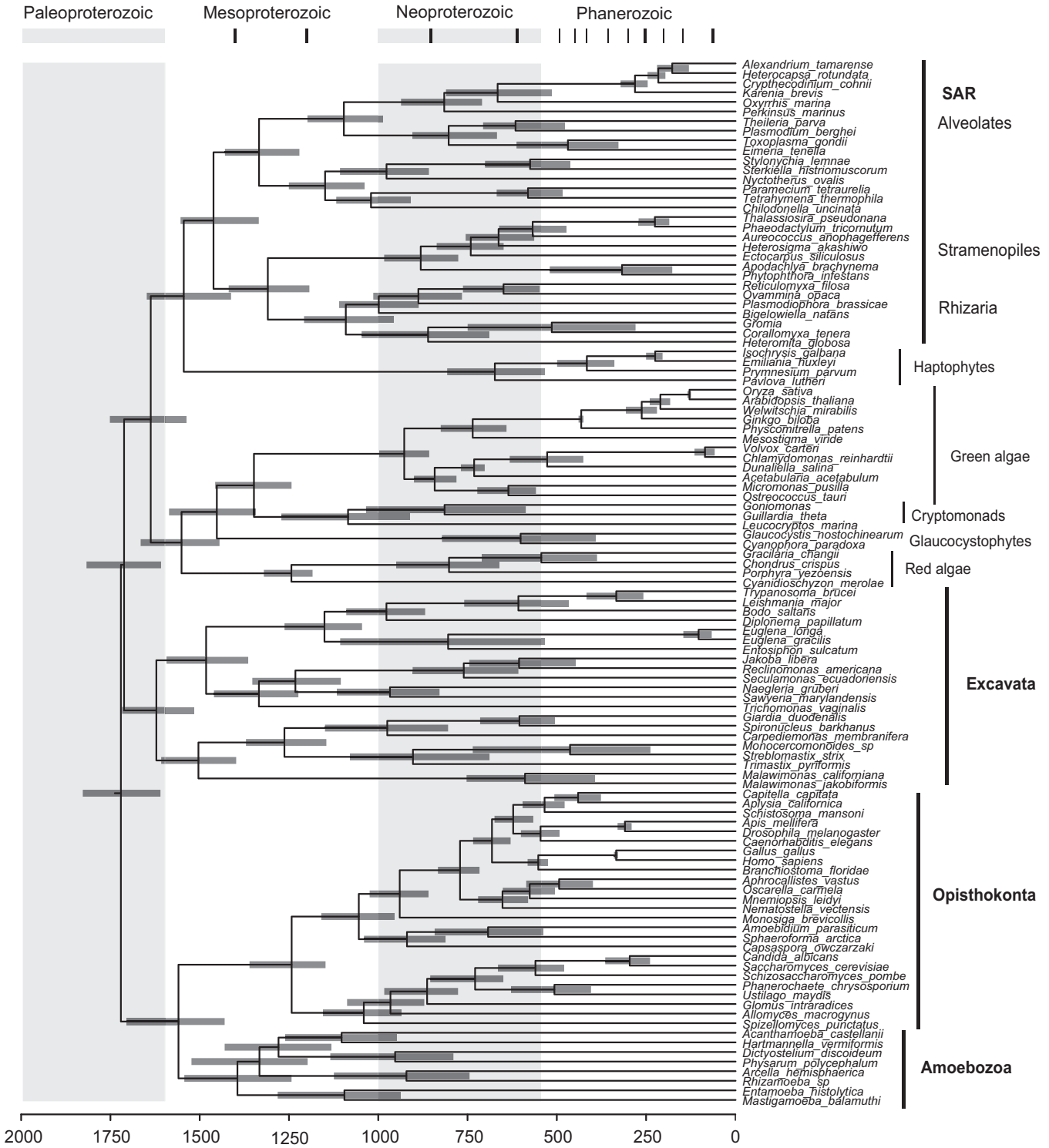

Fig. S6. Time-calibrated tree of eukaryotes using *All* calibration points, 109 taxa, rooted on "Unikonta" and constructed in BEAST (analysis *g*). Other notes as in Fig. S1.

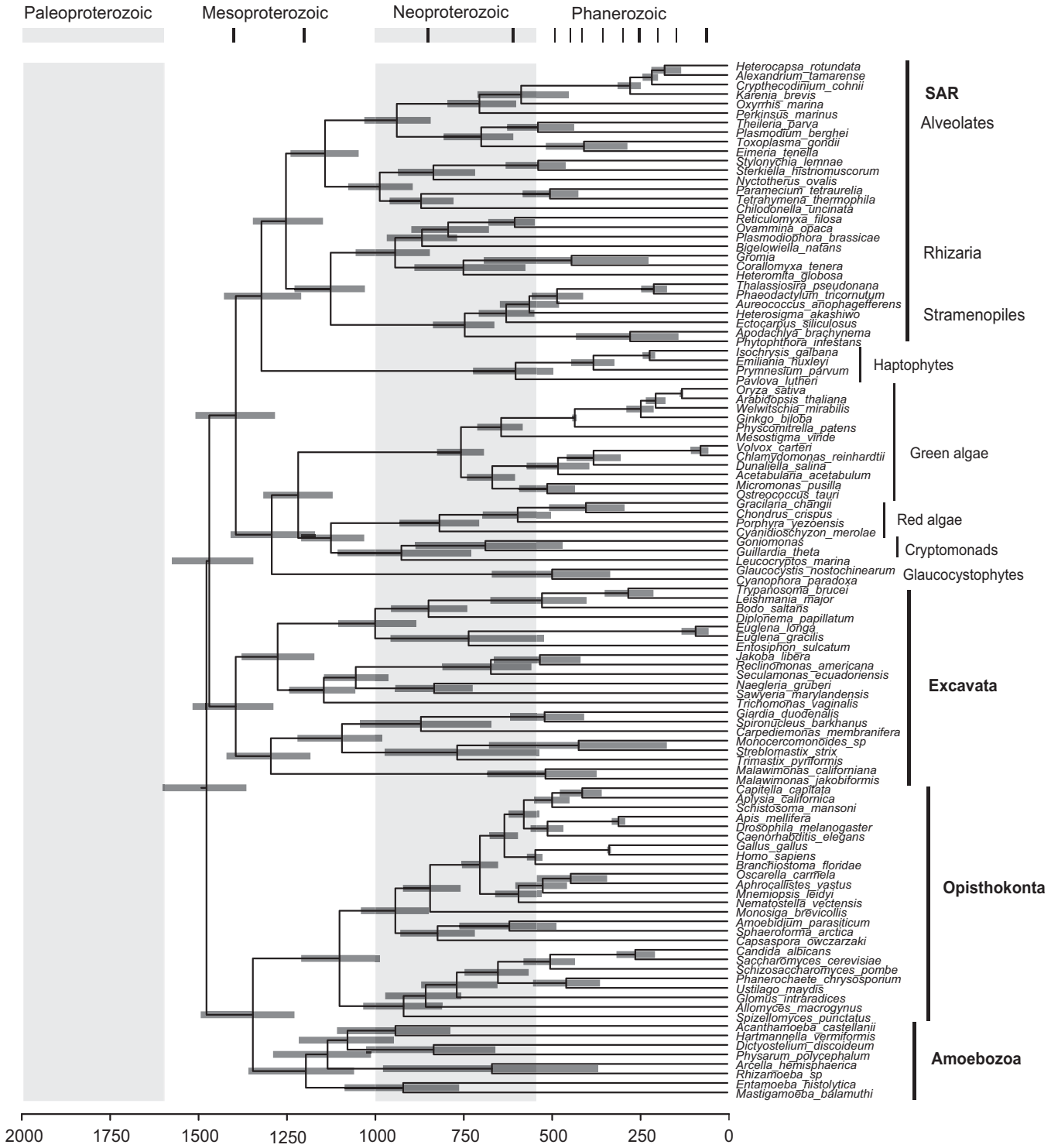

Fig. S7. Time-calibrated tree of eukaryotes using Phanerozoic calibration points, 109 taxa, rooted on "Unikonta" and constructed in BEAST (analysis *h*). Other notes as in Fig. S1.

Table S1. Estimates of dates for the last common ancestor of extant eukaryotes across analyses

Analysis	Taxa	CCs	Root	Root age, Ma		Model	Program	Tree
				Mean	Range			
a	109	*All*	Opis	1774	1632–1911	UCL	BEAST	Fig. 2
b	109	*Phan*	Opis	1478	1362–1595	UCL	BEAST	Fig. S1
c	109	*All 720*	Opis	1679	1548–1797	UCL	BEAST	Fig. S2
d	91	*All*	Opis	1837	1725–1954	UCL	BEAST	Fig. S3
e	109	*All*	Estim	1784	1639–1939	UCL	BEAST	Fig. S4
f	109	*Phan*	Estim	1506	1365–1643	UCL	BEAST	Fig. S5
g	109	*All*	Uni	1717	1601–1819	UCL	BEAST	Fig. S6
h	109	*Phan*	Uni	1471	1347–1604	UCL	BEAST	Fig. S7
i	109	*All*	Opis	1866	1569–2235	UGAM	PhyloBayes	—
j	109	*Phan*	Opis	1594	1288–1979	UGAM	PhyloBayes	—
k	109	*All*	Uni	1810	1549–2161	UGAM	PhyloBayes	—
l	109	*Phan*	Uni	1561	1268–1886	UGAM	PhyloBayes	—
m	109	*All*	Opis	1798	1441–2133	CIR	PhyloBayes	—
n	109	*Phan*	Opis	1038	889–1350	CIR	PhyloBayes	—
o	109	*All*	Uni	1691	1048–2357	CIR	PhyloBayes	—
p	109	*Phan*	Uni	1180	897–1839	CIR	PhyloBayes	—

Root age range is the 95% HPD for BEAST analyses and minimum and maximum ages of 95% confidence interval for PhyloBayes. See Table S2 for details of taxon sampling, and Table 1 for calibration constraints. All trees are available in Dataset S1. *All*, 22 calibration points of Phanerozoic and Proterozoic age included; *All 720*, *Bangiomorpha* CC set to 720 Ma; CCs, calibration constraints; CIR, autocorrelated CIR model; Estim, root estimated by BEAST; model, molecular clock model; Opis, root constrained to Opisthokonta; *Phan*, calibration points of Phanerozoic age included; root, position of the root; UCL, uncorrelated log normal; UGAM, uncorrelated gamma model; Uni, root constrained to "Unikonta".

Table S2. Details of gene and taxon sampling

Lineage	Taxon*	14–3-3	40S	Actin	αtub	βtub	Ef1α	Ef2	Enolase	Grc5	Hsp70cyt	Hsp90	MetK	Rps22a	Rps23a	Tsec61	Sum
Alveolates	***Alexandrium tamarense***	1	1	1	1			1	1	1	1	1	1	1	1	1	14
Alveolates	*Chilodonella uncinata*	1		1	1	1	1			1		1		1	1		10
Alveolates	***Crypthecodinium cohnii***			1		1					1	1					5
Alveolates	***Eimeria tenella***	1		1		1	1		1			1	1	1	1		10
Alveolates	***Heterocapsa rotundata***	1		1	1	1										1	6
Alveolates	***Karenia brevis***	1		1	1	1		1	1		1	1	1			1	11
Alveolates	*Nyctotherus ovalis*	1	1	1	1	1	1	1		1		1			1	1	12
Alveolates	*Oxyrrhis marina*			1	1	1						1			1		6
Alveolates	***Paramecium tetraurelia***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Alveolates	***Perkinsus marinus***	1	1	1	1	1		1	1	1	1	1	1	1	1	1	15
Alveolates	***Plasmodium berghei***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Alveolates	***Sterkiella histriomuscorum***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Alveolates	***Stylonychia lemnae***			1	1	1	1				1						6
Alveolates	***Tetrahymena thermophila***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Alveolates	***Theileria parva***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Alveolates	***Toxoplasma gondii***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Amoebozoa	***Acanthamoeba castellanii***	1	1	1	1	1	1	1	1	1	1	1	1	1	1		15
Amoebozoa	***Arcella hemisphaerica***			1	1												3
Amoebozoa	***Dictyostelium discoideum***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Amoebozoa	***Entamoeba histolytica***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Amoebozoa	***Hartmannella vermiformis***	1		1	1	1	1	1		1	1	1	1	1	1	1	14
Amoebozoa	***Mastigamoeba balamuthi***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Amoebozoa	***Physarum polycephalum***	1	1	1	1	1	1	1	1	1		1	1	1	1	1	15
Amoebozoa	*Rhizamoeba* sp. ATCC 50933			1	1	1											4
Animals	*Aphrocallistes vastus*				1	1					1	1					5
Animals	***Apis mellifera***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Animals	***Aplysia californica***	1	1	1	1	1	1		1	1		1	1	1	1	1	14
Animals	***Branchiostoma floridae***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Animals	***Caenorhabditis elegans***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Animals	***Capitella capitata***	1	1	1					1	1	1		1	1	1	1	11
Animals	***Drosophila melanogaster***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Animals	***Gallus gallus***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Animals	***Homo sapiens***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Animals	***Mnemiopsis leidyi***	1	1	1	1	1	1			1		1		1	1	1	12
Animals	***Nematostella vectensis***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Animals	***Oscarella carmela***	1	1	1	1	1	1	1		1	1	1				1	12
Animals	***Schistosoma mansoni***	1	1	1	1		1		1	1	1	1	1	1	1	1	14
Choanoflagellida	***Monosiga brevicollis***	1	1	1	1	1		1	1	1	1	1	1	1	1	1	15
Cryptophyta	***Goniomonas*†**		1	1	1	1	1	1		1		1			1	1	11
Cryptophyta	***Guillardia theta***	1		1	1	1			1		1	1		1		1	10
Euglenozoa	***Bodo saltans***			1	1	1	1				1	1			1		8
Euglenozoa	***Diplonema papillatum***	1	1	1	1	1		1		1	1	1	1	1	1	1	14
Euglenozoa	***Entosiphon sulcatum***				1	1						1					4
Euglenozoa	***Euglena gracilis***	1		1	1	1	1	1	1	1	1	1		1	1	1	14
Euglenozoa	***Euglena longa***	1	1	1	1	1	1	1	1	1	1	1	1		1		14
Euglenozoa	*Leishmania major*	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16

Table S2. Cont.

Lineage	Taxon*	14–3-3	40S	Actin	αtub	βtub	Ef1α	Ef2	Enolase	Grc5	Hsp70cyt	Hsp90	MetK	Rps22a	Rps23a	Tsec61	Sum
Euglenozoa	*Trypanosoma brucei*	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Fornicata	*Carpediemonas membranifera*				1	1					1	1					5
Fornicata	*Giardia duodenalis* ATCC 50803	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Fornicata	*Spironucleus barkhanus*		1		1	1	1	1	1	1	1	1		1	1		12
Fungi	***Allomyces macrogynus***	1	1	1	1	1			1	1	1	1		1	1	1	13
Fungi	***Candida albicans***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Fungi	***Glomus intraradices***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Fungi	***Phanerochaete chrysosporium***	1	1	1	1	1	1	1	1	1	1	1	1		1	1	15
Fungi	***Saccharomyces cerevisiae***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Fungi	***Schizosaccharomyces pombe***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Fungi	***Spizellomyces punctatus***		1	1	1	1		1	1		1	1	1	1	1		12
Fungi	***Ustilago maydis***	1	1	1	1	1	1		1	1	1	1	1	1	1	1	15
Glaucophytes	***Cyanophora paradoxa***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Glaucophytes	***Glaucocystis nostochinearum***	1	1	1	1	1	1	1		1		1		1	1	1	13
Haptophytes	***Emiliania huxleyi***	1	1	1	1	1		1	1	1	1	1	1	1	1	1	15
Haptophytes	***Isochrysis galbana***	1	1	1	1	1		1	1	1	1	1	1	1	1	1	15
Haptophytes	***Pavlova lutheri***	1	1	1	1	1		1	1	1	1	1			1	1	13
Haptophytes	***Prymnesium parvum***	1	1	1	1	1		1	1	1	1	1	1		1		13
Heterolobosea	***Naegleria gruberi***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Heterolobosea	***Sawyeria marylandensis***	1		1	1		1		1	1	1	1	1	1	1	1	13
Ichthyosporea	*Amoebidium parasiticum*	1	1	1			1	1	1	1	1			1		1	11
Ichthyosporea	***Capsaspora owczarzaki***	1	1	1	1		1	1	1	1	1	1	1	1	1	1	15
Ichthyosporea	***Sphaeroforma arctica***	1	1	1	1			1	1	1	1	1	1	1	1	1	14
Jakodidae	***Jakoba libera***	1	1	1	1	1	1	1				1	1		1	1	12
Jakodidae	***Reclinomonas americana***		1	1	1	1	1	1	1	1	1	1	1	1	1	1	15
Jakodidae	***'Seculamonas ecuadoriensis'***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Kathablepharidae	*Leucocryptos marina*			1	1	1		1			1	1					7
Malawimonas	***Malawimonas californiana***	1	1	1	1	1	1	1		1		1	1	1	1		12
Malawimonas	***Malawimonas jakobiformis***	1	1	1	1	1	1	1		1	1	1	1	1	1	1	15
Parabasalidea	*Trichomonas vaginalis*	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Preaxosytla	***Monocercomonoides* sp.**				1	1			1		1	1					6
Preaxosytla	***Streblomastix strix***				1	1	1					1					5
Preaxosytla	***Trimastix pyriformis***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Rhizaria	***Bigelowiella natans***	1	1	1	1	1		1		1		1		1	1	1	12
Rhizaria	***Corallomyxa tenera***			1	1	1	1	1		1							7
Rhizaria	*Gromia*[‡]			1		1											3
Rhizaria	***Heteromita***[§]	1	1	1	1	1		1		1				1	1		10
Rhizaria	***Ovammina opaca***			1	1	1											4
Rhizaria	*Plasmodiophora brassicae*			1		1										1	4
Rhizaria	***Reticulomyxa filosa***	1		1	1	1			1	1	1	1	1				10
Red algae	***Chondrus crispus***	1	1	1	1	1		1	1	1	1	1	1			1	13
Red algae	*Cyanidioschyzon merolae*			1	1	1	1	1				1					7
Red algae	***Gracilaria changii***	1	1	1	1	1		1	1	1	1	1	1	1	1	1	14
Red algae	***Porphyra yezoensis***	1	1	1	1	1	1	1	1	1			1	1	1	1	14
Stramenopiles	*Apodachlya brachynema*			1	1	1	1		1			1					7
Stramenopiles	***Aureococcus anophagefferens***	1	1	1	1	1	1	1	1	1	1	1	1		1	1	15

Table S2. Cont.

Lineage	Taxon*	14–3-3	40S	Actin	αtub	βtub	Ef1α	Ef2	Enolase	Grc5	Hsp70cyt	Hsp90	MetK	Rps22a	Rps23a	Tsec61	Sum
Stramenopiles	***Ectocarpus siliculosus***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Stramenopiles	*Heterosigma akashiwo*			1	1	1			1		1	1					7
Stramenopiles	***Phaeodactylum tricornutum***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Stramenopiles	***Phytophthora infestans***	1	1	1	1	1	1	1	1	1	1	1	1	1	1		15
Stramenopiles	***Thalassiosira pseudonana***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Green algae	***Acetabularia acetabulum***	1	1	1	1	1	1			1		1				1	10
Green algae	***Arabidopsis thaliana***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Green algae	***Chlamydomonas reinhardtii***	1	1	1	1	1		1	1	1	1	1	1	1	1	1	15
Green algae	***Dunaliella salina***	1	1		1	1		1	1	1	1		1	1	1		12
Green algae	***Ginkgo biloba***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Green algae	***Mesostigma viride***	1	1	1	1	1		1	1	1	1	1	1	1	1		14
Green algae	***Micromonas pusilla***	1	1	1	1	1		1	1	1	1	1	1	1	1	1	15
Green algae	***Oryza sativa***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Green algae	***Ostreococcus tauri***	1	1	1	1	1			1	1	1	1	1		1	1	13
Green algae	***Physcomitrella patens***	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	16
Green algae	***Volvox carteri***	1	1	1	1	1		1	1	1	1	1	1	1	1	1	15
Green algae	***Welwitschia mirabilis***	1	1	1	1	1	1	1		1	1	1	1		1	1	14

*Taxa in bold are included in both the 91-taxon and 109-taxon analyses.
†Composite of *Goniomonas truncata* and *Goniomonas* cf. *pacifica*.
‡Composite of *G. oviformis* and *Gromia* sp. *Antarctica*.
§Composite of. *H. globosa* and *Heteromita* sp. ATCC PRA-74.

Table S3. PhyloBayes calibrations

Taxon*	Node specification		Calibration†	
	Species 1	Species 2	Max	Min
Amniota	*Gallus gallus*	*Homo sapiens*	400	328.3
Angiosperms	*Arabidopsis thaliana*	*Welwitschia mirabilis*	425	133.9
Ascomycetes	*Schizosaccharomyces pombe*	*Phanerochaete chrysosporium*	1,000	400
Coccolithophores	*Emiliania huxleyi*	*Isochrysis galbana*	260	203.6
Diatoms	*Aureococcus anophagefferens*	*Thalassiosira pseudonana*	550	133.9
Dinoflagellates	*Karenia brevis*	*Crypthecodinium cohnii*	300	240
Embryophytes	*Mesostigma viride*	*Oryza sativa*	600	471
Endopterygota	*Apis mellifera*	*Drosophila melanogaster*	350	284.4
Eudicots	*Arabidopsis thaliana*	*Oryza sativa*	133.9	125
Euglenids	*Entosiphon sulcatum*	*Euglena gracilis*	3,000	450
Foraminifera	*Ovammina opaca*	*Reticulomyxa filosa*	3,000	542
Gonyaulacales	*Alexandrium tamarense*	*Crypthecodinium cohnii*	240	196
Pennate diatoms	*Phaeodactylum tricornutum*	*Thalassiosira pseudonana*	110	80
Spirotrichs	*Sterkiella histriomuscorum*	*Stylonychia lemnae*	3,000	444
Trachaeophytes	*Physcomitrella patens*	*Arabidopsis thaliana*	471	425
Vertebrates	*Branchiostoma floridae*	*Homo sapiens*	555	520
Animals	*Nematostella vectensis*	*Capitella capitata*	3,000	632
Arcellinida	*Arcella hemisphaerica*	*Rhizamoeba sp*	3,000	736
Bilateria	*Branchiostoma floridae*	*Capitella capitata*	630	555
Chlorophytes	*Acetabularia acetabulum*	*Volvox carteri*	3,000	700
Ciliates	*Paramecium tetraurelia*	*Chilodonella uncinata*	3,000	736
Florideophyceae	*Chondrus crispus*	*Porphyra yezoensis*	3,000	550
Red algae	*Cyanidioschyzon merolae*	*Chondrus crispus*	3,000	1,174

*Taxon is same as in Table 1; see Table 1 for other notes.
†Calibrations in PhyloBayes are specified as a uniform distribution with minimum and maximum dates, and were run with soft bounds.

Other Supporting Information Files

Dataset S1 (XLS)