

# Primary endosymbiosis events date to the later Proterozoic with cross-calibrated phylogenetic dating of duplicated ATPase proteins

Patrick M. Shih<sup>a,1,2</sup> and Nicholas J. Matzke<sup>b,c,1,2</sup>

Departments of <sup>a</sup>Plant and Microbial Biology and <sup>b</sup>Integrative Biology, and <sup>c</sup>Center for Theoretical Evolutionary Genomics, University of California, Berkeley, CA, 94720

Edited by Joseph Felsenstein, University of Washington, Seattle, WA, and approved May 22, 2013 (received for review March 26, 2013)

**Chloroplasts and mitochondria descended from bacterial ancestors, but the dating of these primary endosymbiosis events remains very uncertain, despite their importance for our understanding of the evolution of both bacteria and eukaryotes. All phylogenetic dating in the Proterozoic and before is difficult: Significant debates surround potential fossil calibration points based on the interpretation of the Precambrian microbial fossil record, and strict molecular clock methods cannot be expected to yield accurate dates over such vast timescales because of strong heterogeneity in rates. Even with more sophisticated relaxed-clock analyses, nodes that are distant from fossil calibrations will have a very high uncertainty in dating. However, endosymbiosis events and gene duplications provide some additional information that has never been exploited in dating; namely, that certain nodes on a gene tree must represent the same events, and thus must have the same or very similar dates, even if the exact date is uncertain. We devised techniques to exploit this information: cross-calibration, in which node date calibrations are reused across a phylogeny, and cross-bracing, in which node date calibrations are formally linked in a hierarchical Bayesian model. We apply these methods to proteins with ancient duplications that have remained associated and originated from plastid and mitochondrial endosymbionts: the  $\alpha$  and  $\beta$  subunits of ATP synthase and its relatives, and the elongation factor thermo unstable. The methods yield reductions in dating uncertainty of 14–26% while only using date calibrations derived from phylogenetically unambiguous Phanerozoic fossils of multicellular plants and animals. Our results suggest that primary plastid endosymbiosis occurred ~900 Mya and mitochondrial endosymbiosis occurred ~1,200 Mya.**

Biologists have often attempted to estimate when key events on the Tree of Life (TOL) occurred. This approach has experienced substantial success when used for dating events in the Phanerozoic [543–0 Mya], but when trying to date deep events on the TOL, such as endosymbiosis events in the Proterozoic (2,500–543 Mya), it becomes increasingly difficult to find reliable fossil calibrations. Molecular dating analysis is performed by calibrating a phylogenetic tree with known dates, usually based on fossil calibration points. Ideally, the dating of phylogenetic events deep in the Precambrian would be well-constrained by fossil calibrations; however, many of the fossil calibrations that have been proposed for Precambrian microorganisms have been controversial because of the difficulty in identifying the clade memberships of these groups.

Although the timing of the origin of eukaryotes is heavily studied and debated, the endosymbiosis events involved in the origin and diversification of many eukaryotic lineages are arguably equally contentious. Fossil records for eukaryotes have been claimed up to 2,700 Mya (1), and others have speculated that “Snowball Earth” events postponed the origin and/or diversification of eukaryotes until as recently as 850–580 Mya (2–4). Interpretation of microfossils is inherently difficult because of difficult preservation, taphonomic, and interpretive issues (e.g., refs. 5 and 6). A less-recognized problem is that fossil calibrations are best done via a phylogenetic analysis of characters, which allows objective placement of fossils on a tree and measurement of the uncertainty of

this placement (7). General similarity to an extant group is an insufficient basis for using a fossil as a date calibration: Characters must place the fossil in the crown group rather than a stem group [which is sometimes an insufficiently appreciated distinction (8)] to constrain the date of the last common ancestor of the crown group (7). However, microfossils typically have a very small number of diagnosable characters (9), thus running the risk of misclassification, especially as a result of homoplasy.

Chemical biomarkers, another strategy that is much used to date Precambrian lineages, are equally problematic because fundamentally, each biomarker constitutes a single character unassociated with other fossil characters. To be used for dating, it must be assumed that the character only evolved once and is unique to one extant clade, but this is not always a safe assumption, as demonstrated by the recent finding that the methylhopane biomarker, once used specifically for cyanobacteria (10), can also be found in a broad range of other bacterial phyla (11, 12).

Apart from uncertainty in fossil calibrations, molecular dating imposes additional uncertainties. Early attempts at molecular dating, starting with Zuckerkandl and Pauling (13), assumed a strict molecular clock to date divergences. Subsequent attempts to date deep nodes in the TOL have given wildly varying results, many of which clearly do not agree with fossil, let alone geological, histories primarily because of rate variation not accounted for by strict clock models (14, 15). More sophisticated models allow for rate variation, and thus provide a more realistic assessment of uncertainty. However, the uncertainty that results can be vast, as the origin of crown eukaryotes has been dated between 3,970 and 1,100 Mya throughout various studies (16).

Uncorrelated relaxed-clock methods, available in Bayesian phylogenetic dating methods, allow the rate of evolution on each branch to be drawn from a common distribution, the parameters of which are themselves estimated during the analysis. One advantage of Bayesian analysis is that it takes into account diverse sources of known prior information. Another technique used in several studies relies on the concatenation of protein sequences to increase phylogenetic signal for estimations of deeply rooted events. However, this strategy does nothing to remedy the problem of scarce and ambiguous fossil calibrations for deep nodes.

Given the difficulty of dating deep nodes in the Proterozoic as well as the lack of studies dating Precambrian events with newer methods, it is useful to explore possible improvements in relaxed clock analyses. We hypothesize that better estimates of rates and rate variability, and thus better estimates of dates and dating

Author contributions: P.M.S. and N.J.M. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

<sup>1</sup>P.M.S. and N.J.M. contributed equally to this work.

<sup>2</sup>To whom correspondence may be addressed. E-mail: pmsih@berkeley.edu or matzke@berkeley.edu.

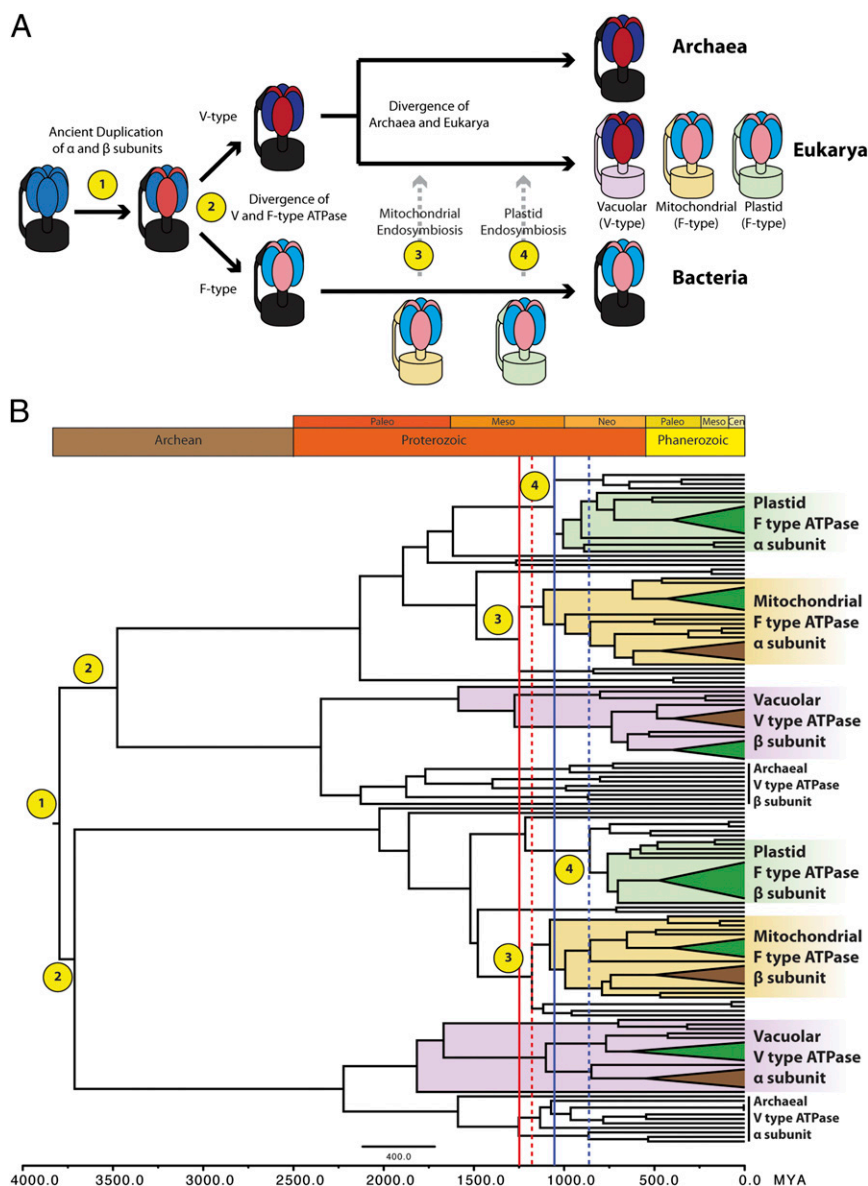
This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1305813110/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1305813110/-DCSupplemental).

uncertainty, would occur if more prior information and more date calibrations were input into analyses. Date calibrations are typically scarce, but we suggest they can be multiplied in cases in which one or more ancient duplications has been universally or near-universally inherited. In such cases, a single fossil calibration can date not just one node in the tree, but several. An example where this is possible is the protein family of ATP synthases (ATPases) found within the  $F_1$  portion of the  $F_1F_0$ -ATPase system and its relatives, the vacuolar  $V_1V_0$ -ATPases and archaeal  $A_1A_0$ -ATPases (17). The  $\alpha$  and  $\beta$  subunits of  $F_1$ -ATPase duplicated before the last universal common ancestor (18, 19) and have been almost universally inherited as a pair since then (Fig. 1A). Furthermore, the core function of the ATPases in energy production has resulted in high conservation and a lower probability of extreme rate variation.

The fact that mitochondria and plastids have retained these ATPase proteins (whether they are encoded by the organellar or the nuclear genome) means that many homologs may coexist in a single organism. For example, plant genomes contain six homologous copies of this ATPase subunit: both homologous  $\alpha$  and  $\beta$  subunits targeted to the mitochondria, chloroplasts, and vacuoles.

Therefore, a single plant fossil, which calibrates the date of the divergence of monocots and eudicots, can actually provide calibration dates for up to six nodes on the ATPase  $\alpha$  and  $\beta$  subunit phylogeny. We propose two methods for use of these calibrations (SI Appendix, Fig. S1). In the first strategy, which we dub cross-calibration, the date calibrations are simply reused at each node, and the dates of these nodes are subsequently sampled independently during the Markov chain Monte Carlo (MCMC) search. Cross-calibration is simple to implement but neglects the fact that nodes representing the same event should have the same date, even if that date is uncertain. We therefore also propose a second strategy, cross-bracing, in which the dates of calibrated nodes representing the same speciation events are linked, and thus covary during MCMC sampling. This is a more accurate representation of our prior knowledge that a single speciation event led to the simultaneous divergence of the nuclear, mitochondrial, plastid ATPase genes (although some variability could be caused by lineage sorting processes).

Iwabe et al. (18) and Gogarten et al. (19) attempted to use ancient duplicated genes in inferring distant evolutionary relationships between the three domains of life, using  $\alpha$  and  $\beta$  subunits of



**Fig. 1.** Evolutionary history of the ATPase  $\alpha$ - and  $\beta$ -subunits and divergence time estimates inferred from cross-calibration analysis. (A) Cartoon schematic that demonstrates the common origin of both  $\alpha$ - and  $\beta$ -subunits, followed by both the mitochondrial and plastid endosymbiosis events, all of which enable the use of cross-calibration methods. Evolutionary events of interest are numbered and labeled onto the subsequent chronogram generated from cross-calibration of the  $\alpha$ - and  $\beta$ -subunits. (B) Time-scale phylogeny generated from Bayesian analysis of cross-calibrated ATPase  $\alpha$ - and  $\beta$ -subunits (SI Appendix, Fig. S8). Blue lines denote the dates estimated for the primary plastid endosymbiosis event. Red lines denote the dates estimated for mitochondrial endosymbiosis. Solid lines represent dates that were inferred from the  $\alpha$ -subunit subsection of the phylogeny; dashed lines were inferred from the  $\beta$ -subunit subclade.



braced node ages tended to be lower by a fixed amount of 37–65 My (*SI Appendix, Table S2 and Fig. S8*).

To further investigate the effect of reducing the number of prior calibration dates, the  $\alpha$ -cross-calibrated analysis using all node calibrations was compared with the  $\alpha$ -cross-calibrated analyses using fewer calibration priors (*SI Appendix, Table S3*). Uncertainty in node age was not dramatically different between the  $\alpha$ -cross-calibrated data set with all date calibrations and subsets of these calibrations (*SI Appendix, Tables S2 and S3*). However, when the heteroscedasticity between node age and uncertainty is accounted for by calculating the CV, comparison of CVs showed a significant decrease in CV (23–44%) when all calibration nodes were used, suggesting that increasing the number of calibration points decreases relative uncertainty in the estimates of node age in  $\alpha$ -only analyses. Moreover, branch rate uncertainty significantly increased for runs with fewer calibrations except run 5 (*SI Appendix, Table S4 and Fig. S9*). Further comparisons of all runs, including the  $\alpha/\beta$ -cross-calibrated and  $\alpha/\beta$ -cross-braced runs, are summarized in *SI Appendix*.

**Dating Symbiosis Events: ATPases.** Because the  $\alpha/\beta$ -cross-calibrated and  $\alpha/\beta$ -cross-braced runs were shown to decrease rate and age uncertainty, but neither method yielded significantly more robust results when compared with the other, for simplicity, we will henceforth refer to only the  $\alpha/\beta$ -cross-calibrated analysis (summarized in Table 1).

The timing of plastid endosymbiosis has been as contentious as dating the rise of eukaryotes. The hypothesis that cyanobacteria are responsible for the Great Oxidation Event (GOE) has led to many studies extrapolating divergence points for a broad range of uses, from dating endosymbiosis events to events of multicellularity (25–28). However, this approach assumes that all crown cyanobacterial lineages emerged at the time of the GOE (29). Our study was aimed at dating the plastid endosymbiosis event agnostic of the GOE, microfossils, or biomarker data, and instead calibrated only by well-accepted Phanerozoic divergence events. Our cross-calibrated analysis estimates primary plastid endosymbiosis and the birth of the Archaeplastida lineage at 857 and 1,055 Mya (857/1,055 Mya), based on F-type  $\alpha$  and  $\beta$  subunits of the tree, respectively. These dates are remarkably similar to the dates estimated by Douzery et al., who predicted that the plastid endosymbiosis occurred between 825 and 1,162 Mya, using 129 concatenated protein sequences, as well as to other previous large-scale and broadly sampled molecular clock studies (30).

Although younger than other predicted estimated divergence dates (31, 32), our dates present a plausible scenario for the changing geochemical properties of the ocean. The rise of photosynthetic eukaryotes through the acquisition of plastids ~900 Mya most likely dramatically added to primary productivity in the sea, which may have significantly contributed to the conversion of euxinic oceans during the Neoproterozoic to its oxygenated state, which persists today (33). This is further supported by the dramatic increase in atmospheric oxygen between 1,005 and 640 Mya (34). Our analysis suggests that the diversification of Archaeplastida occurred near or during the time of the transformation of euxinic

conditions to its modern-day properties and that there was very little lag time between the origin and diversification of photosynthetic eukaryotes.

Numerous phylogenetic studies have placed the plastid endosymbiosis event near the base of the extant cyanobacterial tree (35–37). Assuming that crown cyanobacteria were responsible for the GOE, this would place the plastid endosymbiosis near the time of the GOE. This is in contradiction to our study and many concatenated, multiloci molecular clock studies (30–32), which have conservatively dated the origin of crown eukaryotes well after 2 Gya. It is therefore difficult to reconcile these dates, as plastid endosymbiosis could not have occurred before the origin of eukaryotes. Moreover, all bacterial phyla in our analysis (including cyanobacteria) have diversified after the GOE, suggesting that extant crown cyanobacteria were not responsible for the GOE. Our findings are in contrast with those of Schirmer et al. (28) who date the origin of crown cyanobacteria before the GOE. These findings are attributable to their assignment of ancient (>2 Gya) cyanobacterial-like fossils to extant clades, despite the possibility that the few available morphological characters may be homoplastic and may have evolved several times convergently. Assuming the GOE was of biological origin, our results imply that crown cyanobacteria may not have been responsible for the GOE. However, this does not rule out the possibility of its origin from stem group cyanobacteria, which may have gone extinct during the major transition from euxinic to oxic oceans (33). In line with this idea, the phylogeny of crown cyanobacteria has been interpreted as a large radiation event (35, 37), which may have occurred after the extinction of stem groups and the adaption of crown lineages to the changing ocean surfaces. These extinct lineages may be the Proterozoic cyanobacterial-like fossils described in previous studies (27, 38–40) and used as fossil calibrations by Schirmer et al. (28). Our analysis reflects the controversial nature of contrasting molecular and fossil studies, and thus emphasizes the need to improve existing phylogenetic techniques to more accurately examine the dating of these Precambrian events.

Our cross-calibration analysis dates the rise of modern-day mitochondria through the endosymbiosis of an  $\alpha$ -proteobacterium to be 1,176/1,248 Mya. Although the vacuolar subclades display an earlier date for the last common ancestor of eukaryotes, our interest was in dating the actual divergence between bacteria and mitochondria; other dates in the analyses were treated as nuisance variables. Given that the most recent common ancestor of eukaryotes most likely is younger than the mitochondrial endosymbiosis, we recognize the contradiction between the dates in the two parts of the tree, which is probably caused by fewer calibration points and an accelerated rate of evolution at the base of the V-ATPase tree. However, the only methodological remedy would be to use the cross-bracing technique on those nodes we want to infer, whereas in this study we are examining the potential of cross-linking date calibration nodes. Cross-bracing nodes with dates that are to be inferred rather than used as calibrations should be explored in the future, but issues of extended autocorrelation in the posterior distribution and of low estimated sample size become much more pressing if the nodes targeted for inference are cross-braced.

Parfrey et al. estimate the last common eukaryotic ancestor to be more than 1,600 Mya (31), which is notably older than our analysis. However, when excluding Proterozoic fossil calibrations, they observed shifts in all major clades to be 300 My younger, which is nearly comparable with our results. The effects of excluding Proterozoic microfossil calibrations may explain the incongruence in estimated dates between studies; however, for the purposes of our study, our focus on cross-calibration methods was to increase the amount of dating prior information with younger and less controversial Phanerozoic fossils. Finally, our analysis does not find evidence for the hypothesis that crown eukaryotes originated ~850 Mya and postdate the hypothesized Snowball Earth.

**Table 1. Divergence-time estimates (in millions of years ago) for major endosymbiosis or domain divergence events**

Divergence event and cross-calibrated ATPase $\alpha/\beta$ subunits	Cross-calibrated elongation factor Tu
Plastid endosymbiosis	
$\alpha$ subunit: 1,055 (1,278–913)	1,188 (896–1,613)
$\beta$ subunit: 857 (1,098–720)	
Mitochondrial endosymbiosis	
$\alpha$ subunit: 1,248 (1,838–1,217)	1,196 (909–1,551)
$\beta$ subunit: 1,176 (1,524–1,053)	

Dates in parentheses denote the 95% highest posterior density.

Although earlier Proterozoic and Archean events are not the primary focus of this study, and uncertainties this far back are large, we observe long branches leading to the Eukarya/Archaea split, followed by a radiation of extant Eukarya/Archaea (V-type ATPases) and Eubacteria (F-type ATPases) around 2,000–2,500 Mya. Because the rise in molecular oxygen in the atmosphere occurred around the same time, it is tempting to speculate that this synchronized radiation of extant life across all three kingdoms was somehow facilitated by the GOE and that all extant life-forms are the descendants of lineages that most successfully adapted to the changing biogeochemistry in ocean surfaces.

**Dating Symbiosis Events: Ef-Tu.** Because there may be inherent biases between particular markers used for any phylogenetic analysis, we extended our cross-calibration study to Ef-Tu because of its similar evolutionary history to ATPases, which allows for cross-calibration. Bacterial Ef-Tu and its eukaryotic/archaeal homolog, translation elongation factor 1 $\alpha$  (EF-1 $\alpha$ ), allow for entry of aminoacyl tRNAs into the ribosome, and thus are considered conserved, slowly evolving proteins, decreasing the chance of saturation and high rate variability. The dates estimated from the Ef-Tu chronogram were similar to the dates attained from the ATPase analysis: 1,188 Mya for plastid endosymbiosis and 1,196 Mya for mitochondrial endosymbiosis (Table 1; *SI Appendix, Fig. S10*). Estimations of deeper nodes such as the split between Archaea and Eukarya (1,528 Mya) differed from the ATPase results by almost 800 Mya. This is not surprising, as many of these nodes may inherently be difficult to estimate because of the lack of signal from a saturation of amino acid substitutions (20).

## Conclusion

Cross-calibration and cross-bracing, using duplication or endosymbiosis events, provide useful advantages compared with conventional molecular dating. First, they increase the sampling and sequence data used, which improves accuracy of the dating of internal nodes (41, 42). Second, by increasing the number of sequences that are cross-calibrated, they decrease the chance of artifacts being introduced by underestimated rate variation. Just as there are multiple calibration points for a given divergence event, a divergence event will be estimated multiple times on the tree. Third, the increase in calibration points allows for the use of more well-accepted calibration points closer to the tips of the tree, rather than relying on older and more contentious microscopic, Precambrian fossils.

The flexibility of the BEAST XML input allows unconventional strategies such as ours to be used. However, the cross-bracing technique could be improved. Future efforts should develop algorithms that redesign the MCMC tree search such that nodes with linked dates can be specified and linked nodes can be allowed to share identical dates during sampling. This should eliminate all or most of the need for longer runs to account for increased autocorrelation in the posterior sample. The cross-bracing strategy might also improve inference in another way: nodes with dates that are unknown, but that represent the same event, could be linked, as we have done here for calibration nodes. For example, the nodes representing the divergence of the chloroplasts should have the same or nearly the same date between the  $\alpha$ - and  $\beta$ -subunit gene trees, instead of two individually estimated dates. Further refinements could include linking rates for genes when they are inhabiting the same species, which would avoid the assumption, made here by necessity, that rates and rate variation are independent across the tree.

It is important to note that our approach is different from the common technique of concatenation of gene duplicates into a larger alignment. For example, if a researcher were only interested in dating nodes within plants, to increase signal they might concatenate the  $\alpha$ - and  $\beta$ -subunit sequences from vacuolar, chloroplast, and mitochondrial ATPases. However, this conventional strategy would be useless when the goal is to date nodes in the gene tree that are not represented by nodes in the species tree; for

example, the date of a gene duplication itself, or as in this study, the date of endosymbiosis events.

Although we observed similar dates between ATPase and Ef-Tu, it will be interesting to determine whether other molecular markers that have undergone duplications or endosymbiotic transfers and can be used in cross-calibration will also yield similar dates. Possible examples include aminoacyl-tRNA synthetases (43), translation initiation factors (44), and phytochrome (45) data sets. Cross-calibration could also be extended to large concatenated data sets if all proteins display similar histories.

Regardless of the detailed method used, we argue that because of the difficulty in estimating the timing of Precambrian events, every possible source of information should be included. As we show here, this information is not merely found in the dates of fossil calibrations, it can also include linkages between nodes that represent the same speciation or duplication events. Information about the relative timing of events could also be included; for example, the origin of crown chloroplasts must equal or postdate the origin of crown eukaryotes. Hierarchical Bayesian models excel in the incorporation of such diverse sources of information and should be exploited wherever possible, along with other attempts to ameliorate dependence on controversial date calibrations based on ancient, microscopic fossils that are difficult to interpret and rigorously place on phylogenies.

## Materials and Methods

**Alignments.** ATPase  $\alpha$  and  $\beta$  subunit and Ef-Tu/1 $\alpha$  protein sequences were all gathered from the Uniprot database and are listed in *SI Appendix, Table S5*. Sequences were chosen to cover a broad range of bacterial, archaeal, and eukaryotic phyla. Alignments were generated using the *-maxiterate* strategy in the MAFFT program (46).

**Dating Programs.** Estimation of dated phylogenies was conducted with BEAST 1.7.3 (23, 24). BEAST XML input files were started using BEAUTi 1.7.3 (23, 24), but our novel calibration strategies, described below, required custom modifications to the XML code. The WAG model was chosen as the best-fitting amino acid substitution matrix available in BEAST, based on ProtTest analysis for all data sets (47). Production of the final BEAST XML files for the different combinations of data sets and calibration methods was done via custom programs in R 2.15 (48). BEAST XML files implementing the cross-calibration and cross-bracing methods are available in *SI Appendix, Materials and Methods*. All BEAST runs were inspected for convergence and completeness of sampling the posterior distribution, using Tracer (49).

**Node Date Calibrations.** Dating calibration distributions were based on macroscopic fossils of Phanerozoic plants and animals that provide well-accepted calibration points used in previous molecular dating studies of Phanerozoic groups (50, 51) (*SI Appendix, Table S6*). Although the origin of crown angiosperms estimated by Smith et al. (50) was older than previous studies and fossil records (52), we found the discrepancy of  $\sim$ 80 Mya negligible in comparison with the divergence estimates we were focused on in this study. More important, the other estimated dates used as calibration points from Smith et al. aligned well with the current estimates of divergences within land plants (53, 54). Plant calibration points were used for the plant vacuolar, mitochondrial, and plastid ATPases. The human/chicken and fly/mosquito divergences were used as metazoan calibration points (51). To maintain maximum agnosticism about the date of the last common ancestor and the divergence of the ATPase  $\alpha$  and  $\beta$  subunits, which occurred before the last common ancestor, a uniform distribution before between 3,800 and 2,500 Mya was set at the base of the tree (the split between  $\alpha$  and  $\beta$  subunits), assuming a biological origin of the GOE of 2,500 Mya (55) and that life most likely could not have begun before the Late Heavy Bombardment of Earth, ca. 3,800 Mya (56).

**Cross-Calibration and Cross-Bracing Methods.** In the cross-calibration method, each node in the gene tree corresponding to the same speciation event is assigned the same prior distribution on the date (i.e., the distribution given in *SI Appendix, Table S6*). These distributions are cross-calibrated, or “unlinked”; that is, during MCMC sampling, the date of each node is sampled independent from the prior distribution.

As with cross-calibration, in the cross-bracing method, each node in the gene tree corresponding to the same speciation event is assigned the same prior distribution on the date. However, in the cross-bracing method, the dates of nodes corresponding to the same speciation event are “linked.” As

BEAST cannot formally do joint sampling of node dates, we achieved the same effect by coding into the BEAST XML an additional prior on the differences between the dates of linked nodes and the mean of the linked nodes. This prior was a normal distribution, with a mean of 0 (as any prior on the difference from the mean must have) and a SD set to 1% of the mean of the prior distribution of the date of the speciation event. Thus, although BEAST samples each linked node independently during the actual MCMC sampling, samples in which the linked nodes are far apart will have a low posterior probability and will be rejected more often than in the cross-calibration approach. Inspection of linked node dates in Tracer (57) showed that they were indeed highly correlated to each other, unlike in the cross-calibration approach.

The 1% SD value on the distribution of differences from the mean date was chosen to indicate our prior high confidence that nodes corresponding to the same speciation event should have approximately the same date. The

distribution on differences from the mean date was not set even more tightly, for two reasons. First, lineage-sorting processes can cause some degree of difference in the divergence dates of gene trees during speciation. Second, it was important to give BEAST's MCMC sampler "breathing room" to sample the date of one linked node, then another, then another, and so on, without too many of these moves being rejected, so that the full posterior distribution could be explored. Further analysis was conducted as described in *SI Appendix, Materials and Methods and Supplemental Analysis of BEAST Runs*.

**ACKNOWLEDGMENTS.** P.M.S. was supported by National Science Foundation Grant MCB-0851070 (to Cheryl Kerfeld) and Howard Hughes Medical Institute/Gordon and Betty Moore Foundation Grant GBMF3070 (to Krishna Niyogi). N.J.M. was supported by National Science Foundation Grant DEB-0919451, a UC Berkeley Wang Fellowship, and a UC Berkeley Tien Fellowship.

- Brocks JJ, Logan GA, Buick R, Summons RE (1999) Archean molecular fossils and the early rise of eukaryotes. *Science* 285(5430):1033–1036.
- Hoffman PF, Kaufman AJ, Halverson GP, Schrag DP (1998) A Neoproterozoic snowball earth. *Science* 281(5381):1342–1346.
- Cavalier-Smith T (2006) Cell evolution and Earth history: Stasis and revolution. *Philos Trans R Soc Lond B Biol Sci* 361(1470):969–1006.
- Cavalier-Smith T (2010) Deep phylogeny, ancestral groups and the four ages of life. *Philos Trans R Soc Lond B Biol Sci* 365(1537):111–132.
- Schopf JW, Kudryavtsev AB (2012) Biogenicity of Earth's earliest fossils: A resolution of the controversy. *Gondwana Res* 22(3–4):761–771.
- Brasier MD, et al. (2002) Questioning the evidence for Earth's oldest fossils. *Nature* 416(6876):76–81.
- Parham JF, et al. (2012) Best practices for justifying fossil calibrations. *Syst Biol* 61(2):346–359.
- Budd GE (2003) The cambrian fossil record and the origin of the phyla. *Integr Comp Biol* 43(1):157–165.
- Diver W, Peat C (1979) On the interpretation and classification of Precambrian organic-walled microfossils. *Geology* 7(8):401–404.
- Summons RE, Jahnke LL, Hope JM, Logan GA (1999) 2-Methylhopanoids as biomarkers for cyanobacterial oxygenic photosynthesis. *Nature* 400(6744):554–557.
- Rashby SE, Sessions AL, Summons RE, Newman DK (2007) Biosynthesis of 2-methylbacteriohopanepolyols by an anoxygenic phototroph. *Proc Natl Acad Sci USA* 104(38):15099–15104.
- Welander PV, Coleman ML, Sessions AL, Summons RE, Newman DK (2010) Identification of a methylase required for 2-methylhopanoid production and implications for the interpretation of sedimentary hopanes. *Proc Natl Acad Sci USA* 107(19):8537–8542.
- Zuckermandl E, Pauling L (1965) Molecules as documents of evolutionary history. *J Theor Biol* 8(2):357–366.
- Martin W, Gierl A, Saedler H (1989) Molecular evidence for pre-Cretaceous angiosperm origins. *Nature* 339(6219):46–48.
- Doolittle RF (1992) Reconstructing history with amino acid sequences. *Protein Sci* 1(2):191–200.
- 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(1470):1039–1054.
- Mulkidjanian AY, Makarova KS, Galperin MY, Koonin EV (2007) Inventing the dynamo machine: The evolution of the F-type and V-type ATPases. *Nat Rev Microbiol* 5(11):892–899.
- Iwabe N, Kuma K, Hasegawa M, Osawa S, Miyata T (1989) Evolutionary relationship of archaeobacteria, eubacteria, and eukaryotes inferred from phylogenetic trees of duplicated genes. *Proc Natl Acad Sci USA* 86(23):9355–9359.
- Gogarten JP, et al. (1989) Evolution of the vacuolar H<sup>+</sup>-ATPase: Implications for the origin of eukaryotes. *Proc Natl Acad Sci USA* 86(17):6661–6665.
- Philippe H, Forterre P (1999) The rooting of the universal tree of life is not reliable. *J Mol Evol* 49(4):509–523.
- Doolittle WF, Baptiste E (2007) Pattern pluralism and the Tree of Life hypothesis. *Proc Natl Acad Sci USA* 104(7):2043–2049.
- Hilario E, Gogarten JP (1993) Horizontal transfer of ATPase genes—the tree of life becomes a net of life. *Biosystems* 31(2–3):111–119.
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7(1):214.
- Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* 29(8):1969–1973.
- Falcón LI, Magallón S, Castillo A (2010) Dating the cyanobacterial ancestor of the chloroplast. *ISME J* 4(6):777–783.
- Schirmer BE, Antonelli A, Bagheri HC (2011) The origin of multicellularity in cyanobacteria. *BMC Evol Biol* 11(1):45.
- Tomitani A, Knoll AH, Cavanaugh CM, Ohno T (2006) The evolutionary diversification of cyanobacteria: Molecular-phylogenetic and paleontological perspectives. *Proc Natl Acad Sci USA* 103(14):5442–5447.
- Schirmer BE, de Vos JM, Antonelli A, Bagheri HC (2013) Evolution of multicellularity coincided with increased diversification of cyanobacteria and the Great Oxidation Event. *Proc Natl Acad Sci USA* 110(5):1791–1796.
- Sato N (2006) Origin and Evolution of Plastids: Genomic View on the Unification and Diversity of Plastids. *The Structure and Function of Plastids, Advances in Photosynthesis and Respiration*, eds Wise R, Hooper JK (Springer, The Netherlands), Vol 23, pp 75–102.
- Douzery EJP, Snell EA, Baptiste 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(43):15386–15391.
- Parfrey LW, Lahr DJG, Knoll AH, Katz LA (2011) Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc Natl Acad Sci USA* 108(33):13624–13629.
- Yoon HS, Hackett JD, Ciniglia C, Pinto G, Bhattacharya D (2004) A molecular timeline for the origin of photosynthetic eukaryotes. *Mol Biol Evol* 21(5):809–818.
- Johnston DT, Wolfe-Simon F, Pearson A, Knoll AH (2009) Anoxygenic photosynthesis modulated Proterozoic oxygen and sustained Earth's middle age. *Proc Natl Acad Sci USA* 106(40):16925–16929.
- Carnfield DE, Teske A (1996) Late Proterozoic rise in atmospheric oxygen concentration inferred from phylogenetic and sulphur-isotope studies. *Nature* 382(6587):127–132.
- Criscuolo A, Gribaldo S (2011) Large-scale phylogenomic analyses indicate a deep origin of primary plastids within cyanobacteria. *Mol Biol Evol* 28(11):3019–3032.
- Turner S, Pryer KM, Miao VPW, Palmer JD (1999) Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *J Eukaryot Microbiol* 46(4):327–338.
- Shih PM, et al. (2013) Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc Natl Acad Sci USA* 110(3):1053–1058.
- Schopf JW (1993) Microfossils of the Early Archean Apex chert: New evidence of the antiquity of life. *Science* 260(5108):640–646.
- Schopf JW (2012) The Fossil Record of Cyanobacteria. *Ecology of Cyanobacteria II: Their Diversity in Space and Time*, ed Whitton BA (Springer, The Netherlands), pp 15–36.
- Hofmann HJ (1976) Precambrian microflora, Belcher Islands, Canada; significance and systematics. *J Paleontol* 50(6):1040–1073.
- Zwickl DJ, Hillis DM (2002) Increased taxon sampling greatly reduces phylogenetic error. *Syst Biol* 51(4):588–598.
- Wertheim JO, Sanderson MJ (2011) Estimating diversification rates: How useful are divergence times? *Evolution* 65(2):309–320.
- Brown JR, Doolittle WF (1995) Root of the universal tree of life based on ancient aminoacyl-tRNA synthetase gene duplications. *Proc Natl Acad Sci USA* 92(7):2441–2445.
- Keeling PJ, Fast NM, McFadden GI (1998) Evolutionary relationship between translation initiation factor eIF-2γ and selenocysteine-specific elongation factor SELB: Change of function in translation factors. *J Mol Evol* 47(6):649–655.
- Mathews S, Clements MD, Beilstein MA (2010) A duplicate gene rooting of seed plants and the phylogenetic position of flowering plants. *Philos Trans R Soc Lond B Biol Sci* 365(1539):383–395.
- Katoh K, Kuma K-i, Toh H, Miyata T (2005) MAFFT version 5: Improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* 33(2):511–518.
- Abascal F, Zardoya R, Posada D (2005) ProtTest: Selection of best-fit models of protein evolution. *Bioinformatics* 21(9):2104–2105.
- R Development Core Team (2012) *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria).
- Rambaut A, Drummond AJ (2007) Tracer version 1.5. Available at <http://beast.bio.ed.ac.uk/Tracer>.
- Smith SA, Beaulieu JM, Donoghue MJ (2010) An uncorrelated relaxed-clock analysis suggests an earlier origin for flowering plants. *Proc Natl Acad Sci USA* 107(13):5897–5902.
- Berbee ML, Taylor JW (2010) Dating the molecular clock in fungi – how close are we? *Fungal Biol Rev* 24(1–2):1–16.
- Sanderson MJ, Thorne JL, Wikström N, Bremer K (2004) Molecular evidence on plant divergence times. *Am J Bot* 91(10):1656–1665.
- Wellman CH, Gray J (2000) The microfossil record of early land plants. *Philos Trans R Soc Lond B Biol Sci* 355(1398):717–731, discussion 731–732.
- Doyle JA (1998) Molecules, morphology, fossils, and the relationship of angiosperms and Gnetales. *Mol Phylogenet Evol* 9(3):448–462.
- Kopp RE, Kirschvink JL, Hilburn IA, Nash CZ (2005) The Paleoproterozoic snowball Earth: A climate disaster triggered by the evolution of oxygenic photosynthesis. *Proc Natl Acad Sci USA* 102(32):11131–11136.
- Cohen BA, Swindle TD, Kring DA (2000) Support for the lunar cataclysm hypothesis from lunar meteorite impact melt ages. *Science* 290(5497):1754–1756.
- Rambaut A, Drummond A (2007) Tracer v1.4. Available at <http://beast.bio.ed.ac.uk/Tracer>.

## Supporting Information Appendix

### **Primary endosymbiosis events date to the later Proterozoic with cross-calibrated phylogenetic dating of duplicated ATPase proteins**

Patrick M. Shih<sup>a,1,2</sup> and Nicholas J. Matzke<sup>b,c,1,2</sup>

<sup>a</sup> Department of Plant and Microbial Biology, University of California, Berkeley, California, 94720, USA.

<sup>b</sup> Department of Integrative Biology, University of California, Berkeley, California, 94720, USA.

<sup>c</sup> Center for Theoretical Evolutionary Genomics, University of California, Berkeley, California, 94720, USA.

<sup>1</sup> Authors contributed equally to this work

<sup>2</sup> To whom correspondence should be addressed, email: [matzke@berkeley.edu](mailto:matzke@berkeley.edu) or [pmsih@berkeley.edu](mailto:pmsih@berkeley.edu)

## Table of Contents

Supplemental Materials and Methods .....	3-4
Supplemental Analysis on BEAST Runs .....	4-7
Supplemental Figures S1-S9 .....	8-17
Supplemental Tables S1-S6 .....	18-27

### **Supplemental Material and Methods:**

**Application of the methods and comparison of results.** In order to measure the effect of cross-calibration vs. cross-bracing on an overall dating analysis and the effect of different amounts of prior dating information, nine separate relaxed-clock dating analyses (Table S1) were performed using the program BEAST. The first three runs were done with only the  $\alpha$ -subunits from the overall alignment (run 1), only  $\beta$ -subunits (run 2), and both  $\alpha$  and  $\beta$  subunits, calibrated under the cross-calibration method (run 3, Figure 1). All available node date calibrations were used in these analyses. To examine the effect of systematically removing certain categories of calibration points, runs 4-8 used only  $\alpha$ -subunits, with various categories of subunits removed from the calibrations (Table S1). Specifically, run 4 only included metazoan calibration points, and run 5 only included plant calibration points. We also tested the effect of using calibration points that were more broadly sampled from across the  $\alpha$ -subunit gene tree, but were symbiont/nuclear-specific; thus run 6 used only the calibration points in the chloroplast subclade, run 7 the mitochondrial subclade, and run 8 used only the calibration points from the vacuolar subclade. Finally, run 9 used the same sequence data as run 3 (both  $\alpha$  and  $\beta$  subunits in one large gene tree), but used the cross-bracing method.

All runs except the last were sampled for ten million generations, with samples collected every 1000 generations, and with the first 50% discarded as burnin. The cross-bracing approach induced additional autocorrelation in the runs, requiring much longer runs to assure convergence and adequate ESS (estimated sample size) values for all parameters. Therefore, for cross-bracing, four independent BEAST runs of approximately 40 million generations each were conducted. In each case, the first 20% was discarded as burn-in (as this appeared to be well past the burn-in period), and the remaining samples were concatenated. This resulted in 142,555 samples representing 142.5 million generations of post-burnin sampling.

All runs were inspected in Tracer for convergence and adequate estimated sample size (ESS) values. Sampling was judged to be adequate in all cases. As expected, linked node dates are highly correlated in the cross-bracing approach, resulting in lower ESS values (~50-100) for the linked dates. However, this was not of great concern, as the dates for these nodes are specified in prior distributions, and are not the target of inference, and sample sizes of 50-100 are still adequate to indicate reasonable sampling of the overall distribution. This was confirmed by inspection of trace plots for each parameter, confirming that a reasonable range of values was being stochastically sampled, rather than having a parameter “stuck” on a particular value.

Majority-rule dated consensus trees were generated by using TreeAnnotator v1.73 on the posterior sample of trees for each analysis. The node date and branch rate estimates from each of the nine TreeAnnotator-derived consensus trees were compared to each other using linear regression procedures in R. In addition to comparing the estimates between

runs, the uncertainty of these estimates was compared, as measured by the 95% highest posterior density (HPD width) on the dates and rates. As the uncertainty in dates and rates is typically heteroscedastic (i.e. nodes with higher absolute ages will typically have higher absolute uncertainty in age), the runs were also compared by relative uncertainty in dates and rates as measured by CV (coefficient of variation). CV is equal to the standard deviation divided by the mean; here, standard deviation (SD) was approximated as  $SD = ((95\% \text{ HPD width}) / (2 \times 1.96))$ , an approximation which applies well as almost all nodes had approximately normal distributions (the only exception were the LUCA-related nodes, but these were not a target of inference in our study).

In each comparison of two BEAST runs (comparing node ages, branch rates, uncertainty measures, etc.) the null hypothesis was that the linear regression between the two would have a slope of 1:1 and an intercept of 0. This null expectation would be the guaranteed result if two BEAST runs produced e.g. identical estimates of node age for all nodes, or if node ages from a run were regressed against themselves. Bonferroni corrections for multiple testing were applied, multiplying  $p$ -values by 5 tests for the comparisons between standard and cross-calibration/cross-bracing runs (Table S2).

As BEAST analyses are complex and highly parameterized, priors may interact in unexpected ways to influence results, and this might particularly be an issue as our analyses contain many nodes with dating priors and fixed topology. To ensure that the data rather than the combination of dating priors and tree prior was dominating the inference of node ages for non-calibration nodes deeper in the tree, an alignment lacking data, with all amino acids changed to gaps, was generated and run in BEAST with all calibration points. The no-data run yielded calibration node dates closely following the prior specifications, and non-calibration nodes estimated by the no-data dataset had either dramatically different dates or were not resolved at all, giving strong evidence that the amino acid sequence data are strongly influencing our results.

### **Supplemental Analysis of BEAST runs:**

**Node Age Uncertainty.** The intercept terms on the regressions were not significantly different than zero (Table S2). Comparing the cross-calibration and cross-bracing analyses indicated that cross-bracing may provide slightly lower uncertainty (3%) than the cross-calibration method (Table S2); however, the effect was not statistically significant ( $p=0.598$ ).

**Inference of Coefficient of Variation.** Examination of the CV in node age shows that the reduction in node age uncertainty was not due to mere reductions in the average node age (a concern because node age and node age uncertainty are strongly correlated in BEAST analyses).  $\alpha/\beta$ -calibrated analyses had CVs 14% lower than standard analyses, with strongly significant  $p$ -values (all  $p$ -values  $<0.00023$ ). The regression of the CVs of

the  $\alpha$ -calibrated and  $\beta$ -calibrated analyses against the CVs of cross-braced showed an even greater deviation from a 1:1 slope (22% and 18% lower, respectively), however, this is more than compensated for by a significant positive intercept. A similar, but weaker, effect was found when comparing the  $\alpha/\beta$ -braced and  $\alpha/\beta$ -calibrated runs. In other words, when CVs from a cross-calibrated analysis are used to predict CVs from a cross-braced analysis, the cross-bracing-derived CVs are typically higher by a fixed amount (the intercept), but this effect declines at very high CVs due to a slope below 1:1. The typically slightly higher CVs in the cross-bracing analysis compared to the cross-calibration analysis can be attributed mostly to the fact that the uncertainty in node age declines little between the two methods, but node ages are consistently slightly lower in the cross-bracing analysis.

**Inference of node dates.** There is no significant difference between the ages estimated by the  $\alpha$ -calibrated or  $\beta$ -calibrated analyses (runs 1 and 2) and the cross-calibrated analysis (run 3; Table S2); the slope between mean node ages estimated by the two methods is not significantly different from 1:1.

**Inference of branch rates.** Only small differences were found in the estimation of branch substitution rates (Table S2) between the all cross-calibrated runs. Intercepts were not significantly different from zero, and the slope was indistinguishable from 1:1 for  $\beta$ -calibrated vs. cross-calibrated analysis. There was a significant difference in the slope between  $\alpha$ -calibrated and  $\alpha/\beta$ -calibrated analysis, but the size of the effect was small. Comparison of the rate estimates from the  $\alpha$ -calibrated and  $\beta$ -calibrated runs and the  $\alpha/\beta$ -calibrated run with the  $\alpha/\beta$ -braced run showed significantly negative slopes, however the regression also contains a highly significant and positive intercept with a large effect size; the estimated means of substitution rate on each branch of the cross-bracing analysis are very often higher than for the other analyses.

**Uncertainty in branch rates.** Comparing simple and complex analyses in their estimates of uncertainty in mean per-branch substitution rate estimates (Table S2) appear to show a violation of the linear model used in regression, with the relationship between rate uncertainty in the  $\alpha$ -calibrated and  $\beta$ -calibrated runs and rate uncertainty in the  $\alpha/\beta$ -calibrated and  $\alpha/\beta$ -braced runs showing a linear relationship at lower uncertainties, but flattening out at mid-to-high rate uncertainties. This is expressed in the regression statistics as highly significant p-values ( $p < 1.7E-07$  for all comparisons) and large effect size with slopes 40-60% below 1:1, and intercepts representing approximately 25% of the overall mean rate uncertainty. When the two more complex analyses ( $\alpha/\beta$ -calibrated and  $\alpha/\beta$ -braced) are compared, it is evident that the cross-bracing analysis produces higher absolute estimates of rate uncertainty than those of the cross-calibration estimate by a substantial margin, mostly due to the large intercept ( $p = 7.43E-13$ ).

The higher absolute uncertainty in the rate estimates for cross-bracing may simply be due to the higher rates estimated by cross-bracing, and heteroscedasticity in rate uncertainty. Examination of the regressions based on rate CV (Table S2) shows that this is indeed the case, but that rate uncertainties increased proportionately less than the rate means in the cross-bracing analysis. Therefore, rate CV is typically lower for cross-braced estimates than for either  $\alpha$ -calibrated,  $\beta$ -calibrated, or  $\alpha/\beta$ -calibrated analyses. As with estimates of branch rate uncertainty, there appears to be some evidence of nonlinearity, with the relationship between rate CVs from simpler and more complex analyses flattening out at higher rate CVs. In summary, cross-braced analyses have less relative uncertainty in their estimates of branch substitution rates when rates and rate uncertainty are high.

**Effect of reducing the number of node age calibrations.**  $\alpha$ -calibrated runs using less calibration dates, runs 4-8 (Table S3), were compared against the  $\alpha/\beta$ -calibrated and  $\alpha/\beta$ -braced runs. Datasets with fewer calibration points had systematically slightly lower estimates of node ages than the  $\alpha$ -calibrated and  $\alpha/\beta$ -calibrated runs. The differences were statistically significant (Table S3), with slope and intercept both positively elevated. Interestingly, the difference in node age estimates closely matched the difference in node age estimates between  $\alpha/\beta$ -calibrated and  $\alpha/\beta$ -braced runs, such that the node age differences between runs 4-8 and the cross-bracing run are minimal and mostly insignificant.

Uncertainty in node age was not dramatically different between the  $\alpha$ -calibrated dataset with all calibration node ages versus subsets of these calibrations (Table S3), with all of the statistical tests being non-significant at the  $p=0.05$  level, or barely significant with small effect. Comparison of the reduced-calibrations ( $\alpha$ -calibrated) runs to the  $\alpha/\beta$ -calibrated analysis showed that runs 7 and 8 (only plant-based calibrations, and only vacuolar calibrations) had significantly higher uncertainty on average (slopes respectively 12% and 15% higher than 1:1) than the  $\alpha/\beta$ -calibrated analysis. When compared to the  $\alpha/\beta$ -braced analysis, all reduced-calibrations runs had significantly higher uncertainty, but as found above, higher uncertainty is expected even with the complete set of calibrations used on the  $\alpha$ -calibrated dataset (Table S2).

Comparison of the node age CVs from runs 4-8 to the  $\alpha$ -calibrated,  $\alpha/\beta$ -calibrated, and  $\alpha/\beta$ -braced runs indicated consistently significantly higher CVs in the reduced-calibration runs. In the case of comparison to  $\alpha$ -calibrated and  $\alpha/\beta$ -calibrated runs, this effect is likely due to the consistently lower estimates of node ages in the reduced-calibrations runs, plus perhaps slightly increased absolute uncertainties in node age. The fact that the comparison of node age CV to between runs 4-8 and the cross-bracing run shows highly significant difference in slope (all  $p$ -values  $< 0.001$ ), with the reduced-calibrations having 35%-49% higher CVs, indicates that the effect cannot be solely due to differences in node age. Inspection of the node age CV regression plots shows that certain nodes are dramatically different in CV between the reduced calibrations runs and the all-

calibrations runs. These nodes showing a large difference are the ones which are calibrated in one analysis and not in the other. In other words, when a node that was calibrated in one run is uncalibrated in another, the uncertainty in its age increases dramatically, while the mean estimate of its age changes comparatively little.

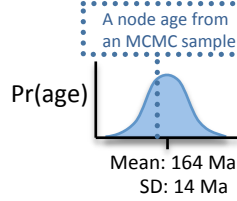
Similar comparisons of branch rate estimates and uncertainty are shown in Table S4. Branches typically have slightly higher rates in the reduced-calibrations runs, explaining the slightly lower average node ages, however, unlike in the node age comparisons, the differences are not significant (Table S4) in the comparison to  $\alpha$ -calibrated (run 1), probably because of greater scatter in the rate estimates between runs compared to the age estimates.

Absolute uncertainty in branch rates is almost always higher in the reduced-calibration runs; of course, this is a partial product of the higher rate estimates in these runs. Comparison to the  $\alpha/\beta$ -braced analysis, which has similar mean estimates of node ages and branch rates, shows dramatically increased CV in the reduced-calibrations runs, indicating that removing calibration points increases relative uncertainty; however, comparison to the  $\alpha$ -calibrated run shows no significant differences in rate CV. Overall, these results show that uncertainty is reduced the most when calibration points from both alpha and beta subunits are utilized with a cross-calibration approach, and is reduced further if the cross-bracing approach is used to tie the node dates together. However, the effect of removing node calibrations within subgroups of the alpha subtree is minimal on the overall estimates, although the effect of changing an individual calibrated node to an uncalibrated one can be a significant increase in age uncertainty. Node age uncertainty is strongly affected by nearby date calibrations, but overall mean node ages are determined by the estimates of branch rates. In a BEAST uncorrelated relaxed-clock analysis, the rates of each branch are drawn from a common lognormal distribution, which should be robust to the inclusion or exclusion of small groups of calibration points.

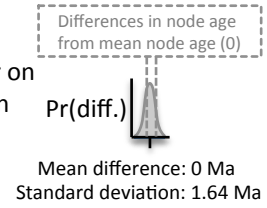
### **Supplemental Figures S1-S10:**

## Prior distributions

A. Traditional prior on node date, e.g. prior distribution on date of the monocot-dicot split:



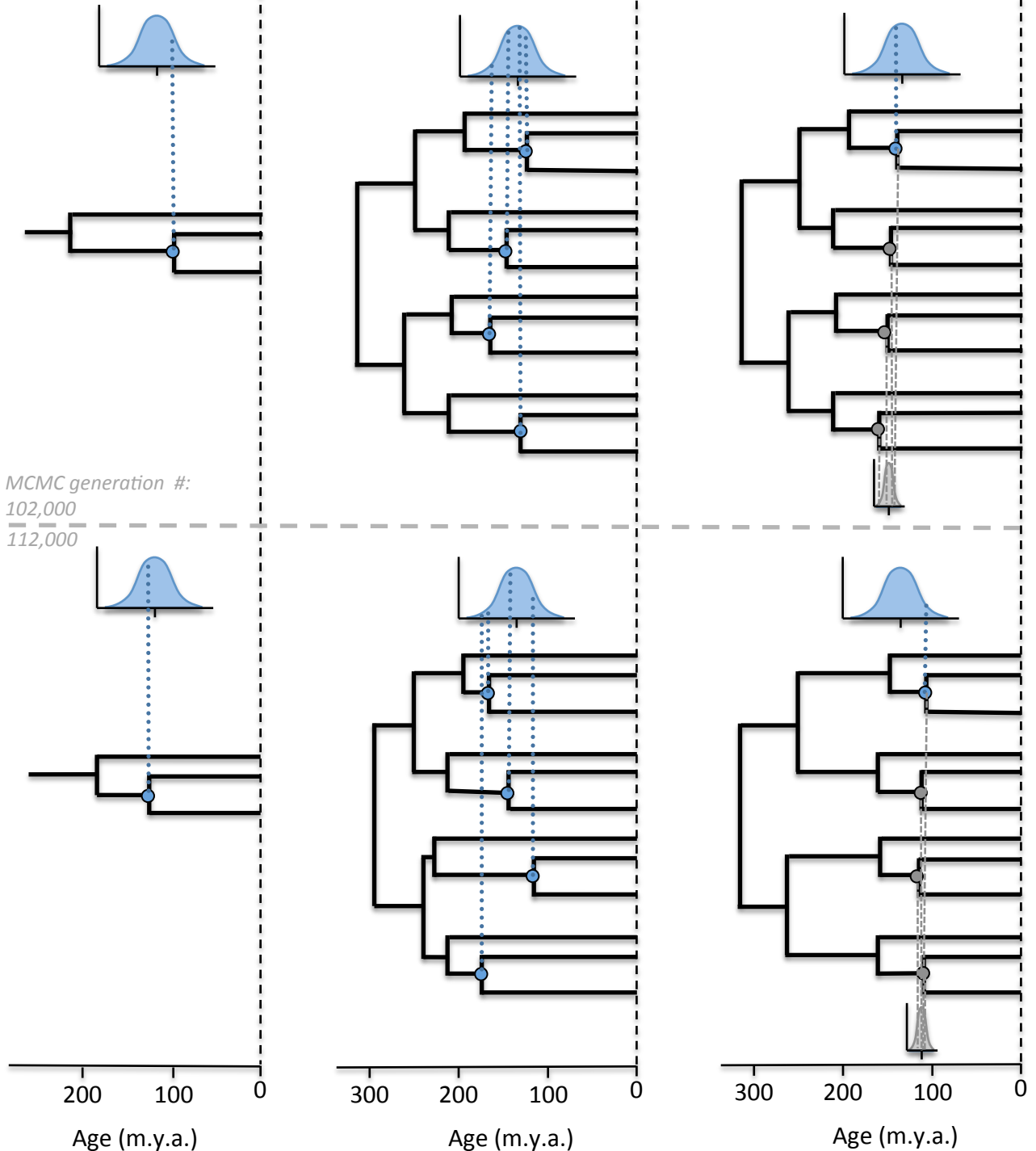
B. New prior on differences in node ages:



### 1. Traditional dating analysis

### 2. Cross-calibration

### 3. Cross-bracing



**Figure S1.** Summary of the cross-calibration and cross-bracing strategies. See main text for details.

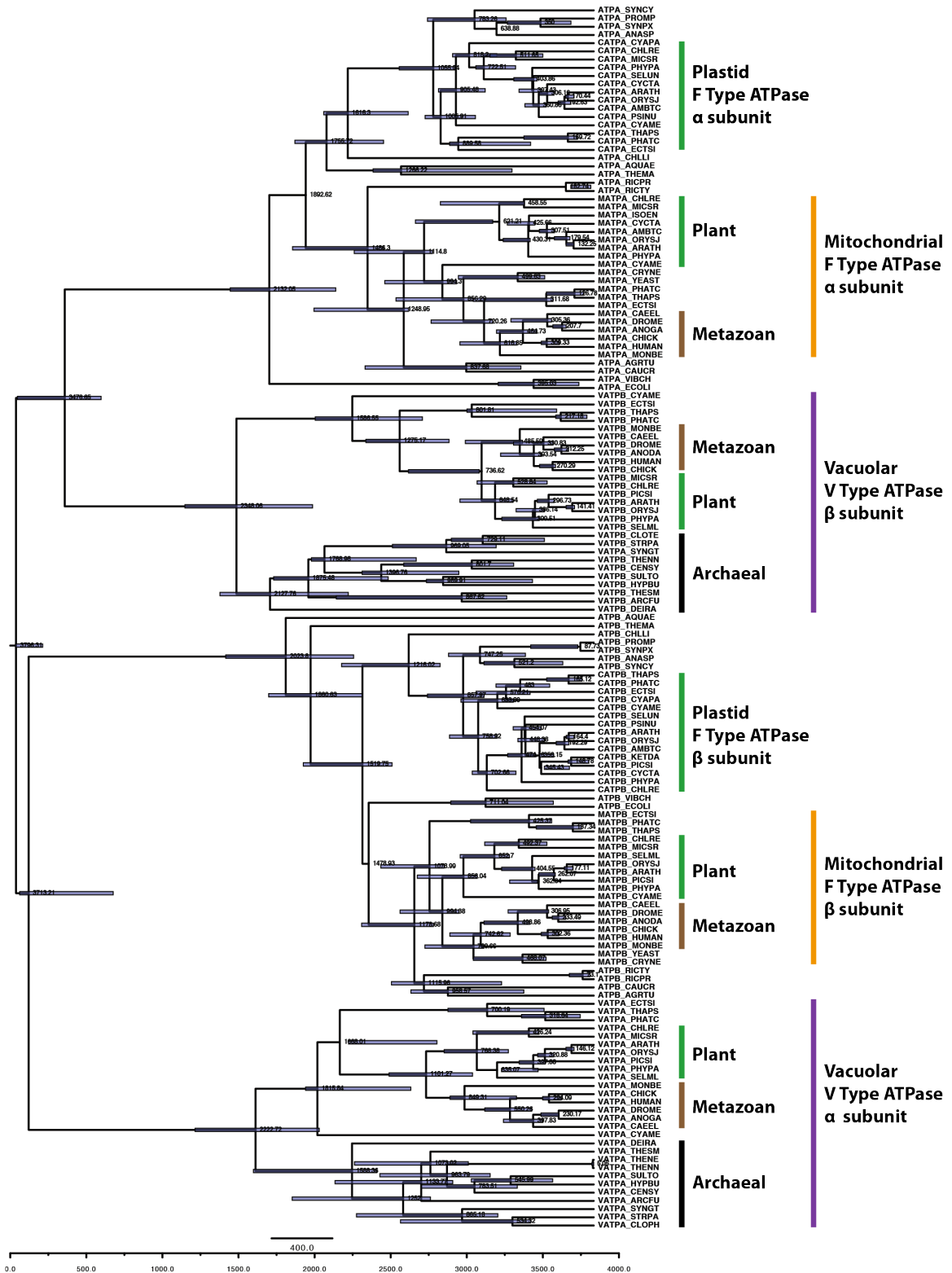


Figure S2: ATPase  $\alpha$  and  $\beta$  subunits cross-calibrated chronogram.



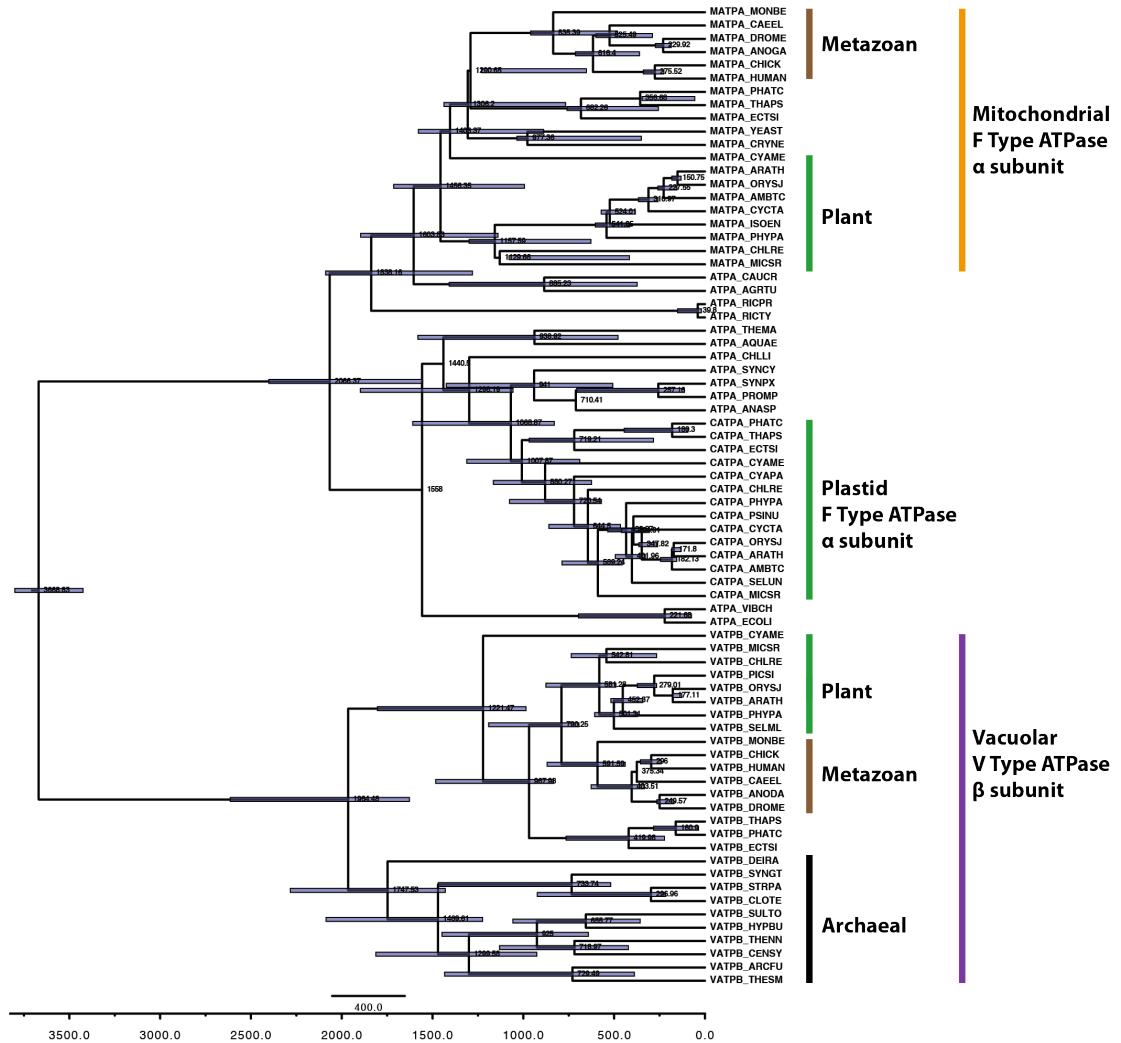


Figure S4. ATPase  $\alpha$ -subunit cross-calibrated chronogram.

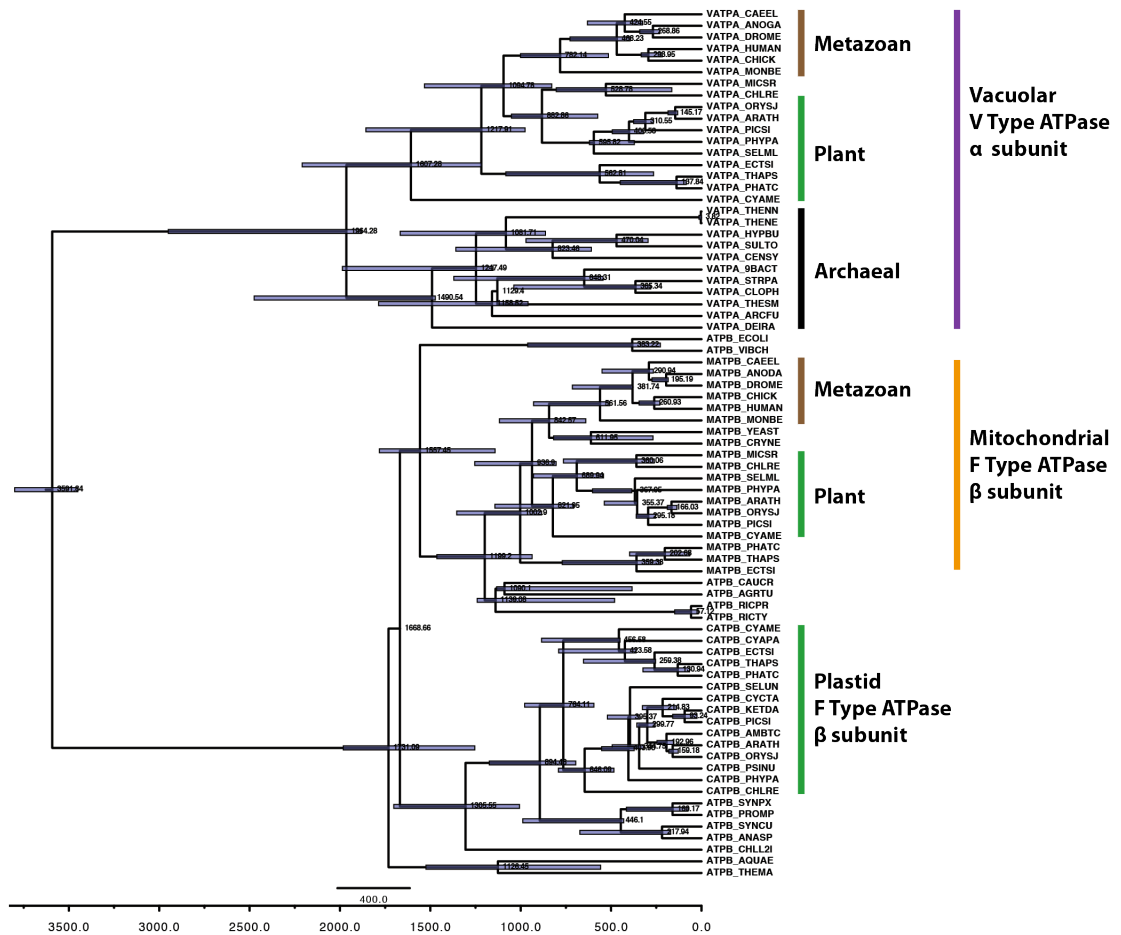
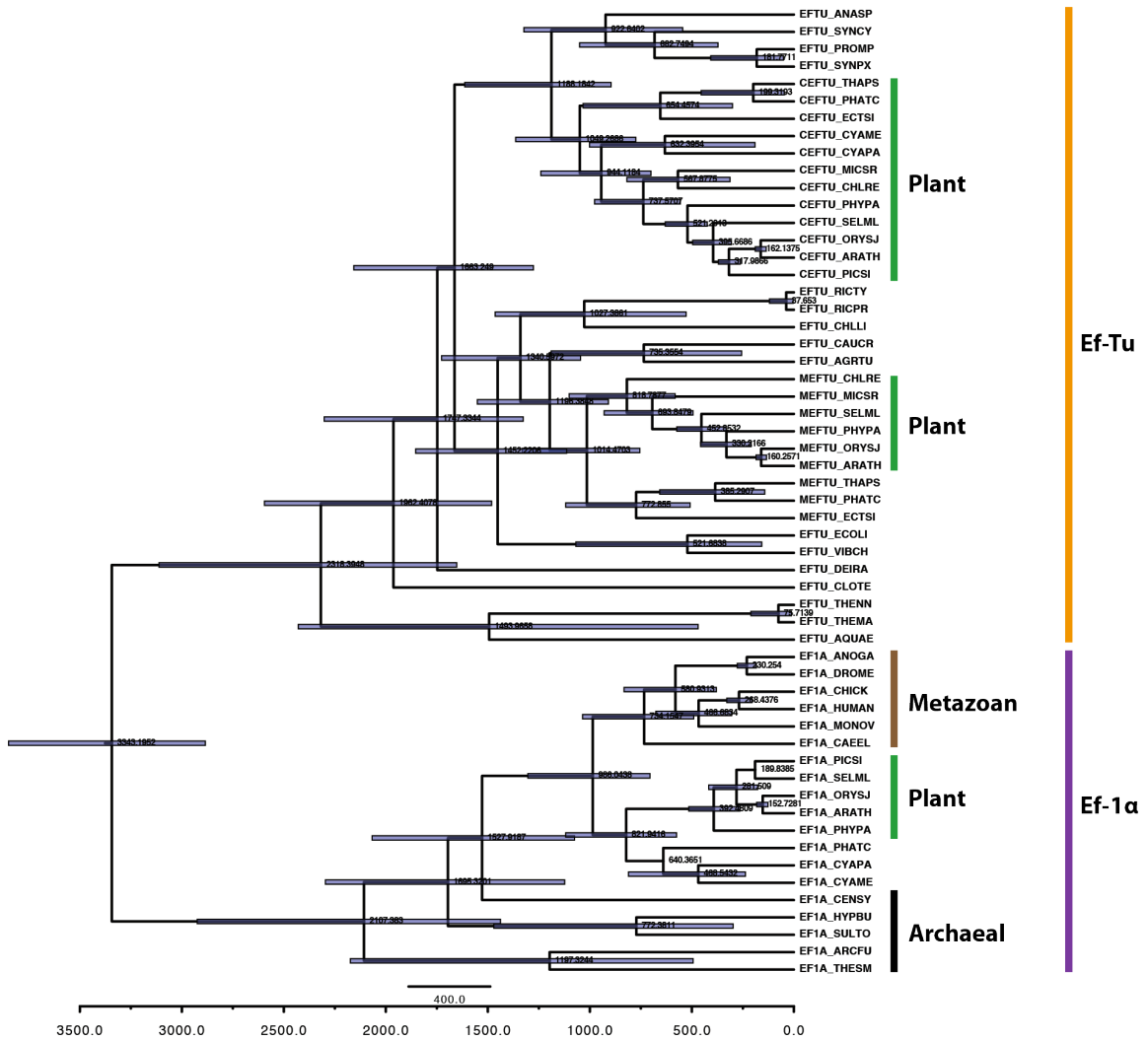
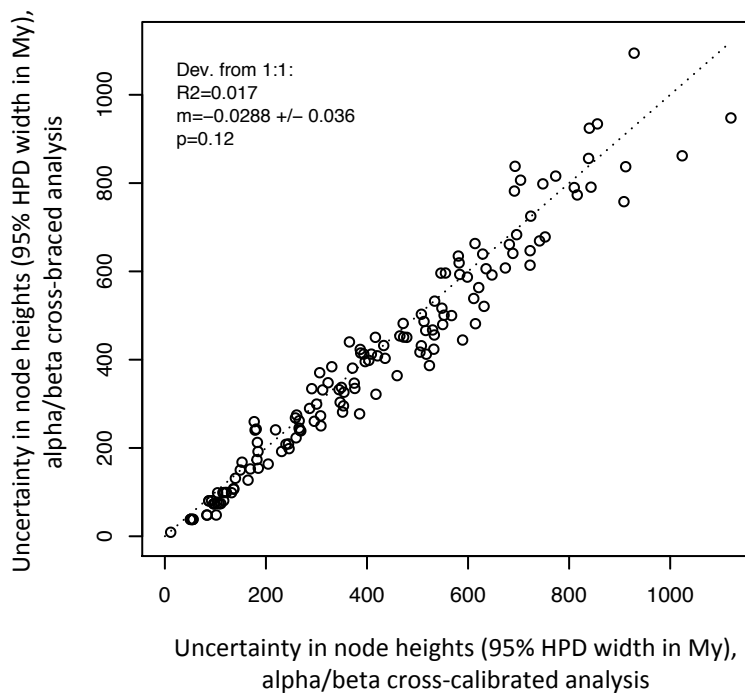


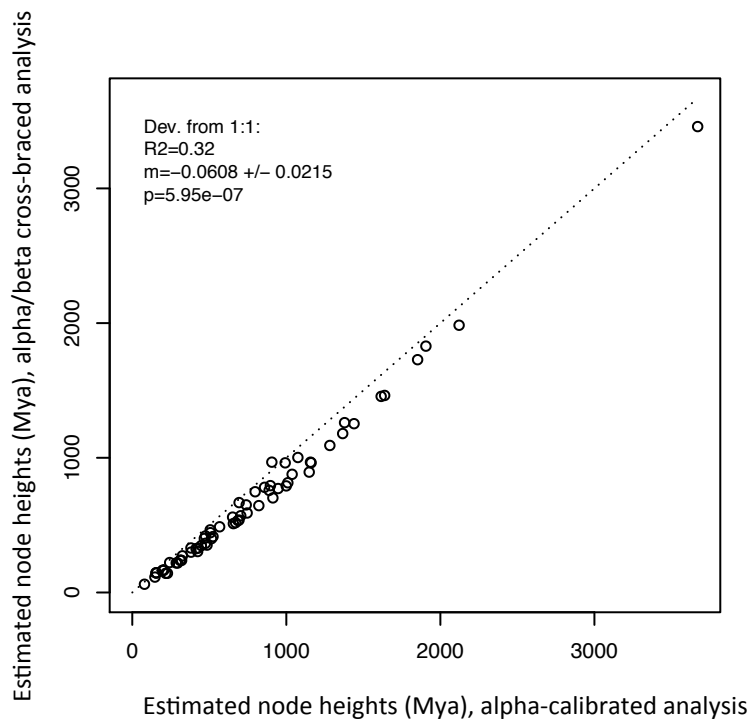
Figure S5: ATPase beta-subunit cross-calibrated chronogram.



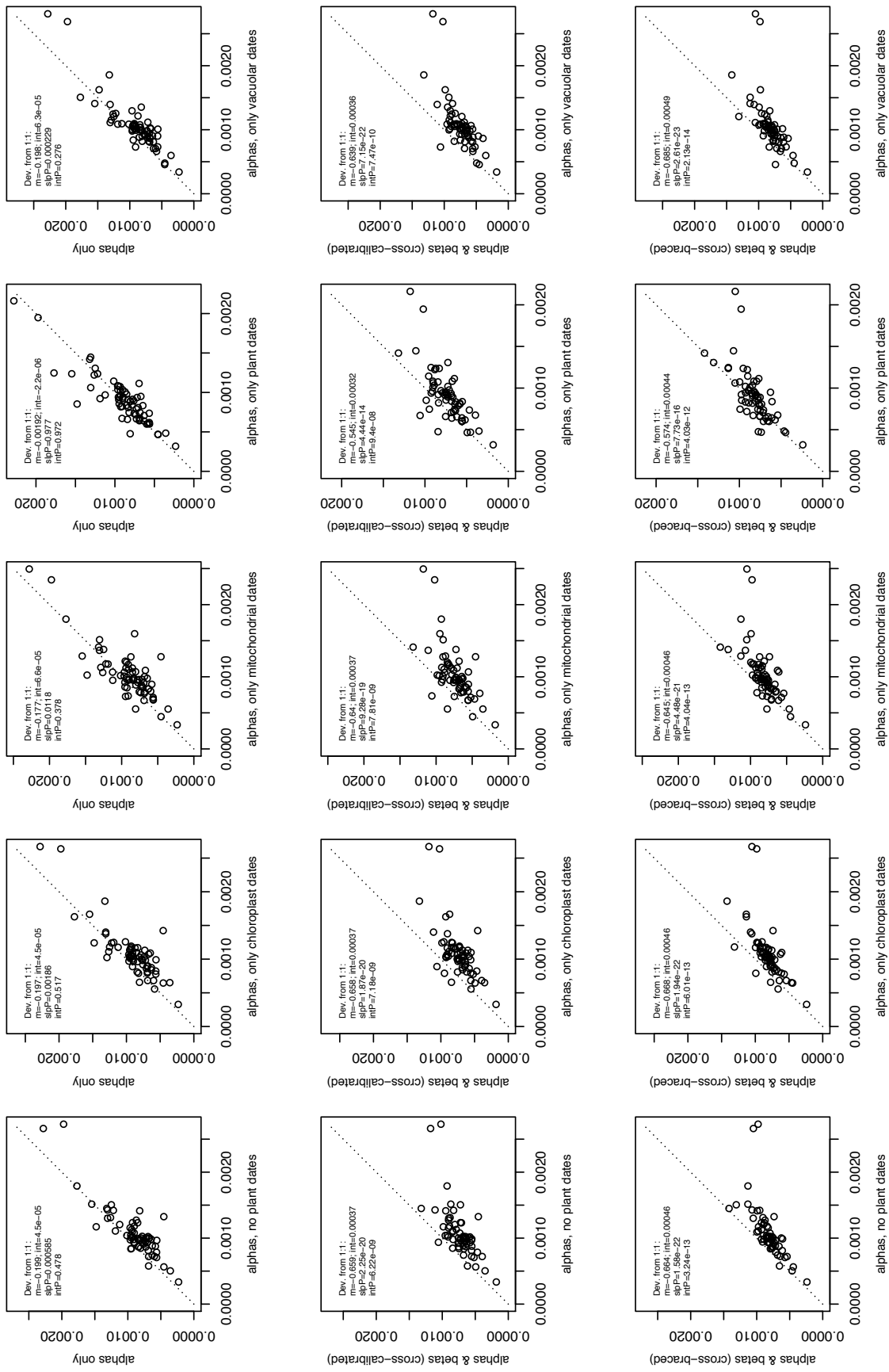
**Figure S6:** ATPase  $\beta$ -subunit cross-calibrated chronogram



**Figure S7:** No significant difference in uncertainty (as measured by 95% HPD width of node height estimates) between alpha/beta cross-calibrated and cross-braced analyses. Regression statistics are for deviation from 1:1 line.



**Figure S8:** Node heights as estimated by alpha/beta cross-bracing tend to be slightly lower than with other methods. Regression statistics are for deviation from 1:1 line.



**Figure S9:** Uncertainty in branch rate estimates as measured by 95% HPD width of branch rates. Five analysis with alpha-only subunits and reduced numbers of calibration points (runs 1 in row 1 (alpha-only, all calibration points), with run 3 in row 2 (alpha/beta cross-calibrated, all calibration points), and with run 9 in row 3 (alpha/beta cross-braced, all calibration points). The dotted line shows the 1:1 slope between the analyses; points below the line indicate lower uncertainty in the analyses using all calibration points (runs 1, 3, and 9). Regression statistics are calculated for deviations from the 1:1 line.

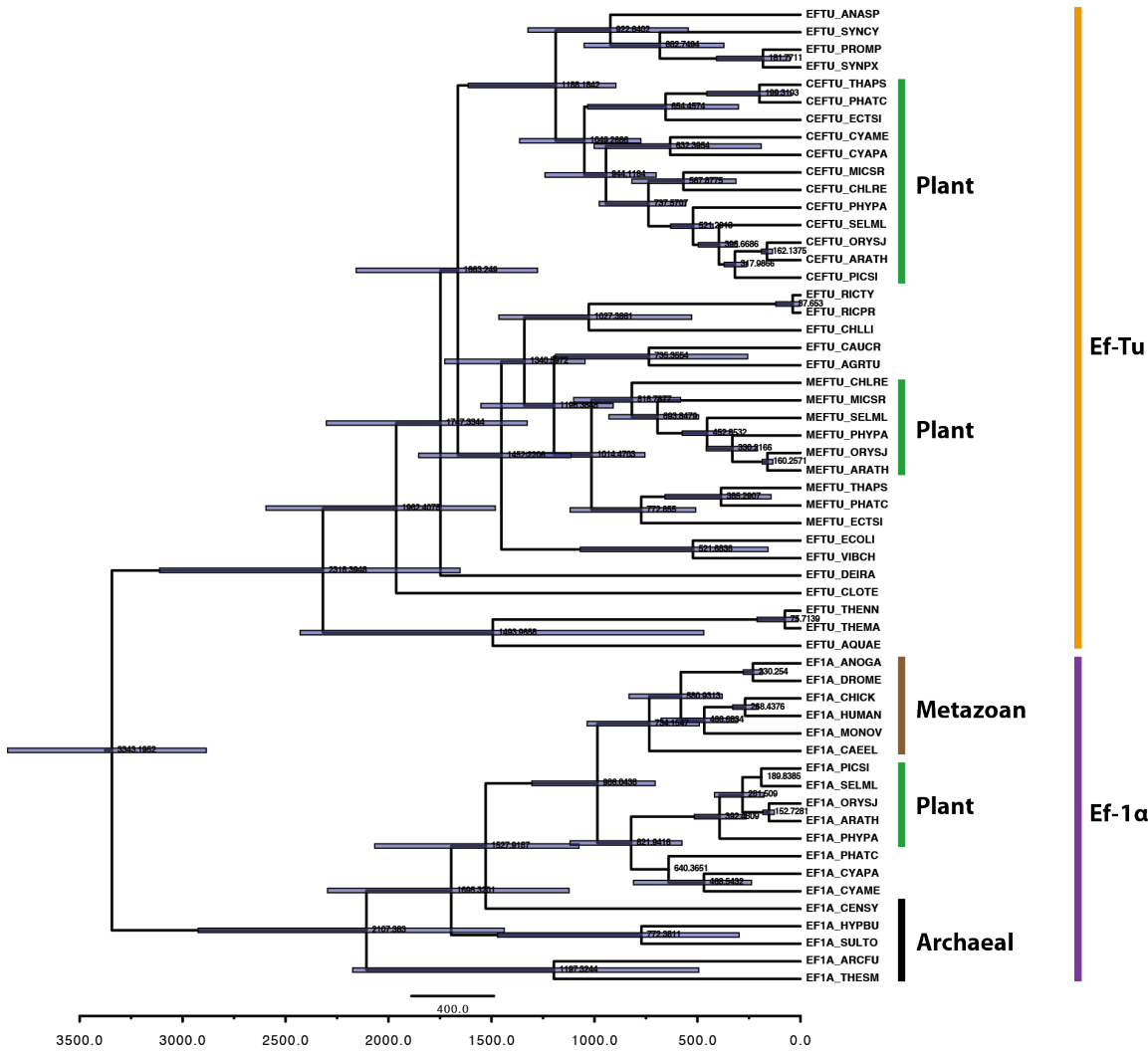


Figure S10: Ef-Tu/ Ef-1α cross-calibrated chronogram.

## **Supplemental Tables S1-S6:**

**Table S1. Description of ATPase BEAST runs with varying levels of calibration priors**

<b>Run</b>	<b>Name</b>	<b>Calibration Type</b>	<b>Description</b>
1	$\alpha$ -cross-calibrated	cross-calibrated	chronogram of only $\alpha$ subunits; all calibration points are used
2	$\beta$ -cross-calibrated	cross-calibrated	chronogram of only $\beta$ subunits; all calibration points are used
3	$\alpha/\beta$ -cross-calibrated	cross-calibrated	chronogram of both $\alpha$ and $\beta$ subunits; all calibration points are used
4	$\alpha$ -metazoan-cross-calibrated	cross-calibrated	chronogram of only $\alpha$ subunits; only metazoan calibration points are used
5	$\alpha$ -plant-cross-calibrated	cross-calibrated	chronogram of only $\alpha$ subunits; only plant calibration points are used
6	$\alpha$ -chloroplast-cross-calibrated	cross-calibrated	chronogram of only $\alpha$ subunits; only plastid calibration points are used
7	$\alpha$ -mitochondria-cross-calibrated	cross-calibrated	chronogram of only $\alpha$ subunits; only mitochondrial calibration points are used
8	$\alpha$ -vacuole-cross-calibrated	cross-calibrated	chronogram of only $\alpha$ subunits; only vacuolar calibration points are used
9	$\alpha/\beta$ -cross-braced	cross-braced	chronogram of both $\alpha$ and $\beta$ subunits; all calibration points are used

**Table S2.** Statistical tests for deviations from a 1:1 relationship between calibration analyses. Linear models were built using a simpler analysis as a predictor (x-axis), and a more complex analysis as a response (y-axis). To remove the 1:1 relationship, the response variable was detrended by subtracting the predictor variable. Thus, if the relationship between e.g. node age uncertainty from a simpler analysis and node age uncertainty in a more complex analysis is truly 1:1, then the detrended response variable will be exactly flat (slope and intercept = 0). A negative slope of -0.05 would indicate that uncertainty in node age is on average 5% lower in the more complex analysis (if the intercept is close to 0). The results show that the mean estimates of node age in cross-calibrated analyses are not significantly lower than in non-calibrated, alpha-alone analyses. *P*-values were corrected by Bonferroni correction for 5 tests. \*=*p*<0.05; \*\*=*p*<0.01; \*\*\*=*p*<0.001.

Comparison	Slope of difference from 1:1 line			Intercept		
	Slope of difference from 1:1 line	95% CI	p	Intercept	95% CI	p
<b>age mean</b>						
alpha vs. cross-calibrated	0.02	(-0.01,0.04)	0.723	2.5	(-20.1,25.2)	4.13585
beta vs. cross-calibrated	-0.04	(-0.08,-0.01)	0.054	31.7	(2,61.5)	0.20073
alpha vs. cross-braced	-0.06	(-0.08,-0.04)	2.97E-06 ***	-57.2	(-77.9,-36.5)	4.66E-06 ***
beta vs. cross-braced	-0.07	(-0.1,-0.04)	1.19E-05 ***	-37.3	(-61.6,-13)	0.01836 *
cross-calibrated vs. cross-braced	-0.05	(-0.07,-0.03)	5.75E-08 ***	-65.0	(-81.4,-48.5)	8.64E-12 ***
<b>uncertainty in node age (HPD width)</b>						
alpha vs. cross-calibrated	-0.22	(-0.3,-0.15)	1.48E-06 ***	30.8	(-12.6,74.2)	0.84297
beta vs. cross-calibrated	-0.16	(-0.22,-0.09)	9.50E-05 ***	26.9	(-6.9,60.7)	0.61656
alpha vs. cross-braced	-0.26	(-0.32,-0.19)	3.61E-10 ***	16.6	(-20.6,53.8)	1.92425
beta vs. cross-braced	-0.14	(-0.22,-0.07)	2.54E-03 **	3.4	(-35.2,42.1)	4.31047
cross-calibrated vs. cross-braced	-0.03	(-0.06,0.01)	0.598	-7.9	(-25,9.3)	1.84405
<b>age coefficient of variation (CV: std. dev. / mean)</b>						
alpha vs. cross-calibrated	-0.14	(-0.2,-0.08)	2.32E-04 ***	-0.001	(-0.014,0.012)	4.55538
beta vs. cross-calibrated	-0.14	(-0.17,-0.11)	2.96E-11 ***	0.007	(0,0.015)	0.31751
alpha vs. cross-braced	-0.22	(-0.28,-0.15)	4.44E-08 ***	0.024	(0.01,0.037)	0.00453 **
beta vs. cross-braced	-0.18	(-0.22,-0.15)	2.48E-13 ***	0.028	(0.019,0.036)	2.00E-07 ***
cross-calibrated vs. cross-braced	-0.08	(-0.11,-0.04)	3.67E-05 ***	0.024	(0.018,0.03)	9.21E-12 ***
<b>rate mean</b>						
alpha vs. cross-calibrated	-0.12	(-0.16,-0.08)	3.79E-06 ***	0.000031	(0.00001,0.00006)	0.10687
beta vs. cross-calibrated	-0.06	(-0.12,-0.01)	0.138	0.000035	(0.00001,0.00007)	0.12761
alpha vs. cross-braced	-0.24	(-0.3,-0.19)	1.75E-12 ***	0.000167	(0.00014,0.0002)	1.13E-14 ***
beta vs. cross-braced	-0.23	(-0.3,-0.16)	2.09E-08 ***	0.000171	(0.00014,0.00021)	8.95E-13 ***
cross-calibrated vs. cross-braced	-0.16	(-0.2,-0.13)	1.53E-16 ***	0.000142	(0.00012,0.00016)	9.54E-31 ***
<b>uncertainty in branch rate (HPD width)</b>						
alpha vs. cross-calibrated	-0.57	(-0.66,-0.48)	1.59E-18 ***	0.000351	(0.00027,0.00044)	1.43E-10 ***
beta vs. cross-calibrated	-0.42	(-0.52,-0.32)	5.56E-11 ***	0.000269	(0.00018,0.00035)	1.76E-07 ***
alpha vs. cross-braced	-0.61	(-0.7,-0.52)	3.25E-20 ***	0.000470	(0.00039,0.00055)	1.46E-15 ***
beta vs. cross-braced	-0.51	(-0.61,-0.41)	6.07E-14 ***	0.000398	(0.00031,0.00048)	8.35E-13 ***
cross-calibrated vs. cross-braced	-0.27	(-0.36,-0.19)	8.75E-09 ***	0.000272	(0.00021,0.00034)	7.43E-13 ***
<b>rate coefficient of variation (CV: std. dev. / mean)</b>						
alpha vs. cross-calibrated	-0.29	(-0.43,-0.14)	1.50E-03 **	0.078	(0.01,0.15)	0.13667
beta vs. cross-calibrated	-0.14	(-0.28,0)	0.309	0.026	(-0.04,0.09)	2.10995
alpha vs. cross-braced	-0.47	(-0.6,-0.33)	1.14E-08 ***	0.142	(0.08,0.2)	1.04E-04 ***
beta vs. cross-braced	-0.22	(-0.33,-0.12)	3.98E-04 ***	0.046	(0,0.09)	0.29495
cross-calibrated vs. cross-braced	-0.30	(-0.38,-0.23)	1.88E-12 ***	0.106	(0.07,0.14)	2.33E-09 ***

**Table S3:** Tests for deviations from a 1:1 relationship between the age-related node statistics from BEAST analyses based on subsets of alpha node calibrations, and (a) the complete list of alpha node calibrations, (b) the cross-calibration method, and (c) the cross-bracing method.

Comparison	Slope of difference from 1:1 line			Intercept	95% CI		p
	95% CI	p	95% CI		p		
<b>age mean</b>							
<i>(a) comparisons to alpha with all calibrations</i>							
age mean: alphas, no plant dates vs. alphas only	0.06	(0.04,0.09)	4.849E-07 ***	65.34	(45.95,84.72)	8.034E-09 ***	
age mean: alphas, only chloroplast dates vs. alphas only	0.09	(0.05,0.12)	6.160E-06 ***	103.43	(75.3,131.55)	7.455E-10 ***	
age mean: alphas, only mitochondrial dates vs. alphas only	0.06	(0.04,0.09)	2.051E-05 ***	82.66	(60.08,105.24)	6.870E-10 ***	
age mean: alphas, only plant dates vs. alphas only	0.03	(0.01,0.05)	2.099E-03 **	37.02	(18.82,55.21)	1.695E-04 ***	
age mean: alphas, only vacuolar dates vs. alphas only	0.06	(0.03,0.08)	8.355E-06 ***	69.23	(49.57,88.89)	2.264E-09 ***	
<i>(b) comparison to cross-calibration</i>							
age mean: alphas, no plant dates vs. alphas & betas (cross-calibrated)	0.08	(0.04,0.12)	4.051E-04 ***	71.72	(37.71,105.72)	1.073E-04 ***	
age mean: alphas, only chloroplast dates vs. alphas & betas (cross-calibrated)	0.10	(0.06,0.15)	3.137E-05 ***	108.46	(71.3,145.63)	2.806E-07 ***	
age mean: alphas, only mitochondrial dates vs. alphas & betas (cross-calibrated)	0.08	(0.04,0.12)	1.309E-04 ***	85.95	(51.58,120.31)	6.476E-06 ***	
age mean: alphas, only plant dates vs. alphas & betas (cross-calibrated)	0.05	(0.01,0.08)	1.033E-02 *	42.54	(11.65,73.43)	0.009 **	
age mean: alphas, only vacuolar dates vs. alphas & betas (cross-calibrated)	0.07	(0.04,0.11)	8.403E-05 ***	70.81	(41,100.61)	1.671E-05 ***	
<i>(c) comparison to cross-bracing</i>							
age mean: alphas, no plant dates vs. alphas & betas (cross-braced)	0.00	(-0.03,0.03)	0.996	3.11	(-20.12,26.35)	0.794	
age mean: alphas, only chloroplast dates vs. alphas & betas (cross-braced)	0.02	(0,0.05)	0.095	36.61	(13.72,59.51)	2.572E-03 **	
age mean: alphas, only mitochondrial dates vs. alphas & betas (cross-braced)	0.01	(-0.02,0.03)	0.634	15.60	(-3.32,34.52)	0.111	
age mean: alphas, only plant dates vs. alphas & betas (cross-braced)	-0.03	(-0.06,0)	0.031 *	-22.33	(-45.95,1.29)	0.068	
age mean: alphas, only vacuolar dates vs. alphas & betas (cross-braced)	-0.01	(-0.03,0.02)	0.662	4.75	(-15.53,25.04)	0.647	
<b>uncertainty in node age (HPD width)</b>							
<i>(a) comparisons to alpha with all calibrations</i>							
age HPD width: alphas, no plant dates vs. alphas only	0.11	(0.03,0.2)	0.012 *	-40.91	(-87.65,5.83)	0.091	
age HPD width: alphas, only chloroplast dates vs. alphas only	0.10	(-0.01,0.2)	0.074	9.66	(-44.55,63.88)	0.728	
age HPD width: alphas, only mitochondrial dates vs. alphas only	0.12	(0.02,0.22)	0.021 *	-11.13	(-63.45,41.18)	0.678	
age HPD width: alphas, only plant dates vs. alphas only	0.06	(0,0.13)	0.055	0.25	(-34.65,35.15)	0.989	
age HPD width: alphas, only vacuolar dates vs. alphas only	0.01	(-0.08,0.11)	0.771	-7.36	(-64,49.27)	0.800	
<i>(b) comparison to cross-calibration</i>							
age HPD width: alphas, no plant dates vs. alphas & betas (cross-calibrated)	-0.09	(-0.19,0)	0.057	-17.62	(-67.78,32.54)	0.494	
age HPD width: alphas, only chloroplast dates vs. alphas & betas (cross-calibrated)	-0.09	(-0.18,-0.01)	0.035 *	11.74	(-31.61,55.09)	0.597	
age HPD width: alphas, only mitochondrial dates vs. alphas & betas (cross-calibrated)	-0.08	(-0.17,0.01)	0.098	-3.78	(-52,44.43)	0.878	
age HPD width: alphas, only plant dates vs. alphas & betas (cross-calibrated)	-0.12	(-0.19,-0.05)	0.002 **	9.77	(-27.95,47.48)	0.614	
age HPD width: alphas, only vacuolar dates vs. alphas & betas (cross-calibrated)	-0.15	(-0.23,-0.07)	3.222E-04 ***	-8.13	(-52.15,35.88)	0.718	
<i>(c) comparison to cross-bracing</i>							
age HPD width: alphas, no plant dates vs. alphas & betas (cross-braced)	-0.11	(-0.18,-0.05)	1.351E-03 **	-39.82	(-74.92,-4.71)	0.030 *	
age HPD width: alphas, only chloroplast dates vs. alphas & betas (cross-braced)	-0.11	(-0.17,-0.05)	7.710E-04 ***	-8.06	(-38.46,22.34)	0.605	
age HPD width: alphas, only mitochondrial dates vs. alphas & betas (cross-braced)	-0.09	(-0.15,-0.02)	9.042E-03 **	-25.90	(-59.3,7.5)	0.133	
age HPD width: alphas, only plant dates vs. alphas & betas (cross-braced)	-0.16	(-0.21,-0.1)	4.198E-07 ***	-5.23	(-34.26,23.8)	0.725	
age HPD width: alphas, only vacuolar dates vs. alphas & betas (cross-braced)	-0.18	(-0.24,-0.11)	4.466E-07 ***	-22.86	(-57.78,12.06)	0.204	
<b>age coefficient of variation (CV: std. dev. / mean)</b>							
<i>(a) comparisons to alpha with all calibrations</i>							
age CV: alphas, no plant dates vs. alphas only	-0.26	(-0.45,-0.06)	0.014 *	0.01	(-0.033,0.062)	0.548	
age CV: alphas, only chloroplast dates vs. alphas only	-0.44	(-0.64,-0.24)	5.201E-05 ***	0.05	(0.005,0.104)	0.033 *	
age CV: alphas, only mitochondrial dates vs. alphas only	-0.35	(-0.55,-0.15)	8.868E-04 ***	0.04	(-0.006,0.087)	0.089	
age CV: alphas, only plant dates vs. alphas only	-0.28	(-0.43,-0.13)	5.529E-04 ***	0.04	(0.007,0.074)	0.021 *	
age CV: alphas, only vacuolar dates vs. alphas only	-0.23	(-0.41,-0.05)	0.013 *	0.01	(-0.036,0.05)	0.743	
<i>(b) comparison to cross-calibration</i>							
age CV: alphas, no plant dates vs. alphas & betas (cross-calibrated)	-0.30	(-0.47,-0.13)	1.159E-03 **	0.00	(-0.044,0.039)	0.903	
age CV: alphas, only chloroplast dates vs. alphas & betas (cross-calibrated)	-0.42	(-0.58,-0.26)	3.843E-06 ***	0.02	(-0.017,0.063)	0.269	
age CV: alphas, only mitochondrial dates vs. alphas & betas (cross-calibrated)	-0.37	(-0.54,-0.2)	6.878E-05 ***	0.02	(-0.024,0.056)	0.430	
age CV: alphas, only plant dates vs. alphas & betas (cross-calibrated)	-0.34	(-0.47,-0.21)	4.814E-06 ***	0.03	(-0.004,0.056)	0.092	
age CV: alphas, only vacuolar dates vs. alphas & betas (cross-calibrated)	-0.26	(-0.4,-0.11)	9.659E-04 ***	-0.01	(-0.05,0.022)	0.443	
<i>(c) comparison to cross-bracing</i>							
age CV: alphas, no plant dates vs. alphas & betas (cross-braced)	-0.44	(-0.62,-0.27)	6.836E-06 ***	0.04	(-0.002,0.084)	0.065	
age CV: alphas, only chloroplast dates vs. alphas & betas (cross-braced)	-0.49	(-0.65,-0.33)	4.545E-08 ***	0.05	(0.012,0.088)	0.012 *	
age CV: alphas, only mitochondrial dates vs. alphas & betas (cross-braced)	-0.45	(-0.61,-0.29)	7.410E-07 ***	0.05	(0.008,0.085)	0.020 *	
age CV: alphas, only plant dates vs. alphas & betas (cross-braced)	-0.37	(-0.48,-0.25)	3.818E-08 ***	0.04	(0.016,0.068)	2.253E-03 **	
age CV: alphas, only vacuolar dates vs. alphas & betas (cross-braced)	-0.35	(-0.5,-0.21)	1.206E-05 ***	0.02	(-0.017,0.054)	0.322	

**Table S4.** Tests for deviations from a 1:1 relationship between the rate-related node statistics from BEAST analyses based on subsets of alpha node calibrations, and (a) the complete list of alpha node calibrations, (b) the cross-calibration method, and (c) the cross-bracing method.

	Slope of difference from 1:1 line	95% CI	p	Intercept	95% CI	p
<b>rate mean: alphas, no plant dates vs. alphas &amp; betas (cross-braced)</b>						
<b>rate mean</b>						
<i>(a) comparisons to alpha with all calibrations</i>						
rate mean: alphas, no plant dates vs. alphas only	0.04	(-0.03,0.11)	0.267	0.00	(-0.0001,-0.0001)	8.803E-05 ***
rate mean: alphas, only chloroplast dates vs. alphas only	0.02	(-0.06,0.09)	0.682	0.00	(-0.0002,-0.0001)	1.503E-05 ***
rate mean: alphas, only mitochondrial dates vs. alphas only	0.01	(-0.06,0.08)	0.836	0.00	(-0.0001,0)	7.518E-04 ***
rate mean: alphas, only plant dates vs. alphas only	0.06	(-0.01,0.12)	0.084	0.00	(-0.0001,0)	0.014 *
rate mean: alphas, only vacuolar dates vs. alphas only	0.01	(-0.03,0.06)	0.559	0.00	(-0.0001,0)	1.507E-05 ***
<i>(b) comparison to cross-calibration</i>						
rate mean: alphas, no plant dates vs. alphas & betas (cross-calibrated)	-0.07	(-0.15,0)	0.057	0.00	(-0.0001,0)	0.014 *
rate mean: alphas, only chloroplast dates vs. alphas & betas (cross-calibrated)	-0.10	(-0.17,-0.03)	0.010 *	0.00	(-0.0001,0)	2.742E-03 **
rate mean: alphas, only mitochondrial dates vs. alphas & betas (cross-calibrated)	-0.10	(-0.17,-0.03)	5.140E-03 **	0.00	(-0.0001,0)	0.033 *
rate mean: alphas, only plant dates vs. alphas & betas (cross-calibrated)	-0.06	(-0.13,0.01)	0.075	0.00	(-0.0001,0)	0.535
rate mean: alphas, only vacuolar dates vs. alphas & betas (cross-calibrated)	-0.10	(-0.16,-0.04)	9.051E-04 ***	0.00	(-0.0001,0)	0.052
<i>(c) comparison to cross-bracing</i>						
rate mean: alphas, no plant dates vs. alphas & betas (cross-braced)	-0.19	(-0.25,-0.13)	1.034E-07 ***	0.00	(0,0.0001)	4.192E-04 ***
rate mean: alphas, only chloroplast dates vs. alphas & betas (cross-braced)	-0.21	(-0.27,-0.15)	1.080E-08 ***	0.00	(0,0.0001)	7.984E-03 **
rate mean: alphas, only mitochondrial dates vs. alphas & betas (cross-braced)	-0.20	(-0.25,-0.15)	3.719E-11 ***	0.00	(0,0.0001)	9.974E-06 ***
rate mean: alphas, only plant dates vs. alphas & betas (cross-braced)	-0.21	(-0.29,-0.13)	2.225E-06 ***	0.00	(0.0001,0.0002)	4.819E-07 ***
rate mean: alphas, only vacuolar dates vs. alphas & betas (cross-braced)	-0.22	(-0.28,-0.17)	9.393E-11 ***	0.00	(0.0001,0.0001)	5.259E-07 ***
<b>uncertainty in branch rate (HPD width)</b>						
<i>(a) comparisons to alpha with all calibrations</i>						
rate HPD width: alphas, no plant dates vs. alphas only	-0.20	(-0.31,-0.09)	5.854E-04 ***	0.00	(-0.0001,0.0002)	0.478
rate HPD width: alphas, only chloroplast dates vs. alphas only	-0.20	(-0.32,-0.08)	1.865E-03 **	0.00	(-0.0001,0.0002)	0.517
rate HPD width: alphas, only mitochondrial dates vs. alphas only	-0.18	(-0.31,-0.04)	0.012 *	0.00	(-0.0001,0.0002)	0.378
rate HPD width: alphas, only plant dates vs. alphas only	0.00	(-0.13,0.13)	0.977	0.00	(-0.0001,0.0001)	0.972
rate HPD width: alphas, only vacuolar dates vs. alphas only	-0.20	(-0.3,-0.1)	2.288E-04 ***	0.00	(0,0.0002)	0.276
<i>(b) comparison to cross-calibration</i>						
rate HPD width: alphas, no plant dates vs. alphas & betas (cross-calibrated)	-0.66	(-0.75,-0.56)	2.253E-20 ***	0.00	(0.0003,0.0005)	6.216E-09 ***
rate HPD width: alphas, only chloroplast dates vs. alphas & betas (cross-calibrated)	-0.66	(-0.75,-0.56)	1.868E-20 ***	0.00	(0.0003,0.0005)	7.178E-09 ***
rate HPD width: alphas, only mitochondrial dates vs. alphas & betas (cross-calibrated)	-0.64	(-0.74,-0.54)	9.277E-19 ***	0.00	(0.0003,0.0005)	7.807E-09 ***
rate HPD width: alphas, only plant dates vs. alphas & betas (cross-calibrated)	-0.54	(-0.65,-0.44)	4.441E-14 ***	0.00	(0.0002,0.0004)	9.405E-08 ***
rate HPD width: alphas, only vacuolar dates vs. alphas & betas (cross-calibrated)	-0.64	(-0.73,-0.55)	7.153E-22 ***	0.00	(0.0003,0.0005)	7.469E-10 ***
<i>(c) comparison to cross-bracing</i>						
rate HPD width: alphas, no plant dates vs. alphas & betas (cross-braced)	-0.66	(-0.75,-0.58)	1.583E-22 ***	0.00	(0.0004,0.0006)	3.241E-13 ***
rate HPD width: alphas, only chloroplast dates vs. alphas & betas (cross-braced)	-0.67	(-0.76,-0.58)	1.940E-22 ***	0.00	(0.0004,0.0006)	6.011E-13 ***
rate HPD width: alphas, only mitochondrial dates vs. alphas & betas (cross-braced)	-0.65	(-0.74,-0.55)	4.481E-21 ***	0.00	(0.0004,0.0006)	4.036E-13 ***
rate HPD width: alphas, only plant dates vs. alphas & betas (cross-braced)	-0.57	(-0.68,-0.47)	7.726E-16 ***	0.00	(0.0003,0.0005)	4.027E-12 ***
rate HPD width: alphas, only vacuolar dates vs. alphas & betas (cross-braced)	-0.68	(-0.77,-0.6)	2.606E-23 ***	0.00	(0.0004,0.0006)	2.126E-14 ***
<b>rate coefficient of variation (CV: std. dev. / mean)</b>						
<i>(a) comparisons to alpha with all calibrations</i>						
rate CV: alphas, no plant dates vs. alphas only	-0.05	(-0.22,0.12)	0.561	0.01	(-0.07,0.09)	0.743
rate CV: alphas, only chloroplast dates vs. alphas only	0.06	(-0.17,0.29)	0.598	-0.01	(-0.11,0.09)	0.849
rate CV: alphas, only mitochondrial dates vs. alphas only	-0.01	(-0.21,0.19)	0.946	0.02	(-0.07,0.11)	0.710
rate CV: alphas, only plant dates vs. alphas only	-0.06	(-0.22,0.1)	0.458	0.04	(-0.03,0.11)	0.254
rate CV: alphas, only vacuolar dates vs. alphas only	0.01	(-0.19,0.21)	0.915	-0.01	(-0.11,0.08)	0.774
<i>(b) comparison to cross-calibration</i>						
rate CV: alphas, no plant dates vs. alphas & betas (cross-calibrated)	-0.26	(-0.47,-0.06)	0.016 *	0.06	(-0.04,0.16)	0.232
rate CV: alphas, only chloroplast dates vs. alphas & betas (cross-calibrated)	-0.07	(-0.29,0.15)	0.516	0.00	(-0.1,0.09)	0.930
rate CV: alphas, only mitochondrial dates vs. alphas & betas (cross-calibrated)	-0.09	(-0.29,0.1)	0.344	0.00	(-0.08,0.09)	0.940
rate CV: alphas, only plant dates vs. alphas & betas (cross-calibrated)	-0.22	(-0.41,-0.04)	0.023 *	0.06	(-0.02,0.14)	0.156
rate CV: alphas, only vacuolar dates vs. alphas & betas (cross-calibrated)	-0.09	(-0.28,0.11)	0.396	-0.02	(-0.11,0.07)	0.631
<i>(c) comparison to cross-bracing</i>						
rate CV: alphas, no plant dates vs. alphas & betas (cross-braced)	-0.42	(-0.59,-0.25)	1.053E-05 ***	0.12	(0.04,0.2)	6.351E-03 **
rate CV: alphas, only chloroplast dates vs. alphas & betas (cross-braced)	-0.26	(-0.44,-0.08)	6.080E-03 **	0.06	(-0.02,0.14)	0.146
rate CV: alphas, only mitochondrial dates vs. alphas & betas (cross-braced)	-0.30	(-0.47,-0.14)	6.007E-04 ***	0.08	(0,0.15)	0.042 *
rate CV: alphas, only plant dates vs. alphas & betas (cross-braced)	-0.33	(-0.47,-0.18)	2.962E-05 ***	0.09	(0.03,0.15)	6.652E-03 **
rate CV: alphas, only vacuolar dates vs. alphas & betas (cross-braced)	-0.30	(-0.47,-0.13)	9.526E-04 ***	0.06	(-0.02,0.14)	0.148

**Table S5.** List of taxa and UniProt accessions for sequences included in this study.

Species	Taxon Abbreviation	Uniprot Accession
<b>F-type ATPase (<math>\alpha</math> subunit)</b>		
<b>Plastid</b>		
<i>Arabidopsis thaliana</i>	CATPA_ARATH	P56757
<i>Oryza sativa</i> subsp. <i>Japonica</i>	CATPA_ORYSJ	P0C2Z6
<i>Amborella trichopoda</i>	CATPA_AMBTC	Q70XV0
<i>Cycas taitungensis</i>	CATPA_CYCTA	A6H5F1
<i>Psilotum nudum</i>	CATPA_PSINU	Q8WI30
<i>Selaginella uncinata</i>	CATPA_SELUN	Q2WGJ0
<i>Physcomitrella patens</i> subsp. <i>patens</i>	CATPA_PHYPA	Q6YXK3
<i>Chlamydomonas reinhardtii</i>	CATPA_CHLRE	P26526
<i>Micromonas</i> sp. strain RCC299	CATPA_MICSR	C1KR42
<i>Cyanophora paradoxa</i>	CATPA_CYAPA	P48080
<i>Cyanidioschyzon merolae</i>	CATPA_CYAME	Q85FQ8
<i>Thalassiosira pseudonana</i>	CATPA_THAPS	A0T0P4
<i>Ectocarpus siliculosus</i>	CATPA_ECTSI	D1J797
<i>Phaeodactylum tricorutum</i>	CATPA_PHATC	A0T0F1
<b>Mitochondrial</b>		
<i>Arabidopsis thaliana</i>	MATPA_ARATH	P92549
<i>Oryza sativa</i> subsp. <i>japonica</i>	MATPA_ORYSJ	P0C522
<i>Cycas taitungensis</i>	MATPA_CYCTA	B0BLD4
<i>Amborella trichopoda</i>	MATPA_AMBTC	Q9T718
<i>Isoetes engelmannii</i>	MATPA_ISOEN	C6FJG2
<i>Physcomitrella patens</i> subsp. <i>patens</i>	MATPA_PHYPA	Q1XGA4
<i>Chlamydomonas reinhardtii</i>	MATPA_CHLRE	Q96550
<i>Micromonas</i> sp. strain RCC299	MATPA_MICSR	C1EHC0
<i>Cyanidioschyzon merolae</i>	MATPA_CYAME	CMT434C*
<i>Homo sapiens</i>	MATPA_HUMAN	P25705
<i>Caenorhabditis elegans</i>	MATPA_CAEEL	Q9XXK1
<i>Drosophila melanogaster</i>	MATPA_DROME	P35381
<i>Anopheles gambiae</i>	MATPA_ANOGA	P35381
<i>Gallus gallus</i>	MATPA_CHICK	Q8UVX3
<i>Monosiga brevicollis</i>	MATPA_MONBE	A9V9Z0
<i>Saccharomyces cerevisiae</i>	MATPA_YEAST	A9V9Z0
<i>Cryptococcus neoformans</i> var. <i>neoformans</i>	MATPA_CRYNE	P07251
<i>Phaeodactylum tricorutum</i>	MATPA_PHATC	B7G531
<i>Thalassiosira pseudonana</i>	MATPA_THAPS	B8C6C6
<i>Ectocarpus siliculosus</i>	MATPA_ECTSI	D8LJM3
<b>Bacterial</b>		
<i>Synechocystis</i> sp. strain PCC 6803	ATPA_SYNCY	P27179
<i>Nostoc</i> sp. strain PCC 7120	ATPA_ANASP	P12405
<i>Prochlorococcus marinus</i> subsp. <i>pastoris</i>	ATPA_PROMP	Q7V037

strain CCMP1986		
<i>Synechococcus sp.</i> strain WH8102	ATPA_SYNPX	Q7U8W5
<i>Rickettsia prowazekii</i> strain, Madrid E	ATPA_RICPR	O50288
<i>Rickettsia typhi</i> strain ATCC VR-144	ATPA_RICTY	Q68VU6
<i>Caulobacter crescentus</i> strain ATCC 19089	ATPA_CAUCR	Q9A2V7
<i>Agrobacterium tumefaciens</i> strain C58	ATPA_AGRTU	Q8UC74
<i>Escherichia coli</i> strain K12	ATPA_ECOLI	P0ABB0
<i>Vibrio cholerae</i> strain ATCC 39315	ATPA_VIBCH	Q9KNH3
<i>Thermotoga maritima</i> strain ATCC 43589	ATPA_THEMA	Q9X1U7
<i>Chlorobium limicola</i> strain DSM 245	ATPA_CHLLI	B3EHU6
<i>Aquifex aeolicus</i> strain VF5	ATPA_AQUAE	O66907

### **V-type ATPase ( $\beta$ subunit)**

<i>Arabidopsis thaliana</i>	VATPB_ARATH	P11574
<i>Oryza sativa subsp. japonica</i>	VATPB_ORYSJ	Q9ASE0
<i>Picea sitchensis</i>	VATPB_PICSI	A9NVU9
<i>Selaginella moellendorffii</i>	VATPB_SELML	D8SQC5
<i>Physcomitrella patens subsp. patens</i>	VATPB_PHYPA	A9SP56
<i>Chlamydomonas reinhardtii</i>	VATPB_CHLRE	A8IA45
<i>Micromonas sp.</i> strain RCC299	VATPB_MICSR	C1DYK7
<i>Cyanidioschyzon merolae</i>	VATPB_CYAME	Q84KP2
<i>Homo sapiens</i>	VATPB_HUMAN	P15313
<i>Caenorhabditis elegans</i>	VATPB_CAEEL	Q19626
<i>Drosophila melanogaster</i>	VATPB_DROME	P31409
<i>Anopheles darling</i>	VATPB_ANODA	E3XA70
<i>Gallus gallus</i>	VATPB_CHICK	P49712
<i>Monosiga brevicollis</i>	VATPB_MONBE	A9V6U8
<i>Phaeodactylum tricornutum</i>	VATPB_PHATC	B7FQQ8
<i>Thalassiosira pseudonana</i>	VATPB_THAPS	B8C0L1
<i>Ectocarpus siliculosus</i>	VATPB_ECTSI	D8LCT6
<i>Archaeoglobus fulgidus</i> strain ATCC 49558	VATPB_ARCFU	O29100
<i>Thermococcus sibiricus</i> DSM 12597	VATPB_THESM	C6A5E7
<i>Cenarchaeum symbiosum</i> strain A	VATPB_CENSY	A0RXK0
<i>Sulfolobus tokodaii</i> strain DSM 16993	VATPB_SULTO	Q971B6
<i>Hyperthermus butylicus</i> strain DSM 5456	VATPB_HYPBU	A2BKX5
<i>Thermotoga neapolitana</i> strain ATCC 49049	VATPB_THENN	B9K814
<i>Deinococcus radiodurans</i> strain ATCC 13939	VATPB_DEIRA	Q9RWG7
<i>Clostridium tetani</i> strain Massachusetts / E88	VATPB_CLOTE	Q896K3
<i>Streptococcus parasanguinis</i> FW213	VATPB_STRPA	I1ZJ86
<i>Synergistetes bacterium</i> SGP1	VATPB_SYNGT	D4M879

### **F-type ATPase ( $\beta$ subunit)**

#### **Plastid**

<i>Arabidopsis thaliana</i>	CATPB_ARATH	P19366
<i>Oryza sativa subsp. japonica</i>	CATPB_ORYSJ	P12085
<i>Amborella trichopoda</i>	CATPB_AMBTC	Q70XZ6
<i>Cycas taitungensis</i>	CATPB_CYCTA	A6H514
<i>Picea sitchensis</i>	CATPB_PICSI	C1IXH0
<i>Keteleeria davidiana</i>	CATPB_KETDA	B7ZIP2

<i>Psilotum nudum</i>	CATPB_PSINU	O03081
<i>Selaginella uncinata</i>	CATPB_SELUN	Q2WGH4
<i>Physcomitrella patens subsp. patens</i>	CATPB_PHYPA	P80658
<i>Chlamydomonas reinhardtii</i>	CATPB_CHLRE	P06541
<i>Cyanophora paradoxa</i>	CATPB_CYAPA	P48081
<i>Cyanidioschyzon merolae</i>	CATPB_CYAME	Q85FT2
<i>Thalassiosira pseudonana</i>	CATPB_THAPS	A0T0R6
<i>Ectocarpus siliculosus</i>	CATPB_ECTSI	D1J7B4
<i>Phaeodactylum tricornutum</i>	CATPB_PHATC	A0T0D2

### **Mitochondrial**

<i>Arabidopsis thaliana</i>	MATPB_ARATH	P83483
<i>Oryza sativa subsp. japonica</i>	MATPB_ORYSJ	Q01859
<i>Picea sitchensis</i>	MATPB_PICSI	A9NUR7
<i>Selaginella moellendorffii</i>	MATPB_SELML	D8QQX5
<i>Physcomitrella patens subsp. patens</i>	MATPB_PHYPA	A9T281
<i>Chlamydomonas reinhardtii</i>	MATPB_CHLRE	P38482
<i>Micromonas sp. strain RCC299</i>	MATPB_MICSR	C1FE16
<i>Cyanidioschyzon merolae</i>	MATPB_CMH197C	CMH197C*
<i>Homo sapiens</i>	MATPB_HUMAN	P06576
<i>Caenorhabditis elegans</i>	MATPB_CAEEL	P46561
<i>Drosophila melanogaster</i>	MATPB_DROME	Q05825
<i>Anopheles darling</i>	MATPB_ANODA	E3XEC7
<i>Gallus gallus</i>	MATPB_CHICK	Q5ZLC5
<i>Monosiga brevicollis</i>	MATPB_MONBE	A9UYC5
<i>Saccharomyces cerevisiae</i>	MATPB_YEAST	P00830
<i>Cryptococcus neoformans var. neoformans</i>	MATPB_CRYNE	Q5KFU0
<i>Ectocarpus siliculosus</i>	MATPB_ECTSI	D7FXG1
<i>Phaeodactylum tricornutum</i>	MATPB_PHATC	B7FS46
<i>Thalassiosira pseudonana</i>	MATPB_THAPS	B5YP88

### **Bacterial**

<i>Synechocystis sp. strain PCC 6803</i>	ATPB_SYNCY	P26527
<i>Nostoc sp. strain PCC 7120</i>	ATPB_ANASP	P06540
<i>Prochlorococcus marinus subsp. pastoris strain CCMP1986</i>	ATPB_PROMP	Q7V049
<i>Synechococcus sp. strain WH8102</i>	ATPB_SYNPX	Q7U8U7
<i>Rickettsia prowazekii strain Madrid E</i>	ATPB_RICPR	O50290
<i>Rickettsia typhi strain ATCC VR-144</i>	ATPB_RICTY	Q68VU8
<i>Caulobacter crescentus strain ATCC 19089</i>	ATPB_CAUCR	Q9A2V9
<i>Agrobacterium tumefaciens strain C58</i>	ATPB_AGRTU	Q8UC76
<i>Escherichia coli strain K12</i>	ATPB_ECOLI	P0ABB4
<i>Vibrio cholerae strain ATCC 39315</i>	ATPB_VIBCH	Q9KNH5
<i>Thermotoga maritima strain ATCC 43589</i>	ATPB_THEMEA	O50550
<i>Chlorobium limicola strain DSM 245</i>	ATPB_CHLLI	B3EDQ7
<i>Aquifex aeolicus strain VF5</i>	ATPB_AQUAE	O67828

### **V-type ATPase ( $\alpha$ subunit)**

<i>Arabidopsis thaliana</i>	VATPA_ARATH	O23654
<i>Oryza sativa subsp. japonica</i>	VATPA_ORYSJ	Q651T8

<i>Picea sitchensis</i>	VATPA_PICSI	D5A887
<i>Selaginella moellendorffii</i>	VATPA_SELML	D8R3X7
<i>Physcomitrella patens subsp. patens</i>	VATPA_PHYPA	A9RGW5
<i>Chlamydomonas reinhardtii</i>	VATPA_CHLRE	A8I164
<i>Micromonas sp. strain RCC299</i>	VATPA_MICSR	C1E9Q8
<i>Cyanidioschyzon merolae</i>	VATPA_CYAME	Q84KP3
<i>Homo sapiens</i>	VATPA_HUMAN	P38606
<i>Caenorhabditis elegans</i>	VATPA_CAEEL	Q9XW92
<i>Drosophila melanogaster</i>	VATPA_DROME	P48602
<i>Anopheles gambiae</i>	VATPA_ANOGA	Q5TTG1
<i>Gallus gallus</i>	VATPA_CHICK	Q90647
<i>Monosiga brevicollis</i>	VATPA_MONBE	A9V438
<i>Phaeodactylum tricornerutum</i>	VATPA_PHATC	B7G162
<i>Thalassiosira pseudonana</i>	VATPA_THAPS	B8CBV3
<i>Ectocarpus siliculosus</i>	VATPA_ECTSI	D8LGA9
<i>Archaeoglobus fulgidus</i> strain ATCC 49558	VATPA_ARCFU	O29101
<i>Thermococcus sibiricus</i> DSM 12597	VATPA_THESM	C6A5E8
<i>Cenarchaeum symbiosum</i> strain A	VATPA_CENSY	A0RXX1
<i>Sulfolobus tokodaii</i> strain DSM 16993	VATPA_SULTO	Q971B7
<i>Hyperthermus butylicus</i> strain DSM 5456	VATPA_HYPBU	A2BKX6
<i>Thermotoga neapolitana</i> strain ATCC 49049	VATPA_THENN	B9K813
<i>Thermotoga neapolitana</i>	VATPA_THENE	Q8GB11
<i>Deinococcus radiodurans</i> strain ATCC 13939	VATPA_DEIRA	Q9RWG8
<i>Clostridium phytofermentans</i> strain ATCC 700394	VATPA_CLOPH	A9KQV0
<i>Streptococcus parasanguinis</i> strain ATCC 15912	VATPA_STRPA	F8DGA3
<i>Synergistetes bacterium</i> SGP1	VATPA_SYNGT	D4M878

### **Elongation factor Tu (Ef-Tu)**

#### **Plastid**

<i>Arabidopsis thaliana</i>	CEFTU_ARATH	P17745
<i>Oryza sativa subsp. japonica</i>	CEFTU_ORYSJ	Q6ZI53
<i>Picea sitchensis</i>	CEFTU_PICSI	C0PQG8
<i>Selaginella moellendorffii</i>	CEFTU_SELML	D8T8L9
<i>Physcomitrella patens subsp. patens</i>	CEFTU_PHYPA	A9T0S0
<i>Micromonas sp. strain RCC299</i>	CEFTU_MICSR	C1KR64
<i>Chlamydomonas reinhardtii</i>	CEFTU_CHLRE	P17746
<i>Cyanidioschyzon merolae</i>	CEFTU_CYAME	Q85FT7
<i>Cyanophora paradoxa</i>	CEFTU_CYAPA	P17245
<i>Ectocarpus siliculosus</i>	CEFTU_ECTSI	D1J725
<i>Thalassiosira pseudonana</i>	CEFTU_THAPS	A0T100
<i>Phaeodactylum tricornerutum</i>	CEFTU_PHATC	A0T0K6

#### **Mitochondrial**

<i>Arabidopsis thaliana</i>	MEFTU_ARATH	Q9ZT91
<i>Oryza sativa subsp. japonica</i>	MEFTU_ORYSJ	Q851Y8
<i>Selaginella moellendorffii</i>	MEFTU_SELML	D8S6G0
<i>Physcomitrella patens subsp. patens</i>	MEFTU_PHYPA	A9T9Z0

<i>Micromonas</i> sp. strain RCC299	MEFTU_MICSR	C1E231
<i>Chlamydomonas reinhardtii</i>	MEFTU_CHLRE	A8HXR2
<i>Ectocarpus siliculosus</i>	MEFTU_ECTSI	D8LDT2
<i>Thalassiosira pseudonana</i>	MEFTU_THAPS	B8CA96
<i>Phaeodactylum tricornutum</i>	MEFTU_PHATC	B7GA11

### **Bacterial**

<i>Escherichia coli</i> strain K12	EFTU_ECOLI	P0CE47
<i>Caulobacter crescentus</i> strain ATCC 19089	EFTU_CAUCR	Q99QM0
<i>Chlorobium limicola</i> strain DSM 245	EFTU_CHLLI	B3EH93
<i>Vibrio cholerae</i> strain ATCC 39315	EFTU_VIBCH	Q9KV37
<i>Rickettsia typhi</i> strain ATCC VR-144	EFTU_RICTY	Q8KT95
<i>Rickettsia prowazekii</i> strain Madrid E	EFTU_RICPR	P48865
<i>Agrobacterium tumefaciens</i> strain C58	EFTU_AGRTU	Q8UE16
<i>Synechococcus</i> sp. strain WH8102	EFTU_SYNPX	Q7U4D1
<i>Aquifex aeolicus</i> strain VF5	EFTU_AQUAE	O66429
<i>Thermotoga maritima</i> strain ATCC 43589	EFTU_THEMA	P13537
<i>Prochlorococcus marinus</i> subsp. <i>pastoris</i> strain CCMP1986	EFTU_PROMP	Q7UZY7
<i>Synechocystis</i> sp. strain PCC 6803	EFTU_SYNCY	P74227
<i>Nostoc</i> sp. strain PCC 7120	EFTU_ANASP	Q8YP63

### **Elongation factor 1 $\alpha$ (Ef-1 $\alpha$ )**

<i>Sulfolobus tokodaii</i> strain DSM 16993	EF1A_SULTO	Q976B1
<i>Hyperthermus butylicus</i> strain DSM 5456	EF1A_HYPBU	A2BN41
<i>Thermotoga neapolitana</i> strain ATCC 49049	EF1A_THENN	B9K884
<i>Cenarchaeum symbiosum</i> strain A	EF1A_CENSY	A0RUM4
<i>Clostridium tetani</i> strain Massachusetts / E88	EF1A_CLOTE	Q877L9
<i>Deinococcus radiodurans</i> strain ATCC 13939	EF1A_DEIRA	Q9R342
<i>Archaeoglobus fulgidus</i> strain ATCC 49558	EF1A_ARCFU	O29325
<i>Thermococcus sibiricus</i> strain MM 739	EF1A_THESM	C6A4R7
<i>Arabidopsis thaliana</i>	EF1A_ARATH	Q8GTY0
<i>Oryza sativa</i> subsp. <i>japonica</i>	EF1A_ORYSJ	O64937
<i>Picea sitchensis</i>	EF1A_PICSI	C0PSF0
<i>Selaginella moellendorffii</i>	EF1A_SELML	D8RAR5
<i>Physcomitrella patens</i> subsp. <i>patens</i>	EF1A_PHYPA	A9SJB4
<i>Cyanophora paradoxa</i>	EF1A_CYAPA	Q9ZSW2
<i>Cyanidioschyzon merolae</i>	EF1A_CYAME	Q84KQ1
<i>Phaeodactylum tricornutum</i>	EF1A_PHATC	B5Y4J2
<i>Caenorhabditis elegans</i>	EF1A_CAEEL	P53013
<i>Drosophila melanogaster</i>	EF1A_DROME	P08736
<i>Anopheles gambiae</i>	EF1A_ANOGA	Q7PT29
<i>Homo sapiens</i>	EF1A_HUMAN	P68104
<i>Gallus gallus</i>	EF1A_CHICK	Q90835
<i>Monosiga ovate</i>	EF1A_MONOV	Q2TTF7

\*Accessions correspond to Cyanidioschyzon merolae genome sequencing project, <http://merolae.biol.s.u-tokyo.ac.jp> (Matsuzaki et al, 2004 Nature)

**Table S6.** Divergence-time calibration points used in this study

<b>Taxon</b>	<b>Constraint in Mya (<math>\pm</math>std dev)</b>
Monocotyledoneae†	156( $\pm$ 14).
Angiospermae†	217( $\pm$ 40)
Gymnospermatophyta†	327( $\pm$ 30)
Tracheophyta†	432( $\pm$ 30).
Land Plants†	477( $\pm$ 70)
Human/Chicken§	300( $\pm$ 30)
Fly/Mosquito§	235( $\pm$ 24)

†Adapted from SA Smith et al (2009)

§Adapted from ML Berbee and JW Taylor (2010)