Early Photosynthetic Eukaryotes Inhabited Low-Salinity Habitats

Early photosynthetic eukaryotes inhabited low salinity habitats

Patricia Sánchez-Baracaldo[1], John Raven[2], Davide Pisani[1], Andrew Knoll[3]

[1]University of Bristol, [2]University of Dundee, UK, [3]Harvard University

Submitted to Proceedings of the National Academy of Sciences of the United States of America

The early evolutionary history of the chloroplast lineage remains an open question. It is widely accepted that the endosymbiosis which established the chloroplast lineage in eukaryotes can be traced back to a single event in which a cyanobacterium was incorporated into a protistan host. It is still unclear, however, which cyanobacteria are most closely related to the chloroplast, when the plastid lineage first evolved, and in what habitats this endosymbiotic event occurred. We present phylogenomic and molecular clock analyses, including data from cyanobacterial and chloroplast genomes using a Bayesian approach, with the aim of estimating the age for the primary endosymbiotic event, the ages of crown groups for photosynthetic eukaryotes and the independent incorporation of a cyanobacterial endosymbiont by *Paulinella*. Our analyses include both broad taxon-sampling (119 taxa) and eighteen fossil calibrations across all cyanobacteria and photosynthetic eukaryotes. Phylogenomic analyses support the hypothesis that the chloroplast lineage diverged from its closet relative, *Gloeomargarita*, a basal cyanobacterial lineage, ~2.1 billion years ago (Bya). Our analyses suggest that the Archaeplastida, consisting of glaucophytes, red algae, green algae and land plants, share a common ancestor that lived ~1.9 Bya. Whereas crown group Rhodophyta evolved in the Mesoproterozoic Era (1600-1000 million years ago, Mya), crown group Chlorophyta and Streptophyta began to radiate early in the Neoproterozoic (1000–542 Mya). Stochastic mapping analyses indicate that the first endosymbiotic event occurred in low salinity environments. Both red and green algae colonized marine environments early in their histories, with prasinophyte green phytoplankton diversifying 850-650 Mya.

Photosynthetic eukaryotes | chloroplast | cyanobacteria | phylogenomics | relaxed molecular clock

Introduction

Life as we know it would not be possible without oxygenic photosynthesis. Cyanobacteria were the only prokaryotes to evolve this metabolism, fundamentally changing redox chemistry early in Earth history (1, 2). Cyanobacteria also had a huge impact on the biological diversity of Earth's ecosystems, partly due to their ability to establish symbiotic relationships with a number of different hosts (3-6). Photosynthesis in eukaryotic organisms stems from two primary endosymbiotic events involving a cyanobacterium engulfed by a protistan host. The older of these events gave rise to the Archaeplastida, a monophyletic group that includes the Glaucocystophyta (glaucophytes), Rhodophyta (red algae) and Viridiplantae (green algae and land plants). In turn, secondary endosymbioses involving archaeplastid lineages (red or green algae) spread photosynthesis to the haptophytes, cryptophytes, euglenids, chlorarachniophyte rhizarians, dinoflagellates, chromerans, and stramenopiles. A second primary endosymbiotic event established photosynthesis within the rhizarian genus *Paulinella*. As primary producers, photosynthetic eukaryotes now dominate most terrestrial (e.g., embryophytes and green algae) and marine (e.g., diatoms, mixotrophic dinoflagellates and coccolithophores) environments. The timing of the first endosymbiotic event and ensuing divergence dates for the three major archaeplastidan lineages are still debated, with molecular clock estimates for the origin of plastids ranging over 800 million years (7). At the same time, the ecological setting in which this endosymbiotic event occurred has not been fully explored (8), partly due to phylogenetic uncertainties and preservational biases of the fossil record. Phylogenomics and trait evolution analysis have pointed to a freshwater origin for cyanobacteria (9-11), providing a novel approach to address the early diversification of terrestrial biota for which the fossil record is poor or uncertain.

The earliest widely accepted fossil evidence of photosynthetic eukaryotes is *Bangiomorpha*, a red alga deposited ~1.1 Bya (12). However, recent reports of multicellular photosynthetic eukaryotes at ~1.6 Bya provide evidence for an earlier establishment of photosynthesis within the eukaryotes (13). Currently, the oldest reliable evidence for eukaryotes as a whole is found in ~1.7 Bya rocks (14). These cyst-like microfossils occur in low diversity assemblages that potentially include stem group eukaryotes or stem representatives of extant major taxa (14,17). Sterane biomarkers originally viewed as evidence for 2.7 Ga eukaryotes have now been reinterpreted as younger contaminants (15, 16). Only around 750-800 Mya do fossils show a major increase in eukaryotic diversity that includes recognizable green algae (e.g., Cladophorales) (14, 17, 18), radiations possibly related to the evolution of eukaryovores – eukaryotes that eat other eukaryotes (19).

Reconstructing and dating the evolutionary history of early photosynthetic eukaryotes has proven challenging. Most phylogenetic studies place the divergence of the chloroplast lineage near the root of cyanobacteria (20-23), although a few studies insert chloroplasts higher in the tree (8) or nest them within derived clades (e.g., Nostocales (24)). Piecing together

Significance

While it is widely accepted that the chrloroplasts in photosynthetic eukaryotes can be traced back to a single cyanobacterial ancestor, the nature of that ancestor remains debated. Chloroplasts have been proposed to derive from either early- or late-branching cyanobacterial lineages, and similarly the timing and ecological setting of this event remains uncertain. Phylogenomic and Bayesian relaxed molecular clock analyses show that the chloroplast lineage branched deep within the cyanobacterial tree of life ~2.1 billion years ago (Bya), and ancestral trait reconstruction places this event in low salinity environments. The chloroplast took another 200 million years to become established, with most extant groups originating much later. Our analyses help to illuminate the little known evolutionary history of early life on land.

Reserved for Publication Footnotes

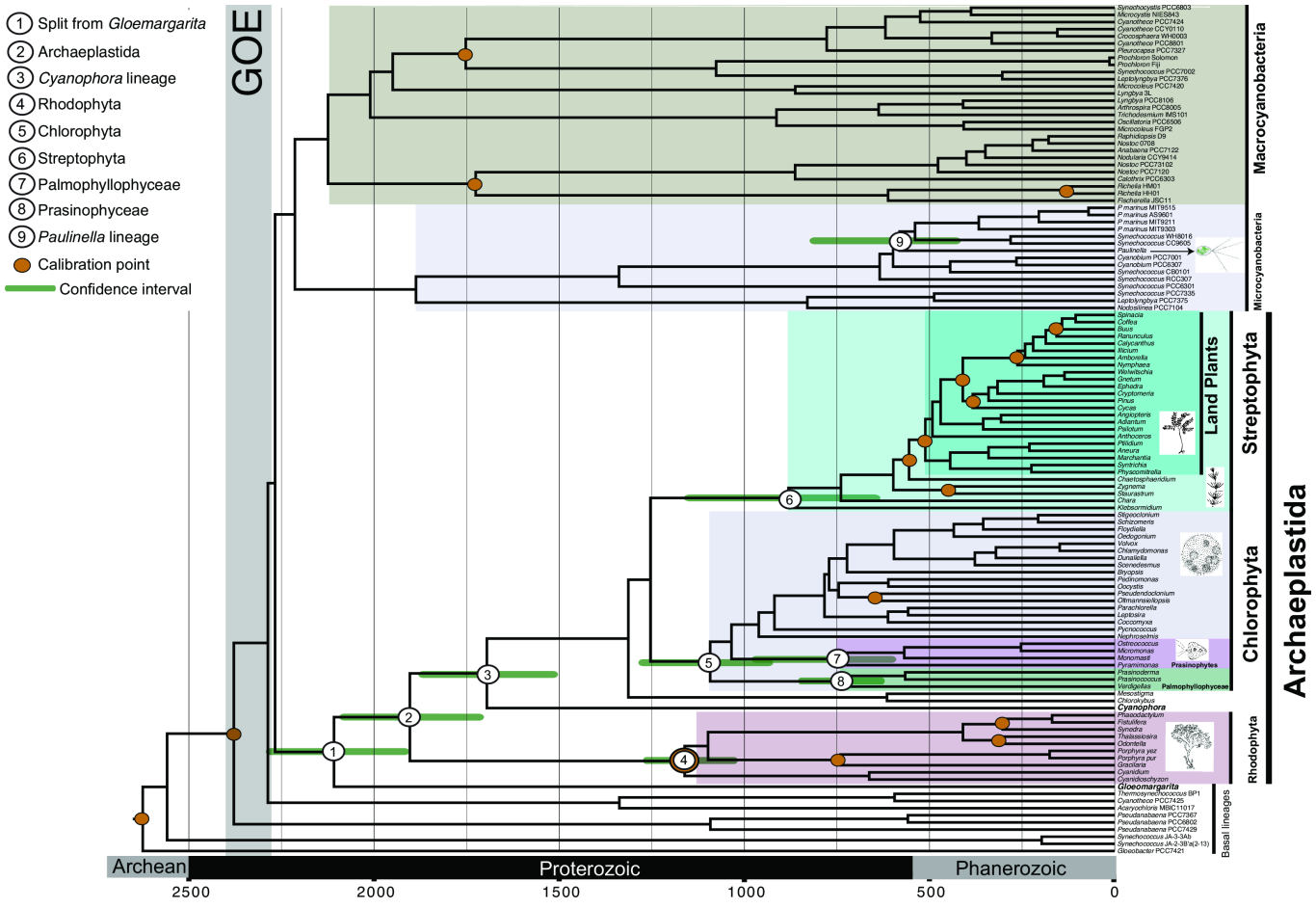

Fig. 1. The origin and diversification of photosynthetic eukaryotes and cyanobacteria as inferred from geologic time. The phylogenetic tree shown was estimated based on twenty-six genes from 117 taxa implementing Phylobayes 1.7a (96). Bayesian relaxed molecular clock analyses were carried out in Phylobayes 4.1 (39) implementing the UGAM (42) and the CAT-GTR substitution model (Table 2). Five calibration points for cyanobacteria and 13 calibrations points for photosynthetic eukaryotes (brown circles) were used (Table 1) for the tree shown and were treated as soft bounds. The root of the tree was set with a maximum age of 2.7 Bya (97) and a minimum age of 2.32 Bya (2). Age estimates for the numbered nodes (1–9) indicated are given in Table 1, which includes the corresponding values for the posterior 95% confidence intervals.

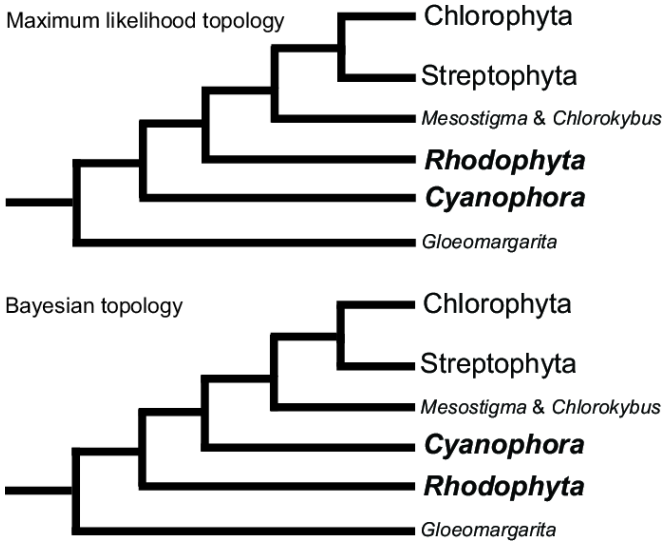

Fig. 2. Alternative hypotheses illustrating the deep-branching relationships within the Archaeplastida: Maximum Likelihood and Bayesian topologies.

the evolution of the chloroplast lineage is challenging because chloroplast genomes have undergone a dramatic reduction in size compared to their cyanobacterial relatives (25, 26). Here, we have implemented a phylogenomic approach to study the early evolutionary history of photosynthetic eukaryotes in the context of cyanobacterial evolution. Genomic data were used to carry out large-scale multi-gene analyses of cyanobacteria and photosynthetic eukaryotes. Molecular clock analyses provide new evidence indicating when the chloroplast lineage and *Paulinella* diverged from their closest cyanobacterial relatives. A Bayesian approach offers insights into the habitat in which the first endosymbiotic event took place during the Proterozoic Eon.

Results

Phylogenomic analyses

Two data sets were analyzed: a genomic dataset including 135 highly conserved proteins (9) compiled from a total of 49 cyanobacterial genomes, and a second dataset including twenty-six genes comprising 119 taxa that include both cyanobacteria and photosynthetic eukaryotes. The first dataset was analysed using Maximum Likelihood (ML) in a two-step process: (1) the 49 cyanobacterial genomes dataset was used to determine the deep-branching relationships of cyanobacteria, and (2) the topology generated in step one, here referred as "genome constraint" (See Supporting Information, including Fig. S1), was used as a backbone constraint for a second ML analysis that used twenty-six

Table 1. Calibration constraints for dating the photosynthetic eukaryotic and cyanobacteria tree of life.

Node/clade	Age constrains Myr		Eon	Reference
Cyanobacteria root	2320	(Min)	Archean	(2)
	2700	(Max)		(97)
	3000	(Max)		(98)
Simple filamentous; Oscillatoria-like filamentous fossils – Belcher supergroup	1900	(Min)	Proterozoic	(52, 105)
Nostocales	1600	(Min)	Proterozoic	(99)
	1900	(Max)		(105, 106)
Pleurocapsales	1700	(Min)	Proterozoic	(100)
	1900	(Max)		(105, 106)
Richelia-Hemiaulus symbiont	110	(Min)	Phanerozoic	(102)
Rhodophyta: Bangiophyceae	1050	(Min)	Proterozoic	(12, 108)
Chlorophyta: Ulvophyceae	635	(Min)	Phanerozoic	(109)
Rhodophyta: Floridiophyceae	600	(Min)	Phanerozoic	(110)
Streptophyta: Zygnemataceae	345	(Min)	Phanerozoic	(111)
Streptophyta: embryophytes or Land Plants	475	(Max)	Phanerozoic	(86)
	501			(17)
Vascular Plants	446	(Min)	Phanerozoic	(112)
Angiosperms-Gymnosperms	385	(Min)	Phanerozoic	(113)
Gymnosperms	385	(Min)	Phanerozoic	(114)
Angiosperms	130	(Min)	Phanerozoic	(115)
Diatoms-Coscinodiscophytina	190	(Min)	Phanerozoic	(107)
	250			
Diatoms-Bacillariophytina	110	(Min)	Phanerozoic	(107)
Dicots	125	(Min)	Phanerozoic	(115)

Table 2. Posterior age estimates in Myr using a Bayesian approach

Node	Clade	UGAM	CIR	Log normal
1	Gloeomargarita + Archaeplastida	2,108 (2,311– 1,907)	2,150 (2,321–1,997)	2,102 (2,273–1,938)
2	Archaeplastida	1,903 (2,117–1,694)	1,939 (2,091–1,808)	1,900 (2,059–1,745)
3	Cyanophora	1,695 (1,931–1,417)	1,781 (1,934–1,646)	1,739 (1,886–1,594)
4	Rhodophyta	1,161 (1,393–990)	1,060 (1,192–969)	1,062 (1,194–950)
5	Chlorophyta	1,092 (1,329–903)	1,086 (1,286–956)	1,037 (1,172 –893).
6	Streptophyta	880 (1,180–663)	999 (1,135 –862)	984 (1,119–845)
7	Palmophyllophyceae	735 (1,031–455)	885 (1,068 –746)	841 (1,004–662)
8	Prasinophytes	745 (1,022 –490)	887 (1,047–746)	841 (977–678)
9	Paulinella + Marine SynPro	634 (888–424)	431 (633–304)	350 (479–235)

Node ID corresponds to those shown in Fig. 1. Age estimates are given for analyses under UGAM, CIR and log normal clock models for the topology generated in Phylobayes. The CAT-GTR replacement model was implemented and the root was set with a maximum age of 2.7 Bya (97) and a minimum age of 2.32 Bya (2).

genes but included an extra set of 70 photosynthetic eukaryotes, for a total of 119 taxa (as previously implemented; see also (9, 10, 27)). The 119 taxa dataset was also analysed using a Bayesian approach without the backbone "genome constraint". Results from the two ML analyses and the Bayesian analysis returned results that are mostly congruent (See Supporting Information, including Fig. S2, S3). All analyses recovered well-supported monophyletic groups consistent with recent phylogenomic studies of cyanobacteria (9, 10, 20, 28, 29). Trees were rooted using *Gloeobacter*, as this taxon has been previously shown to be the deepest branch within cyanobacteria (9, 10, 29), even when including the group's recently discovered non-photosynthetic closest relatives: the Melainabacteria (30, 31), a group of heterotrophic soil and gut bacteria (32).

Previous studies have shown that the chloroplast lineage diverged amongst early branching cyanobacteria lineages (20, 22, 29, 30, 33-36). All analyses strongly support the monophyly of plastids in photosynthetic eukaryotes, with the exclusion of the independently acquired *Paulinella* plastid (See Supporting Information, including Fig. S2, S3). All our analyses also support *Gloeomargarita* as the most closely related cyanobacterium to the Archaeplastida (Fig. 1 and Fig. 2; See also Supporting Information, including Fig. S2, S3), consistent with recent phylogenomic studies (23). *Gloeomargarita* and the Archaeplastida diverged from early branching cyanobacteria prior to the diversification of the Micro- and Macrocyanobacteria (Fig. 1).

Our Maximum Likelihood analyses show *Cyanophora* (Glaucocystophyta) as the sister of all the other Archaeplastida (Fig. 1; see Supporting Information, including Fig. S2), with the Rhodophyta (red algae) and the Viridiplantae (green algae and land plants) as monophyletic sister groups. In contrast, our Bayesian analyses show *Cyanophora* diverging after the Rhodophyta and prior to the emergence of crown group Chlorophyta and Streptophyta (Fig. 1; See Supporting Information, including Fig. S3). Simplified topologies in Figure 2 illustrate the two alternative hypotheses with regard to deep-branching relationships within the Archaeplastida. In our analyses, long branches such as *Chlorokybus* and *Mesostigma* do not emerge as the monophyletic sister group of the Streptophyta (a division including several orders of non-marine green algae and embryophytes), congruent with recent plastid-based phylogenies using a large sampling of cyanobacteria taxa (20), but contrary to

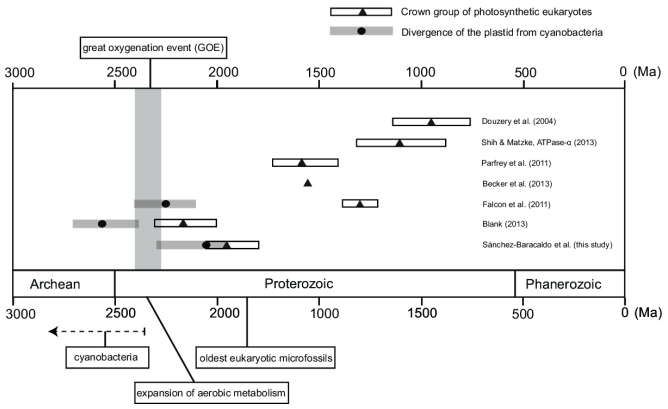

Fig. 3. Comparison of molecular clock estimates for the studies for the divergences of the crown group of photosynthetic eukaryotes and the split of the plastid lineages from their cyanobacteria ancestors. Age divergences for this study correspond to median ages in analyses shown in Figure 1. Circle and triangles are the mean ages for the origin of crown group of photosynthetic eukaryotes, and the common ancestor with cyanobacteria, and the rectangles represent the confidence intervals, respectively.

other previous studies in which these two lineages appear as basal and monophyletic within the Streptophyta (8, 37, 38). Furthermore, our Maximum Likelihood analyses including only eukaryotes indicate that out-group taxon selection does not influence the position of *Chlorokybus* and *Mesostigma* (See Supporting Information, including Fig. S4). By including only *Gloeomargarita,* the position of *Chlorokybus* and *Mesostigma* is also independent of taxon selection, but the placement of *Cyanophora* changes (See Supporting Information, including Fig. S5) – see also Figure 2.

Bayesian relaxed molecular clock and trait evolution analyses

Divergence times were estimated by applying a Bayesian approach in Phylobayes 4.1 (39) to the two alternative tree topologies generated under Maximum Likelihood (See Supporting Information, including Fig. S2) and Bayesian analyses (See Supporting Information, including Fig. S2) - see previous section and methods for details on how these trees were inferred. All analyses applied eighteen calibration points across cyanobacteria and photosynthetic eukaryotes (Table 1). Five calibrations were used for cyanobacteria and eleven for photosynthetic eukaryotes. Divergence times were estimated under three different relaxed molecular clock models: Log-normal (40), CIR (41), and uncorrelated gamma multipliers (UGAM) (42). Results were consistent across all models and trees (Table 2; see also Supporting Information, including Tables S3 and S4). Sensitivity experiments were performed to evaluate the effect of different root ages on our results (i.e., origin of oxygenic photosynthesis at 3 Bya; See Supporting Information, including Table S3). Overall, using an older maximum age root (3 Bya) pushes divergence times further back but does not significantly change our conclusions.

All our analyses indicate that the chloroplast lineage is the sister group of the recently discovered cyanobacterium *Gloeomargarita* (node 1; Fig. 1) from which it diverged ∼ 2.1 Bya, with age estimates, including 95% credibility intervals, ranging from 2.3–1.9 Bya. Similar molecular divergence times have been previously reported based on slow evolving SSU rRNA and *rbcL* (35) (Fig 3). The chloroplast lineage diverged from its cyanobacterial relatives at a time when they exhibited only small cell diameters (10, 11) and likely inhabited low salinity environments characteristic of fresh waters (11) (Fig. 4). The common ancestor of archaeplastids (node 2; Fig. 1) likely evolved during the Paleoproterozoic Era – with 95% credibility interval for this node ranging from 2.1–1.6 Bya; these results agree with other recent studies (8, 35, 43). Previous molecular clock analyses placed the first endosymbiotic event within the late Paleoproterozoic (43), Mesoproterozoic (37, 44), or even Neoproterozoic Era (45); these studies, however, either did not include cyanobacterial lineages or, if they did, relied on only a few cyanobacterial strains (See Fig. 3 for a comparison of previous molecular clock estimates). Bayesian stochastic character mapping analyses reveal that the ancestor of photosynthetic eukaryotes first evolved in low salinity environments (Fig. 4; this conclusion is robust to the phylogenetic placement of *Chlorokybus* and *Mesostigma*).

While the common ancestor the Archaeplastida (node 2) evolved in the Paleoproterozoic Era, crown groups such as the Rhodophyta (node 4), Chlorophyta (node 5), and Streptophyta (node 6) originated later, most likely between 1200 and 900 Mya (Table 2). The *Cyanophora* lineage diverged from other photosynthetic eukaryotes (node 3) more than 1500 Mya. The 95% credibility interval for the ancestors of the Rhodophyta (node 4) range between 1.3–0.9 Bya; these estimates overlap with recent studies including a large number of taxa, encompassing nuclear and plastid genes, from the Rhodophyta (43), and are younger than estimates from studies including nuclear genes estimating their origin during the early Mesoproterozoic (8). Well-supported marine lineages within the Rhodophyta can be paleontologically traced to marginal marine habitats at ∼1.1 Bya (Figs. 1 and 3; Table 2; see Supporting Information, including Table S3, S4). In this study, the age divergences of the Chlorophyta (node 5) and Streptophyta (node 6), excluding *Chlorokybus* and *Mesostigma*, are similar, with 95% credibility intervals ranging between 1.3–0.9 Bya and 1.1–0.67 Bya, respectively; these estimates are younger than previous studies placing the origin of these two major groups in the early Mesoproterozoic Era (8), but overlap with previous studies that propose Neoproterozoic origins (37, 44). Chlorophyte fossils in ca. 800 Mya rocks are consistent with these estimates (18).

Palmophyllophyceae (node 8) is the deepest branch within the Chlorophyta, which is consistent with recent studies; this lineage contains species found in a wide range of marine habitats, including planktonic unicells and loose colonies, as well benthic macroscopic thalli. The 95% credibility interval for the origin of the Palmophyllophyceae (node 8) ranges between 1.0–0.45 Bya. Furthermore, another early branching group of chlorophytes, labeled "prasinophytes" in Fig. 1, consists mainly of marine phytoplankton. Whilst the species included in this study are monophyletic (Fig. 1), phylogenomic studies including a broader taxonomic sampling of the Chlorophyta show the traditionally recognized Prasinophyceae to be paraphyletic (46). The clade included in our analyses is both ecologically important and paleontologically documented; again, our preferred divergence age of ca. 750 Mya for this group (node 8; 1.0–0.49 Bya, including the 95% credibility interval, is consistent with the fossil record (14). Our analyses indicate that relatives of the Palmophyllophyceae and the prasinophytes would have been some of the first Chlorophyta (green algae) to radiate into marine habitats in Neoproterozoic oceans (Fig. 4). The lineage now represented by the plastid of *Paulinella* diverged, as well, from marine *SynPro* clades during the Neoproterozoic Era.

Discussion

Deep branching relationships and the early evolution of photosynthetic eukaryotes

The lack of a terrestrial fossil record has contributed to our lack of understanding about life in early non-marine habitats. However, novel hypotheses about life in early terrestrial environments have emerged based on phylogenomic and trait evolution analyses. Recently sequenced cyanobacterial genomes have improved the resolution of phylogenomic analyses and the deep branching relationships within cyanobacteria (11, 20, 29, 47). Our phylogenomic analyses, including a wide range of plastids from

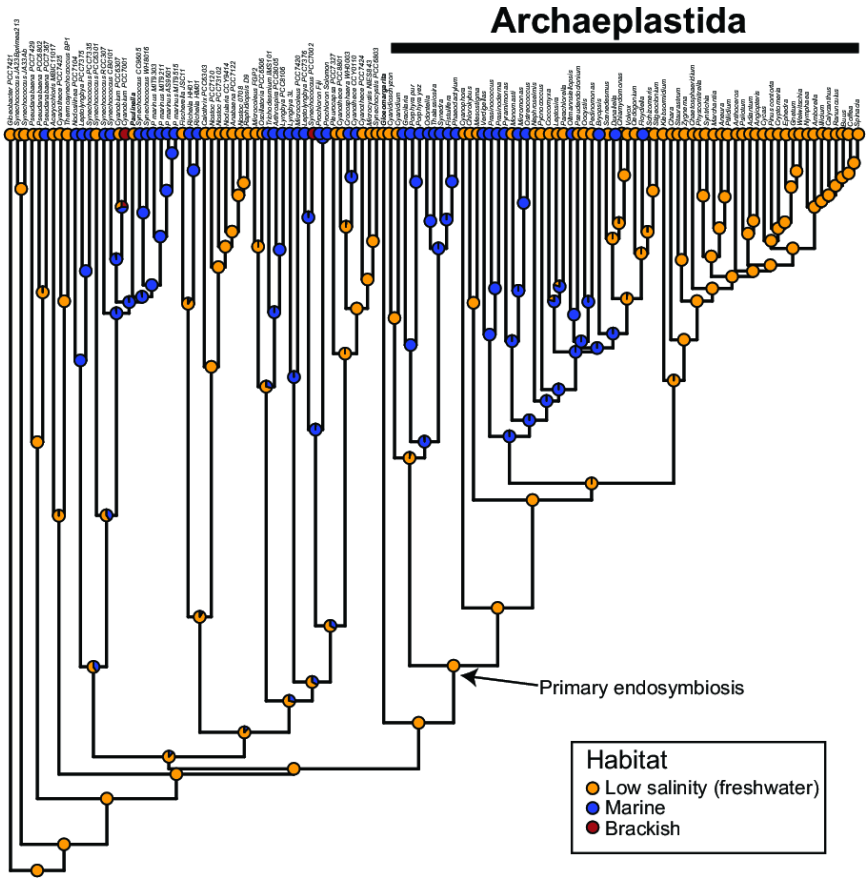

Fig. 4. Stochastic mapping analyses of the evolution of habitat. Phylogenetic tree including 119 taxa was estimated in Phylobayes 1.7a (96). A pie chart at each node indicates the Likelihoods for three character states as follows: blue circle = marine; yellow circle = freshwater, and red = brackish. Character states and full scientific names for abbreviated taxon names, shown here, are listed in Supplementary Table S2.

photosynthetic eukaryotes, are in agreement with some of the most recent large-scale multi-gene phylogenies (11, 20, 22, 29). All of our analyses agree with previous studies showing that the cyanobacteria lineage that led to the chloroplast diverged early in cyanobacterial history (Fig. 1) (20, 22, 30, 33, 48).

Molecular phylogenetics, trait evolution and relaxed molecular clock analyses have recently helped with the interpretation of the early fossil (9, 10, 47) and geochemical records (27). Broad taxonomic studies of prokaryotes imply a terrestrial/freshwater ancestry for cyanobacteria (49, 50) and their early diverging lineages (9, 11, 51). The plastid lineage, at its time of divergence, likely exhibited small cell diameters (1-2 μm), since the radiation leading to the Macrocyanobacteria (a group of cyanobacteria exhibiting cell diameter > 3 μm and observable in the fossil record by 1.9 Bya ((52) Fig. 1) had yet to occur (11). Our Bayesian stochastic mapping analyses provide strong support for the hypothesis that the common ancestor of cyanobacteria and the plastid lineage (node 1; Fig 1) inhabited low salinity environments (Fig, 3). Interestingly, most deeply diverging cyanobacterial lineages use sucrose to cope with changing environmental conditions (e.g., mild changes in osmotic pressure, desiccation tolerance) (53, 54). A similar mechanism was likely inherited by the plastid lineage at the time of the first endosymbiotic event (Fig. 1). Compatible solute genes (i.e., glucosylglycerol, glucosylglycerate and glycine-betaine) involved in regulating osmotic stress in high salinity environments (53, 54) evolved in marine lineages (11, 55) after the divergence of the chloroplast lineage.

All of our analyses strongly support the monophyly of plastids (except for that of *Paulinella*) in photosynthetic eukaryotes (See Supporting Information, including Fig. S2, S3), consistent with previous studies reporting a single endosymbiotic event in the ancestor of the Archaeplastida (34, 37, 44, 56-58). Our analyses provide two alternative hypotheses for deep branching relationships amongst the Archaeplastida (Fig. 2). Like most previous studies, our Maximum Likelihood analyses suggest that *Cyanophora* (Glaucocystophyta) is the sister of all other plastids (Fig. S2), followed by the divergence of the Rhodophyta and the Viridiplantae (green algae and land plants). In contrast, our Bayesian analyses show *Cyanophora* branching after the Rhodophyta (Fig. S3). It is worth highlighting that while most previous phylogenetic studies place *Cyanophora* as the deepest branch with the Archaeplastida (20, 23, 37, 46), other studies have postulated alternative phylogenetic placements for this lineage (8, 35, 36).

Our Maximum Likelihood hypothesis shown in Fig. 2 is consistent with ancestral traits (e.g., peptidoglycan) being retained by the Glaucophytes and later lost by all other algae and land plants (37). Furthermore, this topology supports an evolutionary scenario in which phycobilisomes, the light harvesting antennae anchored in the thylakoid, are an ancestral trait present in cyanobacteria that has been retained in the Glaucophytes and red algae, but lost in green algae and land plants. Given this hypothesis another likely ancestral feature is the turgor-resistant peptidoglycan wall round the 'cyanelle' in the Glaucocystophytes (59). This feature helps minimize energy cost for active water efflux involved in volume regulation by creating low osmolarity in the cytosol (60), and higher osmolarity in the cyanelle related to the higher concentrations of HCO_3^- required for functioning carboxysomes in an inorganic carbon concentrating mechanism (CCM). While the glaucocystophyte *Cyanophora paradoxa* has a CCM (61), its complete genome sequence shows no carboxysome-related genes (62). This is consistent with the hypothesis proposed by Raven (59) for the function of the peptidoglycan wall of the cyanelle, requiring the Rubisco-containing a central body of the cyanelle with high HCO_3^- concentrations for its CCM function

despite being a pyrenoid rather than a carboxysome. However, the Bayesian archaeoplastid phylogeny in Fig.2 shows glaucocystophytes branching after the rhodophytes, so if the implicit assumption is correct (57) that the glaucocystophyte CCM is ancestral within the archaeplastids, this trait must have been lost independently in ancestral red and green algae. A further problem with the ancestral CCM suggestion (57) is the absence of any molecular genetic markers for the glaucocystophyte 'pyrenoid' (63), of knowledge of the enzymes kinetics of their glaucocystophyte Form IB Rubisco (ribulose-1.5-bisphosphate carboxylase-oxygenase), or uncertainties CO_2 and O_2 availability at the time of chloroplast origin (64). In any event, our inferences about the environmental setting of archaeplastid origins are robust to the placement of *Cyanophora*.

Within the Viridiplantae, our analyses show that *Chlorokybus* and *Mesostigma* branched before the Streptophyta and Chlorophyta, as previously found by Shih et al. (20); however, other studies based on plastid genes show relatively low support for this node (8, 36). For our plastid data set, it seems that the position of *Chlorokybus* and *Mesostigma* is independent of out-group taxon selection (See Supporting Information, including Fig. S4). In contrast, nuclear gene-based phylogenies provide high statistical support for the inclusion of *Chlorokybus* and *Mesostigma* within the Streptophyta (38, 65). *Chlorokybus* and *Mesostigma* are long branched taxa and are therefore likely prone to long-branch attraction, and this might explain incongruence in their branching relationships. More genomic studies are needed to fully resolved the deep branching relationships of the charophyte green algae (66).

Timing of evolutionary events

It is worth emphasizing the distinction between the moment in which the plastid ancestor diverged from other lineages of cyanobacteria (node 1) and the last common ancestor of extant plastids (node 2). Our Bayesian relaxed molecular clock analyses suggest that the plastid lineage diverged from other cyanobacteria as early as ∼2.1 Bya (node 1, Fig. 1; Table 2); all analyses including clock models converge to similar age estimates. This date is broadly similar to previous studies, including values in the order of ∼2.4–2.7 Bya based on conserved encoded proteins (8) and ∼2.1 or 2.4 Bya based on SSU RNA data (35) (Fig. 2). In contrast, the median age for the primary endosymbiotic event that led to the origin of the archaeplastid chloroplast is in the order of ∼1.9 Bya (node 2; Figs. 1 and 2). Assuming that no plastid-bearing archaeplastid lineages have gone extinct, i.e. that the last common archaeplastid ancestor was also the first photosynthetic archaeplastid, our analyses suggest a time interval of ∼ 0.2 Byr between the split of the plastid lineage from *Gloeomargarita* (node 1) and the origin of the first photosynthetic eukaryote (node 2). Alternatively, photosynthesis may first have become established in now extinct stem group archaeplastids.

Molecular clock studies have generated a large range of age estimates, spanning over a billion years, for the origin of the Archaeplastida (Fig. 3). Our analyses, which include five cyanobacteria calibration points, suggest that photosynthetic eukaryotes originated relatively early, perhaps ∼1.9 Bya (with 95% credibility interval for this node ranging from 2.1–1.6 Bya). Molecular clock studies vary widely in terms of types of data (e.g., nuclear or plastid genes, rRNA, proteins) and taxa selection. Some estimates based on protein data suggest divergence at ∼0.9 Bya (56); others based on the α and β subunits of ATP synthase and the elongation factor thermo unstable (45, 67) propose ∼1.1 Bya; and by implementing the SSU rRNA and rbcL, still others predict ∼1.3 or 1.2 Bya (35) (Fig 2). Some of the oldest ages are in the order of ∼2.1 Bya, using a set of conserved proteins and RNA genes (8). It is difficult, if not impossible to compare molecular clock studies because discrepancies might be due to taxon sampling, types of substitution and clock models used, fossil calibrations, and accuracy in the implementation of molecular clock methodologies. However, extensive taxon sampling is a significant determinant of accurate phylogenetic estimation (68); here, by including *Gloeomargarita,* we have more accurately constrained the age of the Archaeplastida. In general, studies that include cyanobacterial relatives provide older ages for photosynthetic eukaryotes (Fig. 2) (8, 35).

Within photosynthetic eukaryotes, there is a deep divergence between Viridiplantae and the Rhodophyta (node 3), this split has been estimated to have occurred ∼1.9 Bya. Other estimates for this node have been on the order of ∼1.4 Bya based on DNA and 1.67 Bya for protein (58). There is a significant lag between the common ancestor of the Archaeplastida (node 2) and the origin of major clades of photosynthetic eukaryotes such as the Rhodophyta (node 4), the Chlorophyta (node 5), and the Streptophyta (node 6; excluding *Chlorokybus* and *Mesostigma*). Although basal lineages of extant red algae are non-marine, both fossils and our molecular clocks indicate that reds had colonized marine environments by ∼1.1 Bya (Fig. 1). Moreover previous studies have further tested that inclusion and exclusion of *Bangiomorpha* as a calibration point has little effect on age estimates for the rhodophytes (43, 44). Our molecular clock analyses and fossils further agree that chlorophyte green algae colonized the marine realm during the Neoproterozoic Era, with several lineages subsequently regaining terrestrial and freshwater aquatic environments. More specifically, fossils and molecular clock analyses indicate that prasinophytes emerged as ecologically important constituents of the marine phytoplankton late in the Neoproterozoic Era.

Broadly, then, our study suggests that while the plastid lineage diverged from other cyanobacteria during the Paleoproterozoic Era, it took nearly a billion years for major extant clades to diversify and gain ecological prominence. The younger divergence age of *Paulinella* ∼ 0.5 Bya from its closest relatives is consistent with suggestions that its significantly greater genome retention indicates a much earlier stage of endosymbiosis (69).

Early evolution of the eukaryote lineage

Recent years have seen significant improvements in our understanding of eukaryotic origins. It is now evident that eukaryotes do not constitute a third primary lineage of life, at least not as originally proposed by Woese and Fox (70). Instead, phylogenomic analyses minimally agree that the eukaryotic lineage evolved from within the Archaea (71-74). Further, it is uncontroversial that the establishment of the crown eukaryotes involved a stable endosymbiosis between an archaeon (related to the recently discovered Lokiarchaeota (75) – the cellular host), and an alpha proteobacterium (the ancestor of the mitochondrion (72-74, 76)). A topic of current debate in eukaryotic evolution is whether an amitochondriate proto-eukaryotic lineage ever existed. Bioenergetic arguments have been marshalled to disfavor this view (77), suggesting that only a mitochondriate cell can acquire the complexity observed in living eukaryotes, but these arguments are not universally accepted (78, 79). A recent genomic study (80) suggested that the mitochondrion might have entered the proto-eukaryotic lineage late, when the stem eukaryotic cell was already cytologically rather complex (80). However, these results have been shown to represent a methodological artifact (81). Irrespective of how eukaryote-like the stem eukaryotes were, all crown eukaryotes are mitochondriate and so their last common ancestor must also have contained mitochondria (82). It is not yet clear when the protomitochondrial endosymbiosis was established (76). However, as the free–living ancestor of the mitochondrion was an alpha-proteobacterium, at the very least, this event must have postdated the separation of the alpha-Proteobacteria from their sister lineage their sister lineage (the group including the beta- and gamma-Proteobacteria and possibly other less well known lineages like the Acidithiobacillia and

Zetaproteobacteria (72). Given the capacity of mitochrondria for aerobic respiration, this might well have happened after the GOE.

Our analyses help to constrain the origin of eukaryotes in so far as they suggest that the endosymbiotic event between a crown eukaryote and a cyanobacterium resulted in the origin of the Archaeplastida by ~1.9 Bya. This is consistent with currently available fossil evidence, and with the view that eukaryotes, minimally, must postdate the diversification of Archaea and the node separating the alpha-Proteobacteria from their sister lineage. How much older the base of the eukaryotes might be from the base of the Archaeplastida is unknown, and will in part depend on the topology of the eukaryotic tree, which is still not fully resolved (83).

The early land and marine record

Little is known about early life on land. Non-marine environments are generally sites of erosion rather than deposition, and so are relatively uncommon in all but the youngest sedimentary basins. Moreover, non-marine sedimentary rocks are not easily differentiated in the absence of environmentally diagnostic plant or animal fossils, and those facies most reliably recognized as non-marine on the basis of physical features (alluvial fans, braided stream deposits) tend to be poorly fossiliferous (84). Lacustrine and floodplain environments are known from Archean successions, and stromatolites and ^{12}C-depleted organic matter in these rocks documents early microbial ecosystems driven by photosynthesis (e.g., (85)). Significantly, shales associated with alluvial and fluviatile sandstones in late Mesoproterozoic successions from Scotland and mid-continent North America (86) contain abundant and modestly diverse microfossils reasonably interpreted as eukaryotic (87). The taxonomic affinities of these fossils are unclear, and photosynthetic microorganisms may or may not be present in the assemblage. And at least for the Scotland example, the non-marine origin of fossiliferous shales remains to be demonstrated. Nonetheless, it is reasonable to conclude that eukaryotic organisms had gained a firm foothold in non-marine environments by 1000 million years ago. Thus, the meager fossil record of non-marine environments is consistent with phylogenomic inferences, without adding independent evidence for non-marine algal origins. Paleoenvironmental studies may help more, suggesting as they do that eukaryotes in the terrestrial realm may have been shielded, at least in part, from the persistent subsurface anoxia (especially in shallow lakes and streams, but see (88)), episodic euxinia, and trace metal limitation (89) that potentially limited eukaryotic diversification in Mesoproterozoic oceans. Our analyses are consistent with paleontological data suggesting that while archaeplastids originated in non-marine environments, both red (node 6) and green (node 7) algae radiated in the oceans during the late Mesoproterozoic and early Neoproterozoic intervals, respectively (Fig. 1).

Concluding remarks

Our molecular clock analyses, calibrated by both Phanerozoic and Proterozoic fossils, and stochastic mapping analyses suggest that the common ancestor of the chloroplast lineage and its closest relative *Gloeomargarita* (an early branching lineage) can be traced back to the early Paleoproterozoic Era, ~ 2.1 Bya, in freshwater environments. While photosynthetic eukaryotes appear to have originated ~1.9 Bya, age estimates based on molecular clock studies show a lag between the earliest photosynthetic eukaryotes and the origin of modern crown groups such as the Rhodophyta, Chlorophyta, and Streptophyta. Median ages for the plastid associated with *Paulinella* suggest that this lineage diverged during the Neoproterozoic Era. Phylogenomic and molecular clock analyses are consistent with the modest fossil diversity of eukaryotic clades in earlier Proterozoic oceans and provide an additional line of evidence for interpreting the early evolution of photosynthetic eukaryotes.

Materials and Methods

Alignment and taxon sampling

All sequence data were obtained from GenBank (http://www.ncbi.nlm.nih.gov). Sequence data come from 49 cyanobacterial genomes, as previously described in Blank and Sánchez-Baracaldo (9) and Sánchez-Baracaldo (11). To establish the deep-branching relationships of cyanobacteria, our data set comprised of 135 proteins and two ribosomal RNAs (SSU and LSU); molecular markers included are evolutionarily conserved, had a minimum number of gene duplications and were present in all cyanobacterial taxa (10, 90). A second alignment with a broader taxonomic sampling photosynthetic eukaryotes (70 taxa) and cyanobacteria (49 taxa) were used to infer the evolutionary relationships of cyanobacteria and photosynthetic eukaryotes; this data set included twenty-six genes. Table S1 contains a list of genes shared across all taxonomic groups included in this study. Each gene was aligned independently using SATé 2.2.3 (91), a multiple sequence alignment and phylogenetic reconstruction program. Single gene alignments generated in SATé were imported into Mesquite v. 2.75 (92) to obtain 'nexus' and 'phylip' format files for subsequent analyses. Single alignments were later concatenated into a single nexus format file using Sequence Matrix v 100.0 (93). With the concatenated matrices generated here, we performed two sets of analyses describe below.

Phylogenetic analyses

Maximum Likelihood and Bayesian analyses were implemented to determine the phylogenetic relationships between cyanobacteria and photosynthetic eukaryotes. For the first set of analyses, we performed Maximum Likelihood analyses in two stages. First, we generated genome analyses from 49 cyanobacteria genomes to determine the deep-branching relationships of cyanobacteria in RAxML GUI v.1.1 (94). The topology generated at this stage was later implemented as genome constraint in RAxML GUI v.1.1 (94) for the second data set assembled including twenty-six genes and a total of 70 photosynthetic eukaryotes. A total of 135 protein-coding genes (with 56,251 aa) and two ribosomal RNAs (with a total of 4,555 bp) were used to establish phylogenetic relationships of cyanobacteria. ProTest v.2.4 (95) was used to estimate the best model of evolution for the protein set. To analyse the protein sequences we implemented the LG +G model (gamma-distribution with 4 rate categories).

A second data set was analyzed under the CAT-GTR+G model in Phylobayes MPI 1.7a (96). This data set consisted of twenty-six genes shared across 49 cyanobacteria and 70 photosynthetic eukaryotes. Convergence of Bayesian analyses was tested using the software *Tracecomp* (in Phylobayes) to estimate effective sample sizes and relative differences of key parameters, and *Bpcomp* to calculate mean distance between in trees from independent chains. To test whether taxon selection of the cyanobacteria out-groups had any effect on the branching relationships of the Archaeplastida, we performed Maximum Likelihood analyses including only eukaryotes and the closest known cyanobacteria (*Gloeomargarita*).

Fossil constraints

We used eighteen fossil calibrations across all cyanobacteria and photosynthetic eukaryotes (Table 1). Cyanobacteria arguably have the best fossil record of any bacterial group (47). In this study we applied cyanobacterial fossils that have been previously implemented (9, 11, 33). For the origin of cyanobacteria, two maximum ages were implemented: 2.7 Byr (97) and 3 Byr (98). The rise in atmospheric oxygen at 2.32 Byr was set as the minimum age for the cyanobacterial root (2). The first simple filamentous fossils of cyanobacteria, comparable to *Oscillatoria*, at 1.9 Byr (52); placement of this occurrence was implemented based on previous phylogenomic and trait evolution (i.e., filamentous vs unicellular) studies including *Pseudanabaena* genomes (11, 29). Furthermore by the fossil record recorded larger cyanobacterial microfossils described from 1.9 Bya cherts (52), which is consistent with other evolutionary studies of cyanobacteria showing that 1.9 Byr the macrocyanobacteria, sensu Sánchez-Baracaldo (11), had diverged by this time. Akinetes (99) and multiple fission (100) were be assigned to distinct groups such as the Nostocales and the Pleurocapsales, respectively. Larger phylogenomic studies of cyanobacteria including two *Pleurocapsa* genomes (PCC 7319 and PCC 7327) have established that whilst these two strains do not form a monophyletic group, they belong to a well-supported monophyletic group (11); this broader clade was calibrated here using preserved microfossils. Fossil calibrations across photosynthetic eukaryotes (Table 1) are relatively well characterised, adding an important further constraint to molecular clocks of cyanobacterial evolution.

Bayesian divergence time estimation

Ages were estimated for the topologies generated by the Maximum Likelihood and Bayesian analyses as described above using a subset of genes from the twenty-six genes as follows: AtpA, AtpB, PetB, PsaC, PsbA, PsbD, RbcL, S12. Divergence times were estimated implementing three different relaxed molecular clock models: Log-normal (40), CIR (41) and uncorrelated gamma multiplier model (39) in Phylobayes 4.1 (39). As in the Bayesian analyses substitutions were modelled using the CAT-GTR+G replacement model. For all non-calibrated nodes, we used a birth-death prior on divergence times, and soft bounded calibrations – allowing 0.05% of the prior density to fall outside the minimum-maximum interval of each calibration. Analyses were performed using two different root priors. The first was a Gamma distributed prior allowing the 95% of the prior distribution to fall between 2.32 – 2.7 Bya, the second had the 95% of the prior distribution falling between 2.32 – 3 Bya. The second analysis was performed as a sensitivity test

(see Supporting Information, including Table S4). Convergence of molecular clock analyses was tested using the software Tracecomp. All analyses were also run without data to visualize the marginal priors on node ages implied by our set of fossil calibrations.

Bayesian inference of character evolution

To infer the evolution of habitat, we used Bayesian stochastic character mapping (101). Character states were obtained from the Bergey's Manual of Systematic Bacteriology (102), AlgaeBase (www.algaebase.org), taxa description of genome taxa from the Joint Genome Institute (http://www.jgi.doe.gov), previous studies of trait evolution of cyanobacteria (11), and other cyanobacteria studies Habitat was coded using multistate discrete character states as follows 0 = freshwater (0-0.5 ppt parts per thousand), 1 = marine (from 30 up to 50 ppt parts per thousand), and 2 = brackish (0.5 – 30 ppt parts per thousand). Table S2 contains the characters states used in this study. Analyses were implemented in SIMMAP v1.5 (103) and used time calibrated trees generated in Phylobayes 4 (39) for both Maximum Likelihood and Bayesian analyses (Fig. 4). Sensitivity tests were performed treating *Parachlorella* and *Pedinomonas* as marine or freshwater within the chlorophyta (see Supporting Information, including Fig. S6, S7). Branch lengths were used a direct estimate of rate of evolution, rather than a prior on the rate parameter. For multi-state characters the bias parameter between states is specified as simply 1/κ, where k is the number of states. The overall rate of substitution for both of these classes is a branch length multiplier drawn from a prior gamma distribution. Analyses performed in SIMMAP v1.5 (103) were summarised in R using phylotools 0.1.2 (104).

Data availability

Data associated with this paper are available to download from Dryad htpp: XXXX. The uploaded data includes aligned protein sequences for phylogenetic and molecular clock estimates, calibration files, phylogenetic trees and Phylobayes scripts.

Acknowledgements

We thank Martin Brasier for in depth discussions on the early fossil record, John Huelsenbeck for discussion on the implementation of molecular clocks, and Sam Price for advice on stochastic mapping analyses. Phylogenetic analyses were performed at the High Performance Computer facility (BlueCrystal 3) at the University of Bristol. Funding support for this work came from a Royal Society Dorothy Hodgkin Fellowship for PS-B. The University of Dundee is a registered Scottish charity, No SC015096. AHK acknowledges support from the NASA Astrobiology Institute. **Additional information Supplementary information** accompanies this paper. **Competing interests:** We declare no competing financial interests.

1. Falkowski PG, Fenchel T, & Delong EF (2008) The microbial engines that drive Earth's biogeochemical cycles. *Science* 320(5879):1034-1039.
2. Bekker A, *et al.* (2004) Dating the rise of atmospheric oxygen. *Nature* 427(6970):117-120.
3. Raven JA & Allen JF (2003) Genomics and chloroplast evolution: what did cyanobacteria do for plants? *Genome Biol* 4(3):209.
4. Cornejo-Castillo FM, *et al.* (2016) Cyanobacterial symbionts diverged in the late Cretaceous towards lineage-specific nitrogen fixation factories in single-celled phytoplankton. *Nat Commun* 7:11071.
5. Donia MS. & al. e (2011) Complex microbiome underlying secondary and primary metabolism in the tunicate-*Prochloron* symbiosis. *Proc Natl Acad Sci USA* 108:E1423–E1432.
6. Foster RA, *et al.* (2011) Nitrogen fixation and transfer in open ocean diatom-cyanobacterial symbioses. *ISME J* 5(9):1484-1493.
7. Javaux EJ, and & A.H. Knoll (2017) Micropaleontology of the Lower Mesoproterozoic Roper Group, Australia. *J Paleontol* 91:199-229.
8. Blank CE (2013) Origin and Early Evolution of Photosynthetic Eukaryotes in Freshwater Environments: Reinterpreting Proterozoic Paleobiology and Biogeochemical Processes in Light of Trait Evolution. *J Phycol* 49(6):1040-1055.
9. Blank CE & Sánchez-Baracaldo P (2010) Timing of morphological and ecological innovations in the cyanobacteria--a key to understanding the rise in atmospheric oxygen. *Geobiology* 8(1):1-23.
10. Sánchez-Baracaldo P, Hayes PK, & Blank CE (2005) Morphological and habitat evolution in the Cyanobacteria using a compartmentalization approach. *Geobiology* 3:145–165.
11. Sánchez-Baracaldo P (2015) Origin of marine planktonic cyanobacteria. *Sci Rep-Uk* 5,(17418).
12. 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(3):386-404.
13. Bengtson S, Sallstedt T, Belivanova V, & Whitehouse M (2017) Three-dimensional preservation of cellular and subcellular structures suggests 1.6 billion-year-old crown-group red algae. *PLoS Biol* 15(3):e2000735.
14. Knoll AH (2014) Paleobiological perspectives on early eukaryotic evolution. *Cold Spring Harb Perspect Biol* 6(1).
15. Rasmussen B, Fletcher IR, Brocks JJ, & Kilburn MR (2008) Reassessing the first appearance of eukaryotes and cyanobacteria. *Nature* 455(7216):1101-U1109.
16. French KL, *et al.* (2015) Reappraisal of hydrocarbon biomarkers in Archean rocks. *Proc Natl Acad Sci U S A* 112(19):5915-5920.
17. Graham LE, *et al.* (2013) Resistance of Filamentous Chlorophycean, Ulvophycean, and Xanthophycean Algae to Acetolysis: Testing Proterozoic and Paleozoic Microfossil Attributions. *Int J Plant Sci* 174(6):947-957.
18. Butterfield NJ (1994) Burgess Shale-Type Fossils from a Lower Cambrian Shallow-Shelf Sequence in Northwestern Canada. *Nature* 369(6480):477-479.
19. Knoll AH & Lahr DJG (2016) Fossils, feeding and the evolution of complex multicellularity. *The Origins and Consequences of Multicellularity*, ed K. Niklas and S. Neumann e (MIT Press), pp 3–16.
20. Shih PM, *et al.* (2013) Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *P Natl Acad Sci USA* 110(3):1053-1058.
21. Reyes-Prieto A & Bhattacharya D (2007) Phylogeny of nuclear-encoded plastid-targeted proteins supports an early divergence of glaucophytes within plantae. *Mol Biol Evol* 24(11):2358-2361.
22. 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.
23. Ponce-Toledo RI, *et al.* (2017) An Early-Branching Freshwater Cyanobacterium at the Origin of Plastids. *Curr Biol* 27(3):386-391.
24. Ochoa de Alda JA, Esteban R, Diago ML, & Houmard J (2014) The plastid ancestor originated among one of the major cyanobacterial lineages. *Nat Commun* 5:4937.
25. Martin W, *et al.* (1998) Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393(6681):162-165.
26. Timmis JN, Ayliffe MA, Huang CY, & Martin W (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet* 5(2):123-135.
27. Sánchez-Baracaldo P, Ridgwell A, & Raven JA (2014) A Neoproterozoic transition in the marine nitrogen cycle. *Curr Biol* 24(6):652-657.
28. Latysheva N, Junker VL, Palmer WJ, Codd GA, & Barker D (2012) The evolution of nitrogen fixation in cyanobacteria. *Bioinformatics* 28:603–606
29. Schirrmeister BE, Gugger M, & Donoghue PCJ (2015) Cyanobacteria and the Great Oxidation Event: evidence from genes and fossils. *Palaeontology* 58:769–785.
30. Shih PM, Hemp J, Ward LM, Matzke NJ, & Fischer WW (2016) Crown group Oxyphotobacteria postdate the rise of oxygen. *Geobio* doi: 10.1111/gbi.12200.
31. Di Rienzi SC, *et al.* (2013) The human gut and groundwater harbor non-photosynthetic bacteria belonging to a new candidate phylum sibling to Cyanobacteria. *Elife* 2:e01102.
32. Soo RM, Hemp J, Parks DH, Fischer WW, & Hugenholtz P (2017) On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. *Science* 355(6332):1436-1440.
33. Schirrmeister BE, de Vos J. M., Antonelli A., & C. BH (2013) Evolution of multicellularity coincided with increased diversification of cyanobacteria and the Great Oxidation Event. *Proc Natl Acad Sci USA* 110:1791–1796.
34. Rodriguez-Ezpeleta N, *et al.* (2005) Monophyly of primary photosynthetic eukaryotes: Green plants, red algae, and glaucophytes. *Curr Bio* 15(14):1325-1330.
35. Falcon LI, Magallon S, & Castillo A (2010) Dating the cyanobacterial ancestor of the chloroplast. *ISME J* 4(6):777-783.
36. Marin B, Nowack ECM, & Melkonian M (2005) A plastid in the making: Evidence for a second primary endosymbiosis. *Protist* 156(4):425-432.
37. Becker B (2013) Snow ball earth and the split of Streptophyta and Chlorophyta. *Trends Plant Sci* 18(4):180-183.
38. Finet C, Timme RE, Delwiche CF, & Marleta F (2010) Multigene Phylogeny of the Green Lineage Reveals the Origin and Diversification of Land Plants. *Curr Bio* 20(24):2217-2222.
39. Lartillot N., Lepage T., & S. B (PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* 25:2286–2288.
40. Thorne JL, Kishino H, & Painter IS (1998) Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol* 15(12):1647-1657.
41. Lepage T, Bryant D, Philippe H, & Lartillot N (2007) A general comparison of relaxed molecular clock models. *Mol Biol Evol* 24(12):2669-2680.
42. Drummond AJ, Ho SYW, Phillips MJ, & Rambaut A (2006) Relaxed Phylogenetics and Dating with Confidence. *Plos Biol.* 4:e88.
43. Yang EC, *et al.* (2016) Divergence time estimates and the evolution of major lineages in the florideophyte red algae. *Sci Rep-Uk* 6.
44. Parfrey LW, Lahr DJ, Knoll AH, & Katz LA (2011) Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc Natl Acad Sci U S A* 108(33):13624-13629.
45. Shih PM & Matzke NJ (2013) Primary endosymbiosis events date to the later Proterozoic with cross-calibrated phylogenetic dating of duplicated ATPase proteins. *P Natl Acad Sci USA* 110(30):12355-12360.
46. Leliaert F, *et al.* (2016) Chloroplast phylogenomic analyses reveal the deepest-branching lineage of the Chlorophyta, Palmophyllophyceae class. nov. *Sci Rep-Uk* 6.
47. Schirrmeister BE, Sánchez-Baracaldo P, & Wacey D (2016) Cyanobacterial evolution during the Precambrian. *Int J Astrobiol*:1-18.
48. Turner S, Pryer KM, Miao VP, & Palmer JD (1999) Investigating deep phylogenetic relationships among cyanobacteria and plastids by small subunit rRNA sequence analysis. *J Eukaryot Microbiol* 46(4):327-338.
49. Battistuzzi FU, Feijao A, & Hedges SB (2004) A genomic timescale of prokaryote evolution: insights into the origin of methanogenesis, phototrophy, and the colonization of land. *Bmc Evol Biol* 4:44.
50. Sleep NH & Bird DK (2008) Evolutionary ecology during the rise of dioxygen in the Earth's atmosphere. *Philos T R Soc B* 363(1504):2651-2664.
51. Dagan T & al. e (2013) Genomes of Stigonematalean Cyanobacteria (Subsection V) and the Evolution of Oxygenic Photosynthesis from Prokaryotes to Plastids. *Genome Biol Evol* 5:31–44.
52. Hofmann HJ (1976) Precambrian microflora, Belcher Islands, Canada: significance and systematics. *J. Paleontol.* 50:1040-1073.
53. Hagemann M (2011) Molecular biology of cyanobacterial salt acclimation. *FEMS Microbiol Rev* 35(1):87-123.
54. Hagemann M (2013) *Genomics of Salt Acclimation: Synthesis of Compatible Solutes among Cyanobacteria*. 35(1):87-123.
55. Blank CE (2013) Phylogenetic Distribution of Compatible Solute Synthesis Genes Support

a Freshwater Origin for Cyanobacteria. *J Phycol* 49(5):880-895.
56. Delwiche CF, Kuhsel M, & Palmer JD (1995) Phylogenetic Analysis of Tufa Sequences Indicates a Cyanobacterial Origin of All Plastids. *Mol Biol Evol* 4(2):110-128.
57. Price DC, *et al.* (2012) Cyanophora paradoxa Genome Elucidates Origin of Photosynthesis in Algae and Plants. *Science* 335(6070):843-847.
58. 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.
59. Raven JA (2003) Carboxysomes and peptidoglycan walls of cyanelles: possible physiological functions. *Eur J Phycol* 38(1):47-53.
60. Raven JA & Doblin MA (2014) Active water transport in unicellular algae: where, why, and how. *J Exp Bot* 65(22):6279-6292.
61. Meyer M & Griffiths H (2013) Origins and diversity of eukaryotic CO2-concentrating mechanisms: lessons for the future. *J Exp Bot* 64(3):769-786.
62. Hagemann M, *et al.* (2016) Evolution of photorespiration from cyanobacteria to land plants, considering protein phylogenies and acquisition of carbon concentrating mechanisms. *J Exp Bot* 67(10):2963-2976.
63. Löffelhardt W (2010) Low CO2 stress: glaucocystophytes may found a unique solution. *Symbiosis and Strees: Joint Ventures in Biology* (Springer, Heidelberg), Vol 17, pp 83-94.
64. Raven JA, Beardall J, & Sánchez-Baracaldo P (2017) The possible evolution, and future, of CO2 concentrating mechanism. *J Exp Bot* in press.
65. Wodniok S, *et al.* (2011) Origin of land plants: Do conjugating green algae hold the key? *Bmc Evol Biol* 11.
66. Delwiche CF & Cooper ED (2015) The Evolutionary Origin of a Terrestrial Flora. Curr Biol 25(19):R899-910.
67. 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? *P Natl Acad Sci USA* 101(43):15386-15391.
68. Heath TA, Zwickl DJ, Kim J, & Hillis DM (2008) Taxon sampling affects inferences of macroevolutionary processes from phylogenetic trees. *Syst Biol* 57(1):160-166.
69. Reyes-Prieto A, *et al.* (2010) Differential gene retention in plastids of common recent origin. *Mol Biol Evol* 27(7):1530-1537.
70. Woese CR & Fox GE (1977) Phylogenetic structure of the prokaryotic domain: the primary kingdoms. *Proc Natl Acad Sci U S A* 74(11):5088-5090.
71. Pisani D, Cotton JA, & McInerney JO (2007) Supertrees disentangle the chimerical origin of eukaryotic genomes. *Mol Biol Evol* 24(8):1752-1760.
72. Hug LA, *et al.* (2016) A new view of the tree of life. *Nat Microbiol* 1:16048.
73. Williams TA, Foster PG, Cox CJ, & Embley TM (2013) An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* 504(7479):231-236.
74. Williams TA, Foster PG, Nye TM, Cox CJ, & Embley TM (2012) A congruent phylogenomic signal places eukaryotes within the Archaea. *Proc Biol Sci* 279(1749):4870-4879.
75. Spang A, *et al.* (2015) Complex archaea that bridge the gap between prokaryotes and eukaryotes. *Nature* 521(7551):173-179.
76. McInerney JO, O'Connell MJ, & Pisani D (2014) The hybrid nature of the Eukaryota and a consilient view of life on Earth. *Nat Rev Microbiol* 12(6):449-455.
77. Lane N & Martin W (2010) The energetics of genome complexity. *Nature* 467(7318):929-934.
78. Lynch M & Marinov GK (2015) The bioenergetic costs of a gene. *P Natl Acad Sci USA* 112(51):15690-15695.
79. Lynch M & Marinov GK (2017) Membranes, energetics, and evolution across the prokaryote-eukaryote divide. *Elife* 6.
80. Pittis AA & Gabaldon T (2016) Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. *Nature* 531(7592):101-104.
81. Martin WF, *et al.* (2016) Late mitochondrial origin is pure artefact. *bioRxiv*. https://doi.org/10.1101/055368.
82. Embley TM & Martin W (2006) Eukaryotic evolution, changes and challenges. *Nature* 440(7084):623-630.
83. Burki F (2014) The eukaryotic tree of life from a global phylogenomic perspective. *Cold Spring Harb Perspect Biol* 6(5):a016147.
84. Brasier MD (2013) Green algae (Chlorophyta) and the question of freshwater symbiogenesis in the early proterozoic. *J Phycol* 49:1036–1039.
85. Buick R (1992) The antiquity of oxygenic photosynthesis: evidence from stromatolites in sulphate-deficient Archaean lakes. *Science* 255(5040):74-77.
86. Strother PK & Wellman CH (2016) Palaeoecology of a billion-year-old non-marine cyanobacterium from the Torridon Group and Nonesuch Formation. *Palaeontology* 59(1):89-108.
87. Strother PK, Battison L, Brasier MD, & Wellman CH (2011) Earth's earliest non-marine eukaryotes. *Nature* 473(7348):505-509.
88. Cumming VM, Poulton SW, Rooney AD, & Selby D (2013) Anoxia in the terrestrial environment during the late Mesoproterozoic. *Geology* 41(5):583-586.
89. Parnell J, Spinks S, Andrews S, Thayalan W, & Bowden S (2015) High Molybdenum availability for evolution in a Mesoproterozoic lacustrine environment. *Nat Commun* 6:6996.
90. Shi T & Falkowski PG (2008) Genome evolution in cyanobacteria: the stable core and the variable shell. *Proc Natl Acad Sci U S A* 105(7):2510-2515.
91. Liu K, Raghavan S, Nelesen S, Linder CR, & Warnow T (2009) Rapid and accurate large-scale coestimation of sequence alignments and phylogenetic trees. *Science* 324(5934):1561-1564.
92. Maddison W & Maddison D (2009) *Mesquite: a modular system for evolutionary analysis*. Version 2.75. Available from http://mesquiteproject.org.
93. Vaidya G, Lohman DJ, & Meier R (2011) SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics* 27(2):171-180.
94. Silvestro D & Michalak I (2012) raxmlGUI: a graphical front-end for RAxML. *Org Divers Evol* 12(4):335-337.
95. Abascal F, Zardoya R, & Posada D (2005) ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21(9):2104-2105.
96. Lartillot N, Rodrigue N, Stubbs D, & Richer J (2013) PhyloBayes MPI: Phylogenetic Reconstruction with Infinite Mixtures of Profiles in a Parallel Environment. *Syst Biol* 62(4):611-615.
97. Brocks JJ, Buick R, Summons RE, & Logan GA (2003) A reconstruction of Archean biological diversity based on molecular fossils from the 2.78 to 2.45 billion-year-old Mount Bruce Supergroup, Hamersley Basin, Western Australia. *Geochim Cosmochim Acta* 67:4321–4335.
98. Crowe SA, *et al.* (2013) Atmospheric oxygenation three billion years ago. *Nature* 501(7468):535-+.
99. Golubic S, Sergeev VN, & Knoll AH (1995) Mesoproterozoic Archaeoellipsoides: Akinetes of heterocystous cyanobacteria. *Lethaia* 28(4):285-298.
100. Zhang Y, and Golubic, S. (1987) Endolithic microfossils (cyanophyta) from early Proterozoic stromatolites, Hebei, China. *Acta Micropaleontol Sin* 4:1–3.
101. Huelsenbeck JP, Nielsen R, & Bollback JP (2003) Stochastic mapping of morphological characters. *Syst Biol* 52(2):131-158.
102. Castenholz RW (2001) *Phylum BX. Cyanobacteria. In: Bergey's Manual® of Systematic Bacteriology, Springer New York* New York Ed pp 473–599.
103. Bollback JP (2006) SIMMAP: stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics* 7:88.
104. Revell LJ (2012) phytools: an R package for phylogenetic comparative biology (and other things). 3:217–223.
105. Golubic S & Seong-Joo L (1999) Early cyanobacterial fossil record: preservation, palaeoenvironments and identification. *Eur J Phycol* 34:339–348.
106. Sergeev VN, Gerasimenko LM, & Zavarzin GA (2002) Proterozoic history and present state of cyanobacteria. *Mikrobiologiia* 71:725–740.
107. Sims PA, Mann DG, & Medlin LK (2006) Evolution of the diatoms: insights from fossil, biological and molecular data. *Phycologia* 45:361–402.
108. Knoll AH, Wörndle S, & Kah L (2013) Covariance of microfossil assemblages and microbialite textures across a late Mesoproterozoic carbonate platform. *Palaios* 28:453-470.
109. Verbruggen H, *et al.* (2009) A multi-locus time-calibrated phylogeny of the siphonous green algae. *Mol Phylogenet Evol* 50(3):642-653.
110. Xiao S, Knoll AH, Yuan X, & Pueschel CM (2004) Phosphatized multicellular algae in the Neoproterozoic Doushantuo Formation, China, and the early evolution of florideophyte red algae. *Am J Bot* 91(2):214-227.
111. Mullins GL & Servais T (2008) The diversity of the Carboniferous phytoplankton. *Rev Palaeobot Palyno* 149(1-2):29-49.
112. Steemans P, *et al.* (2009) Origin and Radiation of the Earliest Vascular Land Plants. *Science* 324(5925):353-353.
113. Turnau E, Zavialova N, & Prejbisz A (2009) Wall ultrastructure in some dispersed megaspores and seed-megaspores from the Middle Devonian of northern Poland. *Rev Palaeobot Palyno* 156(1-2):14-33.
114. Gerrienne P, Meyer-Berthaud B, Fairon-Demaret M, Streel M, & Steemans P (2004) Runcaria, a middle Devonian seed plant precursor. *Science* 306(5697):856-858.
115. Friis EM, Pedersen KR, & Crane PR (2006) Cretaceous anglosperin flowers: Innovation and evolution in plant reproduction. *Palaeogeogr Palaeocl* 232(2-4):251-293.