A molecular timescale for the origin of red algal-derived plastids

Jürgen F. H. Strassert1,2,7, Iker Irisarri1,3,4,7, Tom A. Williams5, Fabien Burki1,6,*

1 Department of Organismal Biology, Program in Systematic Biology, Uppsala University, Norbyvägen 18D, 75 236 Uppsala, Sweden
2 Department of Biology, Chemistry, Pharmacy, Institute of Biology, Free University of Berlin, Königin-Luise-Straße 1–3, 14 195 Berlin, Germany
3 Department of Biodiversity and Evolutionary Biology, Museo Nacional de Ciencias Naturales (MNCN-CSIC), José Gutiérrez Abascal 2, 28 006 Madrid, Spain
4 Current address: Department of Applied Bioinformatics, Institute for Microbiology and Genetics, University of Göttingen, Goldschmidtstraße 1, 370 77 Göttingen, Germany
5 School of Biological Sciences, University of Bristol, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK
6 Science for Life Laboratory, Uppsala University, Uppsala, Sweden
7 These authors contributed equally

*Corresponding author:
Fabien Burki
Department of Organismal Biology, Program in Systematic Biology, Uppsala University,
Norbyvägen 18D, 752 36 Uppsala, Sweden
Phone: +46 (0)18 471 2779
E-mail: fabien.burki@ebc.uu.se

Jürgen F. H. Strassert: https://orcid.org/0000-0001-6786-7563
Iker Irisarri: https://orcid.org/0000-0002-3628-1137
Tom A. Williams: https://orcid.org/0000-0003-1072-0223
Fabien Burki: https://orcid.org/0000-0002-8248-8462
Abstract

In modern oceans, eukaryotic phytoplankton is dominated by lineages with red algal-derived plastids such as diatoms, dinoflagellates, and coccolithophores. These lineages and countless others representing a huge diversity of forms and lifestyles all belong to four algal groups: cryptophytes, ochrophytes, haptophytes, and myzozoans. Despite the ecological importance of these groups, we still lack a comprehensive understanding of their evolution and how they obtained their plastids. Over the last years, new hypotheses have emerged to explain the acquisition of red algal-derived plastids by serial endosymbiosis, but the chronology of these putative independent plastid acquisitions remains untested. Here, we have established a timeframe for the origin of red algal-derived plastids under scenarios of serial endosymbiosis, using a taxon- and gene-rich phylogenomic dataset combined to Bayesian molecular clock analyses. We find that the hypotheses of serial endosymbiosis are chronologically possible, as the stem lineages of all red plastid-containing groups overlapped in time. This period in the Mesozoic and Neoproterozoic Eras set the stage for the later expansion to dominance of red algal-derived primary production in the contemporary oceans, which has profoundly altered the global geochemical and ecological conditions of the Earth.

Introduction

Plastids (= chloroplasts) are organelles that allow eukaryotes to perform oxygenic photosynthesis. Oxygenic photosynthesis (hereafter simply photosynthesis) evolved in cyanobacteria around 2.4 billion years ago (bya), leading to the Great Oxidation Event—a rise of oxygen that profoundly transformed the Earth’s atmosphere and shallow ocean. Eukaryotes later acquired the capacity to photosynthesize with the establishment of plastids by endosymbiosis. Plastids originated from primary endosymbiosis between a cyanobacterium and a heterotrophic eukaryotic host, leading to primary plastids in the first photosynthetic eukaryotes. There are three main lineages with primary plastids: red algae, green algae (including land plants), and glaucophytes—altogether forming a large group known as Archaeplastida. Subsequently to the primary endosymbiosis, plastids spread to other eukaryote groups from green and red algae by eukaryote-to-eukaryote endosymbioses, i.e., the uptake of primary plastid-containing algae by eukaryotic hosts. These higher order endosymbioses resulted in complex plastids surrounded by additional membranes, some even retaining the endosymbiont nucleus (the nucleomorph), and led to the diversification of many photosynthetic lineages of global ecological importance, especially those with red algal-derived plastids (e.g., diatoms, dinoflagellates, and apicomplexan parasites).

The evolution of complex red algal-derived plastids has been difficult to decipher, mainly because the phylogeny of host lineages does not straightforwardly track the phylogeny of plastids. From the plastid perspective, phylogenetic and cell biological evidence strongly support a common origin of all complex red plastids. This is at the centre of the chromalveolate hypothesis, which proposed that the series of events needed to establish a plastid are better explained by a single secondary endosymbiosis in the common ancestor of alveolates, stramenoples, cryptophytes, and haptophytes: the four major groups known to harbour complex red plastids. From the host side, however, the phylogenetic relationships of these four groups have become increasingly difficult to reconcile with a single origin of all complex red algal-derived plastids in a common ancestor. Indeed, over a decade of phylogenomic investigations...
have consistently shown that all red plastid-containing lineages are most closely related to a series of plastid-lacking lineages, often representing several paraphyletic taxa, which would require extensive plastid losses under the chromalveolate hypothesis (at least ten). This situation is further complicated by the fact that no cases of complete plastid loss have been demonstrated, except in a few parasitic taxa. Several models compatible with the rhodoplex hypothesis have been proposed, differing in the specifics of the plastid donor and recipient lineages. However, these models of serial endosymbiosis remain highly speculative, in particular because we do not know if they are chronologically possible—did the plastid donor and recipient lineages co-exist? Addressing this important issue requires a reliable timeframe for eukaryote evolution, which has been challenging to obtain owing to a combination of complicating factors, notably: 1) uncertain phylogenetic relationships among the major eukaryote lineages; 2) the lack of genome-scale data for the few microbial groups with a robust fossil record; and 3) a generally poor understanding of methodological choices on the dates estimated for early eukaryote evolution.

Recent molecular clock analyses placed the origin of primary plastids in an ancestor of Archaeplastida in the Paleoproterozoic Era, between 2.1–1.6 bya. The origin of red algae has been estimated to be in the late Mesoproterozoic to early Neoproterozoic (1.3–0.9 bya), after a relatively long lag following the emergence of Archaeplastida. However, an earlier appearance in the late Paleoproterozoic has also been proposed based on molecular analyses. The earliest widely accepted fossil for the crown-group Archaeplastida is the multicellular filamentous red alga Bangiomorpha deposited ~1.2 bya. Recently, older filamentous fossils (Rafatazmia and Ramathallus) interpreted as crown-group red algae were recovered in the ~1.6 Ga old Vindhyan formation in central India. This taxonomic interpretation pushed back the oldest commonly accepted red algal fossil record by 400 million years, and consequently the origin of red algae and Archaeplastida well into the Paleoproterozoic Era. Evidence for the diversification of red algal plastid-containing lineages comes much later in the fossil record, which is apparent in the well-documented Phanerozoic continuous microfossil records starting from about 300 million years ago (mya). This period marks the diversification of some of the most ecologically important algae in modern oceans such as diatoms (ochrophytes), dinoflagellates (myzozoans), and coccolithophorids (haptophytes). If the fossil record is taken at face value, there is therefore a gap of over one billion years between the first appearance of crown-group eukaryotes interpreted as red algae in the early Mesoproterozoic and the later rise to ecological prominence of red algal-derived plastid-containing lineages.

In this study, we combined phylogenomics and molecular clock analyses to investigate the chronology of the origin and spread of complex red plastids among distantly-related eukaryote lineages in order to test the general rhodoplex hypothesis. We assembled a broad gene- and taxon-rich dataset (320 nuclear genes, 733 taxa), incorporating 33 well-established fossil calibrations, to estimate the timing of early eukaryote diversification. We explored the effect of a range of Bayesian molecular clock implementations, relaxed clock models and prior
calibration densities, as well as two alternative roots for the eukaryote tree. Our analyses show that the hypotheses of serial endosymbiosis are chronologically possible, as most red algal plastid acquisitions likely occurred in an overlapping timeframe during the Mesoproterozoic and Neoproterozoic Eras, setting the stage for the subsequent evolution of the most successful algae on Earth.

Results

The phylogeny of eukaryotes. Molecular clock analyses rely on robust tree topologies. To obtain our reference topology, we derived two sub-datasets from the full dataset of 320 protein-coding genes and 733 taxa (Supplementary Fig. 1) to allow computationally intensive analyses: a 136-OTU dataset and a 63-OTU dataset (see Material and Methods). The 136-OTU dataset was used in Maximum Likelihood (ML) inference using the best fitting site-heterogeneous LG+C60+G+F-PMF model (hereafter simply ML-c60) based on a concatenated alignment, as well as with a supertree method consistent with the Multi Species Coalescent model (MSC) in a version of the alignment where the taxa within OTUs were retained individually. The concatenated 63-OTU dataset was analysed by Bayesian inference using the site heterogeneous CAT+GTR+G (hereafter simply catgtr) and LG+C60+G (BI-c60) models, as well as in posterior predictive analyses (PPA) to compare the fit of both models.

The ML-c60 and MSC-derived trees are in good overall agreement with current knowledge of the broad eukaryote evolution. The ML-c60 tree recovered many existing supergroups with maximal bootstrap support (Supplementary Fig. 2), namely the TSAR assemblage, Haptista, Cryptista, Discoba, Amoebozoa and Obazoa. Archaeplastida was also recovered monophyletic, albeit with lower bootstrap support (86%), but this supergroup has previously lacked support in phylogenomic analyses (see Discussion). The relationships among these supergroups are also consistent with published work, most notably the recurrent affinity between Cryptista and Archaeplastida (CA clade), the branching of Haptista with Ancoracysta twista in ML analyses, and the placement of this group deep in the tree (here sister to the CA clade with 95% bootstrap support). The MSC analyses mostly recapitulate the same observations (Supplementary Fig. 3), although with the exceptions of TSAR and Archaeplastida due to the unresolved positions of telonemids as well as red algae and Cryptista, respectively. Taken together, the MSC analyses either supported the results of the concatenated ML analysis, or were inconclusive rather than conflicting.

The catgtr tree (Supplementary Fig. 4) received full posterior probabilities (PP) for all bifurcations and is globally consistent with the ML-c60 analyses (Supplementary Fig. 2), albeit with one important difference for understanding the evolution of plastids. In this tree, the positions of Haptista and A. twista were paraphyletic with respect to TSAR. A Bayesian reanalysis of the 63-OTU dataset under the BI-c60 model recapitulated the ML-c60 topology, suggesting that these topological differences derived from the chosen evolutionary model and not from the use of Bayesian or ML inference (Supplementary Fig. 5). Posterior predictive tests showed that the catgtr model better accounts for compositional heterogeneity than BI-c60 (although neither models did it fully; Supplementary Table 1) and the catgtr topology could not be rejected by ML (approximately unbiased test; p-AU = 0.182). Taken together, these results favoured the relationships described under the catgtr model. Therefore, the backbone topology...
for the divergence time estimation was inferred under a set of constraints defined by the best
"catgtr" topology but using the 136-OTU dataset, which contains a broader taxon sampling
allowing to more precisely place fossil calibrations.

**Timescale for eukaryote evolution.** Our molecular clock analyses revealed congruent dates
inferred under comparable analytical conditions (i.e., clock models, prior calibration densities,
and root positions; see below and Material and Methods) and by all tested implementations
(MCMCTree, PhyloBayes, BEAST). The fossils used as calibrations were chosen to span a wide
diversity of lineages (Table 1), and we followed a conservative approach when interpreting the
fossil record by choosing only fossils widely accepted by the paleontological community. We
also included one generally uncontested biomarker to constrain the emergence of extant Metazoa
in order to expand the otherwise sparse Proterozoic fossil record (24-isopropyl cholestane)\(^{29}\). Our
analyses placed the root of the eukaryote tree well into the Paleoproterozoic Era (Fig. 2). This Era
also saw the origin of primary plastids in the common ancestor of Archaeplastida, which likely
took place between 2,137 and 1,807 mya. Crown group red algae were inferred in the
Paleoproterozoic between 1,984 and 1,732 mya. These age ranges are provided as a conservative
approach encompassing all performed analyses (with the exception of those using t-cauchy
distributions with long tails, see below).

From red algae, plastids then spread to distantly-related groups of eukaryotes by at least
one secondary endosymbiosis. Plastid phylogenies have consistently shown that this red algal
donor lineage belonged to a stem lineage of Rhodophyta\(^{30}\), i.e., it lived after the split of
Cyanidiophyceae and the rest of red algae (Fig. 2). Based on Fig. 2, we inferred a time range for
this donor lineage to be between 1,675 and 1,281 mya. The origination period for the lineages
currently harbouring red algal-derived plastids were inferred as follows: cryptophytes between
1,658 and 440 mya; ochrophytes between 1,298 and 622 mya; haptophytes between 1,943 and
579 mya; myzozoans between 1,520 and 696 mya (dates refer to the 95% HPD intervals in Fig.
2). Thus, the stems of all extant lineages containing red algal plastids—along which these plastids
were acquired—overlap chronologically. This overlap was consistent across all performed
analyses and defines time windows during which endosymbiotic transfers, as proposed by the
serial endosymbiosis hypotheses, could reconcile plastid and nuclear phylogenies (Figs. 2 and 3;
Supplementary Figs. 6, 7, and Supplementary Data 1). Note that PhyloBayes analyses with the
autocorrelated clock model displayed the shortest confidence intervals, while the MCMCTree
analyses with the same clock model recovered generally younger ages for the origin of
ochrophytes and myzozoans (Fig. 3), but these differences did not alter the observation of time
overlap for the putative plastid acquisitions.

To better understand the effect of methodological choices on divergence times, we
performed a battery of sensitivity analyses with MCMCTree that tested different combinations of
clock models, prior calibration densities, and the position of the eukaryote root (32 sensitivity
analyses in total; Supplementary Data 1 and Material and Methods). These analyses indicated
that the prior calibration distributions had the strongest effect on the inferred divergence times,
followed by the clock model, and lastly the root position. Prior calibrations model the uncertainty
of fossil ages and their proximity to the cladogenetic events being calibrated, and thus different
distributions can be understood as more literal (skew-normal), loose (t-cauchy), or conservative
(uniform) interpretations of the fossil record\(^{31}\). As expected from the prior distributions, we
observed younger overall ages with skew-normal calibrations (mean of 905 Ma) compared to uniform (mean of 926 Ma) or t-cauchy distribution with short or long tails (means of 1,105 and 1,591 Ma, respectively; Supplementary Data 1). The excessively old ages and wide 95% HPD intervals (mean of 517 vs. 306 to 372 Ma) inferred with long-tailed t-cauchy distribution were considered biologically implausible, and thus disregarded in the following. The clock model had a modest impact on the posterior dates, with the autocorrelated clock model generally producing slightly younger ages and narrower intervals (mean ages 1,138 Ma; 95% HPD width 356 Ma) than the uncorrelated clock (mean age 1,125 Ma and 95% HPD width 416 Ma). The younger ages inferred under the autocorrelated clock model were more apparent when uniform calibrations were applied (mean ages of 904 vs. 947 Ma) and this trend was inverted in the case of skew-normal distributions (913 vs. 896 Ma). CorrTest indicated that branch lengths are most likely correlated (CorrScore = 0.99808, p < 0.001), suggesting that autocorrelated models might better model our dataset. Finally, the use of the two alternative roots, either on Amorphea or on Discoba, had the smallest effect on the posterior ages. Only marginal differences were observed on the overall node mean times of (1,134 vs. 1,132 Ma for the Amorphea and Discoba roots, respectively) and interval widths (381 vs. 388 Ma). The only exceptions were basal relationships within Discoba, which were noticeably older rooting the tree on this group (Fig. 2; Supplementary Fig. 6).

Further sensitivity analyses were performed with PhyloBayes to confirm the effects of the clock model choice (Supplementary Data 1). Additionally, we tested the effect of the substitution model by comparing LG+G with the catgtr mixture model. We observed a higher impact of the clock model than the evolutionary model. In this case, the uncorrelated clock model led to younger ages and wider HPD intervals than the autocorrelated clock (mean age of 942 vs. 1024 Ma, and 95% HPD width of 420 vs. 321 Ma, respectively). The catgtr model tended to infer wider intervals than the LG+G model (e.g., 95% HPD width of 356 vs. 321 Ma under the autocorrelated clock model). Finally, we tested a third commonly used Bayesian implementation (BEAST) using the LG+G substitution model, the Amorphea root, and uniform prior calibrations. The obtained mean posterior ages and 95% HPD intervals were in overall agreement with those obtained by MCMCTree and PhyloBayes (Supplementary Data 1) corroborating our inferred dates in these analyses.

Discussion

The eukaryote tree of life and the rhodoplex hypothesis. In the last 15 years, the tree of eukaryotes has been extensively rearranged based on phylogenomics. We assembled a dataset containing a dense eukaryotic-wide taxon sampling with 733 taxa and analysed subsamples of it with a variety of mixture models to resolve several important uncertainties that are key for understanding plastid evolution. Notably, we recovered the monophyly of Archaeplastida, which has previously been strongly supported by plastid evidence e.g.,11,32 but not by host (nuclear) phylogenetic markers e.g., 33–36. This topology is consistent with a recent exhaustive phylogenomic study of nuclear markers4 as well as with the long-held view of a single point of entry of photosynthesis in eukaryotes from cyanobacteria through the establishment of primary plastids (with the exception of the chromatophores in Paulinella). The supergroup Cryptista, which includes the red algal plastid-containing cryptophytes, was placed as sister to Archaeplastida. The
association of Cryptista and Archaeplastida has been suggested before, often even disrupting the
monophyly of Archaeplastida\textsuperscript{35,36}, but our analyses robustly recovered the sister relationship of
these two major eukaryote clades. Another contentious placement has concerned the supergroup
Haptista\textsuperscript{26,35,36,39}, which includes the red algal plastid-containing haptophytes, as well as its
possible relationship to the orphan lineage \emph{Ancoracysta twista}\textsuperscript{40}. The best-fitting \texttt{catgtr} model
favoured the position of Haptista as sister to TSAR and placed \emph{A. twista} deeper in the tree,
unrelated to Haptista. We observed that the model (\texttt{catgtr} vs \texttt{lge60}) rather than the method or
program is determinant in placing \emph{A. twista} relative to Haptista—\texttt{lge60} always placed these two
lineages together, both in ML and Bayesian, but \texttt{catgtr} never did—and that variation in the taxon-
sampling did not drastically modify this association.

Given this well-resolved tree of eukaryotes, we then mapped the position of the groups
with red algal-derived plastids (Fig. 2). The rhodoplex hypothesis explains the distribution of
these plastids by a series of endosymbioses, positing that plastids were separately acquired in the
stem lineages of cryptophytes, ochrophytes, haptophytes, and myzozoans\textsuperscript{15}. While the rhodoplex
hypothesis remains speculative, it provides the benefit of reconciling the accumulating
discrepancies between plastid and host phylogenies that existed under the previously prevalent
chromalveolate hypothesis\textsuperscript{12}. In the chromalveolate hypothesis, a red plastid was acquired by
secondary endosymbiosis in the common ancestor of all red plastid-bearing lineages, and
multiple subsequent losses were invoked to explain the patchy distribution of red plastids across
the eukaryote tree. Our phylogeny is consistent with other recent analyses in rendering this
scenario impossible\textsuperscript{11,19,26,35,36}: since Cryptista branch as sister to Archaeplastida, but the other red
plastid lineages branch elsewhere in the tree, the common ancestor of red plastid-bearing lineages
corresponds to one of the earliest nodes in the tree (Fig. 2), which was ancestral to the cell that
acquired primary plastids at the origin of Archaeplastida. As has been pointed out before\textsuperscript{35}, this
scenario would require a red alga to travel backwards in time to be engulfed by one of its distant
ancestors. In contrast, our analyses indicate that the inferred red algal donor lineage (stem
Rhodophytina) was contemporaneous with stem cryptophytes, haptophytes, and myzozoans, but
red plastids were unlikely to have been acquired directly by ochrophytes, which in all likelihood
had not yet diverged from other stramenopiles at the time (although their 95\% HPDs marginally
overlap). Moreover, all red algal plastid-containing lineages are most closely related to lineages
without plastids (Fig. 2), and for which conclusive evidence for a past photosynthetic history
does not exist. This situation would require many plastid losses if red plastids had been
established early and vertically transmitted, indicating that the chromalveolate hypothesis no
longer provides a compelling explanation for the distribution of red plastids in extant eukaryotes.

\textbf{Timing the early evolution of eukaryotes.} We present a detailed molecular clock analysis
providing a timeframe for eukaryote evolution. Our estimates inferred an age for the Last
Eukaryote Common Ancestor (LECA) between 2,386 and 1,958 mya, which is generally older
than in other molecular clock analyses based on phylogenomic data\textsuperscript{41–43}. However, an early
Paleoproterozoic origin of extant eukaryotes fits with the oldest definitive crown-eukaryote
fossils, the putative red algae \emph{Rafatazmia chitrakootensis} and \emph{Ramathallus lobatus} from 1,600
mya\textsuperscript{24}, which imply that eukaryotes must have originated before this time. It also fits with the
recent discoveries of multicellular eukaryotes in different groups of Proterozoic fossils indicating
that eukaryotes were already complex in deep times. For example, the chlorophyte fossil
Proterocladus antiquus in ~1,000 Ma old rocks, taken as evidence for a much earlier appearance of multicellularity in this group of green algae, is in line with our results (Fig. 2). Similarly, multicellular organic-walled microfossils with affinity to fungi were recently reported in the 1–0.9 Ga old Grassy Bay Formation, which pushes back the emergence of fungi by 500 Ma compared to previous studies. In our analyses, fungi were estimated to originate even before (1,759 to 1,078 Ma), consistently with the presence of multicellular organisms around 1 bya. More generally, an early Paleoproterozoic origin of eukaryotes would also be in line with records of aggregative multicellularity appearing more than 2 bya, although the eukaryote affiliation of these fossils is debated.

Our estimated dates placed the common ancestor of Archaeplastida also in the Paleoproterozoic Era, suggesting that eukaryotes acquired primary plastids, and thus photosynthesis, early on. This early origin of plastids is consistent with the putative photosynthetic crown-Archaebactista acritarch Tappania, which offers circumstantial evidence that eukaryotes exploited photosynthesis from the onset of eukaryote evolution. It is also in agreement with a recent molecular clock analysis of early photosynthetic eukaryotes that placed the origin of Archaebactista at ~1,900 mya. Furthermore, total group red algae are generally thought to contain some of the best minimum constraints for the existence of fully photosynthetic eukaryotes, most convincingly the late Mesoproterozoic Bangiomorpha resembling extant large bangiophytes, but also the earlier R. chitrakootensis and R. lobatus, probably also multicellular from 1,600 Ma. Thus, both the fossil record and molecular clock inferences support a Paleoproterozoic origin of primary plastids, and an early Mesoproterozoic origin of red algae.

For any endosymbiotic relationship to be established, the endosymbiont and the host must live at the same time and in the same place in order to interact. While the time intervals for plastid acquisition are sometimes relatively wide in our analyses (Fig. 2), they can be further constrained by overlaying the direction of plastid transfers proposed in the different models congruent with the rhodoplex hypothesis (Fig. 1). Of these models, only those by Stiller et al. and Bodył et al. are compatible with our tree (Fig. 2), because the other models assumed the monophyly of haptophytes with cryptophytes and the acquisition of plastids in an ancestor of that group. In both compatible models, the secondary engulfment of a red alga by a stem cryptophyte can be constrained by the minimum age of the plastid donor lineage, i.e., the age of the last Rhodophytina common ancestor (Fig. 2). This would indicate a rather early plastid acquisition by cryptophytes and a relatively long time span before the diversification of extant cryptophytes. In fact, if a red plastid-bearing cryptophyte was subsequently acquired by a stem ochronyte as both models proposed, this acquisition likely took place at latest 819 mya (95% HPD: 633–1,017), thus providing a maximum constraint for the age of crown cryptophytes. The model of Bodył et al. also proposed that cryptophytes were acquired by an ancestor of haptophytes in a parallel tertiary endosymbiosis, thus resulting in further constraints on the cryptophyte stem lineage. However, phylogenetic analyses of plastid-targeted proteins have instead supported an ochronyte origin for the haptophyte plastids. Following the same logic, the most likely time ranges of plastid acquisitions in all four red algal plastid-containing lineages and for the models and data compatible with our tree are presented in Table 2. Strikingly, this approach to chronologically constraining red plastid endosymbioses revealed that the serial transfers all took place in a relatively short time window between 650 and 1,079 million years from the initial secondary endosymbiosis in a stem cryptophyte to the establishment of plastids in
the ancestors of all modern-day red algal plastid-containing lineages (narrowest and widest 95% HPD widths for the totality of the four endosymbioses; Table 2).

**The origin and rise of algae.** The fossil and biomarker records document the increasing abundance and rise to ecological prominence of primary endosymbiotic algae in the oceans, including red algae, ~659–645 mya. Red plastid-bearing algae, notably diatoms, coccolithophores and dinoflagellates, started to expand later after the Permo-Triassic transition ~250 mya and have since remained major primary producers in the oceans. By contrast, our molecular clock analyses indicate that the events underlying the evolutionary origin of these algae, including the primary plastid endosymbiosis, the origin of red algae, and the subsequent spread of red plastids across the eukaryote tree, all pre-dated the ecological expansion of these groups by at least ~0.5–1 Ga. This long time lag suggests that the selective pressures underpinning the establishment of primary and complex plastids in the Palaeo- and Mesoproterozoic Eras are distinct from those that drove their expansion ~1 Ga later.

The Proterozoic ocean was poor in essential inorganic nutrients, such as phosphate and nitrogen, which may have limited the expansion of larger eukaryotic algae into the open marine realms. Yet, these oligotrophic environments may have been favourable to the origin and early evolution of plastid-containing lineages. Predatory behaviours have been demonstrated in green algae and non-photosynthetic direct relatives of red algae, suggesting that mixotrophy was a key intermediate stage in early evolution of Archaeplastida. More generally, mixotrophy is increasingly recognised as the default lifestyle for many, perhaps most single-cell algae, and is clearly advantageous over both autotrophy or heterotrophy in communities limited by nutrients. Thus, oligotrophic Proterozoic waters could have selected for the increased fitness that mixotrophy provided, but these nutrient-poor environments could not sustain a radiation of eukaryotic algae. The rise to ecological dominance of algae with red plastids during the Mesozoic might have been driven by profound environmental changes such as the increase in coastlines associated with the breakup of the supercontinent Pangea, providing newly flooded continental margins with high-nutrient habitats. Importantly, these environmental changes would have happened after ~1 Ga of evolution since the origin of red algal-derived plastids, providing ample opportunities for the evolution of a genetic toolkit that would prove beneficial with the rise of more favourable habitats. One such example of beneficial genetic innovation is cell protection by a variety of armour plating, which is a convergent feature of many ecologically successful algae with complex red plastids. These armours aid the phytoplankton to protect from grazing and thus represent an additional condition that may have favoured the late Mesozoic expansion to ecological dominance of some groups with red algal-derived plastids.

**Conclusion**

Algae powered by red algal-derived plastids are among the most evolutionary and ecologically successful eukaryotes on Earth. Yet, we still lack a comprehensive understanding of how, and how many times, red plastids were established. In recent years, hypotheses of serial endosymbiosis have flourished to explain how disparate groups of eukaryotes obtained their red plastids. In the present study, we used molecular clocks applied to a broad phylogenomic dataset to test whether the serial endosymbiosis hypotheses are chronologically possible. Our results
indicate that all putative plastid donor and recipient lineages most likely overlapped during Earth history, thus in principle allowing plastids to be passed between distantly related hosts. Furthermore, we showed that the timeframe from the initial secondary endosymbiosis with a red alga to the establishment of all complex red plastids was relatively short, likely spanning between 650 to 1,079 million years mainly during the Mesoproterozoic Era. This relatively short timeframe represents a novel insight into the diversification of photosynthetic eukaryotes during the Mesoproterozoic and the origin of the most ecologically important modern-day algae. More generally, specific serial endosymbiosis hypotheses, if validated, will provide useful relative constraints for better understanding the overall timescale of eukaryote diversification in future paleobiological studies.

Methods

Phylogenomic dataset construction. Two publicly available phylogenomic datasets were used as starting points in this study: a 263 genes, 234 taxa dataset\textsuperscript{26,35}, and a 351 genes, 64 taxa dataset\textsuperscript{58}. Non-overlapping genes between these two datasets (134 genes) were identified by BLAST\textsuperscript{59} analyses, allowing the merging of both datasets to bring the total number of initial genes to 397. We expanded the sampling of species with publicly available genomic/transcriptomic data to obtain a comprehensive eukaryote-wide dataset, with particular attention to taxa most relevant to this study (sources: ensemblgenomes.org, imicrobe.us/#/projects/104, ncbi.nlm.nih.gov, onekp.com, and few publications that did not provide a link to a public sequence database; now all available in EukProt\textsuperscript{60}). The procedure to add taxa was as follows: 1) For each taxon, protein sequences were clustered with CD-HIT\textsuperscript{61} using an identity threshold of 85%; 2) Homologous sequences were retrieved by BLASTP searches using all 397 genes as queries (e-value: 1e-20; coverage cutoff: 0.5); 3) In three rounds, gene trees were constructed and carefully inspected in order to detect and remove putative paralogs and contaminants. For that, sequences were aligned with MAFFT v. 7.310\textsuperscript{62} using either the -auto option (first round) or MAFFT L-INS-i with default settings (second and third round). Ambiguously aligned positions were filtered using trimAL v. 1.4\textsubscript{63} with a gap threshold of 0.8 (all three rounds), followed by Maximum Likelihood (ML) single-gene tree reconstruction with either FastTree v. 2.1.10\textsuperscript{64} using -lg -gamma plus options for more accurate performances (first round) or RAxML v. 8.2.10\textsuberscript{65} with PROTGAMMALGF and 100 rapid bootstrap searches (second and third round). To facilitate the detection of contaminants and paralogs, all taxa were renamed following NCBI’s taxonomy (manually refined—the custom taxonomy is available in Supplementary Fig. 1; Supplementary Data 2), and colour-coded using in-house scripts. Multiple copies from the same taxon were assigned to a unique colour allowing to more easily detect contaminants and paralogs in FigTree v. 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/). After the three rounds of gene tree inspection, we discarded 77 genes (74 out of the 134 genes added from Brown et al.\textsuperscript{58}) due to suspicious clustering of major groups (e.g., duplication of the entire Sar clade in FTSJ1). The resulting dataset comprised 320 genes and 733 eukaryote taxa with ≥5% data; Supplementary Data 2.

For each curated gene, sequence stretches without clear homology (e.g., poor quality stretches of amino acids, or leftover untranslated regions) were removed with PREQUAL v. 1.01\textsubscript{66} employing a posterior probability threshold of 0.95 (ignoring some fast-evolving taxa).
Sequences were aligned using MAFFT G-INS-i with a variable scoring matrix to avoid over-alignment (-unaliglevel 0.6) and trimmed with BMGE v. 1.12 using -g 0.2, -b 5, -m BLOSUM75 parameters. Partial sequences belonging to the same taxon that did not show evidence for paralogy or contamination on the gene trees were merged. All 320 trimmed gene alignments were concatenated with SCaFos v. 1.25 into a supermatrix of 733 taxa and 62,723 aligned amino acid positions (62,552 distinct patterns; gaps and undetermined characters: ~35%; Supplementary Data 2). This dataset was subjected to ML analysis in IQ-TREE69 with the site-homogeneous model LG+G+F and ultrafast bootstrap approximation (UFBoot70; 1,000 replicates) employing the -bb and -bnni flags (IQ-TREE versions 1.6.3 to 1.6.9 have been used in this study). This large tree (Supplementary Fig. 1) was used to select a reduced taxon-sampling strategy that maintained phylogenetic diversity but allowed downstream analyses with more sophisticated models. The taxa selection aimed to 1) retain all major lineages of eukaryotes; 2) preferentially keep slowly-evolving (shorter branches) representatives; 3) preferentially discard taxa with more missing data; and 4) allow the precise taxonomic placement of fossil calibrations. In order to increase sequence coverage, some monophyletic strains or species/genera complexes were combined to form a chimeric operational taxonomic unit (OTU; Supplementary Data 2). This strategy led to a dataset containing 136 OTUs, for which the raw sequences were again filtered, aligned, trimmed, and concatenated as described above, forming a supermatrix with 73,460 aligned amino acid positions (71,540 distinct patterns; gaps and undetermined characters: ~17%; Supplementary Data 2).

**Phylogenomic analyses.** The reduced dataset (136 OTUs) was subjected to ML analysis in IQ-TREE using the best-fit site-heterogeneous model LG+C60+G+F with the PMSF approach to calculate non-parametric bootstrap support (100 replicates; Supplementary Fig. 2). This dataset was also analysed under the Multi Species Coalescent (MSC) approach implemented in ASTRAL-III71 to account for incomplete lineage sorting (here, taxa were not combined into OTUs). In order to improve the phylogenetic signal in the single-gene trees used in ASTRAL, a partial filtering method (i.e., Divvier72 using the -partial option) was applied followed by trimming of highly incomplete positions with trimAL (-gt 0.05). ML single-gene trees were inferred with IQ-TREE under BIC-selected models including site-homogeneous models (such as LG) and empirical profile models (C10-C60). Branch support was inferred with 1,000 replicates of ultrafast bootstrap (-bbni) and branches with <10 support were collapsed. Branch support for the MSC tree (Supplementary Fig. 3) was calculated as quartet supports73, and multilocus bootstrapping (1,000 UFboot2 bootstraps per gene).

Finally, we created a taxon-subset of the 136-OTU dataset containing only 63 OTUs with 73,460 aligned amino acid positions (67,700 distinct patterns; gaps and undetermined characters: ~13%; Supplementary Data 2), to be able to obtain chains convergence with the computationally very demanding CAT+GTR+G model in PhyloBayes-MPI v. 1.8. Three independent Markov Chain Monte Carlo (MCMC) chains were run for 2,600 cycles (all sampled). The initial 500 cycles were removed (as burnin) from each chain before generating a consensus tree using the bpcomp option. Global convergence was achieved in all combinations of the chains (Supplementary Fig. 4) with maxdiff reaching 0. The 63-OTU dataset was also used in posterior predictive analyses (Supplementary Table 1) to informatively select the best topology. For that, the CAT+GTR+G tree was compared to a tree obtained using PhyloBayes-MPI v. 1.8 under
LG+C60+G+F; Supplementary Fig. 5). The reference topology used in the dating analyses corresponded to a ML analyses of the 136-OTU dataset under LG+C60+G+F but constrained by the relationships obtained in the CAT+GTR+G tree.

**Molecular dating.** Bayesian molecular dating was performed with MCMCTree within the PAML package v. 4.9h, PhyloBayes-MPI v. 1.8, and BEAST v. 1.10.4. We used a total of 33 node calibrations based on fossil evidence (retrieved April 2019; Table 1) and the tree topology of Supplementary Fig. 4. MCMCTree was used to perform a set of sensitivity analyses (two chains for each experimental condition) in order to understand the effect of different clock models (uncorrelated or autocorrelated), tree roots, and prior calibration densities. A uniform birth-death tree prior was assumed and analyses were run with either (i) uncorrelated (clock = 2) or (ii) autocorrelated (clock = 3) relaxed clock models. The tree was rooted either on (i) Amorphea or (ii) Discoba (Excavata). Four prior calibration distributions were tested, following dos Reis et al.: (a) uniform (i.e., maxima and minima), (b) skew-normal ($\alpha = 10$ and $\beta$ scale parameters chosen so that the 97.5% cumulative probabilities coincide with the maxima; calculated with MCMCTreeR https://github.com/PuttickMacroevolution/MCMCtreeR), and truncated-cauchy distributions with either (c) short ($p = 0$; $c = 0.1$; $pL = 0.01$) or (d) long ($p = 0$; $c = 10$; $pL = 0.01$) distribution tails. Skew-normal distributions represent ‘literal’ interpretations of the fossil record, assuming minima are close to the real node ages, whereas truncated-cauchy distributions represent more ‘loose’ interpretations that assume older divergences than minimum bounds. In all cases, the root was calibrated using a uniform distribution 1.6–3.2 Ga. All maxima and minima were treated as soft bounds with a default 2.5% prior probability beyond their limits. MCMCTree analyses were run on the entire concatenated alignment using approximate likelihood calculations. Data were analysed as a single partition under the LG+G model. The prior on the mean (or ancestral) rates (‘rgene\_gamma’) were set as diffuse gamma Dirichlet priors indicating severe among-lineage rate heterogeneity and mean rates of 0.02625 and 0.0275 amino acid replacements site-1 10^8 Myr-1, for respectively the Amorphea ($\alpha = 2$, $\beta = 72.73$) and Excavata ($\alpha = 2$, $\beta = 76.19$) roots. The average rate was calculated as mean root-to-tip paths on the corresponding ML trees with the two roots. The rate drift parameter (‘sigma2\_gamma’) was set to indicate considerable rate heterogeneity across lineages ($\alpha = 2$, $\beta = 2$). A 100 million years time unit was assumed. Two independent MCMC chains were run for each analysis, each consisting of 20.2 million generations, of which the first 200,000 were excluded as burnin. Convergence of chains was checked a posteriori using Tracer v. 1.7.1 and all parameters reached ESS > 200. A total of 32 analyses were run, corresponding to all possible combinations of two root positions, two clock models, four calibration distributions, and two MCMC chains per experimental condition.

For computational tractability, PhyloBayes and BEAST were run on a subset of the 10 most clock-like genes (selected with SortaDates1). PhyloBayes analyses were run under (i) uncorrelated and (ii) autocorrelated relaxed clock models and using either the (i) site-homogeneous LG+G or (ii) site-heterogeneous CAT+GTR models. In this case, calibrations were set as uniform priors with soft bounds and assumed a birth-death tree prior and the same tree topology (rooted on Amorphea). Two independent MCMC chains were run until convergence, assessed by PhyloBayes’ built-in tools (bpcomp and tracecomp). For comparative purposes, we run an additional BEAST analysis with similar parameterizations (among those available in
BEAST): the uncorrelated relaxed clock, a fixed tree topology rooted on Amorphea, uniform prior calibrations, a Yule tree prior and the LG+G evolutionary model. Two independent MCMC chains were run for 200 million generations, the first 10% being discarded as burnin. Convergence of chains was checked with Tracer and all parameters reached ESS > 200. CorrTest was used to test the autocorrelation of branch lengths.

References

1. Holland, H. D. Volcanic gases, black smokers, and the great oxidation event. Geochim Cosmochim Acta 66, 3811–3826 (2002).
2. Holland, H. D. The oxygenation of the atmosphere and oceans. Philos Trans R Soc B Biol Sci 361, 903–915 (2006).
3. Adl, S. M. et al. Revisions to the classification, nomenclature, and diversity of eukaryotes. J Eukaryot Microbiol 66, 4–119 (2019).
4. Irisarri, I., Strassert, J. F. H. & Burki, F. Phylogenomic insights into the origin of primary plastids. bioRxiv 2020.08.03.231043 (2020). doi:10.1101/2020.08.03.231043
5. Burki, F. Chapter one - The convoluted evolution of eukaryotes with complex plastids. in Secondary Endosymbioses (ed. Hirakawa, Y. B. T.-A. in B. R.) 84, 1–30 (Academic Press, 2017).
6. Felsner, G. et al. ERAD components in organisms with complex red plastids suggest recruitment of a preexisting protein transport pathway for the periplastid membrane. Genome Biol Evol 3, 140–150 (2010).
7. Hempel, F., Felsner, G. & Maier, U. G. New mechanistic insights into pre-protein transport across the second outermost plastid membrane of diatoms. Mol Microbiol 76, 793–801 (2010).
8. Yoon, H. S., Hackett, J. D., Pinto, G. & Bhattacharya, D. The single, ancient origin of chromist plastids. Proc Natl Acad Sci U S A 99, 15507–15512 (2002).
9. Archibald, J. M. Genomic perspectives on the birth and spread of plastids. Proc Natl Acad Sci U S A 112, 10147–10153 (2015).
10. Timmis, J. N., Ayliffe, M. A., Huang, C. Y. & Martin, W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. Nat Rev Genet 5, 123–135 (2004).
11. Janouškovec, J., Horák, A., Oborník, M., Lukeš, J. & Keeling, P. J. A common red algal origin of the apicomplexan, dinoflagellate, and heterokont plastids. Proc Natl Acad Sci U S A 107, 10949–10954 (2010).
12. Cavalier-Smith, T. Principles of protein and lipid targeting in secondary symbiogenesis: euglenoid, dinoflagellate, and sporozoan plastid origins and the eukaryote family tree. J Eukaryot Microbiol 46, 347–366 (1999).
13. Abrahamsen, M. S. et al. Complete genome sequence of the apicomplexan, Cryptosporidium parvum. Science (80- ) 304, 441–444 (2004).
14. Gornik, S. G. et al. Endosymbiosis undone by stepwise elimination of the plastid in a parasitic dinoflagellate. Proc Natl Acad Sci U S A 112, 5767–5772 (2015).
15. Petersen, J. et al. Chromera velia, endosymbioses and the rhodoplex hypothesis—plastid evolution in cryptophytes, alveolates, stramenopiles, and haptophytes (CASH Lineages). Genome Biol Evol 6, 666–684 (2014).
16. Dorrell, R. G. et al. Chimeric origins of ochrophytes and haptophytes revealed through an ancient plastid proteome. Elife 6, e23717 (2017).
17. Miller, J. J. & Delwiche, C. F. Phylogenomic analysis of Emiliania huxleyi provides evidence for haptophyte–stramenopile association and a chimeric haptophyte nuclear genome. Mar Genomics 21, 31–42 (2015).
18. Stiller, J. W. et al. The evolution of photosynthesis in chromist algae through serial endosymbioses. *Nat Commun* **5**, 5764 (2014).

19. Ševčíková, T. et al. Updating algal evolutionary relationships through plastid genome sequencing: did alveolate plastids emerge through endosymbiosis of an ochrophyte? *Sci Rep* **5**, 10134 (2015).

20. Sánchez-Baracaldo, P., Raven, J. A., Pisani, D. & Knoll, A. H. Early photosynthetic eukaryotes inhabited low-salinity habitats. *Proc Natl Acad Sci U S A* **114**, E7737–E7745 (2017).

21. Blank, C. E. Origin and early evolution of photosynthetic eukaryotes in freshwater environments: reinterpreting proterozoic paleobiology and biogeochemical processes in light of trait evolution. *J Phycol* **49**, 1040–1055 (2013).

22. Yang, E. C. et al. Divergence time estimates and the evolution of major lineages in the florideophyte red algae. *Sci Rep* **6**, 21361 (2016).

23. Butterfield, N. J. Bangiomorpha pubescens n. gen., n. sp.: implications for the evolution of sex, multicellularity, and the Mesoproterozoic/Neoproterozoic radiation of eukaryotes. *Paleobiology* **26**, 386–404 (2000).

24. Bengtson, S., Sallstedt, T., Belivanova, V. & Whitehouse, M. Three-dimensional preservation of cellular and subcellular structures suggests 1.6 billion-year-old crown-group red algae. *PLoS Biol* **15**, e2000735 (2017).

25. Butterfield, N. J. Early evolution of the Eukarya. *Palaeontology* **58**, 5–17 (2015).

26. Strassert, J. F. H., Jamy, M., Mylnikov, A. P., Tikhonenkov, D. V. & Burki, F. New phylogenomic analysis of the enigmatic phylum Telonemia further resolves the eukaryote tree of life. *Mol Biol Evol* **36**, 757–765 (2019).

27. Burki, F., Roger, A. J., Brown, M. W. & Simpson, A. G. B. The new tree of eukaryotes. *Trends Ecol Evol* **35**, 43–55 (2020).

28. Keeling, P. J. & Burki, F. Progress towards the tree of eukaryotes. *Curr Biol* **29**, R808–R817 (2019).

29. Love, G. D. et al. Fossil steroids record the appearance of Demospongiae during the Cryogenian period. *Nature* **457**, 718–721 (2009).

30. Muñoz-Gómez, S. A. et al. The new red algal subphylum proteorhodophytina comprises the largest and most divergent plastid genomes known. *Curr Biol* **27**, 1677-1684.e4 (2017).

31. dos Reis, M. et al. Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. *Curr Biol* **25**, 2939–2950 (2015).

32. Gould, S. B., Maier, U.-G. & Martin, W. F. Protein import and the origin of red complex plastids. *Curr Biol* **25**, R515–R521 (2015).

33. Nozaki, H. et al. Phylogenetic positions of Glaucophyta, green plants (Archaeplastida) and Haptophyta (Chromalveolata) as deduced from slowly evolving nuclear genes. *Mol PhylogenET Evol* **53**, 872–880 (2009).

34. Kim, E. & Graham, L. E. EEF2 analysis challenges the monophyly of Archaeplastida and Chromalveolata. *PLoS One* **3**, e2621 (2008).

35. Burki, F. et al. Untangling the early diversification of eukaryotes: a phylogenomic study of the evolutionary origins of Centrohelida, Haptophyta and Cryptista. *Proc R Soc B Biol Sci* **283**, 20152802 (2016).

36. Baurain, D. et al. Phylogenomic evidence for separate acquisition of plastids in cryptophytes, haptophytes, and stramenopiles. *Mol Biol Evol* **27**, 1698–1709 (2010).

37. Palmer, J. D. A single birth of all plastids? *Nature* **405**, 32–33 (2000).

38. Marin, B., M. Nowack, E. C. & Melkonian, M. A plastid in the making: evidence for a second primary endosymbiosis. *Protist* **156**, 425–432 (2005).

39. Katz, L. A. & Grant, J. R. Taxon-rich phylogenomic analyses resolve the eukaryotic tree
of life and reveal the power of subsampling by sites. Syst Biol 64, 406–415 (2014).

Janouškovec, J. et al. A new lineage of eukaryotes illuminates early mitochondrial genome
reduction. Curr Biol 27, 3717-3724.e5 (2017).

Eme, L., Sharpe, S. C., Brown, M. W. & Roger, A. J. On the age of eukaryotes: evaluating
evidence from fossils and molecular clocks. Cold Spring Harb Perspect Biol 6, a016139
(2014).

Parfrey, L. W., Lahr, D. J. G., Knoll, A. H. & Katz, L. A. Estimating the timing of early
eukaryotic diversification with multigene molecular clocks. Proc Natl Acad Sci U S A 108,
13624–13629 (2011).

Betts, H. C. et al. Integrated genomic and fossil evidence illuminates life’s early evolution
and eukaryote origin. Nat Ecol Evol 2, 1556–1562 (2018).

Tang, Q., Pang, K., Yuan, X. & Xiao, S. A one-billion-year-old multicellular chlorophyte.
Nat Ecol Evol 4, pages543–549 (2020).

Loron, C. C. et al. Early fungi from the Proterozoic era in Arctic Canada. Nature 570,
232–235 (2019).

Redecker, D., Kodner, R. & Graham, L. E. Glomalean fungi from the Ordovician. Science
(80-) 289, 1920 LP – 1921 (2000).

Brasier, M. D., Antcliff, J., Saunders, M. & Wacey, D. Changing the picture of Earth’s
earliest fossils (3.5–1.9 Ga) with new approaches and new discoveries. Proc Natl Acad Sci
U S A 112, 4859–4864 (2015).

Donoghue, P. C. J. Fossil cells. Curr Biol 30, R485–R490 (2020).

Butterfield, N. J. Proterozoic photosynthesis - a critical review. Palaeontology 58, 953–
972 (2015).

Bodył, A., Stiller, J. & Mackiewicz, P. Chromalveolate plastids: direct descent or multiple
endosymbioses? Trends Ecol Evol 24, 119–121 (2009).

Brocks, J. J. et al. The rise of algae in Cryogenian oceans and the emergence of animals.
Nature 548, 578–581 (2017).

Brocks, J. J. The transition from a cyanobacterial to algal world and the emergence of
animals. Emerg Top Life Sci 2, 181–190 (2018).

Maruyama, S. & Kim, E. A modern descendant of early green algal phagotrophs. Curr
Biol 23, 1081–1084 (2013).

Gawryluk, R. M. R. et al. Non-photosynthetic predators are sister to red algae. Nature 572,
240–243 (2019).

Hartmann, M. et al. Mixotrophic basis of Atlantic oligotrophic ecosystems. Proc Natl
Acad Sci U S A 109, 5756–5760 (2012).

Edwards, K. F. Mixotrophy in nanoflagellates across environmental gradients in the ocean.
Proc Natl Acad Sci U S A 116, 6211–6220 (2019).

Katz, M. E., Finkel, Z. V, Grzebyk, D., Knoll, A. H. & Falkowski, P. G. Evolutionary
trajectories and biogeochemical impacts of marine eukaryotic phytoplankton. Annu Rev
Ecol Evol Syst 35, 523–556 (2004).

Brown, M. W. et al. Phylogenomics places orphan protistan lineages in a novel eukaryotic
super-group. Genome Biol Evol 10, 427–433 (2018).

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment
search tool. J Mol Biol 215, 403–410 (1990).

Richter, D. J., Berney, C., Strassert, J. F. H., Burki, F. & de Vargas, C. EukProt: a database
of genome-scale predicted proteins across the diversity of eukaryotic life. bioRxiv
2020.06.30.180687 (2020). doi:10.1101/2020.06.30.180687

Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-
generation sequencing data. Bioinformatics 28, 3150–3152 (2012).

Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7:
improvements in performance and usability. *Mol Biol Evol* **30**, 772–780 (2013).

63. Capella-Gutierrez, S., Silla-Martinez, J. M. & Gabaldon, T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009).

64. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol Evol* **26**, 1641–1650 (2009).

65. Stamatakis, A. RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).

66. Whelan, S., Irisarri, I. & Burki, F. PREQUAL: detecting non-homologous characters in sets of unaligned homologous sequences. *Bioinformatics* **34**, 3929–3930 (2018).

67. Criscuolo, A. & Gribaldo, S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* **10**, (2010).

68. Roure, B., Rodriguez-Ezpeleta, N. & Philippe, H. SCaFoS: a tool for selection, concatenation and fusion of sequences for phylogenomics. *BMC Evol Biol* **7**, (2007).

69. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **32**, 268–274 (2015).

70. Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* **35**, 518–522 (2018).

71. Zhang, C., Rabiee, M., Sayyari, E. & Mirarab, S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**, 153 (2018).

72. Ali, R. H., Bogusz, M. & Whelan, S. Identifying clusters of high confidence homologies in multiple sequence alignments. *Mol Biol Evol* **36**, 2340–2351 (2019).

73. Sayyari, E. & Mirarab, S. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol Biol Evol* **33**, 1654–1668 (2016).

74. dos Reis, M., Donoghue, P. C. J. & Yang, Z. Bayesian molecular clock dating of species divergences in the genomics era. *Nat Rev Genet* **17**, 71–80 (2016).

75. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586–1591 (2007).

76. Suchard, M. A. *et al.* Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol* **4**, vey016 (2018).

77. Derelle, R. *et al.* Bacterial proteins pinpoint a single eukaryotic root. *Proc Natl Acad Sci USA* **112**, E693–E699 (2015).

78. He, D. *et al.* An alternative root for the eukaryote tree of life. *Curr Biol* **24**, 465–470 (2014).

79. dos Reis, M. & Yang, Z. Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Mol Biol Evol* **28**, 2161–2172 (2011).

80. Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst Biol* **67**, 901–904 (2018).

81. Smith, S. A., Brown, J. W. & Walker, J. F. So many genes, so little time: a practical approach to divergence-time estimation in the genome era. *PLoS One* **13**, e0197433 (2018).

82. Tao, Q., Tamura, K., Battistuzzi, F. U. & Kumar, S. A machine learning method for detecting autocorrelation of evolutionary rates in large phylogenies. *Mol Biol Evol* **36**, 811–824 (2019).

83. Sanchez-Puerta, M. V. & Delwiche, C. F. A Hypothesis for plastid evolution in chromalveolates. *J Phycol* **44**, 1097–1107 (2008).

84. Dorrell, R. G. & Smith, A. G. Do red and green make brown?: Perspectives on plastid acquisitions within chromalveolates. *Eukaryot Cell* **10**, 856–868 (2011).
85. Gibson, T. M. et al. Precise age of Bangiomorpha pubescens dates the origin of eukaryotic photosynthesis. Geology 46, 135–138 (2017).

86. Benton, M. J. et al. Constraints on the timescale of animal evolutionary history. Palaeontol Electron 1–107 (2015).

87. Summons, R. E. & Walter, M. R. Molecular fossils and microfossils of prokaryotes and protists from Proterozoic sediments. Am J Sci 290-A, 212–244 (1990).

88. Porter, S. M. & Knoll, A. H. Testate amoebae in the Neoproterozoic Era: evidence from vase-shaped microfossils in the Chuar Group, Grand Canyon. Paleobiology 26, 360–385 (2000).

89. Porter, S. M., Meisterfeld, R. & Knoll, A. H. Vase-shaped microfossils from the Neoproterozoic Chuar Group, Grand Canyon: a classification guided by modern testate amoebae. J Paleontol 77, 409–429 (2003).

90. Fiz-Palacios, O., Leander, B. S. & Heger, T. J. Old lineages in a new ecosystem: diversification of arcellinid amoebae (amoebozoa) and peatland mosses. PLoS One 9, (2014).

91. Butterfield, N. J., Knoll, A. H. & Swett, K. Paleobiology of the Neoproterozoic Svanbergfjellet Formation, Spitsbergen. Lethaia 27, 76 (1994).

92. McIlroy, D., Green, O. R. & Brasier, M. D. Palaeobiology and evolution of the earliest agglutinated Foraminifera: Platysolenites, Spirosolenites and related forms. Lethaia 34, 13–29 (2001).

93. Seilacher, A., Grazhdankin, D. & Legouta, A. Ediacaran biota: the dawn of animal life in the shadow of giant protists. Paleontol Res 7, 43–54 (2003).

94. Sims, P. A., Mann, D. G. & Medlin, L. K. Evolution of the diatoms: insights from fossil, biological and molecular data. Phycologia 45, 361–402 (2006).

95. Clarke, J. T., Warnock, R. C. M. & Donoghue, P. C. J. Establishing a time-scale for plant evolution. New Phytol 192, 266–301 (2011).

96. Steemans, P. et al. Origin and radiation of the earliest vascular land plants. Science 324, 353 (2009).

97. Gray, J. & Boucot, A. J. Is Moyeria a euglenoid? Lethaia 22, 447–456 (1989).

98. Taylor, T. N., Remy, W. & Hass, H. Allomyces in the Devonian. Nature 367, 601 (1994).

99. Taylor, T. N. et al. Fungi from the Rhynie chert: a view from the dark side. Trans R Soc Edinb Earth Sci 94, 457–473 (2004).

100. Strullu-Derrien, C., Kenrick, P., Goral, T. & Knoll, A. H. Testate amoebae in the 407-million-year-old Rhynie Chert. Curr Biol 29, 461–467 (2018).

101. Taylor, T. N., Hass, H. & Kerp, H. The oldest fossil ascomycetes. Nature 399, 648 (1999).

102. Gerrienne, P., Meyer-Berthaud, B., Fairon-Demaret, M., Street, M. & Steemans, P. Runcaria, a middle devonian seed plant precursor. Science 306, 856–858 (2004).

103. Stubblefield, S. P., Taylor, T. N. & Beck, C. B. Studies of Paleozoic fungi. IV. Wood-decaying fungi in Callixylon newberryi from the Upper Devonian. Am J Bot 72, 1765–1774 (1985).

104. Fensome, R. A., Saldarriaga, J. F. & Taylor, M. F. J. R. Dinoflagellate phylogeny revisited: reconciling morphological and molecular based phylogenies. Grana 38, 66–80 (1999).

105. Bown, P. R., Lees, J. A. & Young, J. R. Calcareous nannoplankton evolution and diversity through time. in Coccolithophores (eds. Thierstein, H. R. & Young, J. R.) 481–508 (Springer, 2004). doi:10.1007/978-3-662-06278-4_18

106. Brenner, G. J. Evidence for the earliest stage of angiosperm pollen evolution: a paleoequatorial section from Israel. in Flowering Plant Origin, Evolution and Phylogeny (eds. Taylor, D. W. & Hickey, L. J.) 91–115 (Springer, 1996). doi:10.1007/978-0-585-23095-5_5

107. Berthaud, B., Fairon.
107. Hughes, N. F. & McDougall, A. B. Barremian-Aptian angiospermid pollen records from southern England. *Rev Palaeobot Palynol* **65**, 145–151 (1990).

108. Doyle, J. A. Revised palynological correlations of the lower Potomac group (USA) and the cocobeach sequence of Gabon (Barremian-Aptian). *Cretac Res* **13**, 337–349 (1992).

109. Young, J. R., Bown, P. & Burnett, J. A. Palaeontological perspectives. in *The Haptophyte Algae* (eds. Green, J. C. & Leadbeater, B. S. C.) 379–392 (Oxford University Press, 1994).

110. Fujiwara, S., Tsuzuki, M., Kawachi, M., Minaka, N. & Inouye, I. Molecular phylogeny of the Haptophyta based on the rbcL gene and sequence variation in the spacer region of the RUBISCO operon. *J Phycol* **37**, 121–129 (2001).

111. Knoll, A. H. Paleobiological perspectives on early eukaryotic evolution. *Cold Spring Harb Perspect Biol* **6**, (2014).

**Acknowledgments**

This work was supported by a grant from Science for Life Laboratory available to FB, which covered the salary of JFHS and II. II was also supported by a postdoctoral Juan de la Cierva-Incorporación fellowship (IJCI-2016-29566) from the Spanish Ministry of Economy and Competitiveness (MINECO). TAW was supported by a Royal Society University Research Fellowship and NERC Grant NE/P00251X/1. Computations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX) under Projects 2017-7-65, 2017-7-355, 2018-3-147, 2018-3-288, 2018-8-187, 2018-8-192, and 2019-3-305.

**Author contributions**

FB conceived, designed and supervised the project. JFHS assembled and curated the data, and performed phylogenomic analyses. TAW analysed data under the Multi Species Coalescent model. JFHS compiled the fossil records, and II designed and performed molecular dating analyses. Artwork was made by JFHS (Figs. 1, 2, S1–S6) and II (Figs. 3, S7). All authors drafted the manuscript and read and approved the final version.

**Competing interests**

The authors declare no competing interests.

**Figure legends**

**Fig. 1.** Serial plastid endosymbioses models as proposed by Stiller *et al.*18 (left) and Bodył *et al.*50 (right). The tree topology shown here is based on the results obtained in our study. Further models have been suggested but are not compatible with this topology15,83,84. Numbers denote the level of endosymbiosis events. Note, Myzozoa were not included by Stiller *et al.*18 and the dashed line indicates engulfment of an ochrophyte by the common ancestor of Myzozoa as suggested by Ševčíková *et al.*19.

**Fig. 2.** Time-calibrated phylogeny of extant eukaryotes. Divergence times were inferred with MCMCTree under an autocorrelated relaxed clock model and 33 fossil calibration points as soft-bound uniform priors (Table 1). The tree topology was reconstructed using IQ-TREE under the
LG+C60+G+F model and a constrained tree search following the OTU-reduced CAT+GTR+G topology (Supplementary Fig. 4). Approximate likelihood calculations on the 320 gene concatenation under LG+G and a birth-death tree prior were used. Bars at nodes are 95% HPD. Bars corresponding to the first and last common ancestors of extant red plastid-donating and -containing lineages are highlighted in red and their stems are shaded as indicated. Crowns denote the common ancestors of the extant members of these groups. An absolute time scale in Ma and a geological time scale are shown. The tree depicted here was rooted on Amorphea. An equivalent time-calibrated tree rooted on Excavata is shown in the Supplementary Fig. 6. Cry – Cryptophyta, Hap – Haptophyta, Myz – Myzozoa, Och – Ochrophyta, Rho – Rhodophytina.

**Fig. 3.** Summary of inferred timeframes for the spread of complex red plastids using different software and models. Vertical lines correspond to the lower and upper 95% HPD intervals from the nodes defining the branch of interest and dots indicate their posterior mean divergences. Faded boxes represent the temporal windows for the secondary endosymbioses, constrained by the HPDs and tree topology under the two proposed symbiotic scenarios. Numbers at arrows denote the level of endosymbiosis events (compare with Fig. 1). AC – autocorrelated clock model, UC – uncorrelated clock model.

**Tables**

**Table 1.** Calibrations used for dating the eukaryote tree of life.

| Clade                        | Age constraint (Ma) | Type | Eon       | Refs. |
|------------------------------|---------------------|------|-----------|-------|
| Rhodophyta                  | 1,600               | Min  | Proterozoic | 24    |
| Bangiophyceae/Florideophyceae | 1,600             | Max  | Proterozoic | --b   |
| Bangiophyceae/Florideophyceae | 1,047              | Min  | Proterozoic | 23,85 |
| Metazoa                     | 833                 | Max  | Proterozoic | 86    |
| Ciliophora                  | 740                 | Min  | Proterozoic | 87    |
| Euglyphidae                 | 736                 | Min  | Proterozoic | 88,89 |
| Evesoa/Tubulinea            | 736                 | Min  | Proterozoic | 89,90 |
| Tubulinea                   | 736                 | Max  | Proterozoic | --b   |
| Chlorophyta                 | 700                 | Min  | Proterozoic | 91    |
| Eumetazoa                   | 636                 | Max  | Proterozoic | 86    |
| Deuterostomia               | 636                 | Max  | Proterozoic | 86    |
| Chordata                    | 636                 | Max  | Proterozoic | 86    |
| Bilateria                   | 636                 | Max  | Proterozoic | 86    |
| Arthropoda                  | 636                 | Max  | Proterozoic | 86    |
| Metazoa                     | 635                 | Min  | Proterozoic | 29    |
| Eumetazoa                   | 550                 | Min  | Proterozoic | 86    |
| Bilateria                   | 550                 | Min  | Proterozoic | 86    |
| Mollusca                    | 549                 | Max  | Proterozoic | 86    |
| Foraminifera                | 542                 | Min  | Proterozoic | 92,93 |
| Bacillariophyta             | 541                 | Max  | Proterozoic | 94    |
| Embryophyta                 | 540                 | Max  | Phanerozoic | 20,95 |
| Mollusca                    | 532                 | Min  | Phanerozoic | 86    |
| Deuterostomia               | 515                 | Min  | Phanerozoic | 86    |
| Chordata                    | 514                 | Min  | Phanerozoic | 86    |
| Taxon                                      | Age (Ma) | Time Period     | Reference       |
|-------------------------------------------|----------|-----------------|-----------------|
| Arthropoda                                | 514      | Min Phanerozoic | 86              |
| Embryophyta                               | 470      | Min Phanerozoic | 96              |
| Angiosperms/Gymnosperms                    | 470      | Max Phanerozoic | --b             |
| Euglenales/Eutreptiales<sub>a</sub>      | 450      | Min Phanerozoic | 97              |
| Chytridiomycota<sub>a</sub>               | 410      | Min Phanerozoic | 98, 99          |
| Tubulinea                                 | 405      | Min Phanerozoic | 100             |
| Ascomycota<sub>a</sub>                    | 400      | Min Phanerozoic | 101             |
| Angiosperms/Gymnosperms                    | 385      | Min Phanerozoic | 102             |
| Angiosperms                               | 385      | Max Phanerozoic | --b             |
| Basidiomycota<sub>a</sub>                 | 360      | Min Phanerozoic | 103             |
| Amniota                                   | 333      | Max Phanerozoic | 86              |
| Amniota                                   | 318      | Min Phanerozoic | 86              |
| Core dinoflagellates (excl. Noctilucales) | 300      | Max Phanerozoic | 42              |
| Coccolithales/Isochrysidales              | 260      | Max Phanerozoic | 42              |
| Core dinoflagellates (excl. Noctilucales) | 235      | Min Phanerozoic | 104             |
| Peridiniales                              | 235      | Max Phanerozoic | --b             |
| Gonyaulacales                             | 235      | Max Phanerozoic | --b             |
| Coccolithales/ Isochrysidales             | 225      | Min Phanerozoic | 105             |
| Calcidiscaceae/Coccolithaceae             | 225      | Max Phanerozoic | --b             |
| Peridiniales                              | 210      | Min Phanerozoic | 104             |
| Gonyaulacales                             | 200      | Min Phanerozoic | 104             |
| Bacillariophyta                           | 190      | Min Phanerozoic | 94              |
| Pennales                                  | 190      | Max Phanerozoic | --b             |
| Euarchontoglires                          | 165      | Max Phanerozoic | 86              |
| Angiosperms                               | 130      | Min Phanerozoic | 106             |
| Eudicotyledons (Tricoplates)              | 130      | Max Phanerozoic | --b             |
| Eudicotyledons (Tricoplates)              | 124      | Min Phanerozoic | 107, 108        |
| Aves sensu stricto                        | 87       | Max Phanerozoic | 86              |
| Pennales                                  | 75       | Min Phanerozoic | 94              |
| Aves sensu stricto                        | 66       | Min Phanerozoic | 86              |
| Calcidiscaceae/Coccolithaceae             | 65       | Min Phanerozoic | 109, 110        |
| Euarchontoglires                          | 61       | Min Phanerozoic | 86              |

<sup>a</sup>Max = 1,900 Ma was used; see for example Knoll<sup>111</sup> and Emel<sup>et al.</sup><sup>41</sup>.<br>
<sup>b</sup>Max based on Min age of ancestor.
Table 2. Inferred time intervals for the acquisition of red algal-derived plastids by serial endosymbioses. Time intervals have been calculated from overlapping 95% HPD intervals, constrained by the order of symbioses in the models of Stiller et al.\textsuperscript{18} or Bodyl et al.\textsuperscript{50}. Numbers are in million years. Cry – Cryptophyta, Hap – Haptophyta, Myz – Myzozoa, Och – Ochrophyta, Rho – Rhodophytina.

|        | MCMCTREE |        |        | PHYLOBAYES |        |        | BEAST |        |        |        |
|--------|----------|--------|--------|------------|--------|--------|-------|--------|--------|--------|
|        | AUTOCORRELATED | UNCORRELATED | AUTOCORRELATED | UNCORRELATED | UNCORRELATED | AUTOCORRELATED | UNCORRELATED | AUTOCORRELATED | UNCORRELATED | AUTOCORRELATED | UNCORRELATED |
| Stiller et al. |  |  |  |  |  |  |  |  |  |  |  |
| Rho => Cry       | 1.658  | 1.281  | 377    | 1,453     | 1.158  | 296    | 1,636 | 1,335 | 301    | 1,517  | 1,057  | 461    | 1,502  | 1,210  | 292    |
| Cry => Och       | 1.298  | 623    | 675    | 1,446     | 804    | 642    | 1,518 | 918    | 600    | 1,517  | 747    | 771    | 1,441  | 786    | 654    |
| Och => Hap       | 1.298  | 623    | 675    | 1,446     | 804    | 642    | 1,518 | 918    | 600    | 1,517  | 747    | 771    | 1,441  | 786    | 654    |
| Och => Myz       | 1.298  | 696    | 602    | 1,446     | 1,017  | 428    | 1,518 | 1,138  | 380    | 1,517  | 1,075  | 443    | 1,441  | 1,022  | 419    |

|        |  |  |  |  |  |  |  |  |  |  |  |  |
| Bodyl et al. |  |  |  |  |  |  |  |  |  |  |  |  |
| Rho => Cry       | 1.658  | 1.281  | 377    | 1,453     | 1.158  | 296    | 1,636 | 1,335 | 301    | 1,517  | 1,057  | 461    | 1,502  | 1,210  | 292    |
| Cry => Och       | 1.298  | 623    | 675    | 1,446     | 804    | 642    | 1,518 | 918    | 600    | 1,517  | 747    | 771    | 1,441  | 786    | 654    |
| Cry => Hap       | 1.298  | 579    | 1,079  | 1,453     | 707    | 747    | 1,636 | 1051  | 585    | 1,517  | 573    | 944    | 1,502  | 624    | 878    |
| Hap => Myz       | 1.520  | 696    | 824    | 1,453     | 1,017  | 436    | 1,636 | 1,138  | 498    | 1,517  | 1,075  | 443    | 1,502  | 1,022  | 480    |

Additional information

Supplementary Information: Supplementary Figs. 1–7 and Supplementary Table 1

Supplementary Data 1: Molecular clock results and violin plots

Supplementary Data 2: Sequence data and taxonomy files
**Fig. 1.** Serial plastid endosymbioses models as proposed by Stiller et al.\textsuperscript{18} (left) and Bodył et al.\textsuperscript{50} (right). The tree topology shown here is based on the results obtained in our study. Further models have been suggested but are not compatible with this topology\textsuperscript{15,83,84}. Numbers denote the level of endosymbiosis events. Note, Myzozoa were not included by Stiller et al.\textsuperscript{18} and the dashed line indicates engulfment of an ochrophyte by the common ancestor of Myzozoa as suggested by Ševčíková et al.\textsuperscript{19}.
Fig. 2. Time-calibrated phylogeny of extant eukaryotes. Divergence times were inferred with MCMCTree under an autocorrelated relaxed clock model and 33 fossil calibration points as soft-bound uniform priors (Table 1). The tree topology was reconstructed using IQ-TREE under the LG+C60+G+F model and a constrained tree search following the autocorrelated relaxed clock model and 33 fossil calibration points as soft-bound uniform priors (Table 1). The tree Fig. 2.

2200 1600 1000 540 420 300 140 23

Paleoproterozoic Mesoproterozoic Neoproterozoic Phanerozoic

Stem Rho
Crown Rho

Stem Myz
Crown Myz

Stem Och
Crown Och

Stem Hap
Crown Hap

Stem Cry
Crown Cry

23

Cryptophyta
Haptophyta
Myzozoa
Ochrophyta
Rhodophytina
Fig. 3. Summary of inferred timeframes for the spread of complex red plastids using different software and models. Vertical lines correspond to the lower and upper 95% HPD intervals from the nodes defining the branch of interest and dots indicate their posterior mean divergences. Faded boxes represent the temporal windows for the secondary endosymbioses, constrained by the HPDs and tree topology under the two proposed symbiotic scenarios. Numbers at arrows denote the level of endosymbiosis events (compare with Fig. 1). AC – autocorrelated clock model, UC – uncorrelated clock model.