

Neoproterozoic origin and multiple transitions to macroscopic growth in green seaweeds

Andrea Del Cortona^{a,b,c,d,1}, Christopher J. Jackson^e, François Bucchini^{b,c}, Michiel Van Bel^{b,c}, Sofie D'hondt^a, Pavel Škaloud^f, Charles F. Delwiche^g, Andrew H. Knoll^h, John A. Raven^{i,j,k}, Heroen Verbruggen^e, Klaas Vandepoele^{b,c,d,1,2}, Olivier De Clerck^{a,1,2} Frederik Leliaert^{a,1,2}

^aDepartment of Biology, Phycology Research Group, Ghent University, Krijgslaan 281, 9000 Ghent, Belgium

^bDepartment of Plant Biotechnology and Bioinformatics, Ghent University, Technologiepark 71, 9052 Zwijnaarde, Belgium

^cVIB Center for Plant Systems Biology, Technologiepark 71, 9052 Zwijnaarde, Belgium

^dBioinformatics Institute Ghent, Ghent University, Technologiepark 71, 9052 Zwijnaarde, Belgium

^eSchool of Biosciences, University of Melbourne, Melbourne, Victoria, Australia

^fDepartment of Botany, Faculty of Science, Charles University, Benátská 2, CZ-12800 Prague 2, Czech Republic

^gDepartment of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD 20742, USA

^hDepartment of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, 02138, USA.

ⁱDivision of Plant Sciences, University of Dundee at the James Hutton Institute, Dundee, DD2 5DA, UK

^jSchool of Biological Sciences, University of Western Australia (M048), 35 Stirling Highway, WA 6009, Australia

^kClimate Change Cluster, University of Technology, Ultimo, NSW 2006, Australia

^lMeise Botanic Garden, Nieuwelaan 38, 1860 Meise, Belgium

¹To whom correspondence may be addressed. Email andrea.delcortona@gmail.com, klaas.vandepoele@psb.vib-ugent.be, olivier.declerck@ugent.be or frederik.leliaert@meisebotanicgarden.be.

²K.V., O.D.C and F.L. contributed equally to this work.

Author contributions: A.D.C., H.V., K.V., O.D.C, and F.L. designed research; A.D.C., C.J.J., and S.D. performed research; P.S., and C.F.D. contributed cultures and sequence data; A.D.C., C.J.J., F.B., and M.V.B. analyzed data; and A.D.C., A.H.K., J.A.R., H.V., K.V., O.D.C., and F.L. wrote the paper.

Abstract

The Neoproterozoic Era records the transition from a largely bacterial to a predominantly eukaryotic phototrophic world, creating the foundation for the complex benthic ecosystems that have sustained Metazoa from the Ediacaran Period onward. This study focusses on the evolutionary origins of green seaweeds, which play an important ecological role in the benthos of modern sunlit oceans and likely played a crucial part in the evolution of early animals by structuring benthic habitats and providing novel niches. By applying a phylogenomic approach, we resolve deep relationships of the core Chlorophyta (Ulvophyceae or green seaweeds, and freshwater or terrestrial Chlorophyceae and Trebouxiophyceae) and unveil a rapid radiation of Chlorophyceae and the principal lineages of the Ulvophyceae late in the Neoproterozoic Era. Our time-calibrated tree points to an origin and early diversification of green seaweeds in the late Tonian and Cryogenian periods, an interval marked by two global glaciations, with strong consequent changes in the amount of available marine benthic habitat. We hypothesize that the unicellular and simple multicellular ancestors of green seaweeds survived these extreme climate events in isolated refugia, and diversified following recolonization of benthic environments that became increasingly available as sea ice retreated. An increased supply of nutrients and biotic interactions such as grazing pressure has likely triggered the independent evolution of macroscopic growth via different strategies, including both true multicellularity, and multiple types of giant celled forms.

Key words: green algae, Chlorophyta, phylogenomics, phylogeny, Ulvophyceae

Significance Statement

Green seaweeds are important primary producers along coastlines worldwide, and likely played a key role in the evolution of animals. To understand their origin and diversification, we resolve key relationships among extant green algae using a phylotranscriptomic approach. We calibrate our tree using available fossil data, to reconstruct important evolutionary events such as transitions to benthic environments, and evolution of macroscopic growth. We estimate green seaweeds to have originated in the late Tonian/Cryogenian Period, followed by a marked Ordovician diversification of macroscopic forms. This ancient proliferation of green seaweeds likely modified shallow marine ecosystems, which set off an evolutionary arms race between ever larger seaweeds and grazers.

Introduction

Seaweeds, ecologically important primary producers in marine benthic ecosystems worldwide, have played a prominent role in the global biosphere for many millions of years. Seaweeds comprise red, green, and brown lineages, which evolved independently from unicellular algal ancestors. The red algae are ancient. *Bangiomorpha* and *Raffatazamia*, both reasonably interpreted as red algal fossils, point to an origin of multicellular red algae well into the Mesoproterozoic, 1.0–1.6 Ga ago (1, 2). Brown algae diverged much more recently, with molecular clocks pointing to an Early Jurassic emergence (3, 4). As sister to red algae, microscopic green algae likely originated in Neo- or Mesoproterozoic environments (5, 6), but the lack of a well-supported phylogenetic framework has so far impeded clear interpretation of how many times and when green seaweeds emerged from unicellular ancestors.

There is general consensus that an early split in the evolution of green algae gave rise to two discrete clades. One, the Streptophyta, contains a wide morphological diversity of green algae from freshwater and damp terrestrial habitats, known as charophytes, from which land plants evolved during the Ordovician Period (7, 8). The second clade, the Chlorophyta, diversified as planktonic unicellular organisms, likely in marine as well as freshwater habitats during the late Mesoproterozoic and early Neoproterozoic (9, 10). These ancestral green algae gave rise to several extant lineages of unicellular planktonic marine and freshwater algae, known as the prasinophytes, as well as the core Chlorophyta, which radiated in freshwater, terrestrial, and coastal environments, and evolved a wide diversity of forms ranging from microscopic unicellular and multicellular algae to macroscopic forms (11). Two large core chlorophytan classes, Chlorophyceae and Trebouxiophyceae, are almost entirely restricted to freshwater and terrestrial environments. In contrast, the Ulvophyceae contains the main green seaweed lineages in addition to some smaller microscopic clades, some freshwater species, and the terrestrial Trentepohliales (12, 13). Only two other groups of green seaweeds have phylogenetic affinities outside the Ulvophyceae: the Prasiolales, which belong to the Trebouxiophyceae, and the Palmophyllales, which are allied to prasinophytes (11, 14). These groups are much less diverse compared to the ulvophycean lineages, and clearly evolved independently.

The diversification of Ulvophyceae in marine benthic environments involved the evolution of an astonishing diversity of forms, and most strikingly the evolution of macroscopic, benthic growth forms from small, planktonic unicellular ancestors. Macroscopic growth in green seaweeds presents itself in various forms ranging from multicellular thalli to different types of giant-celled algae with highly specialized cellular and physiological characteristics (15). About ten extant ulvophycean orders are currently recognized, each characterized by a distinctive set of morphological and cytological features (13). Some orders (e.g. Ulvales and Ulotrichales) evolved multicellularity with coupled

mitosis and cytokinesis, resulting in uninucleate cells. The Cladophorales evolved siphonocladous multicellular algae, in which mitosis is uncoupled from cytokinesis, resulting in large multinucleate cells with nuclei organized in fixed cytoplasmic domains. The Dasycladales and Bryopsidales evolved siphonous (acellular) macroscopic forms composed of a single giant tubular cell containing thousands to millions of nuclei, or a single macronucleus, and exhibiting cytoplasmic streaming, which enables transport of RNA transcripts, organelles and nutrients throughout the thallus. Some siphonous algae, including species of *Caulerpa*, reach meters in size, thus qualifying as the largest known cells. Although the available data are limited, it seems that acellular algae have similar maximum photosynthetic and nutrient acquisition rates, light absorptance, and ability to growth at low irradiances, to those of morphologically similar multicellular algae (16). The smaller, non-seaweed orders (e.g. Ignatiales, Scotinosphaerales, and Oltmannsiellopsidales), are morphologically less complex, including microscopic unicellular forms with uninucleate cells.

Understanding the origin and ecological diversification of green seaweeds requires a well-resolved phylogeny of the core Chlorophyta, and reliable estimates of the timing of inferred diversification events. Early studies based on ultrastructural features, such as the fine structure the flagellar apparatus, cytokinesis and mitosis have been instrumental in defining higher level groupings of green algae, but they have been inconclusive in determining the relationships among these groups (11). In addition, monophyly of the Ulvophyceae has been questioned because of the absence of shared derived characters (11). Molecular phylogenetic studies based on nuclear and chloroplast gene data often yielded ambivalent or contradictory results (reviewed in 17, 18). These studies make it clear that resolving relationships within the core Chlorophyta is a difficult task, attributed to the antiquity of the clade and possibly further confounded by the rapidity of the early evolutionary radiations (13, 19). An accurate phylogenetic reconstruction will thus require more elaborate sampling, both in terms of species and genes.

Dating divergence times in the phylogeny of the Chlorophyta has been challenging because of difficulties in interpreting fossils with respect to extant taxa. Microfossils in Paleo- to Neoproterozoic rocks have sometimes been assigned to green algae (20-22), but these interpretations are uncertain as they rely on comparisons of simple morphologies (23). Similarly, the assignment of the middle Neoproterozoic filamentous fossil *Proterocladus* (ca. 750 mya) to the Cladophorales (24, 25), is questioned by some (3, 26). Reliable chlorophytan fossils include resistant outer walls of prasinophyte cysts known as phycmata in Ediacaran and Paleozoic deposits (27, 28) and fossils of siphonous seaweeds (Bryopsidales, Dasycladales) from the Cambro-Ordovician onward (29-33). Although reliable green algal fossils from the Neoproterozoic are scarce, organic biomarkers

(steroids) indicate that green algae persisted throughout the Cryogenian, and rose to abundance between the Sturtian and Marinoan glaciations (659-645 mya) (34-36).

The principal goal of our study was to resolve the evolutionary relationships among the main lineages of core Chlorophyta and reconstruct key evolutionary events, such as transitions to benthic marine environments, and the evolution of macroscopic growth. We use a rigorous phylotranscriptomic approach, thereby increasing nuclear gene sampling by an order of magnitude and produce a tree calibrated in geological time using available fossil data that facilitates interpreting the ecological and evolutionary context of modern green seaweed origins.

Results

Transcriptome data

We collected and analyzed nuclear encoded protein-coding genes from 55 species mined from 15 genomes and 40 transcriptomes (Table S1, S2). 14 transcriptomes were newly generated during this study. Our dataset includes representatives from the major lineages of Streptophyta and Chlorophyta. A denser taxon sampling of the core Chlorophyta, including all main orders of Ulvophyceae (Bryopsidales, Cladophorales, Dasycladales, Ignatiales, Oltmannsiellopsidales, Scotinosphaerales, Trentepohliales, Ulotrichales and Ulvales) was obtained to consolidate evolutionary relationships in this group and advance our understanding of the origin and diversification of green seaweeds.

Eight sequence alignments were assembled for phylogenetic analyses (Table S3). The largest alignment consisted of 539 high-confidence single-copy genes (hereafter referred to as coreGF, see Materials and Methods). From this dataset a subset of 355 genes was selected, with at least one sequence in each ulvophycean order (hereafter referred to as ulvoGF). Partial sequences from the transcriptomes were either scaffolded (scaffolded dataset) or removed (unscaffolded dataset), resulting in a more comprehensive and a more conservative version of the coreGF and ulvoGF datasets (Figure S1). Finally, ambiguously aligned regions were removed to obtain the corresponding trimmed datasets. The eight datasets were analyzed with a range of phylogenetic methods including supermatrix and the coalescence-based analyses. The significance of conflicting topologies was tested to assess the robustness of our findings.

Green algal phylogeny

Trees estimated by the various methods and datasets were globally well-supported and congruent, be it that different topologies were recovered for a few specific relationships. Most differences were due to analysis method (i.e. supermatrix versus coalescence-based) rather than the dataset used (Figure S3). Chlorodendrophyceae and Pedinophyceae were recovered as the two earliest diverging lineages of the core Chlorophyta, although their relative position differed in the different analyses (Figure 1, Figure S3, Figure S4, Table S4). The Trebouxiophyceae and Chlorophyceae were recovered as monophyletic groups with high support, with the Trebouxiophyceae consisting of two distinct clades, the Chlorellales and the core Trebouxiophyceae. The Trebouxiophyceae were recovered sister to the clade containing the Chlorophyceae and Ulvophyceae in all analyses.

The siphonous seaweed order Bryopsidales was resolved as the sister clade of the Chlorophyceae, rendering the Ulvophyceae non-monophyletic, in the supermatrix analyses (Figure 1, Figure 2A, Figure S2). Conversely, in the coalescent-based analyses, the Bryopsidales was sister to the remaining Ulvophyceae, with very short branches separating the Bryopsidales, remaining Ulvophyceae, and Chlorophyceae (Figure 2A). A polytomy test could not reject the null hypothesis that the length of the branches in question equals zero, indicating a hard polytomy (Figure 2B, Figure S6). In addition, AU-tests for the trimmed datasets could not reject the topology constrained to conform the coalescence-based analyses (Ulvophyceae monophyletic). For the untrimmed datasets, however, Ulvophyceae monophyly was significantly rejected (Table S4). Analysis of the gene-wise log-likelihood showed that slightly more than half of the genes supported the monophyly of Ulvophyceae for 7 of the 8 datasets (Figure 2C).

The supermatrix and coalescence-based analyses supported the same overall relationships among the remaining orders of Ulvophyceae. Two major clades were recovered, both containing seaweed and unicellular lineages. The position the Ignatiales was not well-supported in any of the phylogenetic analyses, and similar number of genes supported five different relationships (Figure S4, Table S4). The inclusion of *Ignatius* was also found to cause instability in the phylogenetic analyses, and removing it led to an overall higher support for the relationships among the main ulvophycean clades (Figure S5).

Time-calibrated phylogeny

To estimate the time-frame of diversification, we inferred a chronogram based on 10 of the most clock-like genes, using different molecular clock models, and constraining the analyses with supermatrix- and coalescence-based topologies, respectively. The choice of calibration points largely follows that of Jackson et al. (6). A series of analyses was performed in order to evaluate the effect of

calibration nodes (Table S5, Figure S7). Results indicate that the core Chlorophyta emerged during the Neoproterozoic Era, ca. 1000-700 mya (Figure 4B, Table S7). Conditional on the calibration points, diversification of the main ulvophycean lineages took place just before or during the Cryogenian.

Transition to marine benthic habitats and evolution of macroscopic growth

Our statistical inference of the ancestral states of key ecological and cyto-morphological traits (Figure S8) indicate that the Trebouxiophyceae and Chlorophyceae diversified in freshwater environments mainly. The radiation event at the base of the diversification of Chlorophyceae, Bryopsidales and Ulvophyceae is associated with a major switch back to marine environments early in the evolution of Bryopsidales and Ulvophyceae, populating coastal environments with marine benthic green seaweeds.

Discussion

We present a phylogenetic reconstruction of the core Chlorophyta based on the most comprehensive multigene dataset to date, and contextualize the relationships in light of transition to marine benthic environments and evolution of multicellularity and macroscopic growth. Although this study focused on the evolution of the Chlorophyta, the relationships among the main lineages of Streptophyta are consistent with current phylogenetic consensus, validating the strength of our phylotranscriptomic approach in confidently resolving difficult phylogenetic relationships, such as identifying the closest living relative of the land plants (37, 38). Our analyses resolve some long-standing phylogenetic questions within the Chlorophyta (39, 40), including the evolutionary placement of the unicellular Pedinophyceae and Chlorodendrophyceae as the earliest diverging lineages of the core Chlorophyta, and monophyly of the Trebouxiophyceae (Supporting information). Our analyses suggest a hard polytomy between the Chlorophyceae, siphonous seaweeds of the Bryopsidales and the remaining Ulvophyceae. A roughly equal numbers of genes supported alternative topologies, indicative of ancient incomplete lineage sorting or a non-bifurcating evolutionary history (41, 42). Although coalescence-based analyses have the ability to detect and incorporate conflicting signals in gene-trees in phylogenetic reconstruction, and supermatrix analyses may converge on a wrong species-tree when incomplete lineage sorting is high (43-45), we hypothesize these lineages represent a rapid ancient radiation.

Our phylogenetic results allow us to propose a scenario for the evolutionary history of green seaweeds. Given that several early-branching ulvophycean lineages consist of unicellular algae (including the Ignatiales, Oltmansiellopsidales, Scotinosphaerales, and some lineages of Ulotrionales and Ulvales), the ancestral ulvophycean was likely a small unicellular organism with a single nucleus. Our phylogeny also affirms the hypothesis by Cocquyt et al. (13) that macroscopic growth evolved independently in various lineages of ulvophyceans from ancestral unicellular green algae, and by different mechanisms: multicellularity with coupled mitosis and cytokinesis (independently in the Ulvales-Ulotrionales and Trentepohliales), multicellularity with uncoupled mitosis and cytokinesis (siphonocladous organization in the Cladophorales and *Blastophysa*) and a siphonous architecture (independently in the Bryopsidales and Dasycladales). The sister relationship of Cladophorales (plus *Blastophysa*) and Trentepohliales may indicate a single origin of multicellularity at the base of this clade, followed by the evolution of multinuclear cells in the Cladophorales. However, the phragmoplast-mediated cell division in Trentepohliales, producing plasmodesmata, differs strongly from cell division in the Cladophorales, which takes place by ingrowth of a diaphragm-like cross wall (11). It may therefore be more reasonable to think that multicellularity evolved independently in the Trentepohliales and the Cladophorales. Contrary to previous phylogenetic studies, which indicated a sister relationship between Bryopsidales and Dasycladales (13, 39), the separate positions of these orders point towards an independent evolution of siphonous organization in Bryopsidales and Dasycladales from ancestral unicellular green algae. The distinct position of the Bryopsidales is unexpected from a morphological view-point, but is supported by ultrastructural features (Supporting information), as well as by independent molecular data. The Trentepohliales, Cladophorales, Dasycladales, and Scotinosphaerales are characterized by a non-canonical nuclear genetic code while the Bryopsidales and all other green algae possess the canonical code (46, 47), pointing towards a single origin of a non-canonical nuclear code in the green algae.

Our relaxed molecular clock analysis indicates that the early diversification of core Chlorophyta took place in the Tonian Period, with several green algal lineages persisting throughout the Cryogenian ice ages. The early diversification of ulvophycean lineages took place during this interval, as well. Diversification before and survival during the Cryogenian glacial interval has been inferred for other groups of eukaryotes based on fossil evidence, including complex forms with blade-stipe-holdfast differentiation that have been interpreted as benthic macroalgae (48-50). It is important to note, however, that the scarce fossil record and the uncertain nature of ancient green algal fossils, along with methodological bias impose a limit on the precision of ancient divergence time estimates (51). These unavoidable uncertainties in node age assignments have to be accommodated in our interpretations, and therefore one cannot dismiss the hypothesis that green seaweeds began to

diversify earlier, in the middle or even early Tonian Period. One concrete, well-dated fossil that has been put forward as important to estimate the age of the Ulvophyceae is *Proterocladus*, from ca. 780 Ma shales of the upper Svanbergfjellet Formation in Svalbard, originally interpreted as a *Cladophora*-like alga (25). *Proterocladus* has typical branches that are in cytoplasmic contact with the parent cell, which is as a derived feature in the Cladophorales (52). Assignment of *Proterocladus* to the Cladophorales would imply that the order would be older than 780 my, and would push the divergences of the main ulvophycean lineages even further back in time, inconsistent with our time-calibrated phylogeny, which suggests an Ediacaran or Cambrian origin of the Cladophorales. While *Proterocladus* is certainly eukaryotic, and appears to be coenocytic, we prefer to consider its taxonomic affinity as uncertain based on the late appearance of Cladophorales in our molecular clock analyses.

Ulvophycean diversification could have been promoted by the highly dynamic climatic conditions and habitat variability of the Cryogenian periods (53, 54). Little is known about the paleontological conditions of these periods that would have allowed lineage diversification and survival, but a number of different scenarios can be put forward. The later Tonian Period was a time of evolutionary innovation among both photosynthetic and heterotrophic eukaryotes in a number of major eukaryotic clades (54); thus, inferred divergence of green seaweeds would appear to be part of a broader pattern of Tonian diversification. This has sometimes been related to an increase in atmospheric oxygen levels (e.g., 54), but evidence in support of this hypothesis is mixed, suggesting that any redox change was small relative to later Ediacaran oxygenation (e.g., 55). Alternatively, evolutionary drivers may have included the expansion of eukaryovorous protists in the oceans, supported by both fossils and molecular clock estimates for major eukaryote-eating protistan lineages (56, 57). Experiments show that predation can select for multicellularity in originally unicellular prey populations (58, 59), consistent with the observed and inferred Tonian diversification of multicellular and coenocytic forms, as well as scales and other protistan armor (54, 57).

As already noted, the ensuing Cryogenian Period was characterized by two widespread glaciations, the older, protracted Sturtian ice age and the younger and shorter Marinoan glaciation (53). Among the most extreme climatic events in recorded Earth history, these ice ages, involved the freezing of large parts of the ocean surface for millions of years and dramatic fluctuations in biogeochemical cycling. Snowball Earth conditions were certainly unfavorable for planktonic eukaryotes, resulting in greatly diminished photosynthesis in the marine realm. The fossil record is consistent with this notion, showing low overall eukaryotic diversity during this interval (60). An important key to the success and diversification of ulvophycean green algae during the Cryogenian, thus may have been

an evolutionary transition from a planktonic to benthic lifestyle, a shared feature of all extant ulvophycean lineages. During glaciated conditions, benthic environments may have been among the few suitable habitats left for green algal lineages to persist, and diversification may have been promoted in different ways. First, the Sturtian glaciation (716-659 Ma) lasted a long time, but shows evidence of glacial waxing and waning, resulting in secular variation in suitable habitat availability (53). During this period, when colonisable benthic substrates would have oscillated between rare and uncommon, ulvophycean populations likely experienced prolonged isolation and demographic declines that would amplify the effects of genetic drift. We hypothesize that, in concert with local adaptation, this has promoted early diversification of ulvophycean lineages. Macroscopic compressions of probable algae in shales within the Marinoan Nantuo Tillite demonstrate that macroalgae could and did survive at least locally during times when ice was widespread (50). As ice sheets decayed, habitat space would have increased dramatically, perhaps helped along by a transient increase in nutrients (especially P) associated with high post-glacial weathering fluxes. Arguably, the emergence of new ulvophycean lineages could have been rapid in this new permissive ecology, where competition for newly available habitats would have been low, allowing a broader suite of mutations to persist (61).

We hypothesize that following the transition to a greenhouse world at the end of the Marinoan glaciation, surviving lineages would have been able to diversify, and new morphological experiments ensued. The rapid diversification of eukaryotic life in the early Ediacaran Period is evidenced by molecular biomarkers, which document the rise to global ecological prominence of green algal phytoplankton (36), as well as the diversification of macroscopic seaweeds (both florideophyte reds and greens (62, 63)) and animals (64-66). The transition from predominantly cyanobacteria to green phytoplankton has been hypothesized to reflect increasing nutrient availability, a conjecture supported by a state change in the abundance of phosphorite deposits during the Ediacaran Period (67). Concomitant increase in oxygen availability, recorded by multiple geochemical signatures (55) fulfills another prediction of increased productivity.

The combination of increasing oxygen and food supply likely facilitated the radiation of animals (34, 68), and metazoan evolution, in turn fed back onto seaweed diversification. Indeed, the Paleozoic evolution of multicellularity and macroscopic growth in ulvophyceans may have been triggered by the introduction of novel grazing pressures by animals in the late Cambrian and Ordovician (30, 66). This may have started an evolutionary arm race between grazers and algae, which would have resulted in increasingly complex feeding and defense strategies, such as larger thalli, larger cells, and mineralized skeletons (49, 59). The relatively late appearance of large grazing animals in the the late

Cambrian and Ordovician, such as jawed polychaetes (30, 66) would explain the considerable time lag between the origin of ulvophyceans, and further diversification and origin of crown group ulvophyceans. This scenario is consistent with the predation hypothesis of Stanley (69), who postulated that grazing by benthic animals resulted in a mutual feedback system that drove increased diversity of both macroalgae and animals. The spread of calcium carbonate skeletons in both red and green algae is consistent with the predation hypothesis. Macroscopic growth may also have been facilitated by other factors, including competition for space in benthic habitats, pressure for upright growth to overshadow benthic unicellular algae, a means of increasing nutrient uptake from the water column, or a means of spreading reproductive propagules more widely. A possible mechanism of evolution of attached ulvophycean algae from planktonic ancestors (70) is through intermediary unattached (pleustophytic) macroalgal forms (71), which is consistent with the absence of obvious holdfasts from early ulvophycean fossils (30). In turn, the proliferation of green macroalgae may have resulted in more efficient energy transfer, richer food webs, and the modified shallow marine ecosystems, which may have allowed the evolution of larger and more complex animals (30, 72).

The long-term isolation of ulvophycean lineages in Cryogenian refugia could explain the independent evolution of macroscopic growth in the different clades using radically different mechanisms, as inferred from our phylogeny. Our phylogenetic results, combined with the fossil record indicates that there may have been a time lag of 100 million years or more after the early diversification of ulvophyceans where these microbial ulvophyceans persisted, but never became dominant, followed by transitions to macroscopic growth, which may have occurred in different periods in the different lineages. Our analyses indicate that macroscopic crown group ulvophyceans may have originated between the early Paleozoic (e.g. Bryopsidales) and early Mesozoic (e.g. Cladophorales). This hypothesis is supported by the fossil record, which shows that seaweed morphogroups of the Ediacaran remained largely unchanged during the Cambrian, but that there was a major replacement of this early seaweed flora concomitant with the Great Ordovician Biodiversification Event (30). These observations are also consistent with the fact that the earliest fossils that can be reliably linked to extant ulvophycean clades (Bryopsidales, Dasycladales, and Ulotrichales) are only found from the Ordovician onward (but see discussion on *Proterocladus* above); most (but not all) have calcium carbonate skeletons.

In conclusion, our phylogenetic analysis of the green algae, based on the most comprehensive nuclear gene dataset to date, provides a general picture on the evolution of green seaweeds in which the origin and early diversification of the ulvophyceans likely took place late in the late Tonian and

Cryogenian, followed by a marked Ordovician diversification of macroscopic crown group taxa, including those that produce calcium carbonate skeletons and are well preserved in the fossil record. That this trajectory is similar to that of animals reminds us not to forget these organisms when we think about the environmental circumstances of early green algal evolution. Finally, our analyses provide a phylogenetic framework to study the evolution of the genetic toolkit for multicellularity and macroscopic growth in green seaweeds, a group that has been largely neglected in studies of large-scale gene discovery (73).

Materials and Methods

Dataset retrieval, RNA extraction and sequencing. DNA sequence data were mined from 15 genomes and 40 transcriptomes, 14 transcriptomes were generated for this study based on cultured strains or freshly collected specimens (*SI Appendix, Material and Methods*), while the remaining data were retrieved from publicly available repositories (*SI Appendix, Table S1*). RNA extractions follow (74). RNA quality and quantity were assessed with Qubit and Nanodrop spectrophotometer, and integrity was assessed with a Bioanalyzer 2100. RNA-seq libraries were sequenced as reported in Table S2.

Transcriptome assembly, frameshift error correction and ORF detection. At the time of the experiment, pre-assembled transcriptomes were only available for *Acrosiphonia* sp., *Blastophysa rhizopus* and *Caulerpa taxifolia* (Table S1). All remaining assemblies were performed in house, starting from the raw reads using a custom semi-automated pipeline (*SI Appendix, Material and Methods*). For each of the 40 transcriptomes, transcripts were clustered with CD-HIT-EST v. 4.6.1 (75) with a similarity cut-off of 97.5%, and only the longest transcript was retained for downstream analysis as representative of the cluster. After taxonomic profiling of the transcripts (*SI Appendix, Material and Methods*), only eukaryotic transcripts were retained for downstream analysis, bacterial transcripts and transcripts lacking sequence similarity to known proteins were discarded.

Transcripts with putative frameshift errors were identified after initial processing in TRAPID (76), and transcripts carrying a putative frameshift error were corrected with FrameDP 1.2.2 (77), using the *Chlamydomonas* proteome as reference to guide the frameshift correction step (*SI Appendix, Material and Methods*). Each transcript was translated into the corresponding amino acid sequences with the transeq algorithm from the EMBOSS package, using the appropriate translation table, and added to the proteome data of the 15 genomes, resulting in 1,228,821 amino acid sequences.

Gene Family inference. Sequences were used to build a custom PLAZA 4.0 instance (78), and single-copy families were selected by identifying the 620 picoPLAZA single-copy genes (79) (*SI Appendix, Material and Methods*). To remove potential contaminants in the transcriptomic data from the single-copy gene families, only sequences that were classified as “Viridiplantae” after an additional sequence similarity search with the GhostKOALA (80) webserver were retained for downstream analyses. To further reduce the residual redundancy of the transcriptome datasets in the single-copy gene families, for each gene family the nucleotide sequences of each species were collapsed with CAP3 (81), using stringent parameters to avoid artefactual creation of chimeras: gap penalty 12 (-g) and overlap percent identity cutoff 98% (-p). This set of 620 quasi single-copy genes was used for the downstream phylogenetic analyses.

Sequence alignments and filtering. Amino acid sequences of the 620 single-copy genes were aligned with MAFFT v. 7.187 (82), using accuracy-oriented parameters (--localpair --maxiterate 1,000) and an offset value (--ep) of 0.075. To identify and trim eventual residual in-paralogs, we followed a phylogeny-guided approach, in which alignments were manually curated to retain only full length or fragments of orthologous single-copy genes (*SI Appendix, Material and Methods*). After filtering, 539 high-confidence single-copy genes out of the 620 initial quasi single-copy genes were retained (hereafter referred to as “coreGF”). The coreGF dataset formed the basis for the construction of seven derived datasets based on different filtering approaches (*SI Appendix, Material and Methods, Table S3*).

Phylogenetic Analysis. The eight datasets were analyzed using a supermatrix and a coalescence-based phylogenetic approach. ML supermatrix analyses were performed using IQtree with two settings: 1) a gene-wise partitioned analysis (83) was performed, assigning the best substitution model inferred to each partition; 2) an analysis using mixture models was performed using an LG+F+G plus a C20-profile mixture model of substitution rates (84). To estimate the best substitution model of each partition, ML trees were built with IQtree for each single-copy gene, inferring the best model and rate of heterogeneity across sites. All ML analyses were run with 1,000 ultra-fast bootstrap and SH-aLRT branch test replicates.

Gene trees were used also for the coalescent-based analyses, using ASTRAL v. 5.6.1 (85). First, for each ML gene tree, low support branches (ultra fast BS support < 10) were collapsed with Newick Utilities v. 1.6 (86). Branches contracted in the ML gene trees were removed as well from the pool of the corresponding 1,000 bootstrap trees generated during the ML reconstruction. Then, two independent runs were performed either using the ML tree for each gene (BestML), or using the multilocus bootstrap support (MLBS) approach. For the MLBS analysis, 100 replicates were run (-r)

starting from the 1,000 contracted bootstrap trees for each gene, allowing gene and site resampling (--gene-resampling flag).

Statistical tests for rejecting the null hypothesis of polytomies were performed in ASTRAL (-t 10 flag), following (87). Briefly, for each of the eight datasets, the coalescence-based MLBS tree was tested (-q flag) by random sampling subsets of ML gene trees representing 1-100% of the total gene families in the dataset, with a minimum of 20 ML gene trees per subset. For each dataset and for each subset, ten independent replicates were generated, and analyzed with ASTRAL on BestML mode with the -t 10 flag -q flag to score the MLBS trees. The support for a number of key relationships was analyzed in each subset of the datasets and the median of the p-values for each subset of trees was calculated.

Different topologies for a number of key relationships were tested using Approximately Unbiased (AU) tests (88) implemented in IQtree, with 100,000 RELL resamplings (89). In addition to the AU test, the gene-wise log-likelihood scores and the percentage of genes supporting alternative topologies were calculated as outlined by (90) and (42).

Calibrated phylogenetic tree. Due to the high computational cost, the molecular clock analysis was restricted to the 539 coreGF scaffold trim data set. Clock-likeness of each gene was assessed with the package SortDate (91) against the ML supermatrix and the coalescence-based topologies, scoring the trees on minimal conflict, low root-to-tip variance, and discernible amounts of molecular evolution. The 10 most clock-like genes for each topology were concatenated and subjected to relaxed clock analyses (2,806 and 2,857 amino acid residues for the ML supermatrix and the coalescence-based topology, respectively). Node calibrations were transferred from fossil information and from node age estimates from previous studies (Table S5). All analyses were run with the same set of calibration nodes, except for the UB (*Proterocladus*) and the RT (root age, i.e.: Streptophyta-Chlorophyta split) nodes. The UB node was either constrained (UB₁) or unconstrained in time (UB₀). For the root age, three different priors were used (RT₁-RT₃), or the root was left unconstrained (RT₀) (Table S5).

Relaxed molecular clock analyses were run with PhyloBayes 4.1b (92). Two sets of analyses were run on two fixed topologies (the ML supermatrix and the coalescence-based topologies), using the set of clock-like genes for each topology. Both lognormal autocorrelated clock (-ln flag) and uncorrelated gamma multiplier clock (-ugam flag) models were tested for each dataset, and the models were run with either LG+Γ4 or CATGTR+Γ4 models of amino acid substitutions. In total, 64 different analyses were run (Table S6) to test the influence of different models, and of root and key prior ages on the age estimations. For each analysis, two distinct MCMC chains were run for at least 10,000

generations. The convergence of the log likelihoods and parameters estimates were tested in PhyloBayes. Chains were summarized after discarding the first 750 generations as burn-in.

The ultrametric trees were used to guide the ancestral state reconstruction of the ecological and cyto-morphological traits in Phytools (93). The posterior probabilities of the ancestral state of each node were calculated from summaries of 1,000 replicates of simulated stochastic character map (make.simmap), using empirical Bayes method under the ADR model, which permits backward and forward rates between states to have different values.

Acknowledgements

This work was supported by Ghent University (BOF/01J04813) with infrastructure funded by EMBC Belgium—FWO project GOH3817N (to O.D.C.), the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. H2020-MSCA-ITN-2015-675752, the Australian Research Council (DP150100705 to H.V.), and the National Science Foundation (GRAToL 10136495 to C.F.D.). We thank Endymion Cooper for help with generating transcriptomic data. The University of Dundee is a registered Scottish charity, No. 050196. Part of this work has been included and presented in the PhD dissertation of A.D.C. (94).

References

1. N. J. Butterfield, *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).
2. S. Bengtson, T. Sallstedt, V. Belivanova, M. Whitehouse, Three-dimensional preservation of cellular and subcellular structures suggests 1.6 billion-year-old crown-group red algae. *PLoS Biol.* **15**, e2000735 (2017).
3. C. Berney, J. Pawlowski, A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. *Proc. R. Soc. B* **273**, 1867-1872 (2006).
4. J. W. Brown, U. Sorhannus, A molecular genetic timescale for the diversification of autotrophic stramenopiles (Ochrophyta): substantive underestimation of putative fossil ages. *PLoS One* **5**, e12759 (2010).
5. P. Sánchez-Baracaldo, J. A. Raven, D. Pisani, A. H. Knoll, Early photosynthetic eukaryotes inhabited low-salinity habitats. *Proc. Natl Acad. Sci. U.S.A.* **114**, E7737-E7745 (2017).
6. C. Jackson, A. H. Knoll, C. X. Chan, H. Verbruggen, Plastid phylogenomics with broad taxon sampling further elucidates the distinct evolutionary origins and timing of secondary green plastids. *Sci. Rep.* **8**, 1523 (2018).
7. R. M. McCourt, C. F. Delwiche, K. G. Karol, Charophyte algae and land plant origins. *Trends Ecol. Evol.* **19**, 661-666 (2004).
8. P. Kenrick, P. R. Crane, The origin and early evolution of plants on land. *Nature* **389**, 33-39 (1997).

9. S. M. Porter, "The fossil record of early eukaryotic diversification" in *The Paleontological Society Papers 10: Neoproterozoic-Cambrian Biological Revolutions*, J. H. Lipps, B. Waggoner, Eds. (Paleontological Society, New Haven, 2004), pp. 35-50.
10. A. H. Knoll, E. J. Javaux, D. Hewitt, P. Cohen, Eukaryotic organisms in Proterozoic oceans. *Phil. Trans. R. Soc. B* **361**, 1023-1038 (2006).
11. F. Leliaert *et al.*, Phylogeny and molecular evolution of the green algae. *Crit. Rev. Plant Sci.* **31**, 1-46 (2012).
12. P. Škaloud, F. Rindi, C. Boedeker, F. Leliaert, *Freshwater Flora of Central Europe, Vol 13: Chlorophyta: Ulvophyceae*, Süßwasserflora von Mitteleuropa (Springer Spektrum, Berlin, Heidelberg, 2018), vol. 13, pp. 288.
13. E. Cocquyt, H. Verbruggen, F. Leliaert, O. De Clerck, Evolution and cytological diversification of the green seaweeds (Ulvophyceae). *Mol. Biol. Evol.* **27**, 2052-2061 (2010).
14. F. Leliaert *et al.*, Chloroplast phylogenomic analyses reveal the deepest-branching lineage of the Chlorophyta, Palmophyllophyceae class. nov. *Sci. Rep.* **6**, 25367 (2016).
15. I. Mine, D. Menzel, K. Okuda, Morphogenesis in giant-celled algae. *Int. Rev. Cell Mol. Biol.* **266**, 37-83 (2008).
16. J. A. Raven, C. A. Knight, J. Beardall, Genome and cell size variation across algal taxa. *Perspect. Phycol.* 10.1127/pip/2019/0085, in press (2019).
17. A. Del Cortona, F. Leliaert, Molecular evolution and morphological diversification of ulvophytes (Chlorophyta). *Perspect. Phycol.* **5**, 27-43 (2018).
18. L. Fang, F. Leliaert, Z. H. Zhang, D. Penny, B. J. Zhong, Evolution of the Chlorophyta: Insights from chloroplast phylogenomic analyses. *J. Syst. Evol.* **55**, 322-332 (2017).
19. M. Turmel, C. Otis, C. Lemieux, Divergent copies of the large inverted repeat in the chloroplast genomes of ulvophycean green algae. *Sci. Rep.* **7**, 994 (2017).
20. H. Tappan, *Palaeobiology of Plant Protists* (Freeman, San Francisco, 1980), pp. 1028.
21. C. Loron, M. Moczyłowska, Tonian (Neoproterozoic) eukaryotic and prokaryotic organic-walled microfossils from the upper Visingsö Group, Sweden. *Palynology* **42**, 220-254 (2018).
22. M. Moczyłowska, Algal affinities of Ediacaran and Cambrian organic-walled microfossils with internal reproductive bodies: Tanarium and other morphotypes. *Palynology* **40**, 83-121 (2016).
23. E. J. Javaux, A. H. Knoll, Micropaleontology of the lower Mesoproterozoic Roper Group, Australia, and implications for early eukaryotic evolution. *J. Paleontol.* **91**, 199-229 (2017).
24. N. J. Butterfield, Modes of pre-Ediacaran multicellularity. *Precambrian Res.* **173**, 201-211 (2009).
25. N. J. Butterfield, A. H. Knoll, K. Swett, Paleobiology of the Neoproterozoic Svanbergfjellet Formation, Spitsbergen. *Fossils and Strata* **34**, 1-84 (1994).
26. L. E. Graham, Digging deeper: why we need more Proterozoic algal fossils and how to get them. *J. Phycol.* **55**, 1-6 (2019).
27. G. K. Colbath, H. R. Grenfell, Review of biological affinities of Paleozoic acid-resistant, organic-walled eukaryotic algal microfossils (including "acritarchs"). *Rev. Palaeobot. Palynol.* **86**, 287-314 (1995).
28. K. R. Arouri, P. F. Greenwood, M. R. Walter, Biological affinities of Neoproterozoic acritarchs from Australia: microscopic and chemical characterisation. *Org. Geochem.* **31**, 75-89 (2000).
29. S. T. LoDuca, New Ordovician marine macroalgae from North America, with observations on *Buthograptus*, *Callithamnopsis*, and *Chaetocladus*. *J. Paleontol.* **93**, 197-214 (2018).
30. S. LoDuca, N. Bykova, M. Wu, S. Xiao, Y. Zhao, Seaweed morphology and ecology during the great animal diversification events of the early Paleozoic: a tale of two floras. *Geobiol.* **15**, 588-616 (2017).
31. H. Verbruggen *et al.*, A multi-locus time-calibrated phylogeny of the siphonous green algae. *Mol. Phylogenet. Evol.* **50**, 642-653 (2009).
32. D. F. Satterthwait (1976) Paleobiology and paleoecology of Middle Cambrian algae from western North America. (University of California, Los Angeles), p 120.

33. S. Conway Morris, R. A. Robison, More soft-bodied animals and algae from the Middle Cambrian of Utah and British Columbia. *Kansas Univ. Paleont. Contr.* **122**, 1-48 (1988).
34. J. J. Brocks *et al.*, The rise of algae in Cryogenian oceans and the emergence of animals. *Nature* **548**, 578 (2017).
35. G. D. Love *et al.*, Fossil steroids record the appearance of Demospongiae during the Cryogenian period. *Nature* **457**, 718 (2009).
36. Y. Hoshino *et al.*, Cryogenian evolution of stigmasteroid biosynthesis. *Sci. Adv.* **3**, e1700887 (2017).
37. B. Zhong, L. Liu, Z. Yan, D. Penny, Origin of land plants using the multispecies coalescent model. *Trends Plant Sci.* **18**, 492-495 (2013).
38. N. J. Wickett *et al.*, Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proc. Natl Acad. Sci. U.S.A.* **111**, E4859-E4868 (2014).
39. K. Fučíková *et al.*, New phylogenetic hypotheses for the core Chlorophyta based on chloroplast sequence data. *Front. Ecol. Evol.* **2**, 63 (2014).
40. M. Turmel, C. Lemieux, Evolution of the plastid genome in green algae. *Adv. Bot. Res.* **85**, 157-193 (2018).
41. D. A. Pollard, V. N. Iyer, A. M. Moses, M. B. Eisen, Widespread discordance of gene trees with species tree in *Drosophila*: evidence for incomplete lineage sorting. *PLoS Genet.* **2**, e173-e173 (2006).
42. X.-X. Shen, C. T. Hittinger, A. Rokas, Contentious relationships in phylogenomic studies can be driven by a handful of genes. *Nat. Ecol. Evol.* **1**, 0126 (2017).
43. E. K. Molloy, T. Warnow, To include or not to include: The impact of gene filtering on species tree estimation methods. *Syst. Biol.* **67**, 285-303 (2018).
44. S. Roch, M. Steel, Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. *Theor. Popul. Biol.* **100**, 56-62 (2015).
45. S. Mirarab, M. S. Bayzid, T. Warnow, Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. *Syst. Biol.* **65**, 366-380 (2014).
46. E. Cocquyt *et al.*, Complex phylogenetic distribution of a non-canonical genetic code in green algae. *BMC Evol. Biol.* **10**, 327 (2010).
47. G. H. Gile, P. M. Novis, D. S. Cragg, G. C. Zuccarello, P. J. Keeling, The distribution of Elongation Factor-1 alpha (EF-1 α), Elongation Factor-Like (EFL), and a non-canonical genetic code in the Ulvophyceae: discrete genetic characters support a consistent phylogenetic framework. *J. Eukaryot. Microbiol.* **56**, 367-372 (2009).
48. A. H. Knoll, The multiple origins of complex multicellularity. *Annu. Rev. Earth Planet. Sci.* **39**, 217-239 (2011).
49. A. H. Knoll, Paleobiological perspectives on early eukaryotic evolution. *Cold Spring Harb. Perspect. Biol.* **6**, a016121 (2014).
50. Q. Ye *et al.*, The survival of benthic macroscopic phototrophs on a Neoproterozoic snowball Earth. *Geology* **43**, 507-510 (2015).
51. M. Dos Reis *et al.*, Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. *Curr. Biol.* **25**, 2939-2950 (2015).
52. F. Leliaert, O. De Clerck, H. Verbruggen, C. Boedeker, E. Coppejans, Molecular phylogeny of the Siphonocladales (Chlorophyta: Cladophorophyceae). *Mol. Phylogenet. Evol.* **44**, 1237-1256 (2007).
53. P. F. Hoffman *et al.*, Snowball Earth climate dynamics and Cryogenian geology-geobiology. *Sci. Adv.* **3**, e1600983 (2017).
54. S. Xiao, Q. Tang, After the boring billion and before the freezing millions: evolutionary patterns and innovations in the Tonian Period. *Emerg. Top. Life. Sci.* **2**, 161-171 (2018).
55. T. W. Lyons, C. T. Reinhard, N. J. Planavsky, The rise of oxygen in Earth's early ocean and atmosphere. *Nature* **506**, 307-315 (2014).

56. A. H. Knoll, D. J. Lahr, "Fossils, feeding, and the evolution of complex multicellularity" in *Multicellularity, Origins and Evolution. The Vienna Series in Theoretical Biology.*, K. J. Niklas, S. A. Newman, Eds. (MIT Press, Cambridge, MA, 2016), pp. 1-16.
57. P. A. Cohen, L. A. Riedman, It's a protist-eat-protist world: recalcitrance, predation, and evolution in the Tonian–Cryogenian ocean. *Emerg. Top. Life. Sci.* **2**, 173-180 (2018).
58. M. E. Boraas, D. B. Seale, J. E. Boxhorn, Phagotrophy by a flagellate selects for colonial prey: a possible origin of multicellularity. *Evol. Ecol.* **12**, 153-164 (1998).
59. J. C. Coates, Umm-E-Aiman, B. Charrier, Understanding "green" multicellularity: do seaweeds hold the key? *Front. Plant Sci.* **5** (2015).
60. P. A. Cohen, F. A. Macdonald, The Proterozoic record of eukaryotes. *Paleobiology* **41**, 610-632 (2015).
61. A. H. Knoll, *Life on a Young Planet: The First Three Billion Years of Evolution on Earth* (Princeton University Press, Princeton, NJ, 2003), pp. 277.
62. S. Xiao, X. Yuan, M. Steiner, A. H. Knoll, Macroscopic carbonaceous compressions in a terminal Proterozoic shale: a systematic reassessment of the Miaohu biota, South China. *J. Paleontol.* **76**, 347-376 (2002).
63. S. H. Xiao, A. H. Knoll, X. L. Yuan, C. M. Pueschel, Phosphatized multicellular algae in the Neoproterozoic Doushantuo Formation, China, and the early evolution of florideophyte red algae. *Am. J. Bot.* **91**, 214-227 (2004).
64. X. Yuan, Z. Chen, S. Xiao, C. Zhou, H. Hua, An early Ediacaran assemblage of macroscopic and morphologically differentiated eukaryotes. *Nature* **470**, 390-393 (2011).
65. B. Shen, L. Dong, S. Xiao, M. Kowalewski, The Avalon explosion: evolution of Ediacara morphospace. *Science* **319**, 81-84 (2008).
66. R. Wood *et al.*, Integrated records of environmental change and evolution challenge the Cambrian Explosion. *Nat. Ecol. Evol.* **3**, 528-538 (2019).
67. P. J. Cook, J. H. Shergold, *Phosphate deposits of the world: Volume 1: Proterozoic and Cambrian phosphorites* (Cambridge University Press, Cambridge, UK, 1986), pp. 408
68. A. H. Knoll, Food for early animal evolution. *Nature* **548**, 528-530 (2017).
69. S. M. Stanley, An ecological theory for the sudden origin of multicellular life in the late Precambrian. *Proc. Natl Acad. Sci. U.S.A.* **70**, 1486-1489 (1973).
70. J. Raven, Nutritional strategies of submerged benthic plants: the acquisition of C, N and P by rhizophytes and haptophytes. *New Phytol.* **88**, 1-30 (1981).
71. J. Raven, Blue carbon: past, present and future, with emphasis on macroalgae. *Biol. Lett.* **14**, 20180336 (2018).
72. J. J. Brocks, The transition from a cyanobacterial to algal world and the emergence of animals. *Emerg. Top. Life. Sci.* **2**, 181-190 (2018).
73. O. De Clerck *et al.*, Insights into the evolution of multicellularity from the Sea Lettuce genome. *Curr. Biol.* **28**, 2921-2933.e2925 (2018).
74. J. D. Palmer, Physical and gene mapping of chloroplast DNA from *Atriplex triangularis* and *Cucumis sativa*. *Nucleic Acids Res.* **10**, 1593-1605 (1982).
75. W. Li, A. Godzik, Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658-1659 (2006).
76. M. Van Bel *et al.*, TRAPID: an efficient online tool for the functional and comparative analysis of de novo RNA-Seq transcriptomes. *Genome Biol.* **14**, R134 (2013).
77. J. Gouzy, S. Carrere, T. Schiex, FrameDP: sensitive peptide detection on noisy matured sequences. *Bioinformatics* **25**, 670-671 (2009).
78. M. Van Bel *et al.*, PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res.* **46**, D1190-D1196 (2018).
79. K. Vandepoele *et al.*, pico-PLAZA, a genome database of microbial photosynthetic eukaryotes. *Environ. Microbiol.* **15**, 2147-2153 (2013).
80. M. Kanehisa, Y. Sato, K. Morishima, BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726-731 (2016).

81. X. Huang, A. Madan, CAP3: A DNA sequence assembly program. *Genome Res.* **9**, 868-877 (1999).
82. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30** (2013).
83. O. Chernomor, A. von Haeseler, B. Q. Minh, Terrace aware data structure for phylogenomic inference from supermatrices. *Syst. Biol.* **65**, 997-1008 (2016).
84. L. Si Quang, O. Gascuel, N. Lartillot, Empirical profile mixture models for phylogenetic reconstruction. *Bioinformatics* **24**, 2317-2323 (2008).
85. C. Zhang, M. Rabiee, E. Sayyari, S. Mirarab, ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics* **19**, 153 (2018).
86. T. Junier, E. M. Zdobnov, The Newick utilities: high-throughput phylogenetic tree processing in the Unix shell. *Bioinformatics* **26**, 1669-1670 (2010).
87. E. Sayyari, S. Mirarab, Testing for polytomies in phylogenetic species trees using quartet frequencies. *Genes* **9**, 132 (2018).
88. H. Shimodaira, An approximately unbiased test of phylogenetic tree selection. *Syst. Biol.* **51**, 492-508 (2002).
89. H. Kishino, T. Miyata, M. Hasegawa, Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *J. Mol. Evol.* **31**, 151-160 (1990).
90. Y. Chiari, V. Cahais, N. Galtier, F. Delsuc, Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). *BMC Biol.* **10**, 65 (2012).
91. S. A. Smith, J. W. Brown, J. F. Walker, So many genes, so little time: A practical approach to divergence-time estimation in the genomic era. *PLoS One* **13**, e0197433 (2018).
92. N. Lartillot, T. Lepage, S. Blanquart, PhyloBayes 3: a Bayesian software package for phylogenetic reconstruction and molecular dating. *Bioinformatics* **25**, 2286-2288 (2009).
93. L. J. Revell, phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217-223 (2012).
94. A. Del Cortona (2018) The rise of Algae: molecular evolution of macroscopic growth in green algae. in *Department of Biology, and Department of Plant Biotechnology and Bioinformatics* (Ghent University, Ghent).

Figures

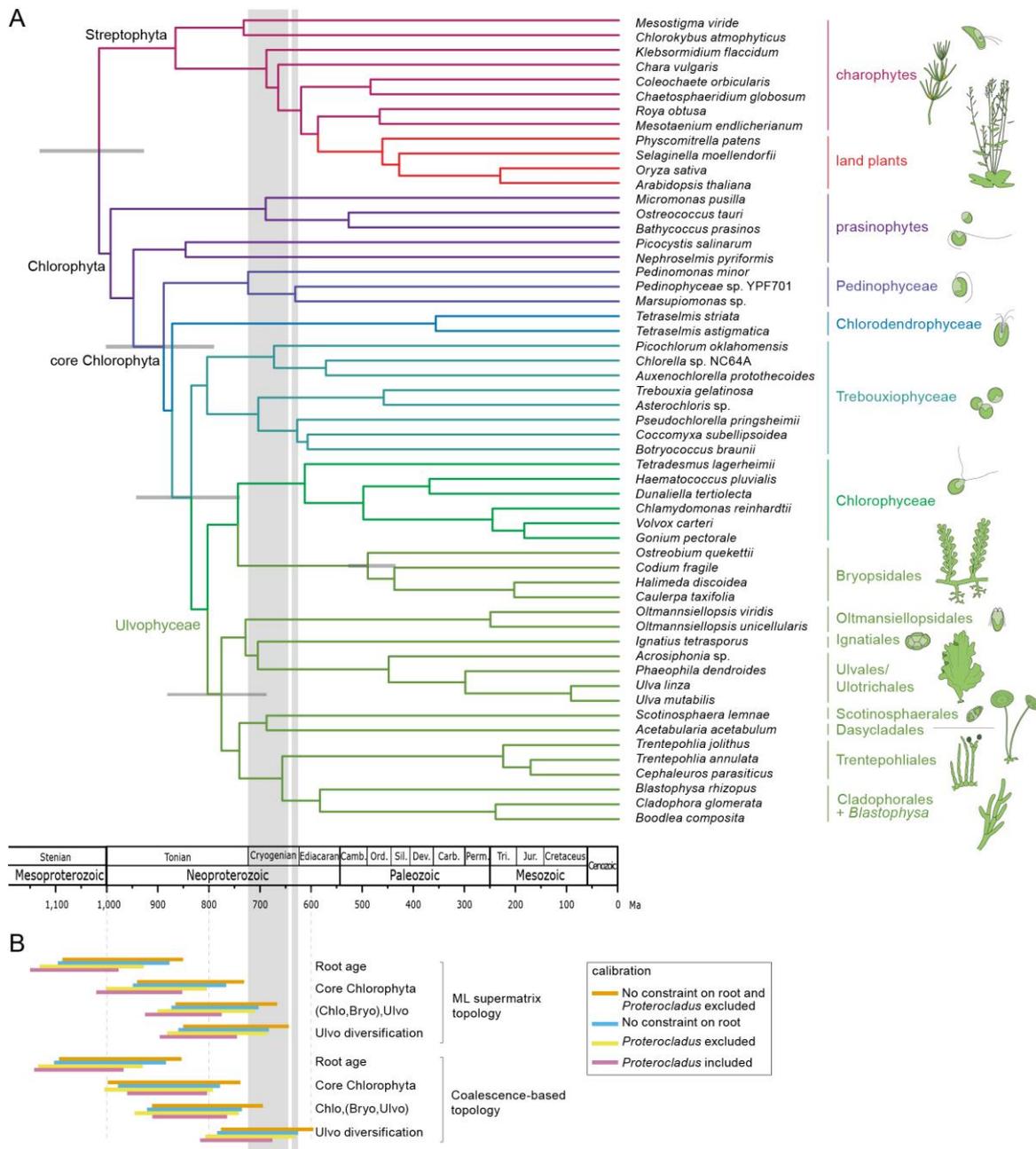


Figure 1. Time-calibrated phylogeny of the green algae. **A.** The topology of the tree is based on the maximum likelihood analysis inferred from a concatenated amino acid alignment of 539 nuclear genes (supermatrix analysis of the coreGF-scaffolded-untrimmed data set, Figure S2). Branch lengths are based on a relaxed molecular clock analysis of the ten most clock-like genes from the scaffold trimmed dataset, and excluding *Proterocladus* as calibration point. Error bars are indicated for a number of key nodes. **B.** Divergence time confidence intervals of key nodes inferred from different analyses (Figure S7, Table S6).

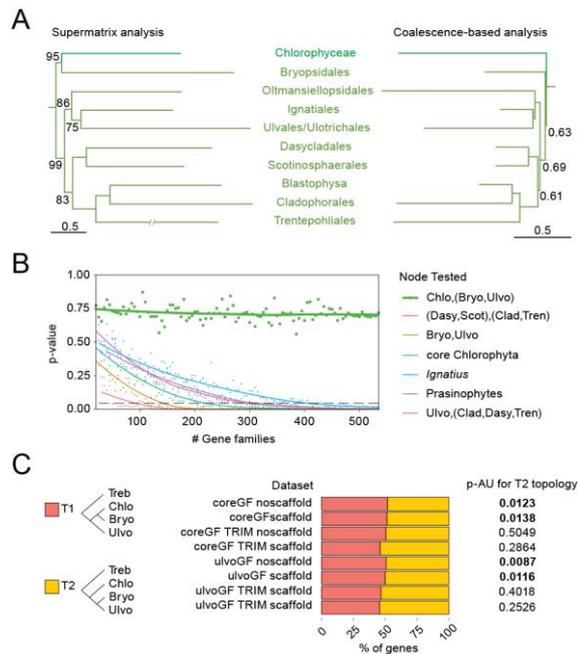


Figure 2. Relationships between Bryopsidales, remaining Ulvophyceae, and Chlorophyceae. **A.** Phylogeny (Chlorophyceae-Ulvophyceae clade pruned) as recovered by the supermatrix (Figure S2) and coalescence-based analysis. **B.** Test for polytomy null-hypothesis for selected relationships based on the coreGF scaffold trimmed dataset (see methods and Figure S6 for details). For the Chlorophyceae, Bryopsidales, and remaining Ulvophyceae (thick green line), increasing gene numbers did not reduce the p-value of rejection. Proportion of genes supporting a sister relationship between Bryopsidales and Chlorophyceae (T1), and a sister relationship between the Bryopsidales and remaining Ulvophyceae (T2), and summary of the AU-test for the constrained topology (T2) (see Table S4 for details). Bryo: Bryopsidales; Chld: Chlorodendrophyceae; Chlo: Chlorophyceae; Clad: Cladophorales+*Blastophysa*; Dasy: Dasycladales; Igna: Ignatiales; Oltm: Oltmansiellopsidales; Pedi: Pedinophyceae; Scot: Scotinosphaerales; Treb: Trebouxiophyceae; Tren: Trentepohliales; Ulvo: Ulvophyceae (excluding Bryopsidales).

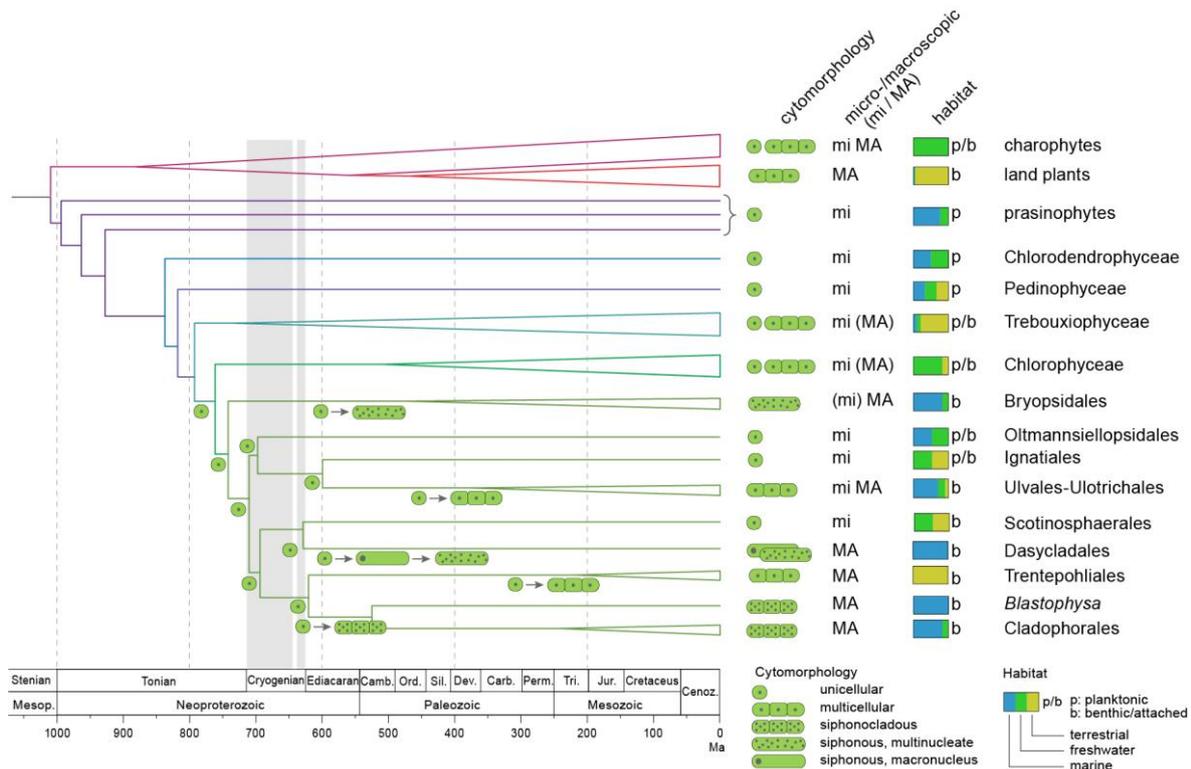


Figure 3. Hypothesis for the evolution of multicellularity and macroscopic growth, and transition to benthic marine habitats in the Ulvophyceae. The topology of the tree is based on the coalescence-based analysis of the coreGF-scaffolded-untrimmed data set. Branch lengths are based on a relaxed molecular clock analysis of the ten most clock-like genes from the coreGF-scaffolded-trimmed dataset, and excluding *Proterocladus* as calibration point.

Supplementary Information for

Neoproterozoic origin and multiple transitions to macroscopic growth in green seaweeds

Andrea Del Cortona, Christopher J. Jackson, François Bucchini, Michiel Van Bel, Sofie D'hondt, Pavel Škaloud, Charles F. Delwiche, Andrew H. Knoll, John A. Raven, Heroen Verbruggen, Klaas Vandepoele, Olivier De Clerck, Frederik Leliaert

Corresponding authors. Email andrea.delcortona@gmail.com, klaas.vandepoele@psb.vib-ugent.be, olivier.declerck@ugent.be or frederik.leliaert@meisebotanicgarden.be.

This PDF file includes:

Supplementary text 1: Material and methods

Supplementary text 2: Discussion on phylogenetic relationships within the core Chlorophyta

Figures S1 to S8

Tables S1 to S6

SI references

Supplementary text 1: Material and methods

Sampling and culture conditions. 14 transcriptomes were generated for this study, based on cultured or freshly collected specimens. *Acetabularia acetabulum*, *Oltmannsiellopsis viridis* and *Scotinosphaera lemnae* cultures were grown at 20 °C and 12-h light/12-h dark cycle in Ace-25 medium (1) and 1.5% agar-solidified Bold Basal medium (2), respectively. *Marsupiomonas* sp., *Pedinomonas minor* and the pedinophyte strain YPF-701 (NIES Microbial Culture Collection strain NIES-2566) were cultured in Guillard's F/2 medium at 20 °C and 14-h light/10-h dark cycle. For *Ostreobium* sp. HV05042, *Halimeda discoidea* and *Codium fragile* freshly collected material was used for RNA extraction. Collection details are given in (3). Unicellular microscopic algae cultures (*Marsupiomonas* sp., *Oltmannsiellopsis viridis*, *Pedinomonas minor*, pedinophyte strain YPF-701 and *Scotinosphaera lemnae*) were harvested during their exponential phase. Whole macroscopic seaweed specimens were harvested from cultures (*Acetabularia acetabulum*) or collected in their natural environment (*Codium fragile*, *Halimeda discoidea* and *Ostreobium* sp. HV05042) and ground in liquid nitrogen for RNA extraction.

Transcriptome assembly and taxonomic profiling. Assemblies were performed from the raw reads using a custom semi-automated pipeline, consisted of the following steps. Quality of the raw reads were assessed with FastQC v.0.10.1 (<http://www.bioinformatics.babraham.ac.uk>, last accessed March 1, 2019). Low-quality reads (average Phreds quality score below 20) and low quality read ends were trimmed with Fastx v.0.0.13 (https://github.com/agordon/fastx_toolkit, last accessed March 1, 2019). Trimmed reads shorter than 30 bp were discarded. Transcriptome *de novo* assembly of *Botryococcus braunii*, *Chlorokybus atmophyticus*, *Ulva linza* (Roche 454 data) were performed with CLC Genomics Workbench version 7.5.1 (<http://www.clcbio.com>, last accessed on March 1, 2019), using a word size of 63 and standard parameters. Transcriptome *de novo* assembly for the remaining species were performed with Trinity 2.1.1 (4), in SE or PE mode were appropriate depending on the RNA-seq library type, after *in silico* reads normalization.

Taxonomic profiling of the transcripts was performed using the following protocol: first, the transcripts were compared to the NCBI non-redundant protein database by sequence similarity searches, using Tera-BLAST DeCypher (Active Motif, USA); then, for each transcript, sequence similarity searches were combined with the NCBI Taxonomy information of the top ten BLAST hits in order to discriminate between eukaryotic and bacterial transcripts or transcripts lacking similarity to known protein-coding genes.

Frameshift errors correction and ORF detection. Transcripts with putative frameshift errors were identified after initial processing in TRAPID (5), using the *Chlamydomonas reinhardtii* proteome as reference database for the sequence similarity searches. Transcripts carrying a putative frameshift error were corrected with FramedP 1.2.2 (6), using the *Chlamydomonas* proteome as reference to guide the frameshift correction step. The longest coding frame was detected with a custom java script plugged into TRAPID using four different translation tables (1, 6, 11 and 16), to take into account alternative nuclear genetic codes (7). Briefly, for each transcript coding frames were predicted with each of the four translation tables. Then, for each transcript, the longest coding sequences detected were retained for the downstream analysis, and the corresponding translation table recorded. Concordance with orthologous sequences predicted by the TRAPID pipeline by sequence similarity searches was also taken into account.

Gene Family inference. Sequences were used to build a custom PLAZA 4.0 instance (8) as follows. All-against-all sequence similarity comparison was executed with DIAMOND v. 0.9.18 (9), and the similarity matrix was used to infer homologous relationships among proteins using the graph-based Markov clustering method implemented in Tribe-MCL v. 10-201 (10) with parameters: -scheme 4 -l 2. This resulted in the clustering of 976,181 protein sequences (79.5% of the total proteins) into 69,462 gene families, leaving 252,640 singleton proteins. A procedure was applied to identify and flag outlier proteins from gene families and subfamilies if they showed similarity only to a minority of all family members (11). Single-copy families were selected by identifying the 620 picoPLAZA

single-copy genes (12). At this point, to remove potential contaminants in the transcriptomic data from the single-copy gene families, only sequences that were classified as “Viridiplantae” after an additional sequence similarity search with the GhostKOALA (13) webserver were retained for downstream analyses. To further reduce the residual redundancy of the transcriptome datasets in the single-copy gene families, for each gene family the nucleotide sequences of each species were collapsed with CAP3 (14), using stringent parameters to avoid artefactual creation of chimeras: gap penalty 12 (-g) and overlap percent identity cutoff 98% (-p). This set of 620 quasi single-copy genes was used for the downstream phylogenetic analyses.

Sequence alignment: identification and trimming of residual in-paralogs. To identify and trim eventual residual in-paralogs, we followed a phylogeny-guided approach. As a first step, a Maximum Likelihood (ML) tree for the resulting alignments was obtained with IQtree v. 1.6.0 (15), inferring the best model and allowing invariable sites and free rate of heterogeneity across sites (16, 17), with parameters: number of ultra-fast bootstrap replicates (-bb) 1,000 (18) and SH-aLRT branch test (19) with 1,000 replicates (-alrt). The resulting trees were visualized in MEGA v. 6 (20) and the corresponding alignments were inspected and processed in Geneious v. 8.0.5 (Biomatters Ltd., <https://www.geneious.com/>).

Alignments were manually curated to retain only full length or fragments of orthologous single-copy genes: (1) except for *Mesostigma viride* and *Chlorokybus atmophyticus*, which were allowed to cluster with either the Streptophyta or the Chlorophyta, Streptophyta and Chlorophyta monophyly was enforced for the remaining species (e.g. *Arabidopsis thaliana* sequences clustering within the Chlorophyta would be excluded) as obvious sign of paralogy or contamination of environmental samples. (2) For a species with two or more overlapping sequence, one in the expected phylogenetic position (according to orthologous sequences of other species in the same taxon), the other one in a conflicting position, the conflicting sequences were removed. In case of overlapping sequences with concordant phylogenetic signal, regardless of their phylogenetic position, but always enforcing Streptophyta and Chlorophyta monophyly, the longest one was retained, if full-length, otherwise both were retained. In case of conflicting but non-overlapping sequences, both sequences were retained and scaffolded. (3) Gene family alignments with sequences from less than 30 species (ca. half of the total number of species) were discarded. (4) Alignments composed by two or more near-identical paralogs (e.g.: ribosomal subunit proteins) and where confident segregation of the paralogs was difficult were discarded.

Sequence alignment: construction of seven derived datasets. The coreGF dataset formed the basis for the construction of seven derived datasets based on different filtering approaches (Table S3): coreGF unscaffolded, composed by 539 single-copy genes, with partial sequences removed; coreGF scaffolded, composed by 539 single-copy genes, with partial sequences scaffolded; coreGF unscaffolded TRIM, as coreGF unscaffolded, but with less conserved regions filtered; coreGF scaffolded TRIM: as coreGF scaffolded, but with less conserved regions filtered; ulvoGF unscaffolded, composed by a 355 single-copy genes subset of coreGF focussing on Ulvophyceae, with partial sequences removed; ulvoGF scaffolded, composed by a 355 single-copy genes subset of coreGF focussing on Ulvophyceae, with partial sequences scaffolded; ulvoGF unscaffolded TRIM, as ulvoGF unscaffolded, but with less conserved regions filtered; and ulvoGF scaffolded TRIM, as ulvoGF scaffolded, but with less conserved regions filtered.

Poorly aligned regions were removed to obtain the corresponding trimmed datasets by trimming the amino acid alignments with TrimAl v. 1.2 (21), with parameters: gapthreshold 0.75 and simthreshold 0.001. Partial sequences from the transcriptomes were either scaffolded (scaffolded dataset) or removed (unscaffolded dataset), resulting in a more comprehensive and a more conservative version of the coreGF and ulvoGF datasets, respectively. Briefly, each gene family with more than one sequence per species (398 gene families), was processed independently for orthology-guided scaffolding with a custom script. First, an all-against-all sequence similarity search was performed with BLASTp v. 2.5.0+ (22), using an e-value cut-off of $10e^{-5}$. The sequence with the highest bitscore was selected as reference for that gene family. Then, for each species with more than one sequence,

the sequences were concatenated according to their relative position to the reference sequence, based on the BLASTp alignments. This resulted in gene families with only one sequence per species. This approach resulted in two types of alignments: a conservative “unscaffold” dataset, where within each gene family, species with more than one sequence were discarded, and a more comprehensive but potentially noisier “scaffold” dataset, where multiple sequences of the same species were scaffolded. Unscaffold and scaffold datasets were re-aligned with MAFFT, using accuracy-oriented parameters (--localpair --maxiterate 1,000 --ep 0.075).

Supplementary information text 2: Discussion on phylogenetic relationships within the core Chlorophyta

Phylogenetic positions of Pedinophyceae and Chlorodendrophyceae. Our phylogenetic analyses recovered the Chlorodendrophyceae and Pedinophyceae as the two earliest diverging lineages of the core Chlorophyta, consistent with some previous phylogenetic analyses based on nuclear- and plastid-encoded rDNA sequences (23) and chloroplast phylogenomic analyses (3, 24, 25). However, the exact phylogenetic position (Chlorodendrophyceae first, or Pedinophyceae first) differed between analyses (supplementary figures). The phylogenetic position of these two lineages, relative to the remaining core chlorophytan lineages is relevant in light of evolution of cell division. In contrast to the prasinophytes where cell division occurs by furrowing, three classes of the core Chlorophyta, including the Chlorophyceae, Trebouxiophyceae and Chlorodendrophyceae, are characterized by a new mode of cell division that is mediated by a phycoplast, which is an array of microtubules oriented parallel to the plane of cell division, determining the formation of a new cell wall (26, 27). A phycoplast is absent in the Pedinophyceae and Ulvophyceae (Trentepohliales have a phragmoplast-like cytokinesis, similar to that found in some Streptophyta). Since our phylogenetic analyses are inconclusive with respect to the earliest diverging lineage of the core Chlorophyta (Pedinophyceae or Chlorodendrophyceae) it remains uncertain whether the phycoplast evolved at the base of the core Chlorophyta and was subsequently lost in the Pedinophyceae, or evolved after the divergence of the Pedinophyceae.

Phylogenetic relationships between the Ulvophyceae, Trebouxiophyceae and Chlorophyceae (UTC). The Trebouxiophyceae and Chlorophyceae were unambiguously recovered as monophyletic groups, with the Trebouxiophyceae sister to a clade containing the Chlorophyceae and Ulvophyceae. Previous phylogenetic studies have not been able to resolve these relationships unambiguously (27). Ultrastructural evidence has been interpreted as either providing support for a sister relation between Chlorophyceae and Trebouxiophyceae based on the shared presence of a phycoplast and a non-persistent mitotic spindle (26), or between Trebouxiophyceae and Ulvophyceae based on a counterclockwise orientation of the flagellar basal bodies (28). Phylogenies based on nuclear rDNA sequences have been inconclusive, with the branching order of the UTC classes depending on taxon sampling, alignment methods and phylogenetic methods. Some early 18S rDNA phylogenies showed a sister relationship between Chlorophyceae and Trebouxiophyceae (e.g., 29) while others revealed a sister relation between the Chlorophyceae and Ulvophyceae (e.g., 30). A 10-gene (2 plastid and 8 nuclear genes) phylogeny (31) showed a sister relationship between the Chlorophyceae and Ulvophyceae, indicating that the ancestor of the UTC clade had a phycoplast and a non-persistent mitotic spindle, and that both characters have been lost in the Ulvophyceae. The ancestral status of a phycoplast and a non-persistent mitotic spindle is congruent with the presence of these features in the early diverging Chlorodendrophyceae (26). Likewise the ancestor of the UTC clade probably had a counterclockwise orientation of the flagellar basal bodies, evolving to a direct opposite or clockwise orientation in the Chlorophyceae. Again this interpretation is congruent with the presence of a counterclockwise basal body orientation in the Chlorodendrophyceae. Chloroplast phylogenomic analyses have not been able to resolve the branching order of the UTC classes

conclusively, with the main uncertainties including monophyly of the Trebouxiophyceae and Ulvophyceae (see below), and the sister clade of the Chlorophyceae (25, 32).

Monophyly of the Ulvophyceae and phylogenetic position of the Bryopsidales. The distinct phylogenetic position of the Bryopsidales in our analyses is unexpected from a morphological viewpoint. The order has long been allied with the Dasycladales based on the shared siphonous structure, and this relationship has been confirmed by small nuclear gene datasets (30, 31, 33), while in chloroplast phylogenomic studies, the position of the Bryopsidales has been unstable (32, 34, 35). Conversely, ultrastructural features of the flagellar apparatus suggested a close relationship between the Dasycladales and Cladophorales. The two orders share a flattened flagellar apparatus, have similar striations on the distal fibers connecting the basal bodies, and lack terminal caps on the anterior surface of the basal bodies (36). In assuming a sister relationship between the Bryopsidales and Dasycladales based on analysis of a 10-gene dataset, Cocquyt et al. (31) hypothesised that the evolution of siphonous growth would have involved two steps, including the development of an enlarged ancestral cell with a macronucleus and cytoplasmic streaming, allowing increased transcription from a single nucleus and distribution of transcripts across the cell, followed by the evolution of multinucleate siphons, which possibly occurred independently in the Bryopsidales and Dasycladales in association with the evolution of complex, macroscopic plants that have outgrown the potential of a single macronucleus. Our phylogenetic results point towards an entirely independent evolution of siphonous organization in Bryopsidales and Dasycladales from an ancestral unicellular green alga. The distinct phylogenetic position of the Bryopsidales, separate from the Trentepohliales, Cladophorales, and Dasycladales is supported by independent molecular data. The latter orders, along with the Scotinosphaerales and *Blastophysa*, are characterized by a non-canonical nuclear genetic code while the Bryopsidales and all other green algae have the canonical code (7, 37), pointing towards a single origin of a non-canonical nuclear code in the green algae. The Bryopsidales and Dasycladales also present fundamental differences in life cycle and cytology, for example in their approaches towards cytoplasmic streaming (microtubules and actin versus actin-only-powered, respectively) (38-43), supporting a more distant relationship between the two orders.

Phylogenetic position of the microscopic ulvophycean orders: Ignatiales, Oltmansiellopsidales and Scotinosphaerales. The phylogenetic affinities of the microscopic and mostly unicellular ulvophycean orders Ignatiales (including non-motile unicellular or loose colonial algae living on soil or freshwater macrophytes), Oltmansiellopsidales (including motile and non-motile unicellular or colonial algae from marine or freshwater habitats), and Scotinosphaerales (including non-motile unicellular algae from freshwater or damp terrestrial habitats) are important to understand the evolution of macroscopic growth and multicellularity in the Ulvophyceae. Species in these orders were originally described in the Chlorophyceae or Trebouxiophyceae based on morphological grounds, but molecular phylogenetic studies firmly placed them in the Ulvophyceae (30, 31, 44, 45). Our analyses are consistent with chloroplast phylogenomic analyses in placing the Ignatiales and Oltmansiellopsidales in a clade together with the Ulvales-Ulotrichales (32), but in contrast to chloroplast-based analyses, the Oltmansiellopsidales was found to branch off first in our analyses. In contrast to a two-gene analysis, which recovered the Scotinosphaerales sister to a clade including the Ignatiales, Oltmansiellopsidales, and Ulvales-Ulotrichales (45), our analyses indicate that the Scotinosphaerales are sister to the Dasycladales. Although taxon sampling and the exact phylogenetic positions of microscopic ulvophycean lineages differed from the 10-gene analysis of Cocquyt et al. (31), our results support the hypothesis that macroscopic growth evolved independently in various lineages of ulvophytes from ancestral unicellular green algae by different mechanisms: by developing multicellularity with coupled mitosis and cytokinesis (independently in the Ulvales-Ulotrichales and Trentepohliales), by developing multicellularity with uncoupled mitosis and cytokinesis (siphonocladous organisation in the Cladophorales and *Blastophysa*) or by developing a siphonous architecture (independently in the Bryopsidales and Dasycladales).

Supplementary tables

Table S1. Transcriptomic and genomic datasets used in this study.

Species	Classification	Publication	Type of data	Source
Chlorophyta				
<i>Caulerpa taxifolia</i>	Ulvophyceae: Bryopsidales	(46)	Transcriptome	SRP041084
<i>Codium fragile</i>	Ulvophyceae: Bryopsidales	This study	Transcriptome	TBA
<i>Halimeda discoidea</i>	Ulvophyceae: Bryopsidales	This study	Transcriptome	TBA
<i>Ostreobium quekettii</i>	Ulvophyceae: Bryopsidales	This study	Transcriptome	TBA
<i>Blastophysa rhizopus</i>	Ulvophyceae: unclassified	(47)	Transcriptome	OneKP
<i>Boodlea composita</i>	Ulvophyceae: Cladophorales	(48)	Transcriptome	SRR5500908
<i>Cladophora glomerata</i>	Ulvophyceae: Cladophorales	(47)	Transcriptome	OneKP
<i>Acetabularia acetabulum</i>	Ulvophyceae: Dasycladales	this study	Transcriptome	TBA
<i>Ignatius tetrasporus</i>	Ulvophyceae: Ignatiales	(47)	Transcriptome	OneKP
<i>Oltmannsiellopsis unicellularis</i>	Ulvophyceae: Oltmannsiellopsidales	N/A	Transcriptome	TBA
<i>Oltmannsiellopsis viridis</i>	Ulvophyceae: Oltmannsiellopsidales	This study	Transcriptome	TBA
<i>Scotinosphaera lemnae</i>	Ulvophyceae: Scotinosphaerales	This study	Transcriptome	TBA
<i>Cephaleuros parasiticus</i>	Ulvophyceae: Trentepohliales	N/A	Transcriptome	TBA
<i>Trentepohlia annulata</i>	Ulvophyceae: Trentepohliales	N/A	Transcriptome	TBA
<i>Trentepohlia jolithus</i>	Ulvophyceae: Trentepohliales	(49)	Transcriptome	SRR1044982
<i>Acrosiphonia</i> sp. SAG-127.80	Ulvophyceae: Ulotrichales	(47)	Transcriptome	OneKP
<i>Phaeophila dendroides</i>	Ulvophyceae: Ulvales	N/A	Transcriptome	TBA
<i>Ulva linza</i>	Ulvophyceae: Ulvales	(50)	Transcriptome	SRR504341
<i>Ulva mutabilis</i>	Ulvophyceae: Ulvales	(51)De Clerck et al. 2018	Genome	ORCAE
<i>Tetradismus lagerheimii</i> SAG 38.81	Chlorophyceae	N/A	Transcriptome	SRR1174737
<i>Chlamydomonas reinhardtii</i>	Chlorophyceae	(52)	Genome	JGI4.0
<i>Dunaliella tertiolecta</i>	Chlorophyceae	(53)	Transcriptome	MMTESP1127
<i>Gonium pectorale</i>	Chlorophyceae	(54)	Genome	NCBI
<i>Haematococcus pluvialis</i>	Chlorophyceae	(55)	Transcriptome	SRR2148810
<i>Volvox carteri</i>	Chlorophyceae	(56)	Genome	JGI1.0
<i>Asterochloris</i> sp. Cgr/DA1pho	Trebouxiophyceae	N/A	Genome	JGI2.0
<i>Botryococcus braunii</i>	Trebouxiophyceae	N/A	Transcriptome	SRR069634
<i>Coccomyxa subellipsoidea</i>	Trebouxiophyceae	(57)	Genome	JGI1.0
<i>Pseudochlorella pringsheimii</i>	Trebouxiophyceae	(58)	Transcriptome	SRR490104
<i>Trebouxia gelatinosa</i>	Trebouxiophyceae	(59)	Transcriptome	SRR988248
<i>Auxenochlorella protothecoides</i>	Trebouxiophyceae: Chlorellales	(60)	Genome	KEGG
<i>Chlorella</i> sp. NC64A	Trebouxiophyceae: Chlorellales	(61)	Genome	JGI1.0
<i>Picochlorum oklahomensis</i>	Trebouxiophyceae: Chlorellales	(53)	Transcriptome	MMETSP1330
<i>Tetraselmis astigmatica</i>	Chlorodendrophyceae	(53)	Transcriptome	MMTESP0804
<i>Tetraselmis striata</i>	Chlorodendrophyceae	(53)	Transcriptome	MMETSP0817
<i>Marsupiomonas</i> sp.	Pedinophyceae	This study	Transcriptome	TBA
<i>Pedinomonas minor</i>	Pedinophyceae	This study	Transcriptome	TBA
Pedinophyceae sp. YPF701	Pedinophyceae	This study	Transcriptome	TBA
<i>Bathycoccus prasinos</i>	prasinophytes	(62)	Genome	ORCAE
<i>Micromonas pusilla</i>	prasinophytes	(63)	Genome	JGI2.0
<i>Nephroselmis pyriformis</i>	prasinophytes	(53)	Transcriptome	MMETSP0034
<i>Ostreococcus tauri</i>	prasinophytes	(64)	Genome	ORCAE
<i>Picocystis salinarum</i>	prasinophytes	(53)	Transcriptome	MMETSP0807
Streptophyta				
<i>Chara vulgaris</i>	Charophyceae	(47)	Transcriptome	ERR364366
<i>Chlorokybus atmophyticus</i>	Chlorokybophyceae	(65)	Transcriptome	SRR064329
<i>Chaetosphaeridium globosum</i>	Coleochaetophyceae	N/A	Transcriptome	ERR364369
<i>Coleochaete orbicularis</i>	Coleochaetophyceae	(66)	Transcriptome	SRR1594679
<i>Arabidopsis thaliana</i>	embryophytes	(67)	Genome	TAIR10
<i>Oryza sativa</i>	embryophytes	(68)	Genome	TIGR6.1
<i>Physcomitrella patens</i>	embryophytes	(69)	Genome	JGI1.1
<i>Selaginella moellendorffii</i>	embryophytes	(70)	Genome	JGI1.0

<i>Klebsormidium flaccidum</i>	Klebsormidiophyceae	(66)	Transcriptome	SRR1594644
<i>Mesostigma viride</i>	Mesostigmatophyceae	(66)	Transcriptome	SRR1594255
<i>Mesotaenium endlicherianum</i>	Zygnematophyceae	N/A	Transcriptome	ERR364377
<i>Roya obtusa</i>	Zygnematophyceae	N/A	Transcriptome	ERR364380

JGI: datasets available at the DOE Joint Genome Institute (<https://genome.jgi.doe.gov/portal/>), KEGG: datasets available at the Kyoto Encyclopedia of Genes and Genomes (<https://www.genome.jp/kegg/genome.html>), MMETSP: datasets available at the Short Read Archive (<https://www.ncbi.nlm.nih.gov/sra/>), NCBI: dataset available at the National Center for Biotechnology Information (<https://www.ncbi.nlm.nih.gov/>), OneKP: datasets available at (http://www.onekp.com/public_read_data.html), ORCAE: datasets available at Ghent University (<http://bioinformatics.psb.ugent.be/orcae/>), SRP, SRR, ERR: datasets available at the Short Read Archive (<https://www.ncbi.nlm.nih.gov/sra/>), TAIR: dataset available at The Arabidopsis Information Resource (<https://www.arabidopsis.org/>), TBA: To be announced, datasets not released yet, TIGR: dataset available at the Rice Genome Annotation Project (<http://rice.plantbiology.msu.edu/>).

Table S2. Sequencing technology and read depth of the transcriptome data.

species	Sequencing technology	depth of sequencing
<i>Acetabularia acetabulum</i>	Illumina NextSeq 500	103M 2x150 bp PE
<i>Acrosiphonia</i> sp. SAG-127.80	N/A	N/A
<i>Tetradasmus lagerheimii</i> SAG 38.81	Illumina HiSeq 2000	21M 2x101 bp PE
<i>Blastophysa rhizopus</i>	N/A	N/A
<i>Boodlea composita</i>	Illumina NextSeq	32M 2x75 bp PE
<i>Botryococcus braunii</i>	454 GS FLX	1.2M
<i>Caulerpa taxifolia</i>	Illumina HiSeq 2000	178M 2x95 bp PE
<i>Cephaleuros parasiticus</i>	Illumina HiSeq 1000	47M 2x100 bp PE
<i>Chaetosphaeridium globosum</i>	Illumina Genome Analyzer II	8M 2x75 bp PE
<i>Chara vulgaris</i>	Illumina Genome Analyzer II	8M 2x75 bp PE
<i>Chlorokybus atmophyticus</i>	454 GS FLX Titanium	444k
<i>Cladophora glomerata</i>	Illumina HiSeq 2000	12M 2x90 bp PE
<i>Codium fragile</i>	Illumina NextSeq	11M 2x150 bp PE
<i>Coleochaete orbicularis</i>	Illumina HiSeq 1000	44M 2x100 bp PE
<i>Dunaliella tertiolecta</i>	Illumina HiSeq 2000	32M 2x50 bp PE
<i>Haematococcus pluvialis</i>	Illumina HiSeq 2500	12M 2x100 bp PE
<i>Halimeda discoidea</i>	Illumina NextSeq	7M 2x150 bp PE
<i>Ignatius tetrasporus</i>	Illumina HiSeq 2000	11M 2x90 bp PE
<i>Klebsormidium flaccidum</i>	Illumina HiSeq 1000	50M 2x100 bp PE
<i>Marsupiomonas</i> sp.	Illumina NextSeq	19M 2x150 bp PE
<i>Mesostigma viride</i>	Illumina HiSeq 1000	47M 2x100 bp PE
<i>Mesotaenium endlicherianum</i>	Illumina HiSeq 2000	13M 2x90 bp PE
<i>Nephroselmis pyriformis</i>	Illumina HiSeq 2000	23M 2x100 bp PE
<i>Oltmannsiellopsis unicellularis</i>	Illumina HiSeq 1000	34M 2x100 bp PE
<i>Oltmannsiellopsis viridis</i>	Illumina NextSeq	26M 2x150 bp PE
<i>Ostreobium quekettii</i>	Illumina NextSeq	24M 2x150 bp PE
<i>Pedinomonas minor</i>	Illumina NextSeq	20M 2x150 bp PE
<i>Phaeophila dendroides</i>	Illumina HiSeq 1000	60M 2x100 bp PE
<i>Picochlorum oklahomensis</i>	Illumina HiSeq 2500	22M 2x50 bp PE
<i>Picocystis salinarum</i>	Illumina HiSeq 2000	16M 2x100 bp PE
<i>Pseudochlorella pringsheimii</i>	Illumina Genome Analyzer IIx	35M 40bp SE
<i>Roya obtusa</i>	Illumina HiSeq 2000	13M 2x90 bp PE
<i>Scotinosphaera lemnae</i>	Illumina NextSeq 500	39M 2x150 bp PE
<i>Tetraselmis astigmatica</i>	Illumina HiSeq 2500	29M 2x50 bp PE
<i>Tetraselmis striata</i>	Illumina HiSeq 2000	20M 2x50 bp PE
<i>Trebouxia gelatinosa</i>	Illumina HiSeq 2000	20M 2x100 bp PE
<i>Trentepohlia annulata</i>	Illumina HiSeq 1000	40M 2x100 bp PE
<i>Trentepohlia jolithus</i>	Illumina HiSeq 2000	26M 2x90 bp PE
<i>Ulva linza</i>	454 GS FLX	251k
<i>Pedinophyceae</i> sp. YPF701	Illumina NextSeq	18M 2x150 bp PE

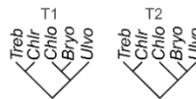
Table S3. Alignments metrics of the eight datasets analyzed.

	coreGF				ulvoGF			
	unscaffold		scaffold		unscaffold		Scaffold	
	untrimmed	trimmed	untrimmed	trimmed	untrimmed	trimmed	untrimmed	Trimmed
# genes	539	539	539	539	355	355	355	355
# sites (nt)	1,751,925	342,501	1,877,544	342,501	1,143,789	252,612	1,244,520	252,612
# sites (aa)	583,975	114,167	625,848	114,167	381,263	84,204	414,840	84,204
% missing data	65.6	27.3	69.7	19.9	43.7	24.0	54.8	15.9

nt = nucleotides, aa = amino acid

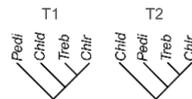
Table S4. Comparing alternative hypotheses using the Shimodaira-Hasegawa (SH) Test and Approximately Unbiased (AU) Test. Constraint trees include: (A) Monophyletic Ulvophyceae (coalescence-based topology), (B) Chlorodendrophyceae diverging first, followed by the Pedinophyceae, (C) Different positions of the Ignatiales. The topological schemas (hypotheses) are depicted. T1 represents the unconstrained tree topology.

A. Monophyletic Ulvophyceae



dataset	Tree	logL	deltaL	p-SH	p-AU
coreGF noscaffold	1	-15,791,662.12	0	1.0000 +	0.9877 +
	2	-15,791,832.34	170.224	0.0135 -	0.0123 -
coreGF scaffold	1	-17,211,164.2	0	1.0000 +	0.9862 +
	2	-17,211,348.3	184.097	0.0136 -	0.0138 -
coreGF noscaffold TRIM	1	-3,848,877.26	1.132	0.4952 +	0.4951 +
	2	-3,848,876.12	0	1.0000 +	0.5049 +
coreGF scaffold TRIM	1	-4,182,782.87	0	1.0000 +	0.7136 +
	2	-4,182,827.07	44.192	0.2920 +	0.2864 +
ulvoGF noscaffold	1	-10,849,295.9	0	1.0000 +	0.9913 +
	2	-10,849,457.1	161.15	0.0085 -	0.0087 -
ulvoGF scaffold	1	-11,946,468	0	1.0000 +	0.9884 +
	2	-11,946,637.9	169.952	0.0110 -	0.0116 -
ulvoGF noscaffold TRIM	1	-2,851,993.21	0	1.0000 +	0.5982 +
	2	-2,852,009.83	16.622	0.4013 +	0.4018 +
ulvoGF scaffold TRIM	1	-3,124,270.99	0	1.0000 +	0.7474 +
	2	-3,124,320.02	49.031	0.2490 +	0.2526 +

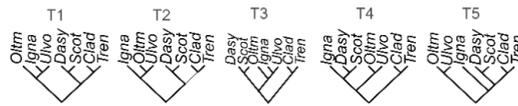
B. Chlorodendrophyceae first diverging, followed by the Pedinophyceae



dataset	Tree	logL	deltaL	p-SH	p-AU
coreGF noscaffold	1	-15,791,662	0	1.0000 +	1.0000 +
	2	-15,791,999	337.044	0.0000 -	0.0000 -
coreGF scaffold	1	-17,211,164.2	0	1.0000 +	1.0000 +
	2	-17,211,561.5	397.272	0.0000 -	0.0000 -
coreGF noscaffold TRIM	1	-3,848,876.18	0	1.0000 +	0.9559 +
	2	-3,848,957.15	80.976	0.0437 -	0.0441 -
coreGF scaffold TRIM	1	-4,182,782.81	0	1.0000 +	0.9012 +
	2	-4,182,848.9	66.085	0.0991 +	0.0988 +
ulvoGF noscaffold	1	-10,849,295.9	0	1.0000 +	0.9988 +
	2	-10,849,499.2	203.306	0.0014 -	0.0012 -
ulvoGF scaffold	1	-11,94,6468	0	1.0000 +	0.9996 +
	2	-11,946,705.9	237.929	0.0006 -	0.0004 -

ulvoGF noscaffold TRIM	1	-2,851,993.2	0	1.0000 +	0.7870 +
	2	-2,852,027.89	34.683	0.2089 +	0.2130 +
ulvoGF scaffold TRIM	1	-3,124,271.12	0	1.0000 +	0.6113 +
	2	-3,124,284.86	13.739	0.3841 +	0.3887 +

C. Different positions of the Ignatiales



dataset	Tree	logL	deltaL	p-SH	p-AU
coreGF noscaffold	1	-15,791,661	0	1.0000 +	0.6439 +
	2	-15,791,765	103.93	0.2229 +	0.1019 +
	3	-15,791,661	0	0.8447 +	0.6531 +
	4	-15,791,780	118.92	0.2255 +	0.1879 +
	5	-15,791,909	247.72	0.0113 -	0.0037 -
coreGF scaffold	1	-17,211,164	0	0.8368 +	0.6500 +
	2	-17,211,270	105.62	0.2733 +	0.1357 +
	3	-17,211,164	0	1.0000 +	0.6597 +
	4	-17,211,320	155.99	0.1703 +	0.1463 +
	5	-17,211,550	385.4	0.0005 -	0.0006 -
coreGF noscaffold TRIM	1	-3,848,877	1.134	0.7028 +	0.5484 +
	2	-3,848,899	22.713	0.7231 +	0.3274 +
	3	-3,848,877	1.134	0.7028 +	0.5488 +
	4	-3,848,876	0	1.0000 +	0.5262 +
	5	-3,849,029	152.94	0.0820 +	0.0005 -
coreGF scaffold TRIM	1	-4,182,822	39.404	0.4750 +	0.4015 +
	2	-4,182,833	49.973	0.4789 +	0.2475 +
	3	-4,182,822	39.404	0.4750 +	0.4016 +
	4	-4,182,783	0	1.0000 +	0.6821 +
	5	-4,183,033	250.47	0.0017 -	0.0000 -
ulvoGF noscaffold	1	-10,849,296	0	0.8502 +	0.6308 +
	2	-10,849,366	70.066	0.3830 +	0.1790 +
	3	-10,849,296	0	1.0000 +	0.6373 +
	4	-10,849,374	78.303	0.3406 +	0.2808 +
	5	-10,849,524	228.58	0.0110 -	0.0019 -
ulvoGF scaffold	1	-11,946,468	0	1.0000 +	0.6450 +
	2	-11,946,534	65.504	0.4695 +	0.2546 +
	3	-11,946,468	0	0.8458 +	0.6365 +
	4	-11,946,567	99.317	0.2998 +	0.2378 +
	5	-11,946,819	350.58	0.0005 -	0.0000 -
ulvoGF noscaffold TRIM	1	-2,852,007	14.265	0.5863 +	0.4358 +
	2	-2,852,006	12.597	0.7725 +	0.4275 +
	3	-2,852,007	14.265	0.5863 +	0.4439 +
	4	-2,851,993	0	1.0000 +	0.5709 +
	5	-2,852,131	137.62	0.0334 -	0.0003 -
ulvoGF scaffold TRIM	1	-3,124,319	48.194	0.4686 +	0.3697 +
	2	-3,124,318	47.546	0.5131 +	0.2771 +
	3	-3,124,313	42.563	0.5007 +	0.3350 +
	4	-3,124,271	0	1.0000 +	0.7340 +
	5	-3,124,489	218.71	0.0112 -	0.0001 -

Bryo = Bryopsidales, Chld = Chlorodendrophyceae, Chlo = Chlorophyceae, Chlr = Chlorellales, Clad = Cladophorales, Dasy = Dasycladales, Igna = Ignatiales, Oltm = Oltmannsielopsidales, Treb = Trebouxiophyceae sensu stricto, Tren = Trentepohliales, Pedi = Pedinophyceae, Scot = Scotinosphaerales, Ulvo = Ulvophyceae.

deltaL: logL difference from the maximal logL in the set.

bp-RELL: bootstrap proportion using REll method.

p-KH: p-value of one sided Kishino-Hasegawa test.

p-SH: p-value of Shimodaira-Hasegawa test.

c-ELW: Expected Likelihood Weight.

p-AU: p-value of approximately unbiased (AU) test.

Plus signs denote the 95% confidence sets. Minus signs denote significant exclusion. All tests performed 100,000 resamplings using the REll method.

Table S5. Node calibrations used for relaxed molecular clock analysis

Node	Node	calibration	Fossil or transfer	period	prior	Reference
C	Characeae	C ₁	Transfer from other study	Devonian	U[416-∞[(71)
L	Land plants (embryophytes)	L ₁	Cryptospore assemblage from the Middle Ordovician (Dapingian)	Ordovician	U[475-∞[(71)
V	Vascular plants	V ₁	<i>Baragwanathia longifolia</i>	Silurian	U[421-∞[(71)
T ₁	<i>Botryococcus</i> stem node	T ₁	<i>Botryococcus</i>	Carboniferous	U[299-∞ [(72)
UA	<i>Ostreobium</i> stem node	UA ₁	Transfer from other study	n/a	U[533-425]	(73)
UB	Ulvoephyceae/Chlorophyceae stem node	UB ₁	<i>Proterocladus</i>	Neoproterozoic	U[716-∞ [(74)
		UB ₀	Absence of calibration	n/a	n/a	n/a
RT	Streptophyta-Chlorophyta split	RT ₁	Transfer from other study	n/a	U[1279-1159]	(75)
		RT ₂	Transfer from other study	n/a	U[1015-863]	(76)
		RT ₃	RT ₁ /RT ₂ hybrid	n/a	U[1279-863]	n/a
		RT ₀	n/a	n/a	n/a	n/a

Table S6. Relaxed molecular clock analyses results.

Tree	clock model	sub model	Root	<i>Proterocladus</i>	root_age (My)	Core Chlorophyta (My)	(Chlo,Bryo),Ulvo (My)	Chlo,(Bryo,Ulvo) (My)
ML	ln	CAT	RT0	UB1	1,021 (1,072-969)	885 (960-810)	801 (879-753)	n/a
ML	ln	CAT	RT1	UB1	1,185 (1,209-1,162)	1,005 (1,079-931)	895 (966-823)	n/a
ML	ln	CAT	RT2	UB1	980 (1,004-956)	858 (904-815)	781 (823-749)	n/a
ML	ln	CAT	RT3	UB1	1,022 (1,073-971)	886 (978-822)	802 (880-753)	n/a
ML	ugam	CAT	RT0	UB1	1,013 (1,088-937)	911 (1,065-814)	834 (974-753)	n/a
ML	ugam	CAT	RT1	UB1	1,198 (1,229-1,168)	1,063 (1,149-978)	967 (1,056-877)	n/a
ML	ugam	CAT	RT2	UB1	961 (995-927)	869 (931-809)	800 (860-749)	n/a
ML	ugam	CAT	RT3	UB1	1,013 (1,086-940)	910 (1,060-815)	835 (971-754)	n/a
ML	ln	LG	RT0	UB1	1,030 (1,081-979)	895 (989-831)	802 (881-755)	n/a
ML	ln	LG	RT1	UB1	1,185 (1,209-1,162)	1,008 (1,081-925)	890 (961-813)	n/a
ML	ln	LG	RT2	UB1	984 (1,007-961)	865 (909-823)	782 (723-751)	n/a
ML	ln	LG	RT3	UB1	1,031 (1,081-980)	896 (989-833)	803 (882-756)	n/a
ML	ugam	LG	RT0	UB1	1,021 (1,096-945)	919 (1,072-820)	837 (976-755)	n/a
ML	ugam	LG	RT1	UB1	1,198 (1,229-1,167)	1,065 (1,152-981)	964 (1,053-876)	n/a
ML	ugam	LG	RT2	UB1	964 (997-931)	872 (931-813)	799 (858-750)	n/a
ML	ugam	LG	RT3	UB1	1,021 (1,094-948)	918 (1,065-821)	836 (969-756)	n/a
CB	ln	CAT	RT0	UB1	1,012 (1,069-955)	833 (931-776)	n/a	759 (845-718)
CB	ln	CAT	RT1	UB1	1,189 (1,215-1,163)	951 (1,023-877)	n/a	854 (927-781)
CB	ln	CAT	RT2	UB1	973 (1,001-945)	809 (855-772)	n/a	740 (785-717)
CB	ln	CAT	RT3	UB1	1,014 (1,072-957)	833 (934-775)	n/a	758 (846-718)
CB	ugam	CAT	RT0	UB1	993 (1,062-925)	851 (987-775)	n/a	776 (890-718)
CB	ugam	CAT	RT1	UB1	1,196 (1,226-1,166)	1,008 (1,094-922)	n/a	912 (1,000-825)
CB	ugam	CAT	RT2	UB1	955 (990-920)	824 (887-772)	n/a	753 (814-717)
CB	ugam	CAT	RT3	UB1	991 (1,058-924)	851 (984-776)	n/a	775 (896-718)
CB	ln	LG	RT0	UB1	1,022 (1,079-965)	836 (935-777)	n/a	760 (847-718)
CB	ln	LG	RT1	UB1	1,190 (1,217-1,163)	948 (1,021-876)	n/a	851 (925-779)
CB	ln	LG	RT2	UB1	977 (1,003-951)	808 (853-772)	n/a	739 (783-717)
CB	ln	LG	RT3	UB1	1,018 (1,075-962)	834 (932-776)	n/a	760 (846-718)
CB	ugam	LG	RT0	UB1	991 (1,059-923)	849 (986-774)	n/a	776 (898-718)
CB	ugam	LG	RT1	UB1	1196 (1,225-1,166)	1,006 (1,092-918)	n/a	910 (999-821)
CB	ugam	LG	RT2	UB1	955 (991-920)	823 (887-771)	n/a	753 (814-718)
CB	ugam	LG	RT3	UB1	992 (1,059-926)	850 (982-775)	n/a	775 (894-718)
ML	ln	CAT	RT0	UB0	972 (1,038-907)	837 (960-724)	756 (863-657)	n/a
ML	ln	CAT	RT1	UB0	1184 (1,207-1,162)	1,004 (1,077-929)	895 (965-823)	n/a
ML	ln	CAT	RT2	UB0	949 (988-910)	817 (890-736)	739 (809-665)	n/a
ML	ln	CAT	RT3	UB0	978 (1,039-916)	841 (961-739)	860 (864-668)	n/a
ML	ugam	CAT	RT0	UB0	943 (1,039-848)	846 (1,031-697)	774 (942-639)	n/a
ML	ugam	CAT	RT1	UB0	1,199 (1,230-1,168)	1,064 (1,151-978)	968 (1,058-876)	n/a
ML	ugam	CAT	RT2	UB0	937 (979-895)	840 (923-760)	768 (852-691)	n/a

ML	ugam	CAT	RT3	UB0	972 (1,049-895)	870 (1,035-764)	975 (948-693)	n/a
ML	ln	LG	RT0	UB0	983 (1,049-917)	845 (967-734)	757 (862-659)	n/a
ML	ln	LG	RT1	UB0	1,185 (1,208-1,162)	1,005 (1,078-931)	888 (957-815)	n/a
ML	ln	LG	RT2	UB0	953 (991-915)	823(896-740)	738 (808-662)	n/a
ML	ln	LG	RT3	UB0	987 (1,049-925)	852 (970-749)	763 (865-670)	n/a
ML	ugam	LG	RT0	UB0	953 (1,048-858)	854 (1,041-708)	777 (945-643)	n/a
ML	ugam	LG	RT1	UB0	1,199 (1,229-1,168)	1,066 (1,150-982)	965 (1,053-876)	n/a
ML	ugam	LG	RT2	UB0	937 (979-896)	840 (921-763)	765 (845-688)	n/a
ML	ugam	LG	RT3	UB0	977 (1,055-899)	876 (1,041-768)	796 (947-693)	n/a
CB	ln	CAT	RT0	UB0	968 (1,038-898)	792 (914-696)	n/a	721 (829-638)
CB	ln	CAT	RT1	UB0	1,190 (1,216-1,163)	952 (1,026-879)	n/a	854 (928-783)
CB	ln	CAT	RT2	UB0	944 (984-905)	775 (844-709)	n/a	707 (772-646)
CB	ln	CAT	RT3	UB0	974 (1,038-910)	797 (913-712)	n/a	725 (828-650)
CB	ugam	CAT	RT0	UB0	931 (1,019-842)	795 (960-663)	n/a	723 (872-605)
CB	ugam	CAT	RT1	UB0	1,196 (1,226-1,166)	1,007 (1,094-919)	n/a	910 (1,001-821)
CB	ugam	CAT	RT2	UB0	933 (975-892)	798 (879-720)	n/a	725 (807-651)
CB	ugam	CAT	RT3	UB0	963 (1,034-892)	822 (970-725)	n/a	747 (811-655)
CB	ln	LG	RT0	UB0	975 (1,045-905)	795 (916-697)	n/a	724 (833-640)
CB	ln	LG	RT1	UB0	1,190 (1,216-1,163)	948 (1,023-877)	n/a	852 (926-780)
CB	ln	LG	RT2	UB0	949 (988-911)	776 (843-708)	n/a	708 (772-647)
CB	ln	LG	RT3	UB0	983 (1,050-917)	800(919-711)	n/a	727 (832-650)
CB	ugam	LG	RT0	UB0	933 (1,020-846)	797 (958-668)	n/a	736 (873-609)
CB	ugam	LG	RT1	UB0	1,196 (1,226-1,166)	1,007 (1,094-922)	n/a	912 (1,001-825)
CB	ugam	LG	RT2	UB0	933 (975-892)	797 (878-721)	n/a	725 (506-652)
CB	ugam	LG	RT3	UB0	962 (1,032-892)	821 (966-725)	n/a	747 (881-657)

ML: Topology inferred by supermatrix analyses

CB: Topology inferred by coalescence-based analyses

Supplementary figures

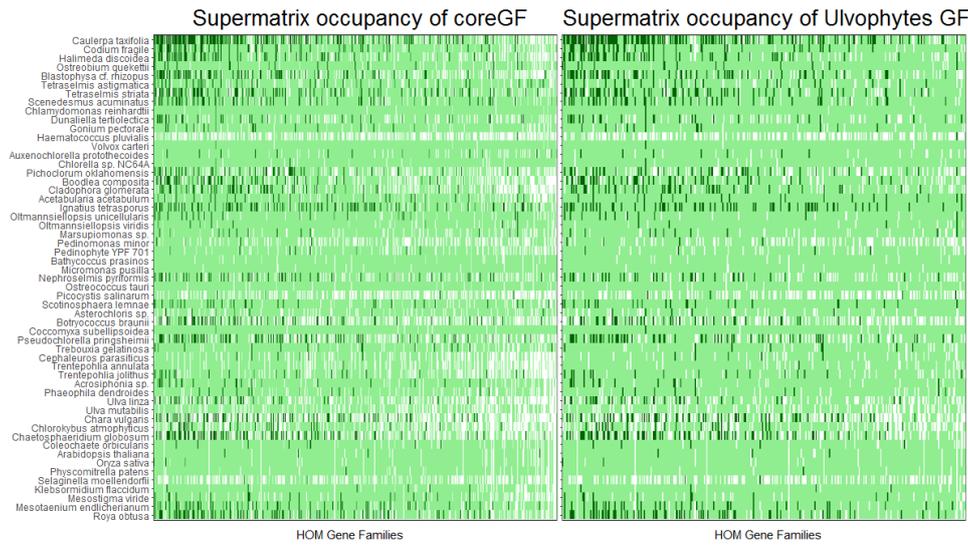


Figure S1. Supermatrix occupancy of the coreGF and ulvoGF datasets. Supermatrix occupancy of the 539 coreGF and 355 ulvoGF single-copy genes. From the left to the right, the genes with highest to lowest taxon occupancy. Dark green: scaffolded sequence (absent from the unscaffold datasets); light green: available sequence; white: missing sequence.

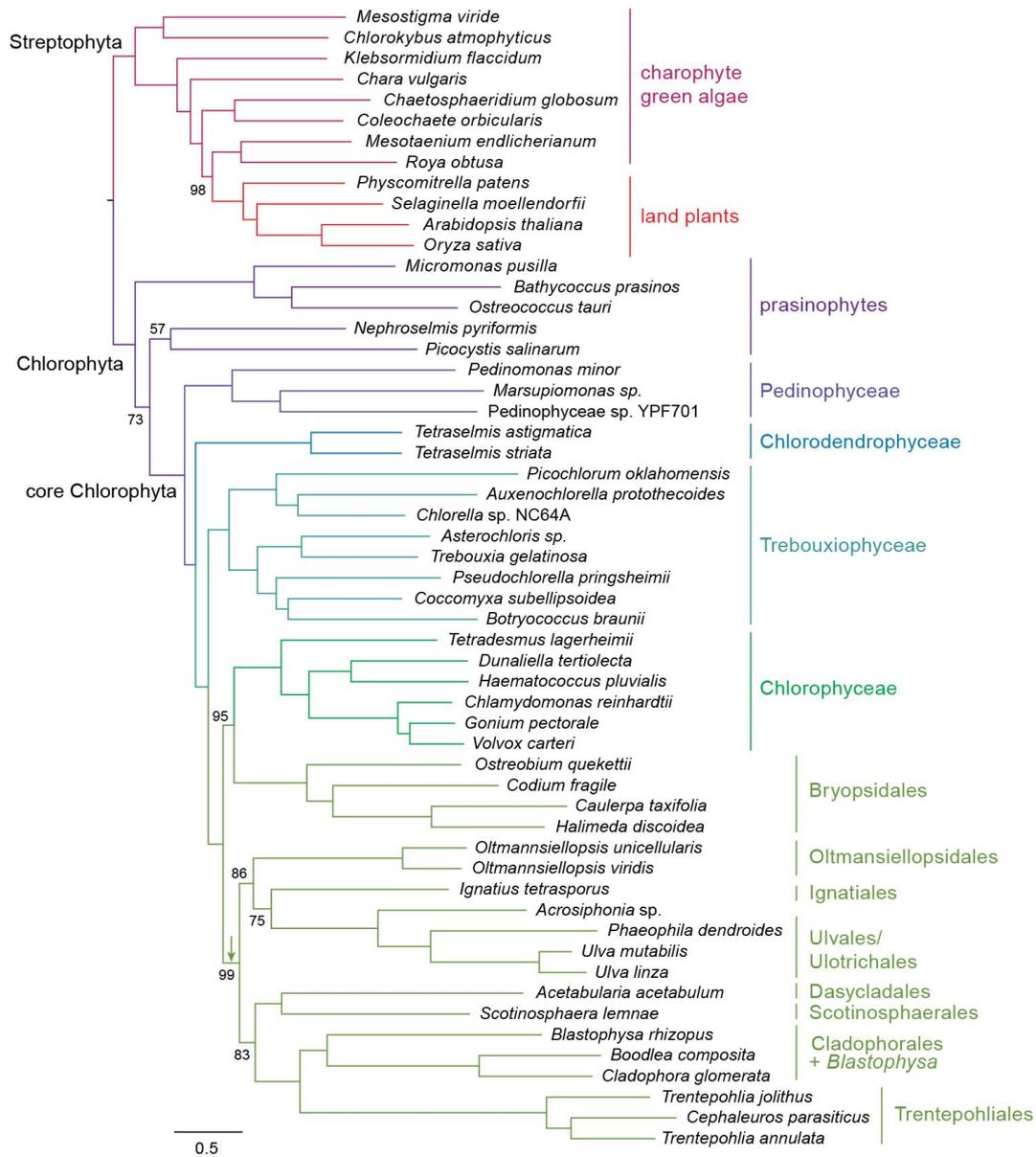


Figure S2. Maximum likelihood phylogenetic tree of the green algae inferred from a concatenated amino acid alignment of 539 nuclear genes (supermatrix analysis of the coreGF-scaffolded-untrimmed data set). Bootstrap support values are shown at the nodes (100% values are omitted). Scale bar indicates estimated number of substitutions per amino acid position. The green arrow indicates the phylogenetic position of the Bryopsidales in the coalescence-based analyses.

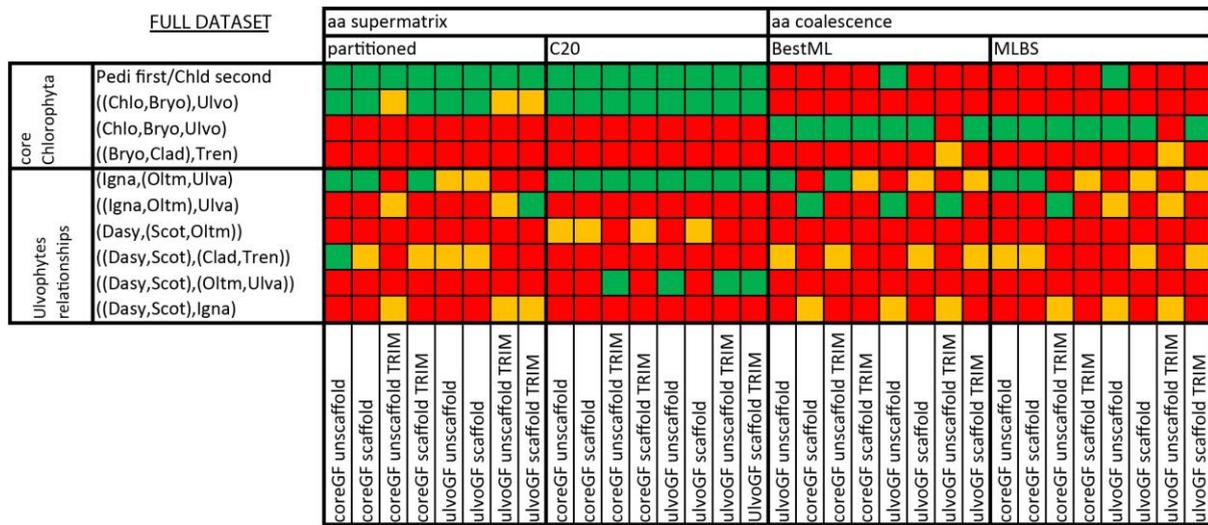


Figure S3. Summary of support for hypotheses of specific chlorophytan relationships, based on supermatrix or coalescence-based analyses of the eight amino acid alignments (Table S3). Relationships are summarized in newick tree format. Green tiles indicate high support (bootstrap values and posterior probabilities > 75% and 0.75, respectively), yellow tiles indicate weak support, while red tiles indicate that the phylogenetic relationship was rejected. Bryo: Bryopsidales; Chld: Chlorodendrophyceae; Chlo: Chlorophyceae; Chlr: Chlorellales; Clad: Cladophorales+*Blastophysa*; Dasy: Dasycladales; Igna: Ignatiales; Oltm: Oltmannsiellopsidales; Pedi: Pedinophyceae; Scot: Scotinosphaerales; Tren: Trentepohliales; Ulva: Ulvales-Ulotrichales; Ulvo: Ulvophyceae excluding Bryopsidales.

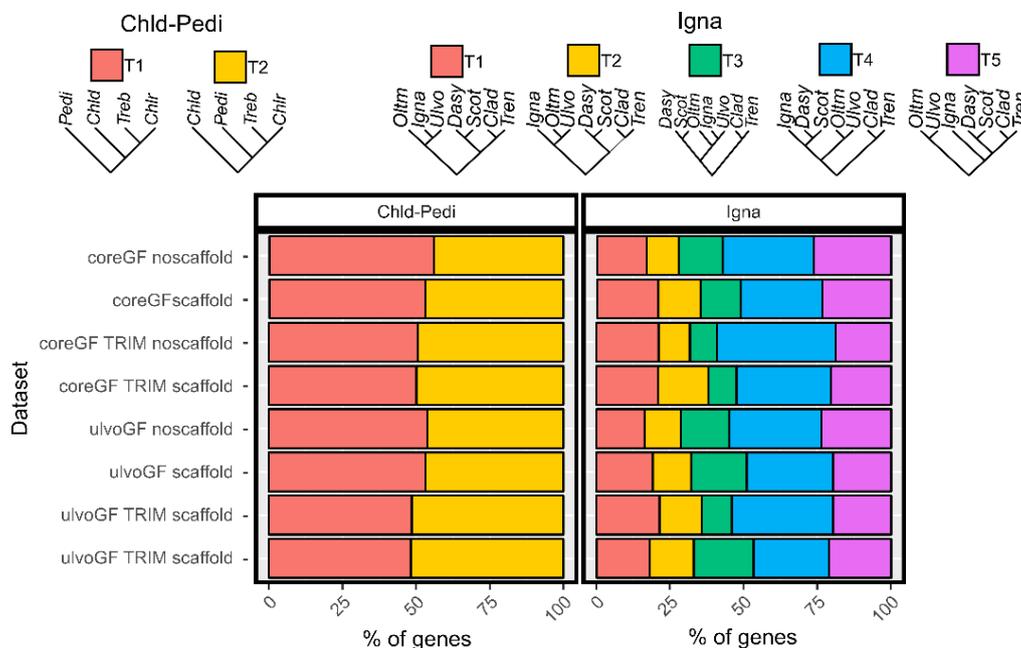


Figure S4. Proportion of genes supporting each of the alternative topologies in pairwise log-likelihood score comparisons. Alternative topologies were also tested using AU-tests (Table S4). Bryo: Bryopsidales; Chld: Chlorodendrophyceae; Chlo: Chlorophyceae; Chlr: Chlorellales; Clad: Cladophorales+*Blastophysa*; Dasy: Dasycladales; Igna: Ignatiales; Oltm: Oltmannsiellopsidales; Pedi: Pedinophyceae; Scot: Scotinosphaerales; Tren: Trentepohliales; Ulvo: Ulvales-Ulotrichales.

		NO IGNATIUS																							
		aa supermatrix								aa coalescence															
core Chlorophyta	Ulvophytes relationships	partitioned								BestML				MLBS											
		Pedi first/Chld second ((Chlo,Bryo),Ulvo) (Chlo,Bryo,Ulvo) ((Bryo,Clad),Tren)	(Dasy,(Scot,Oltm)) ((Dasy,Scot),(Oltm,Ulva)) ((Dasy,Scot),(Oltm,Ulva))	coreGF unscaffold	coreGF scaffold	coreGF unscaffold TRIM	coreGF scaffold TRIM	ulvoGF unscaffold	ulvoGF scaffold	ulvoGF unscaffold TRIM	ulvoGF scaffold TRIM	coreGF unscaffold	coreGF scaffold	coreGF unscaffold TRIM	coreGF scaffold TRIM	ulvoGF unscaffold	ulvoGF scaffold	ulvoGF unscaffold TRIM	ulvoGF scaffold TRIM	coreGF unscaffold	coreGF scaffold	coreGF unscaffold TRIM	coreGF scaffold TRIM	ulvoGF unscaffold	ulvoGF scaffold
coreGF unscaffold	coreGF scaffold			coreGF unscaffold TRIM	coreGF scaffold TRIM	ulvoGF unscaffold	ulvoGF scaffold	ulvoGF unscaffold TRIM	ulvoGF scaffold TRIM	coreGF unscaffold	coreGF scaffold	coreGF unscaffold TRIM	coreGF scaffold TRIM	ulvoGF unscaffold	ulvoGF scaffold	ulvoGF unscaffold TRIM	ulvoGF scaffold TRIM	coreGF unscaffold	coreGF scaffold	coreGF unscaffold TRIM	coreGF scaffold TRIM	ulvoGF unscaffold	ulvoGF scaffold	ulvoGF unscaffold TRIM	ulvoGF scaffold TRIM
coreGF unscaffold	coreGF scaffold			coreGF unscaffold TRIM	coreGF scaffold TRIM	ulvoGF unscaffold	ulvoGF scaffold	ulvoGF unscaffold TRIM	ulvoGF scaffold TRIM	coreGF unscaffold	coreGF scaffold	coreGF unscaffold TRIM	coreGF scaffold TRIM	ulvoGF unscaffold	ulvoGF scaffold	ulvoGF unscaffold TRIM	ulvoGF scaffold TRIM	coreGF unscaffold	coreGF scaffold	coreGF unscaffold TRIM	coreGF scaffold TRIM	ulvoGF unscaffold	ulvoGF scaffold	ulvoGF unscaffold TRIM	ulvoGF scaffold TRIM
coreGF unscaffold	coreGF scaffold			coreGF unscaffold TRIM	coreGF scaffold TRIM	ulvoGF unscaffold	ulvoGF scaffold	ulvoGF unscaffold TRIM	ulvoGF scaffold TRIM	coreGF unscaffold	coreGF scaffold	coreGF unscaffold TRIM	coreGF scaffold TRIM	ulvoGF unscaffold	ulvoGF scaffold	ulvoGF unscaffold TRIM	ulvoGF scaffold TRIM	coreGF unscaffold	coreGF scaffold	coreGF unscaffold TRIM	coreGF scaffold TRIM	ulvoGF unscaffold	ulvoGF scaffold	ulvoGF unscaffold TRIM	ulvoGF scaffold TRIM

Figure S5. Summary of support for hypotheses of Chlorophyta relationships in analyses where *Ignatius* was pruned from the dataset. See legend of Figure S2 for details.

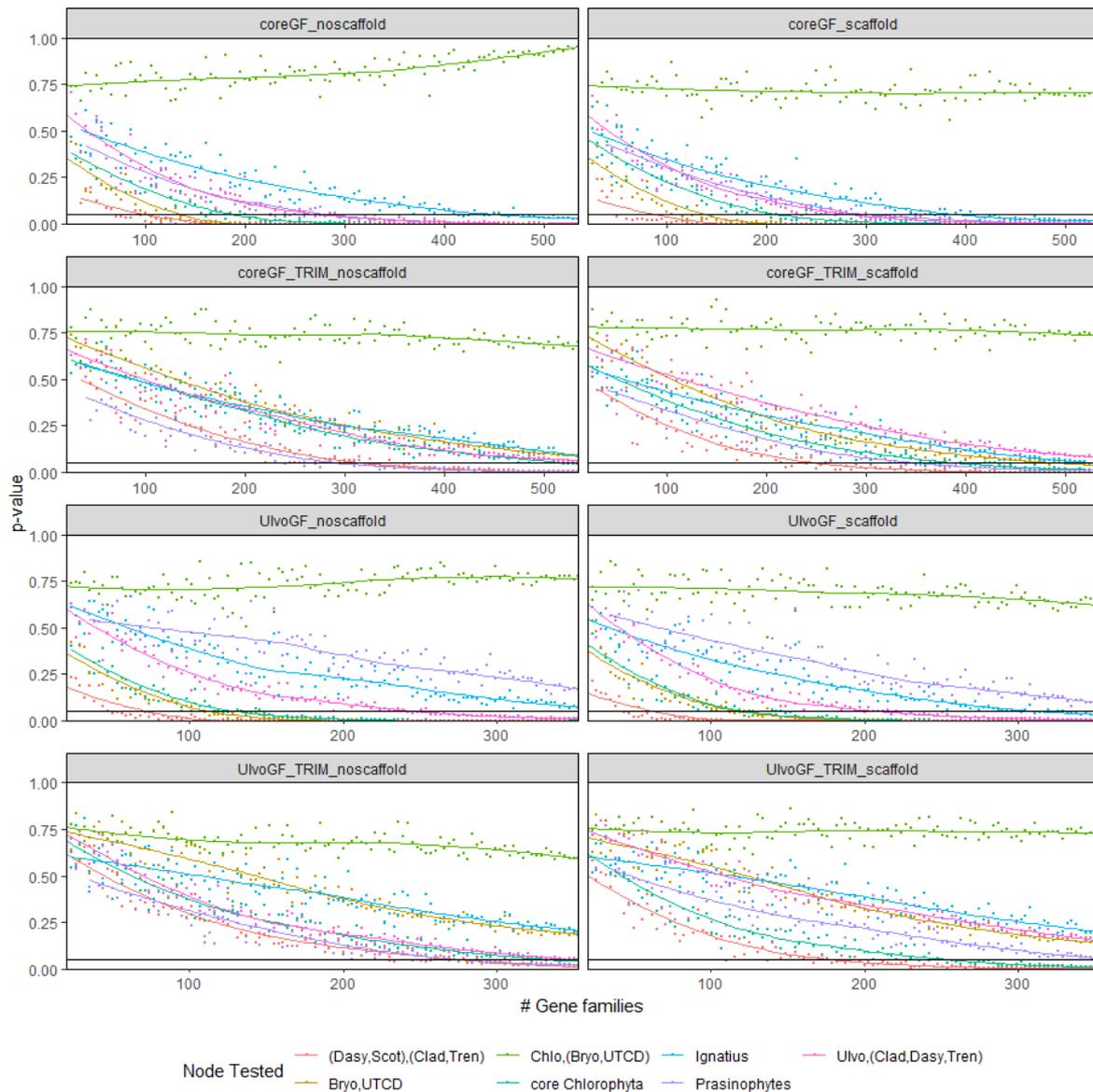


Figure S6. Tests for polytomy null-hypothesis. Polytomy test results for selected branches for each dataset. Gene trees were built with random growing subsets of genes (1%, 2%, ... 100%, but not less than 20 genes), 10 replicates were run for each dataset and each subset. The median of the p-values of selected branches for each subset is plotted on the y-axis, against the number of genes in the subset (x-axis). The horizontal black line indicates p-value = 0.05.

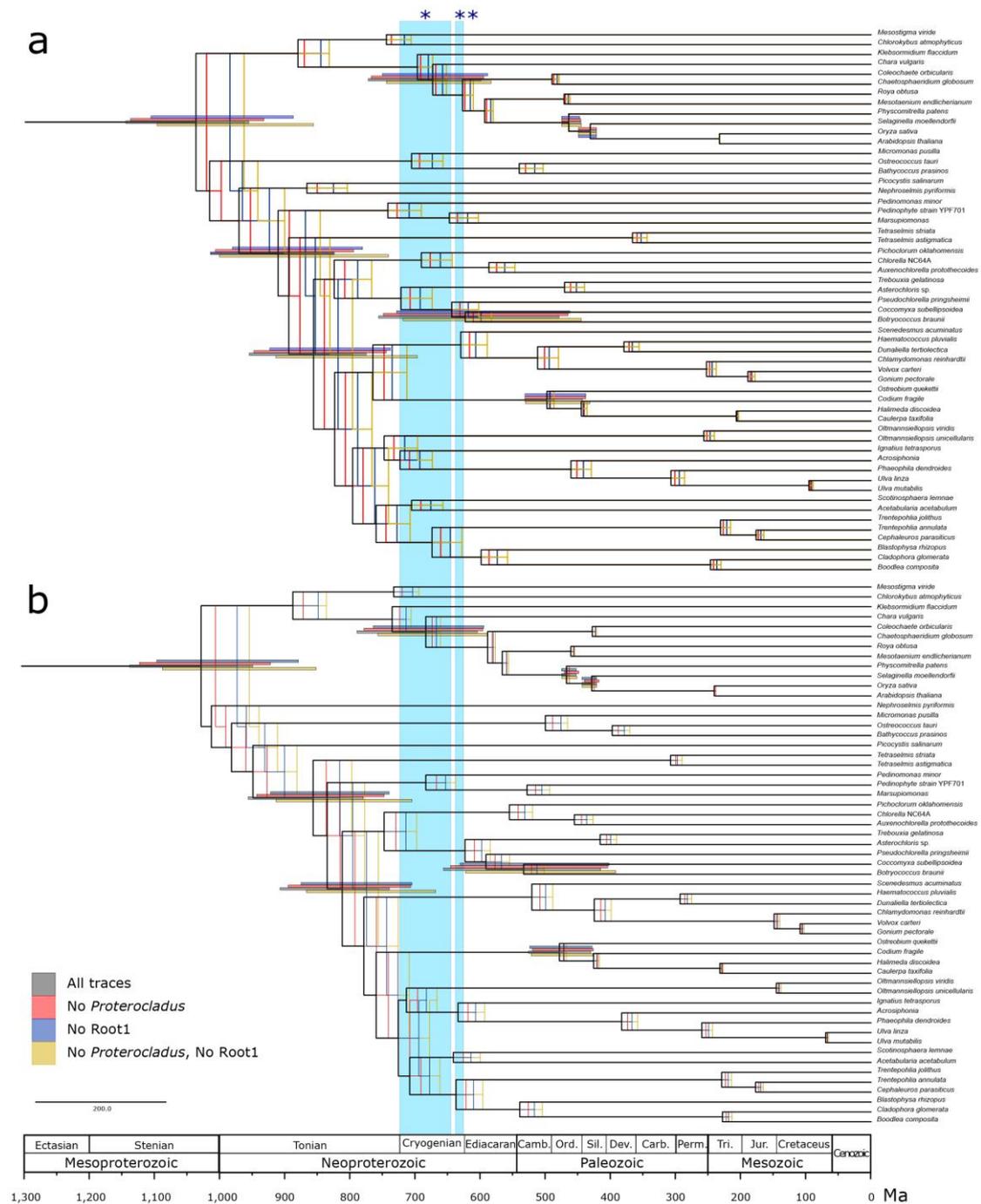


Figure S7. Summary of time-calibrated phylogenetic trees based on different analyses, including all traces, analyses without R1 root, without the *Proterocladus* calibrating point, and without R1 root nor *Proterocladus* (Table S5, S6). Chronograms are based on the 10 most clock-like genes from the coreGF scaffolded trimmed dataset (see methods). Figure A shows the analyses constrained with the ML-based topology (T1), Figure B shows the analyses constrained with the coalescence-based topology (T4). Node values indicate average node ages. The bars represent 95% confidence interval (CI) of the calibration nodes relevant for this study. The two blue bars indicate the Sturtian (*) and the Marinoan (**) glaciations.

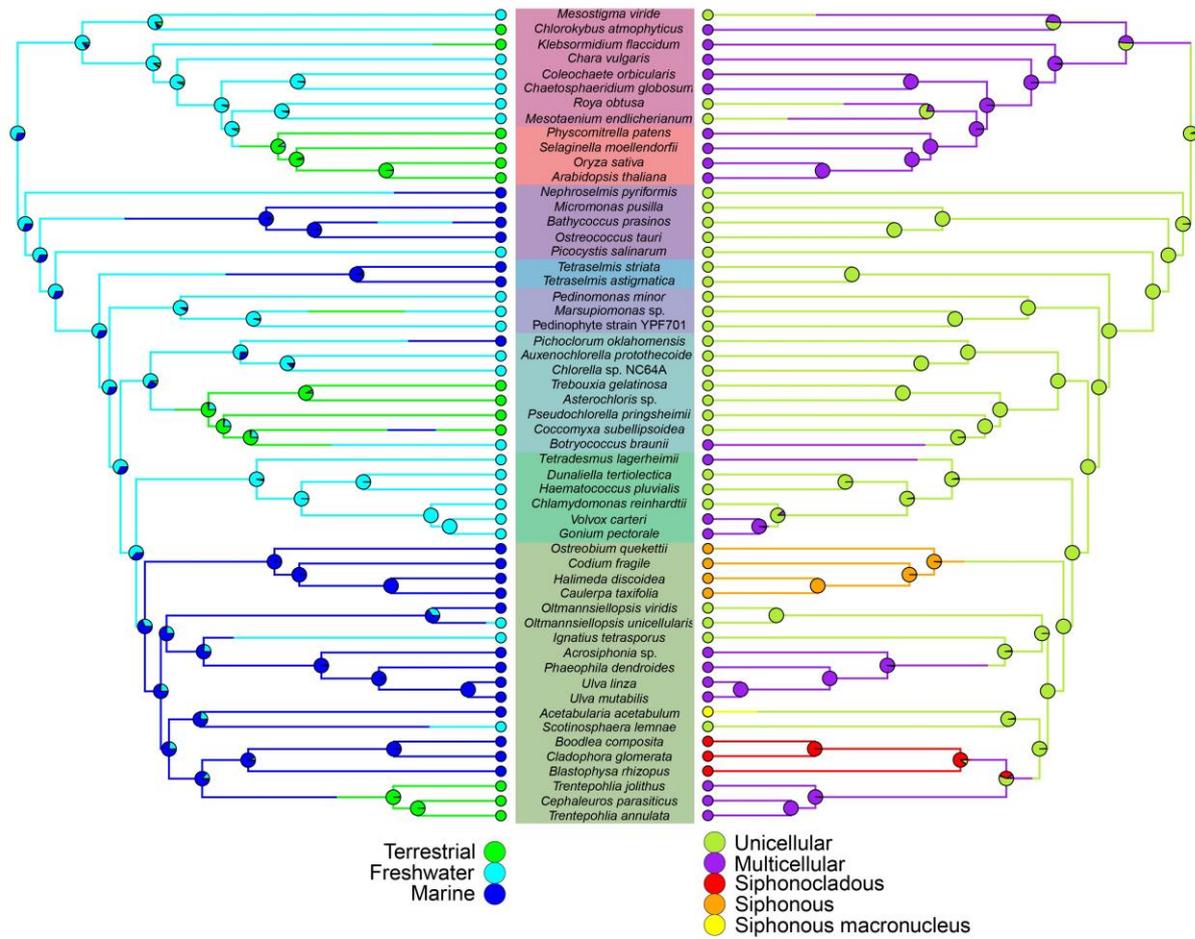


Figure S8. Ancestral state estimation of habitat (left) and cyto-morphological (right) traits, plotted on the ultrametric tree.

SI References

1. B. E. Hunt, D. F. Mandoli, A new, artificial sea water that facilitates growth of large numbers of cells of *Acetabularia acetabulum* (Chlorophyta) and reduces the labor inherent in cell culture. *J. Phycol.* **32**, 483-495 (1996).
2. H. W. Bischoff, H. C. Bold, Phycological Studies IV. Some soil algae from Enchanted Rock and related algal species. *Univ. Texas Publ.* **6318**, 1-95 (1963).
3. C. Jackson, A. H. Knoll, C. X. Chan, H. Verbruggen, Plastid phylogenomics with broad taxon sampling further elucidates the distinct evolutionary origins and timing of secondary green plastids. *Sci. Rep.* **8**, 1523 (2018).
4. M. G. Grabherr *et al.*, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644-U130 (2011).
5. M. Van Bel *et al.*, TRAPID: an efficient online tool for the functional and comparative analysis of de novo RNA-Seq transcriptomes. *Genome Biol.* **14**, R134 (2013).
6. J. Gouzy, S. Carrere, T. Schiex, FrameDP: sensitive peptide detection on noisy matured sequences. *Bioinformatics* **25**, 670-671 (2009).
7. E. Cocquyt *et al.*, Complex phylogenetic distribution of a non-canonical genetic code in green algae. *BMC Evol. Biol.* **10**, 327 (2010).
8. M. Van Bel *et al.*, PLAZA 4.0: an integrative resource for functional, evolutionary and comparative plant genomics. *Nucleic Acids Res.* **46**, D1190-D1196 (2018).
9. B. Buchfink, C. Xie, D. H. Huson, Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* **12**, 59-60 (2014).
10. A. J. Enright, S. Van Dongen, C. A. Ouzounis, An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* **30**, 1575-1584 (2002).
11. S. Proost *et al.*, PLAZA: A Comparative Genomics Resource to Study Gene and Genome Evolution in Plants. *Plant Cell* **21**, 3718-3731 (2009).
12. K. Vandepoele *et al.*, pico-PLAZA, a genome database of microbial photosynthetic eukaryotes. *Environ. Microbiol.* **15**, 2147-2153 (2013).
13. M. Kanehisa, Y. Sato, K. Morishima, BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* **428**, 726-731 (2016).
14. X. Huang, A. Madan, CAP3: A DNA sequence assembly program. *Genome Res.* **9**, 868-877 (1999).
15. L.-T. Nguyen, H. A. Schmidt, A. von Haeseler, B. Q. Minh, IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol. Biol. Evol.* **32**, 268-274 (2015).
16. S. Kalyaanamoorthy, B. Q. Minh, T. K. F. Wong, A. von Haeseler, L. S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587 (2017).
17. J. Soubrier *et al.*, The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Mol. Biol. Evol.* **29**, 3345-3358 (2012).
18. D. T. Hoang, O. Chernomor, A. von Haeseler, B. Q. Minh, L. S. Vinh, UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518-522 (2018).
19. S. Guindon *et al.*, New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307-321 (2010).
20. K. Tamura, G. Stecher, D. Peterson, A. Filipski, S. Kumar, MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* **30**, 2725-2729 (2013).
21. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009).
22. G. M. Boratyn *et al.*, BLAST: a more efficient report with usability improvements. *Nucleic Acids Res.* **41**, W29-W33 (2013).
23. B. Marin, Nested in the Chlorellales or independent class? Phylogeny and classification of the Pedinophyceae (Viridiplantae) revealed by molecular phylogenetic analyses of complete nuclear and plastid-encoded rRNA operons. *Protist* **163**, 778-805 (2012).
24. K. Fučíková *et al.*, New phylogenetic hypotheses for the core Chlorophyta based on chloroplast sequence data. *Front. Ecol. Evol.* **2**, 63 (2014).
25. M. Turmel, J.-C. de Cambiaire, C. Otis, C. Lemieux, Distinctive architecture of the chloroplast genome in the chlorodendrophycean green algae *Scherffelia dubia* and *Tetraselmis* sp. CCMP 881. *PLoS One* **11**, e0148934 (2016).
26. K. R. Mattox, K. D. Stewart, "Classification of the green algae: a concept based on comparative cytology" in Systematics of the green algae, D. E. G. Irvine, D. M. John, Eds. (Academic Press, London, 1984), pp. 29-72.
27. F. Leliaert *et al.*, Phylogeny and molecular evolution of the green algae. *Crit. Rev. Plant Sci.* **31**, 1-46 (2012).

28. H. J. Sluiman, The green algal class Ulvophyceae. An ultrastructural survey and classification. *Crypt. Bot.* **1**, 83-94 (1989).
29. J. M. López-Bautista, R. L. Chapman, Phylogenetic affinities of the Trentepohliales inferred from small-subunit rDNA. *Int. J. Syst. Evol. Microbiol.* **53**, 2099-2106 (2003).
30. S. Watanabe, T. Nakayama, Ultrastructure and phylogenetic relationships of the unicellular green algae *Ignatius tetrasporus* and *Pseudocharacium americanum* (Chlorophyta). *Phycol. Res.* **55**, 1-16 (2007).
31. E. Cocquyt, H. Verbruggen, F. Leliaert, O. De Clerck, Evolution and cytological diversification of the green seaweeds (Ulvophyceae). *Mol. Biol. Evol.* **27**, 2052-2061 (2010).
32. M. Turmel, C. Otis, C. Lemieux, Divergent copies of the large inverted repeat in the chloroplast genomes of ulvophycean green algae. *Sci. Rep.* **7**, 994 (2017).
33. F. Leliaert *et al.*, Systematics of the marine microfilamentous green algae *Uronema curvatum* and *Urospora microscopica* (Chlorophyta). *Eur. J. Phycol.* **44**, 487-496 (2009).
34. F. Leliaert, J. M. Lopez-Bautista, The chloroplast genomes of *Bryopsis plumosa* and *Tydemania expeditionis* (Bryopsidales, Chlorophyta): compact genomes and genes of bacterial origin. *BMC Genomics* **16**, 204 (2015).
35. L. Fang *et al.*, Improving phylogenetic inference of core Chlorophyta using chloroplast sequences with strong phylogenetic signals and heterogeneous models. *Mol. Phylogenet. Evol.* **127**, 248–255 (2018).
36. C. J. O'Kelly, G. L. Floyd, Flagellar apparatus absolute orientations and the phylogeny of the green algae. *Biosystems* **16**, 227-251 (1984).
37. G. H. Gile, P. M. Novis, D. S. Cragg, G. C. Zuccarello, P. J. Keeling, The distribution of Elongation Factor-1 alpha (EF-1 α), Elongation Factor-Like (EFL), and a non-canonical genetic code in the Ulvophyceae: discrete genetic characters support a consistent phylogenetic framework. *J. Eukaryot. Microbiol.* **56**, 367-372 (2009).
38. D. Menzel, Cell differentiation and the cytoskeleton in *Acetabularia*. *New Phytol.* **128**, 369-393 (1994).
39. I. Mine, D. Menzel, K. Okuda, Morphogenesis in giant-celled algae. *Int. Rev. Cell Mol. Biol.* **266**, 37-83 (2008).
40. D. Sabnis, W. Jacobs, Cytoplasmic streaming and microtubules in the coenocytic marine alga, *Caulerpa prolifera*. *J. Cell Sci.* **2**, 465-472 (1967).
41. K. Kuroa, E. Manabe, Microtubule-associated cytoplasmic streaming in *Caulerpa*. *Proceedings of the Japan Academy, Series B* **59**, 131-134 (1983).
42. H. Vogel, G. E. Grieninger, K. H. Zetsche, Differential messenger RNA gradients in the unicellular alga *Acetabularia acetabulum*. Role of the Cytoskeleton. *Plant Physiol.* **129**, 1407-1416 (2002).
43. I. Mine, Y. Aota, D. Menzel, K. Okuda, Poly(A)⁺ RNA and cytoskeleton during cyst formation in the cap ray of *Acetabularia peniculus*. *Protoplasma* **226**, 199-206 (2005).
44. P. Škaloud, F. Rindi, C. Boedeker, F. Leliaert, *Freshwater Flora of Central Europe, Vol 13: Chlorophyta: Ulvophyceae*, Süßwasserflora von Mitteleuropa (Springer Spektrum, Berlin, Heidelberg, 2018), vol. 13, pp. 288.
45. P. Škaloud, T. Kalina, K. Nemjová, O. De Clerck, F. Leliaert, Morphology and phylogenetic position of the freshwater green microalgae *Chlorochytrium* (Chlorophyceae) and *Scotinosphaera* (Scotinosphaerales, ord. nov., Ulvophyceae). *J. Phycol.* **49**, 115-129 (2013).
46. A. Ranjan, B. T. Townsley, Y. Ichihashi, N. R. Sinha, D. H. Chitwood, An intracellular transcriptomic atlas of the giant coenocyte *Caulerpa taxifolia*. *PLoS Genet.* **11**, e1004900 (2015).
47. N. Matasci *et al.*, Data access for the 1,000 Plants (1KP) project. *GigaScience* **3**, 17 (2014).
48. A. Del Cortona *et al.*, The plastid genome in Cladophorales green algae is encoded by hairpin chromosomes. *Curr. Biol.* **27**, 3771-3782.e3776 (2017).
49. Q. Li, J. Liu, L. Zhang, Q. Liu, De novo transcriptome analysis of an aerial microalga *Trentepohlia jolithus*: pathway description and gene discovery for carbon fixation and carotenoid biosynthesis. *PLoS One* **9**, e108488 (2014).
50. X. Zhang *et al.*, De novo sequencing and analysis of the *Ulva linza* transcriptome to discover putative mechanisms associated with its successful colonization of coastal ecosystems. *BMC Genomics* **13** (2012).
51. O. De Clerck *et al.*, Insights into the evolution of multicellularity from the Sea Lettuce genome. *Curr. Biol.* **28**, 2921-2933.e2925 (2018).
52. S. S. Merchant *et al.*, The *Chlamydomonas* Genome Reveals the Evolution of Key Animal and Plant Functions. *Science* **318**, 245 (2007).
53. P. J. Keeling *et al.*, The Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP): illuminating the functional diversity of eukaryotic life in the oceans through transcriptome sequencing. *PLOS Biology* **12** (2014).

54. E. R. Hanschen *et al.*, The *Gonium pectorale* genome demonstrates co-option of cell cycle regulation during the evolution of multicellularity. *Nat. Comm.* **7**, 11370 (2016).
55. Z. Gao *et al.*, Transcriptome Analysis in *Haematococcus pluvialis*: Astaxanthin Induction by Salicylic Acid (SA) and Jasmonic Acid (JA). *PLOS ONE* **10**, e0140609 (2015).
56. S. E. Prochnik *et al.*, Genomic Analysis of Organismal Complexity in the Multicellular Green Alga *Volvox carteri*. *Science* **329**, 223 (2010).
57. G. Blanc *et al.*, The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation. *Genome Biology* **13**, R39 (2012).
58. J. Zhang *et al.*, Overexpression of the soybean transcription factor GmDof4 significantly enhances the lipid content of *Chlorella ellipsoidea*. *Biotechnology for Biofuels* **7**, 128 (2014).
59. F. C. Carniel *et al.*, New features of desiccation tolerance in the lichen photobiont *Trebouxia gelatinosa* are revealed by a transcriptomic approach. *Plant Molecular Biology* **91**, 319-339 (2016).
60. C. Gao *et al.*, Oil accumulation mechanisms of the oleaginous microalga *Chlorella protothecoides* revealed through its genome, transcriptomes, and proteomes. *BMC Genomics* **15**, 582 (2014).
61. G. Blanc *et al.*, The *Chlorella variabilis* NC64A Genome Reveals Adaptation to Photosymbiosis, Coevolution with Viruses, and Cryptic Sex. *The Plant Cell* **22**, 2943 (2010).
62. H. Moreau *et al.*, Gene functionalities and genome structure in *Bathycoccus prasinos* reflect cellular specializations at the base of the green lineage. *Genome Biology* **13**, R74 (2012).
63. A. Z. Worden *et al.*, Green Evolution and Dynamic Adaptations Revealed by Genomes of the Marine Picoeukaryotes *Micromonas*. *Science* **324**, 268 (2009).
64. B. Palenik *et al.*, The tiny eukaryote *Ostreococcus* provides genomic insights into the paradox of plankton speciation. *Proceedings of the National Academy of Sciences* **104**, 7705 (2007).
65. R. E. Timme, T. R. Bachvaroff, C. F. Delwiche, Broad phylogenomic sampling and the sister lineage of land plants. *PLOS ONE* **7** (2012).
66. C. Ju *et al.*, Conservation of ethylene as a plant hormone over 450 million years of evolution. *Nature Plants* **1**, 14004 (2015).
67. The Arabidopsis Genome Initiative, Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796 (2000).
68. International Rice Genome Sequencing Project, The map-based sequence of the rice genome. *Nature* **436**, 793 (2005).
69. S. A. Rensing *et al.*, The *Physcomitrella* Genome Reveals Evolutionary Insights into the Conquest of Land by Plants. *Science* **319**, 64 (2008).
70. J. A. Banks *et al.*, The Selaginella Genome Identifies Genetic Changes Associated with the Evolution of Vascular Plants. *Science* **332**, 960 (2011).
71. S. Magallón, K. W. Hilu, D. Quandt, Land plant evolutionary timeline: gene effects are secondary to fossil constraints in relaxed clock estimation of age and substitution rates. *Am. J. Bot.* **100**, 556-573 (2013).
72. G. K. Colbath, H. R. Grenfell, Review of biological affinities of Paleozoic acid-resistant, organic-walled eukaryotic algal microfossils (including "acritarchs"). *Rev. Palaeobot. Palynol.* **86**, 287-314 (1995).
73. H. Verbruggen *et al.*, A multi-locus time-calibrated phylogeny of the siphonous green algae. *Mol. Phylogenet. Evol.* **50**, 642-653 (2009).
74. N. J. Butterfield, A. H. Knoll, K. Swett, Paleobiology of the Neoproterozoic Svanbergfjellet Formation, Spitsbergen. *Fossils and Strata* **34**, 1-84 (1994).
75. M. D. Herron, J. D. Hackett, F. O. Aylward, R. E. Michod, Triassic origin and early radiation of multicellular volvocine algae. *Proc. Natl Acad. Sci. U.S.A.* **106**, 3254-3258 (2009).
76. L. W. Parfrey, D. J. G. Lahr, A. H. Knoll, L. A. Katz, Estimating the timing of early eukaryotic diversification with multigene molecular clocks. *Proc. Natl Acad. Sci. U.S.A.* **108**, 13624-13629 (2011).