New phylogenetic hypotheses for the core Chlorophyta based on chloroplast sequence data

Karolina Fučíková[1], Frederik Leliaert[2,3], Endymion D. Cooper[4], Pavel Škaloud[5], Sofie D'Hondt[2], Olivier De Clerck[2], Carlos F. D. Gurgel[6], Louise A. Lewis[1], Paul O. Lewis[1], Juan M. Lopez-Bautista[3], Charles F. Delwiche[4] and Heroen Verbruggen[7]*

[1] Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, CT, USA
[2] Phycology Research Group, Biology Department, Ghent University, Ghent, Belgium
[3] Department of Biological Sciences, The University of Alabama, Tuscaloosa, AL, USA
[4] Department of Cell Biology and Molecular Genetics and the Maryland Agricultural Experiment Station, University of Maryland, College Park, MD, USA
[5] Department of Botany, Faculty of Science, Charles University in Prague, Prague, Czech Republic
[6] School of Earth and Environmental Sciences, The University of Adelaide, Adelaide, SA, Australia
[7] School of Botany, University of Melbourne, Melbourne, VIC, Australia

Phylogenetic relationships in the green algal phylum Chlorophyta have long been subject to debate, especially at higher taxonomic ranks (order, class). The relationships among three traditionally defined and well-studied classes, Chlorophyceae, Trebouxiophyceae, and Ulvophyceae are of particular interest, as these groups are species-rich and ecologically important worldwide. Different phylogenetic hypotheses have been proposed over the past two decades and the monophyly of the individual classes has been disputed on occasion. Our study seeks to test these hypotheses by combining high throughput sequencing data from the chloroplast genome with increased taxon sampling. Our results suggest that while many of the deep relationships are still problematic to resolve, the classes Trebouxiophyceae and Ulvophyceae are likely not monophyletic as currently defined. Our results also support relationships among several trebouxiophycean taxa that were previously unresolved. Finally, we propose that the common term for the grouping of the three classes, "UTC clade," be replaced with the term "core Chlorophyta" for the well-supported clade containing Chlorophyceae, taxa belonging to Ulvophyceae and Trebouxiophyceae, and the classes Chlorodendrophyceae and Pedinophyceae.

Keywords: chloroplast DNA, Chlorophyta, fast site removal, green algae, multi-gene phylogeny, molecular systematic, phylogenomics, Viridiplantae
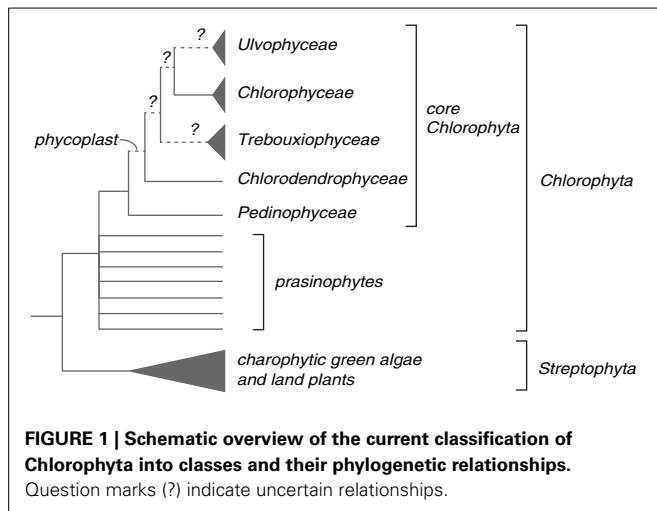
INTRODUCTION

New lineages of green algae are discovered every year as unusual habitats are surveyed and cryptic diversity unveiled. This has resulted in an explosion of new taxonomy. Within the phylum Chlorophyta, most of this taxonomic progress is made on the level of species and genera, whereas studies above this taxonomic level are less frequent and often lack consensus (e.g., Carlile et al., 2011; Boedeker et al., 2012; Neustupa et al., 2013a,b; Fučíková et al., 2014a). As far as the higher classification is concerned, four classes have traditionally been defined using ultrastructural characters, namely flagellar apparatus configuration and features of the cell division process: Ulvophyceae, Trebouxiophyceae, Chlorophyceae and Prasinophyceae (e.g., Mattox and Stewart, 1984; Sluiman, 1989).

Analyses of molecular data, in some cases supported by ultra-structural information, have shown that the unicellular planktonic Prasinophyceae form a paraphyletic group of about ten lineages, only some of which have been formally described as classes (Marin and Melkonian, 2010; Leliaert et al., 2011). Ancestors of the extant prasinophytes gave rise to the morphologically and ecologically diverse core Chlorophyta, which

include three major classes: Ulvophyceae, Trebouxiophyceae and Chlorophyceae (UTC). In addition, two smaller, early diverging lineages of core Chlorophyta have been identified: the Chlorodendrophyceae and Pedinophyceae (Moestrup, 1991; Massjuk, 2006; Leliaert et al., 2012; Marin, 2012) (**Figure 1**). The UTC classes plus the Chlorodendrophyceae are characterized by a novel mode of cell division, mediated by a phycoplast, which is absent in the prasinophytes and Pedinophyceae, and secondarily lost in the Ulvophyceae (**Figure 1**; reviewed in Leliaert et al., 2012; Marin, 2012).

The class Chlorophyceae comprises diverse morphological forms that predominantly inhabit freshwater and terrestrial habitats. Five orders are recognized within the class (Oedogoniales, Chaetophorales, Chaetopeltidales, Volvocales and Sphaeropleales), and the relationships among them are fairly well understood (Booton et al., 1998; Buchheim et al., 2001; Wolf et al., 2002; Turmel et al., 2008; Tippery et al., 2012). By contrast, higher classification within Ulvophyceae and especially Trebouxiophyceae is still unresolved.

Ulvophyceans, most of which are marine multicellular and/or macroscopic algae, are currently divided into seven orders,

FIGURE 1 | Schematic overview of the current classification of Chlorophyta into classes and their phylogenetic relationships. Question marks (?) indicate uncertain relationships.

with several additional genera of unknown affiliation. Most single and multigene phylogenies have provided weak or no support for monophyly of the class, but have instead supported two clades: one containing Oltmannsiellopsidales, Ulvales and Ulotrichales, the other consisting of Trentepohliales, Bryopsidales, Dasycladales and Cladophorales (e.g., Watanabe and Nakayama, 2007; Cocquyt et al., 2009; Lü et al., 2011). The earliest study providing stronger support for a monophyletic Ulvophyceae was based on phylogenetic analysis of eight nuclear and two plastid genes (Cocquyt et al., 2010). Analyses of these data also resulted in good resolution and support for most nodes in the phylogeny of the class.

Trebouxiophycean algae are of special importance to ecologists and evolutionary biologists because of their affinity for terrestrial habitats and symbiotic lifestyles. This group is known for the convoluted taxonomic histories of its taxa, which are probably the result of convergent evolution on morphologically simple forms. Taxon-rich 18S phylogenies often only provide weak support for a monophyletic Trebouxiophyceae (e.g., Neustupa et al., 2011), although analyses of the complete nuclear-encoded rRNA operon have provided somewhat stronger support (Marin, 2012). The orders Chlorellales, Microthamniales, and Trebouxiales are well defined, but most of the trebouxiophycean diversity lacks higher-level classification, i.e., many genera are not affiliated with a particular order or family. Leliaert et al. (2012) recognized seven major trebouxiophycean lineages, but recent publications describing novel genera suggest that there are perhaps as many as 16 distinct lineages in Trebouxiophyceae, and that the relationships among these lineages are difficult to resolve with data from one or a small number of genes (Neustupa et al., 2011, 2013a,b; Gaysina et al., 2013; Fučíková et al., 2014b).

It is obvious from the information above that the monophyly of the UTC classes and the relationships within the core Chlorophyta have been subject to debate (Leliaert et al., 2012). Phylogenetic support for the Chlorophyceae clade is high in most single- and multi-gene molecular phylogenetic studies, but the monophyly of Ulvophyceae and Trebouxiophyceae is generally poorly supported (e.g., Lü et al., 2011; Novis et al., 2013) (Supplement S1). Nevertheless, the three classes have been treated as monophyletic so far, and as their respective relationships were

explored, all three possible relationships among the classes have been proposed, depending on gene and taxon sampling and phylogenetic methods used (Supplement S1). From an ultrastructural perspective, the shared presence of a non-persistent mitotic spindle may be interpreted as providing support for a relationship between Chlorophyceae and Trebouxiophyceae (Mattox and Stewart, 1984), whereas the counter-clockwise orientation of the flagellar apparatus unites the Trebouxiophyceae and Ulvophyceae (Sluiman, 1989).

Thus, far, attempts to reconstruct the deep phylogenetic relationships of Chlorophyta have seldom yielded a well-supported backbone of the phylogeny. Studies with deep divergences in Chlorophyta generally suffer from sparse and/or uneven taxon sampling, or limited gene sampling (Supplement S1). Most likely, an approach combining dense and even taxon sampling (representing all major lineages) with data from multiple loci will be necessary to arrive at a well-supported phylogeny. In the present study, we seek to evaluate the monophyly of Trebouxiophyceae, Ulvophyceae, and Chlorophyceae, the relationships among them, and assess the relationships among major lineages within the classes. In doing so, we generate new phylogenetic hypotheses about the relationships in the core Chlorophyta using chloroplast sequence datasets in addition to the traditionally used 18S gene. Our data sets feature the best balance to date between gene and taxon sampling.

MATERIALS AND METHODS
GENERAL APPROACH
We took a two-pronged phylogenetic approach to address the goals described above. Firstly, we analyzed a dataset of eight genes [7 chloroplast genes + nuclear ribosomal small subunit rRNA gene (18S)] from a taxonomically dense sample of 56 taxa (22 trebouxiophyceans, 11 ulvophyceans and 8 chlorophyceans, as well as 12 prasinophyceans including one representative of Chlorodendrophyceae and two of Pedinophyceae). Second, we used chloroplast genome data to compile a 53-gene dataset for a smaller set of taxa (10 trebouxiophyceans, 8 ulvophyceans, 8 chlorophyceans, and 9 prasinophyceans). In both cases, data from a broad selection of prasinophyte lineages were used to root the phylogeny. The more extensive taxon sampling of the 8-gene dataset served to evaluate the monophyly of the classes and assess within-class relationships. The purpose of the genome-scale dataset was to determine whether or not data from additional genes (although available from fewer taxa) would improve the resolution in the phylogeny, and particularly, increase support for the deeper divergences.

DATA ACQUISITION
For the 8-gene dataset, we obtained *atp*A, *psa*A, *psa*B, *psb*A, *psb*C, *rbc*L, *tuf*A, and 18S rDNA sequences using several techniques and sources. For 12 taxa, the chloroplast genes were determined using PCR and Sanger sequencing (detailed methods in Supplement S2). For 7 taxa, they were extracted from assemblies of high-throughput sequencing data generated by the authors (Supplement S2). For 34 taxa (including prasinophycean outgroups), they were downloaded from Genbank and the Community Cyberinfrastructure for Advanced Microbial

Ecology Research and Analysis (Sun et al., 2011). Supplement S3 shows an overview of the data sources. The 7 plastid genes were chosen because they are among the slowest evolving genes in the chloroplast genome (Pombert et al., 2006) and therefore most appropriate for a study focused on ancient divergences.

The taxon sampling covers eight of the major trebouxiophycean lineages (incl. all seven depicted in Leliaert et al., 2012), six out of 11 ulvophycean lineages (Leliaert et al., 2012), all five chlorophycean orders, and 12 prasinophytes. All 18S sequences were downloaded from Genbank (Supplement S3).

For the genome-scale dataset, we obtained complete or draft chloroplast genome sequences from 7 taxa (Supplement S2). Data for an additional 31 taxa were extracted from chloroplast genomes or genome fragments available on Genbank (Supplement S4). We extracted sequences for 53 chloroplast protein-coding genes (Supplement S4).

SEQUENCE ALIGNMENT

Chloroplast gene sequence data were aligned for each gene separately using the ClustalW translational alignment function (Larkin et al., 2007) in Geneious v.R6 (Biomatters, www.geneious.com) and the 18S gene sequences were aligned using default ClustalW settings. The single-gene alignments resulting from this were concatenated to produce the two initial alignments, i.e., the 8-gene alignment with dense taxon sampling and the genome-scale alignment with sparser taxon sampling.

From these two alignments, we created several derivatives, summarized in **Table 1**. First, we removed the 18S from the 8-gene alignment in order to evaluate the signal in the chloroplast genes without the influence of 18S. We also removed unreliably aligned, hypervariable sites from the genome-scale data set where large variable regions were present. The Gblocks server (Castresana, 2000; http://molevol.cmima.csic.es/castresana/Gblocks_server.html) removed 10,870 of the total 20,177 codons, including the entire *psa*M, *rpl*32, and the large and highly variable *ycf*1. The least stringent settings on the Gblocks server were used: allowing smaller final blocks, gap positions within the final blocks, less strict flanking positions and many contiguous non-conserved positions.

Second, we used a fast site removal approach to improve the signal-to-noise ratio for deep divergences by focusing on

slowly evolving positions that should provide more information about ancient events (Waddell et al., 1999; Delsuc et al., 2005; Rodríguez-Ezpeleta et al., 2007a). The fast site removal procedure consisted of: (1) inferring a preliminary guide phylogeny with RAxML v.7.3.5 (Stamatakis, 2006) using a GTR+I+Γ model partitioned into codon positions for coding genes, (2) calculating site-specific evolutionary rates with the "site rates" standard analysis in HyPhy v.2.2 (Kosakovsky Pond et al., 2005) and using the RAxML guide tree from point 1, (3) creating alignments containing the 50, 55, 60, 65, 70, 75, 80, 85, 90, and 95% slowest sites in the alignments using SiteStripper v.1.01 (Verbruggen, 2012) and the list of site-specific rates calculated at step 2. This procedure was applied separately to the 7-gene, 8-gene and genome-scale datasets.

PHYLOGENETIC INFERENCE

We inferred maximum likelihood trees from the various alignments using RAxML v.7.3.5 (Stamatakis, 2006), specifying a GTR+I+Γ model and a partitioning strategy in which codon positions were separated (3 partitions total). For the 8-gene dataset, the 18S gene had its own partition (4 partitions total). Branch support was assessed by bootstrapping (Felsenstein, 1985), with 1000 replicates for the 8-gene dataset and 500 for the genome-scale dataset.

Additionally, we conducted Bayesian analyses for the 80% complete (20% site-stripped) data sets (7G-80%, 8G-80%, GS-80%) using MrBayes v.3.2.1 (Huelsenbeck and Ronquist, 2001; Ronquist and Huelsenbeck, 2003), running each analysis for 10,000,000 generations with 1 cold and 3 heated chains and sampling each 1000 generations. Two independent runs were performed for each data set. The same partitioning strategies and model of evolution were used as in the ML analyses described above, and default settings were used unless specified otherwise above. The first 10% of samples were discarded as burn-in. Convergence of the runs and stability of parameters were assessed using Tracer v.1.5 (Rambaut and Drummond, 2007). Bayesian analyses using Phycas (Lewis et al., 2011) allowing polytomous trees were also carried out on the 8G-80% and GS-80% data sets. Bayesian analyses using Phycas (www.phycas.org) allowing polytomous trees were also carried out on the 8G-80% and GS-80% data sets. Details on the analysis settings and results are given in Supplement S7.

TOPOLOGY TESTING

To evaluate whether alternative topologies yield significantly worse (lower likelihood) trees than our ML trees, we performed a series of analyses with constraints on the tree topology (**Table 2**) using RAxML. We performed these analyses for each complete data set (7G-100%, 8G-100%, GS-100%) as well as for each 80% complete data set (7G-80%, 8G-80%, GS-80%). The analysis settings used in the constrained analyses were identical to those in the unconstrained analyses (GTR+I+Γ model, partitioned by codon position). The best tree from each constrained analysis was compared to the ML tree from the corresponding unconstrained analysis using the Approximately Unbiased (AU) test in Consel (Shimodaira, 2004) with sitewise likelihoods calculated by RAxML.

Table 1 | Different types of alignments created.

Included genes	Fast site removal
7G = 7-gene	Percentage of slow-evolving sites retained: 100, 95, 90, 85, 80, 75, 70, 65, 60, 55, 50%
8G = 8-gene	
GS = genome-scale	

First, 3 alignments were generated that differed in the number of genes they included (column 1, 3 conditions). Second, fast site removal was applied to generate subalignments retaining the slowest sites (column 2, 11 conditions). Details of the conditions can be found in the main text. In total, 3 × 11 = 33 alignments were generated. We refer to particular alignments by their abbreviation, e.g., 7G-80% indicates the alignment with 7 genes and 80% slowest sites retained.

Table 2 | Results of the AU test investigating the monophyly of Ulvophyceae, Trebouxiophyceae, and the two classes with the exclusion of *Oltmannsiellopsis* and *Choricystis*, respectively.

	7-gene	8-gene	Genome-scale
COMPLETE DATA SETS			
Ulvophyceae	**<0.001**	**0.007**	**0.002**
Trebouxiophyceae	**0.017**	**0.032**	n/a
Ulvophyceae w/o *Oltmannsiellopsis*	0.257	n/a	0.100
Trebouxiophyceae w/o *Choricystis*	0.326	n/a	n/a
Unconstrained	Best	Best	Best
80% COMPLETE DATASETS			
Ulvophyceae	**<0.001**	**0.002**	**<0.001**
Trebouxiophyceae	**<0.001**	**0.007**	n/a
Ulvophyceae w/o *Oltmannsiellopsis*	0.062	0.166	**0.001**
Trebouxiophyceae w/o *Choricystis*	0.308	0.355	0.448
Unconstrained	Best	Best	Best

p-values lower than 0.05 are highlighted in boldface to emphasize a statistically significant result, i.e., the constrained topology being significantly worse than unconstrained. "n/a" indicates cases where the unconstrained topology already indicates monophyly of the group in question, or where the critical taxon (Choricystis) is absent from the data set.

SUMMARIZING BOOTSTRAP SUPPORT

The fast site removal strategy described above led to a total of 33 alignments being subjected to ML analyses (11 each for the 7-gene, 8-gene and genome-scale alignments). In order to investigate the effect of fast site removal on the consistency of the phylogenetic signal for early branching events in the core Chlorophyta, we designed the following procedure. First, a script was written to extract bootstrap values for the branches comprising the backbone of the ingroup, including those branches within the core Chlorophyta that occur deeper down in the tree than the well-recognized lineages (Pedinophyceae, Chlorophyceae, Ulvales-Ulotrichales, Trentepohliales, Dasycladales, Bryopsidales, Trebouxiales, Chlorellaceae, Oocystaceae, *Prasiola* lineage, *Coccomyxa-Hemichloris* lineage, *Watanabea*, *Leptosira*, *Tetraselmis*, *Oltmannsiellopsis*). The script was applied to the 33 ML trees annotated with bootstrap values and the average bootstrap support across the backbone branches was calculated. Second, we calculated the rate of evolution of each site-stripped alignment by obtaining and averaging the rate of evolution for each site using HyPhy (Kosakovsky Pond et al., 2005). Because we wanted rates to be comparable across all our alignments, we (1) calculated them for the taxa in common between our alignments, (2) used a fixed reference phylogeny, and (3) used the HyPhy sitewise rate calculation script used for PhyDesign (Lopez-Giraldez and Townsend, 2011). To achieve the first point, we reduced our 7-gene and 8-gene alignments to the taxa present in the genome-scale alignment. A coarse reference chronogram needed for the procedure was obtained by applying penalized likelihood rate smoothing in APE, with the root age set to 1 (Sanderson, 2002; Paradis et al., 2004). With this information at hand for each alignment, we plotted the average of the obtained bootstrap values as a function of the average site rate in the alignment.

RESULTS
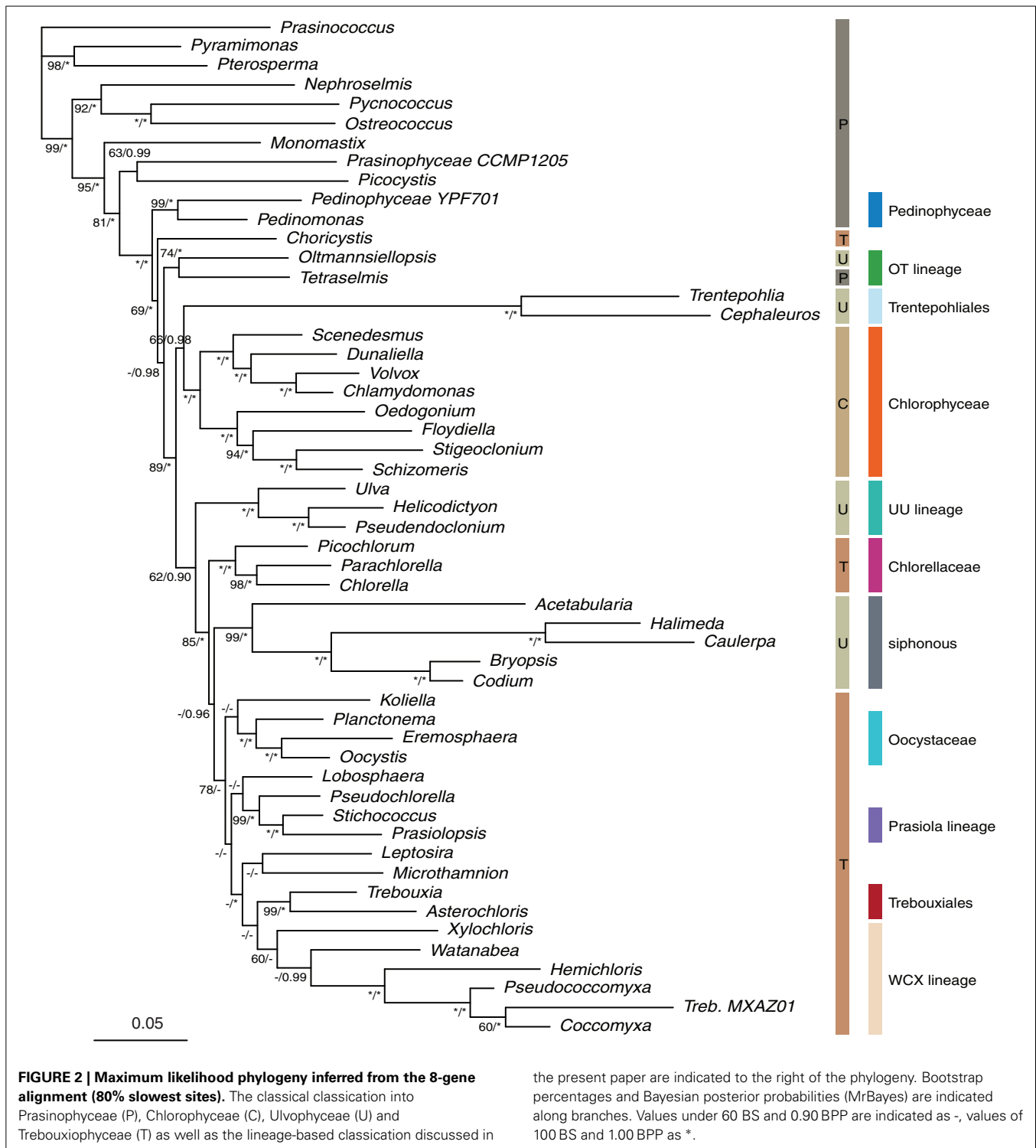
Our 8-gene dataset consisted of 53 taxa and was 91% complete at the gene × taxon level (data distribution summarized in Supplement S3). The genome-scale dataset consisted of 38 taxa and was 75% complete and no pair of taxa had completely non-overlapping sets of sequence data (Supplement S4). The lower completeness of this dataset was in part due to the inclusion of several prasinophyte outgroup taxa for which complete chloroplast genomes were unavailable. Within the ingroup, the matrix was 79% complete. All alignments and inferred ML trees were deposited in TreeBase (study number 16,203). Alignment lengths were 11,286 nucleotides for the 7-gene dataset, 13,301 nucleotides for the 8-gene dataset and 27,921 nucleotides (after GBlocks) for the genome-scale dataset.

All data sets were partitioned by codon position and in the 8-gene data sets 18S was given a separate partition. The GTR+I+Γ model of evolution was used for all subsets.

The phylogenetic trees resulting from ML analyses on the 8G-80% and GS-80% datasets are presented in **Figures 2** and **3**, respectively, as these trees were overall best resolved and supported (**Figure 4**). The remaining trees are presented in Supplement S5 and are considered in our treatment of the results.

None of our analyses recovered the UTC clade as it is currently defined. However, the clade containing all ulvophyceans, trebouxiophyceans, and chlorophyceans together with *Tetraselmis* and representatives of the Pedinophyceae received good support (generally >90 BS and >0.95 BPP, often 100/1.00) and was consistently recovered in all analyses. Pedinophytes were recovered as sister to the remainder of the core Chlorophyta with high support (**Figures 2**, **3**). The class Chlorophyceae was recovered as monophyletic and strongly supported regardless of how the data were filtered or analyzed, and the relationships among orders within the class were also recovered as currently accepted. The classes Ulvophyceae and Trebouxiophyceae were not recovered as monophyletic groups in most analyses. In the GS-100% data set, monophyly of Trebouxiophyceae was weakly supported (50% BS). With 10% or more of the fastest sites being stripped, Chlorellaceae separated from the rest of the trebouxiophycean representatives (Supplements S5, S6), as was the case in the MrBayes analysis of the GS-80% data set. However, Trebouxiophyceae was monophyletic in the Phycas analysis of the GS-80% data set (Supplement S7). Ulvophyceae were not found to be monophyletic in any analysis. The 7G-100% and 7G-80% analyses with denser taxon sampling showed a similar picture without support for the traditional classes Trebouxiophyceae and Ulvophyceae. The addition of 18S to this dataset (8G-80%) produced a phylogeny, in which Trebouxiophyceae (with the exception of *Choricystis*) and Ulvophyceae (with the exception of *Oltmannsiellopsis*) were recovered as monophyletic groups, but with poor branch support (35% BS support in both cases). The MrBayes analysis of 8G-80% dataset yielded a topology identical to the ML analysis with most nodes being well supported.

Constraining Ulvophyceae as monophyletic resulted in a significantly worse tree in all our AU tests (**Table 2**). However, excluding *Oltmannsiellopsis* from the constrained ulvophycean clade yielded a tree that was not significantly worse than the unconstrained tree except in GS-80% (**Table 2**). Similarly,
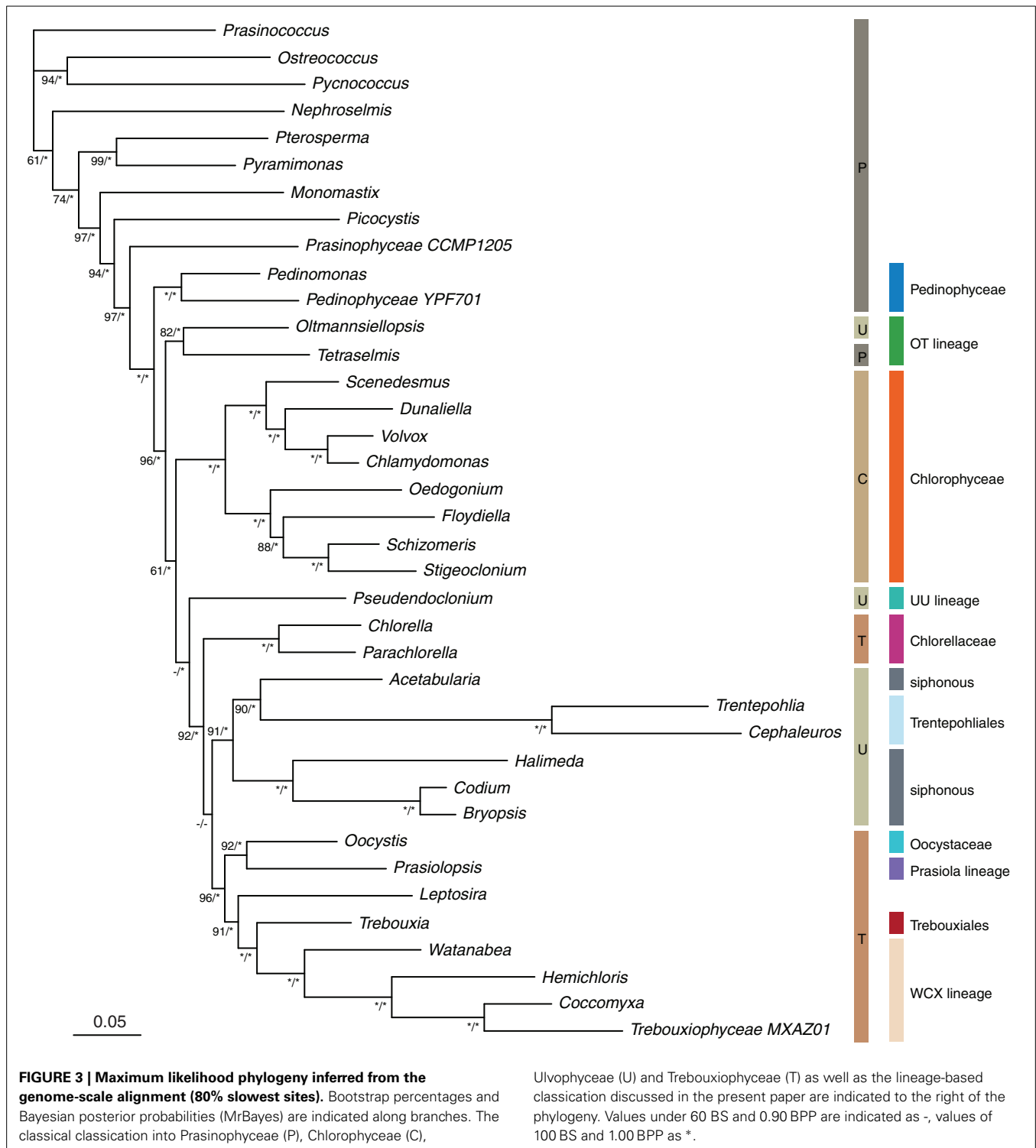
FIGURE 2 | Maximum likelihood phylogeny inferred from the 8-gene alignment (80% slowest sites). The classical classification into Prasinophyceae (P), Chlorophyceae (C), Ulvophyceae (U) and Trebouxiophyceae (T) as well as the lineage-based classification discussed in the present paper are indicated to the right of the phylogeny. Bootstrap percentages and Bayesian posterior probabilities (MrBayes) are indicated along branches. Values under 60 BS and 0.90 BPP are indicated as -, values of 100 BS and 1.00 BPP as *.

enforcing a monophyletic Trebouxiophyceae yielded a significantly worse result for the 7- and 8-gene analyses, but after exclusion of *Choricystis*, the monophyly of the remaining trebouxiophyceans was not rejected (**Table 2**). *Choricystis* was not included in the genome scale analysis.

Among the early-branching lineages of the core Chlorophyta is the *Oltmannsiellopsis + Tetraselmis* pair (labeled as "OT lineage"
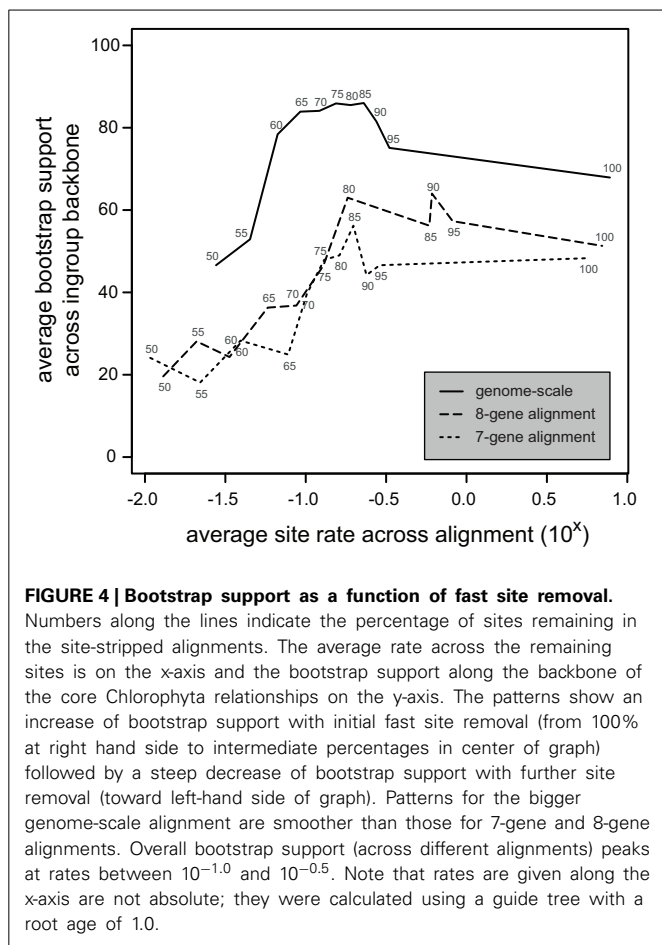
in **Figures 2**, **3**). This lineage was recovered consistently in all analyses, with variable support.

At lower taxonomic levels, several phylogenetic hypotheses emerge from our data. Among trebouxiophyceans, the *Coccomyxa* clade was recovered with *Xylochloris* and *Watanabea*, together labeled as the "WCX lineage" (**Figures 2**, **3**). This clade was very stable and persisted in all analyses except for the highly

FIGURE 3 | Maximum likelihood phylogeny inferred from the genome-scale alignment (80% slowest sites). Bootstrap percentages and Bayesian posterior probabilities (MrBayes) are indicated along branches. The classical classication into Prasinophyceae (P), Chlorophyceae (C), Ulvophyceae (U) and Trebouxiophyceae (T) as well as the lineage-based classication discussed in the present paper are indicated to the right of the phylogeny. Values under 60 BS and 0.90 BPP are indicated as -, values of 100 BS and 1.00 BPP as *.

site-stripped 7- and 8-gene data sets. The unidentified trebouxiophyte strain MX-AZ01, whose organelle genomes were recently published (Servín-Garcidueñas and Martínez-Romero, 2012), is also a member of the WCX lineage. In the genome-scale analyses, Trebouxiales (represented only by *Trebouxia*) was strongly supported as a sister to the WCX lineage. This relationship was

also recovered in 7-gene analyses (90–75% site-stripped) and was present in most 8-gene analyses. A sister relationship between Chlorellaceae and Oocystaceae, commonly inferred by analyses of 18S (e.g., Pažoutová et al., 2010; Neustupa et al., 2013a), was not confirmed by our analyses. Instead, *Oocystis* and *Prasiolopsis* grouped together with moderate to good support (increasing with

FIGURE 4 | Bootstrap support as a function of fast site removal.
Numbers along the lines indicate the percentage of sites remaining in the site-stripped alignments. The average rate across the remaining sites is on the x-axis and the bootstrap support along the backbone of the core Chlorophyta relationships on the y-axis. The patterns show an increase of bootstrap support with initial fast site removal (from 100% at right hand side to intermediate percentages in center of graph) followed by a steep decrease of bootstrap support with further site removal (toward left-hand side of graph). Patterns for the bigger genome-scale alignment are smoother than those for 7-gene and 8-gene alignments. Overall bootstrap support (across different alignments) peaks at rates between $10^{-1.0}$ and $10^{-0.5}$. Note that rates are given along the x-axis are not absolute; they were calculated using a guide tree with a root age of 1.0.

progressive site-stripping and 1.00 BPP in the MrBayes analysis) in the genome-scale analyses. A weakly supported clade containing Oocystaceae and the members of the *Prasiola* clade was also recovered in analyses of moderately site-stripped data sets (7G-95–80%, 8G-100–85%).

Leptosira was recovered sister to *Microthamnion* in the 7- and 8-gene analyses with moderate and good support, respectively. The support for this relationship decreased with increased site-stripping in both cases. The effect of site-stripping on bootstrap support of selected clades in all three data sets is summarized in Supplement S6.

The relationships uncovered among the ulvophycean orders differed markedly between analyses. The only relationship between orders that was consistently recovered was that between the Ulvales and Ulotrichales. The two siphonous orders Dasycladales and Bryopsidales grouped together in some analyses, and even with very high bootstrap support in the 8-gene dataset. The two species of Trentepohliales were subtended by long branches, even in the site-stripped analyses, suggesting that not only fast positions account for the length of the branches. The placement of Trentepohliales was unstable across trees, either being among the early-branching lineages of the core Chlorophyta radiation, or as sister to *Acetabularia* (Dasycladales) in some analyses. The latter relationship achieved high support in the genome-scale dataset with moderate amounts of site stripping

(BS ≥ 90 in GS-65–80% and 1.00 BPP in the MrBayes analysis of GS-80%).

Lastly, *Oltmanniellopsis*, which is currently also classified as an ulvophyte, groups with *Tetraselmis* elsewhere in the tree, and it does this consistently.

The MrBayes analyses of the 7-gene and 8-gene datasets (7G-80%, 8G-80%) generally supported the relationships that were well-supported by the ML analyses (Bayesian posterior probabilities, BPP, on **Figure 2**). The MrBayes analysis of the genome-scale dataset (GS-80%) yielded a tree with a well-supported backbone and supported many of the shallower relationships including the class Chlorophyceae and relationships within it. The Bayesian analyses using the polytomy prior in Phycas (Figure S7) yield polytomies in the 8-gene but not the genomic-scale datasets, further supporting the view that the 8-gene dataset contains less topological information than the genome-scale dataset.

The fast site removal approach showed a gentle increase in bootstrap support across the ingroup backbone for intermediate levels of site removal followed by a steep decrease of bootstrap support with further site removal, indicating an initial improvement of the signal to noise ratio followed by loss of signal as the proportion of removed sites approaches 50% (**Figure 4**). Overall bootstrap values are substantially higher and trends were less erratic for the genome-scale dataset. Backbone bootstrap values are highest for all datasets for alignments with average site rates in the $10^{-0.5}$–$10^{-1.0}$ range. The trees we presented in **Figures 2, 3** are those consisting of the 80% slowest sites of the 8-gene and genome-scale datasets, both of which lie in this range of high signal to noise.

DISCUSSION

Our analyses in combination with other recent phylogenetic analyses of the Chlorophyta provide strong evidence that the traditional classification with four classes does not reflect the evolutionary history of the group. Efforts have already been made in formally classifying some of the early-branching prasinophyte lineages (Cavalier-Smith, 1993; Marin and Melkonian, 2010). Our analyses illustrate that similar efforts will eventually be needed to update the classification of the more derived Chlorophyta. It is clear that some widely accepted concepts, including the "UTC clade" and probably the existing classes are outdated. We propose use of the term "core Chlorophyta" to indicate the previous UTC taxa plus the Pedinophyceae and Chlorodendrophyceae (**Figure 3**). Our analyses agree with those of Marin (2012) and Matsumoto et al. (2011) in recognizing the core Chlorophyta. We recovered this relationship in every analysis.

Our analyses also support a sister relationship between *Tetraselmis* and *Oltmanniellopsis*, which was found branching early in the radiation of the core Chlorophyta. Previous analyses recovered *Tetraselmis*, together with *Scherffelia* (not included in our analyses), as sister to all members of the UTC classes (which include *Oltmanniellopsis*; e.g., Steinkötter et al., 1994; Nakayama et al., 1998). There are a number of ultrastructural similarities between *Tetraselmis* and *Oltmanniellopsis* that provide further indications for their relatedness. Both taxa have four flagella and a counter-clockwise orientation of the flagellar

basal bodies and rootlets (although this may be interpreted as a shared ancestral condition) and prominently striated rhizoplasts (Salisbury et al., 1981; Chihara et al., 1986; Lokhorst and Star, 1993). However, while both taxa have a closed mitosis, *Tetraselmis* performs cell division via a phycoplast, which is characteristic for Trebouxiophyceae and Chlorophyceae but not the Ulvophyceae. The presence or absence of a phycoplast has never been specifically examined in *Oltmannsiellopsis*, a current member of Ulvophyceae. The phylogenetic relatedness of the two genera was previously shown in the figures in Matsumoto et al. (2011) but was not discussed. We cannot exclude the possibility that this relationship is a result of phylogenetic artifact (i.e., long-branch attraction; LBA). We have sampled only one species from the Chlorodendrophyceae (the lineage to which *Tetraselmis* belongs) and one from the Oltmannsiellopsidales. This has led to both taxa being subtended by a long branch, increasing the potential for LBA. However, homoplasy leading to LBA is most likely to occur in fast-evolving sites and given that the *Oltmannsiellopsis-Tetraselmis* clade persisted in our site-stripped analyses makes the LBA explanation less likely. Further studies with additional samples from both these lineages will undoubtedly shed light on this issue.

Our placement of Pedinophyceae as sister to the remainder of core Chlorophyta corroborates the findings of Marin (2012) based on nuclear and plastid rRNA operons and contradicts the analyses of Turmel et al. (2009b) and Pombert and Keeling (2010), which placed *Pedinomonas* in the proximity of Chlorellales (Trebouxiophyceae) based on analyses of chloroplast and mitochondrial genome data. Our analyses hint at one possible explanation for this incongruence. With our expanded sampling, Chlorellaceae (but not Oocystaceae, see section below) are frequently placed outside of the remaining Trebouxiophyceae and instead represent an early-diverging lineage in the core Chlorophyta. Rather than assuming monophyletic Trebouxiophyceae and arguing whether or not pedinophytes are inside or outside of Trebouxiophyceae, we should consider other phylogenetic hypotheses, such as both Chlorellaceae and Pedinophyceae being placed outside of the remaining trebouxiophyceans. Our analyses, however, find no support for a sister relationship between Chlorellaceae and Pedinophyceae.

The example of the Pedinophyceae and Chlorellaceae emphasizes the necessity for studies that sample broadly across traditional class boundaries. To date, systematic studies have often focused on within-class relationships, ignoring the possibility that lineages from other classes may be interspersed between their focal taxa (Cocquyt et al., 2010; Neustupa et al., 2013a,b; Fučíková et al., 2014b). This limited focus is in part due to the very different nature of the groups—Trebouxiophyceae are mostly terrestrial unicells and Ulvophyceae are mostly marine seaweeds—with each requiring different sampling and culturing procedures not commonly found within individual research groups. The results of our study provide a clear signal that these boundaries need to be bridged and that broad sampling across different groups is a prerequisite for deriving a reliable phylogenetic classification of the core Chlorophyta.

Our results disagree not only with several single-gene studies (examples in Supplement S1) but notably also with two recent

multigene studies. The first was Cocquyt et al. (2010), which focused mainly on relationships within Ulvophyceae and used a similar site-removal approach to phylogenetic inference as the present study. In addition to a very different gene selection (2 plastid genes, 18S, and 7 nuclear protein-coding genes), differences in taxon sampling is a plausible explanation for the contradicting findings, as Cocquyt et al. (2010) only included Chlorellales (here shown to form a separate clade from the remaining trebouxiophytes) to represent Trebouxiophyceae and did not include either of the genera identified as most problematic in our study: *Choricystis* and *Oltmannsiellopsis*. The second notable case is Ruhfel et al. (2014), which focused primarily on land plants and used a genome-scale chloroplast data set to infer the phylogeny of Viridiplantae. In this case, only complete published chloroplast genomes were included in the analysis, and the study was therefore missing several critical taxa: among the 26 included Chlorophyta, Chlorodendrophyceae were not represented and the sampling of Trebouxiophyceae was limited to four Chlorellales, *Coccomyxa, Leptosira*, and *Oocystis*. In addition, only seven prasinophytes were included, omitting e.g. deeply branching genera *Picocystis* and *Prasinococcus*. Despite several topological differences from our results, Ruhfel et al. (2014) also recovered neither monophyletic Trebouxiophyceae nor monophyletic Ulvophyceae, but found strong evidence for monophyletic Chlorophyceae.

Our results, especially compared to the recent multigene studies discussed above, demonstrate that large datasets will be needed to resolve the early diversification of the core Chlorophyta and that the tradeoff in sampling intensity between number of loci and taxon density must be carefully considered. The ML support in the backbone of the core Chlorophyta remains fairly low in our analyses despite increased sampling of genes and taxa, with a handful of nodes receiving less than 70% bootstrap support even in our otherwise well-resolved genome-scale analyses (**Figure 3** showing good support for relationships in early Chlorophyta and high BPP support for most nodes). Genome-scale data can now be obtained at moderate cost with high-throughput sequencing (e.g., Glenn, 2011). In this context, the hypotheses formulated here lay the ground for a critical re-evaluation of deep phylogenetic relationships within the core Chlorophyta. The increase in support observed in our genome-scale analysis (compared to the 7- and 8-gene analyses) suggests that sequencing additional chloroplast genomes would be a reasonable strategy (**Figures 2–4**). Considering the discordance between our results and those of the nuclear gene-analyses of Cocquyt et al. (2010), complementing this chloroplast-based approach with transcriptome sequencing to obtain nuclear gene data would be a useful exercise. There is a possibility that the organellar phylogeny is legitimately discordant with the nuclear phylogeny, although introgression at such a high taxonomic level would be extraordinary. Another logical next step in pursuing these questions will be the inclusion of other critical lineages of Trebouxiophyceae and Ulvophyceae (details below), as well as testing and accounting for potential systematic bias in the data (e.g., nucleotide composition).

At a lower taxonomic level, we recovered the trebouxiophycean families Chlorellaceae and Oocystaceae (incl. *Planctonema*) as

strongly monophyletic. Contrary to most previously published phylogenies derived from 18S data (e.g., Pažoutová et al., 2010; Marin, 2012; Neustupa et al., 2013a), however, these two families were not joined as a clade. This finding is consistent with recent *rbc*L-based phylogenies of Neustupa et al. (2013a) and Fučíková et al. (2014b), and it is therefore possible that the inclusion of Oocystaceae in Chlorellales needs to be re-considered and additional data from nuclear and possibly mitochondrial markers may be helpful in making this taxonomic decision. Concatenation of data from 18S and *rbc*L tends to join Chlorellaceae and Oocystaceae into a clade (Neustupa et al., 2013b; Fučíková et al., 2014b), but such a result was not recovered by our 8-gene analyses, likely due to the prevalence of the chloroplast signal.

Our study is the first to confidently place *Watanabea*, *Coccomyxa* (and affiliates), and *Xylochloris* together (here referred to as the WCX lineage). Given the great diversity (especially ecological) of these taxa, it is difficult to identify any uniting characters for this group. Morphologically, the lineage mostly contains inconspicuous unicells, although colonial (*Botryococcus*) and siphonous (*Phyllosiphon*) forms are also found here. Ecologically, this clade spans from free-living to parasitic (*Phyllosiphon*) to symbiotic (*Elliptochloris*), terrestrial to freshwater, although terrestrial habit predominates.

The genus *Leptosira*, while mostly considered trebouxiophycean, has also been placed at the base of Chlorophyceae (Zuccarello et al., 2009; Turmel et al., 2009a,b). Contrary to these studies and consistent with recent publications on trebouxiophycean diversity (e.g., Neustupa et al., 2013a,b), our results place *Leptosira* with the majority of Trebouxiophyceae (**Figures 2**, **3**).

Placement of the genus *Choricystis* is problematic in the sense that it does not group with other Trebouxiophyceae. Constraining *Choricystis* plus the remaining trebouxiophyceans as monophyletic resulted in a significantly worse tree in all our AU tests. However, enforcing monophyly of trebouxiophyceans (but excluding *Choricystis*) produced a constrained tree that is not significantly worse than the unconstrained tree, indicating that monophyly of Trebouxiophyceae, exclusive of *Choricystis*, cannot be rejected. In this context, it is important to note that we only have *Choricystis* in the 7- and 8-gene datasets. In the genome-scale dataset (80% slowest sites), Trebouxiophyceae are recovered as non-monophyletic (**Figure 3**), but monophyly was not rejected by the AU test (**Table 2**). Studies using 18S generally place *Choricystis* in the WCX lineage, a group that also contains the prolific lipid producer *Botryococcus* (e.g., Neustupa et al., 2013a,b). However, our plastid phylogeny indicates that *Choricystis* may be a deeply diverging taxon in the core Chlorophyta, not affiliated with other trebouxiophycean or ulvophycean taxa (**Figure 2**, Supplement S5). The *rbc*L phylogeny presented by Neustupa et al. (2013a) also placed *Choricystis* outside of the *Coccomyxa/Elliptochloris* clade. Concatenation of the 18S and plastid data in Neustupa et al. (2013a) brought *Choricystis* into the proximity of the WCX lineage with moderate support. Analogously to the case of Chlorellaceae and Oocystaceae, the placement of *Choricystis* may represent a true conflict of signals between 18S and plastid data.

Despite being the most comprehensive to date, our study is missing several deeply diverging trebouxiophyte lineages. Because

deep relationships within Trebouxiophyceae remain uncertain (Neustupa et al., 2013a; Fučíková et al., 2014b), few lineages are represented by more than one genus in our study. Future studies should incorporate representatives of *Dictyochloropsis*, *Neocystis*, *Parietochloris*, and the recently described *Leptochlorella*, *Xerochlorella*, and *Eremochloris*, all of which represent deep divergences within the traditionally defined Trebouxiophyceae.

The class Chlorophyceae was strongly supported as a monophyletic group in all our analyses and the relationships among orders were congruent with previously published findings (e.g., Buchheim et al., 2012; Tippery et al., 2012). A group of taxa of uncertain affinities sometimes referred to as "Treubarinia" as well as the genera *Jenufa*, *Microspora*, *Parallela* and *Golenkinia* may form clades outside of the five established orders (Němcová et al., 2011) and should be included in future studies. However, these taxa are unlikely to disrupt the monophyly of Chlorophyceae, as they were previously shown to fall within the class (Němcová et al., 2011) and therefore do not affect the conclusions drawn in the present study.

The instability of the relationships between ulvophycean orders across datasets and analyses does not allow us to draw definitive conclusions about them. Nevertheless, topology tests do provide strong evidence against monophyly of Ulvophyceae as currently defined. This incongruence between the current circumscription and the phylogeny is primarily due to the inclusion of *Oltmannsiellopsis* in Ulvophyceae. As discussed above, we find good support for the placement of *Oltmannsiellopsis* in proximity to *Tetraselmis* (Chlorodendrophyceae). However, even with the exclusion of *Oltmannsiellopsis*, the hypothesis that the remaining Ulvophyceae are monophyletic is still significantly worse than the ML tree in the genome-scale data set (**Table 1**).

The relationships among ulvophycean orders may be affected by systematic bias. Representatives of the siphonous orders Bryopsidales (*Halimeda*, *Caulerpa*, *Codium*, *Bryopsis*) and Dasycladales (*Acetabularia*) are on long branches in our trees, and the branches leading up to the two Trentepohliales species (*Trentepohlia*, *Cephaleuros*) are even longer. In previous analyses based on partial nuclear encoded rDNA sequences (Zechman et al., 1990) as well as combined nuclear and chloroplast data (Cocquyt et al., 2010), the siphonous orders were recovered as more closely related to each other than to the Trentepohliales. Our analyses are split, with some recovering the expected sister relationship between the siphonous orders Dasycladales and Bryopsidales (e.g., **Figure 2**), while others provide strong support for a sister relationship between *Acetabularia* (Dasycladales) and Trentepohliales (**Figure 3**), rendering the siphonous green algae nonmonophyletic. While we cannot reject this topological configuration, we speculate that it is a result of LBA between *Acetabularia* and the Trentepohliales. If it were correct, it could imply a marine origin for the Trentepohliales, which are terrestrial or epiphytic, or possibly the existence of an as-yet uncharacterized lineage of freshwater relatives of the Dasycladales. Interestingly, the addition of 18S to the chloroplast data reduces the attraction of *Acetabularia* and Trentepohliales in some of our analyses (Supplement S6, compare right panel for 8-gene and 7-gene datasets). It will be critical for future studies to increase sampling in the Dasycladales and Trentepohliales and revisit the

relationships between these two orders as well as the Bryopsidales and Cladophorales.

Three major lineages of Ulvophyceae are missing from our datasets and their inclusion could have an impact on the relationships presented here. The Cladophorales (and *Blastophysa*) are an early-branching lineage shown in nuclear gene analyses to be related to Trentepohliales and the siphonous lineage (Cocquyt et al., 2010). The Cladophorales introduce an interesting problem for phylogenetic analysis of plastid datasets due to their atypical plastid genomes (La Claire et al., 1998; La Claire and Wang, 2000). The one trustworthy *rbc*L gene sequence published thus far is highly divergent from those of other Ulvophyceae (Deng et al., 2014), introducing a very long branch that would hinder rather than facilitate accurate reconstruction of the phylogeny. Two other, early-branching lineages, the Scotinosphaerales (Škaloud et al., 2013), and a clade containing the genera *Ignatius* and *Pseudocharacium* (Watanabe and Nakayama, 2007), are also absent from our dataset and need to be included in future analyses in order to arrive at a comprehensive picture of the ulvophycean radiation.

The use of fast site removal has gained popularity for the reconstruction of ancient relationships (e.g., Rodríguez-Ezpeleta et al., 2007a), including studies of the green algae (Lemieux et al., 2007; Rodríguez-Ezpeleta et al., 2007b; Cocquyt et al., 2010). The removal of fast sites aims to reduce the impact of non-phylogenetic signal by eliminating saturated sites (sites at which multiple substitutions have taken place) that are more likely to contain homoplasious patterns. Fast-evolving sites can also deviate from model assumptions more than slow-evolving sites, for example having nucleotide compositions that reflect mutational bias. Depending on the evolutionary depth of the phylogenetic relationships, homoplasy due to saturation or model violation in such sites can mask the true phylogenetic signal that remains present at slower-evolving positions. Typically, a range of site removal fractions are compared in order to identify the optimal proportion of deleted sites in which maximum phylogenetic signal is retained while maximum non-phylogenetic signal is removed. This tradeoff is evident in our analyses, where the bootstrap support values are highest at intermediate levels of fast site removal. Consistent with the observation that stochastic error is more problematic in smaller alignments and systematic error more problematic in larger alignments (Rodríguez-Ezpeleta et al., 2007a), we found that the bootstrap value profiles of the smaller 7-gene and 8-gene datasets did not show as clear a pattern as the genome-scale dataset. However, they did show a slight increase in average branch support toward intermediate levels of site stripping followed by a sharp decrease in average bootstrap. From this comparison it follows that larger alignments provide greater potential to isolate the true phylogenetic signal.

By aligning the site stripping conditions based on their evolutionary rate (**Figure 4**), it became clear that datasets with an average site rate within particular bounds ($10^{-0.5}$–$10^{-1.0}$) yielded the strongest signal about the early diversification of core Chlorophyta, regardless of what dataset the reduced alignments were derived from. This graph also indicates that the percentage of fast sites removed is not indicative of the average

site rate (e.g., rates of the 50% slowest sites in the genome-scale alignment are substantially higher than rates of the 50% slowest sites in the 7-gene alignment). This is a logical consequence of the differing rates of evolution of the genes in these datasets, because the 7-genes of the shorter alignment were specifically chosen for their slower evolutionary rate. Note that the rates given in **Figure 4** can be directly compared between alignments because site rates were estimated using the branch lengths of the guide tree (i.e., the chronogram, without re-estimation of branch lengths by HyPhy). As such, all rates are relative to the same standard and can be directly compared across alignments (7-gene, 8-gene, and genome-scale) and site stripping conditions.

In conclusion, our results underscore the need for critical examination of the assumption that Trebouxiophyceae and Ulvophyceae are monophyletic classes. We find no support in our chloroplast analyses for the monophyly of these classes, and instead find that support for their monophyly decreases with increasing alignment length. Future systematic studies will likely need to include additional early-branching taxa from both classes to ensure adequate taxon representation and avoid potential long-branch attraction. Unlike the strongly monophyletic Chlorophyceae, the remainder of the core Chlorophyta may represent several class-ranked lineages that do not correspond with the presently recognized classes. It is relevant to note here that some authors have advocated to subdivide the Ulvophyceae into five classes (Ulvophyceae *sensu stricto*, Cladophorophyceae, Bryopsidophyceae, Dasycladophyceae, and Trentepohliophyceae) based on apparent differences in thallus architecture, cellular organization, chloroplast morphology, cell wall composition, and life histories (Van Den Hoek et al., 1995). Clearly, the status of the genus *Oltmannsiellopsis* as member of Ulvophyceae will have to be re-evaluated, as it seems possible if not plausible that this genus is instead a relative of the Chlorodendrophyceae (*Tetraselmis* and its relatives). Ultrastructural examination of the cell division process in *Oltmannsiellopsis* may add further information about its evolutionary history and affiliation. Based on our results we conclude that it is time to retire the concept of a UTC lineage consisting of the traditional classes Ulvophyceae, Trebouxiophyceae and Chlorophyceae, and replace the term "UTC clade" with "core Chlorophyta" instead—a group for which evidence of monophyly is strong and undisputed.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://www.frontiersin.org/journal/10.3389/fevo.2014.00063/abstract

REFERENCES

Boedeker, C., O'Kelly, C. J., Star, W., and Leliaert, F. (2012). Molecular phylogeny and taxonomy of the *Aegagropila* clade (Cladophorales, Ulvophyceae), including the description of *Aegagropilopsis* gen. nov. and *Pseudocladophora* gen. nov. *J. Phycol.* 48, 808–825. doi: 10.1111/j.1529-8817.2012.01145.x

Booton, G. C., Floyd, G. L., and Fuerst, P. A. (1998). Origins and affinities of the filamentous green algal orders Chaetophorales and Oedogoniales based on 18S rRNA gene sequences. *J. Phycol.* 34, 312–318. doi: 10.1046/j.1529-8817.1998.340312.x

Buchheim, M. A., Michalopulos, E. A., and Buchheim, J. A. (2001). Phylogeny of the Chlorophyceae with special reference to the Sphaeropleales: a study of 18S and 26S rDNA data. *J. Phycol.* 37, 819–835. doi: 10.1046/j.1529-8817.2001.00162.x

Buchheim, M. A., Sutherland, D. M., Schleicher, T., Förster, F., and Wolf, M. (2012). Phylogeny of Oedogoniales, Chaetophorales and Chaetopeltidales (Chlorophyceae): inferences from sequence-structure analysis of ITS2. *Ann. Bot.* 109, 109–116. doi: 10.1093/aob/mcr275

Carlile, A. L., O'Kelly, C. J., and Sherwood, A. R. (2011). The green algal genus *Cloniophora* represents a novel lineage in the Ulvales: a proposal for Cloniophoraceae fam. nov. *J. Phycol.* 47, 1379–1387. doi: 10.1111/j.1529-8817.2011.01065.x

Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* 17, 540–552. doi: 10.1093/oxfordjournals.molbev.a026334

Cavalier-Smith, T. (1993). "The origin, losses and gains of chloroplasts," in *Origins of Plastids: Symbiogenesis, Prochlorophytes, and the Origins of Chloroplasts,* ed R. A. Lewin (New York, NY: Chapman and Hall), 291–349.

Chihara, M., Inouye, I., and Takahata, N. (1986). *Oltmannsiellopsis,* a new genus of marine flagellate (Dunaliellaceae, Chlorophyceae). *Arch. Protistenk.* 132, 313–324. doi: 10.1016/S0003-9365(86)80026-4

Cocquyt, E., Verbruggen, H., Leliaert, F., and De Clerck, O. (2010). Evolution and cytological diversification of the green seaweeds (Ulvophyceae). *Mol. Biol. Evol.* 27, 2052–2061. doi: 10.1093/molbev/msq091

Cocquyt, E., Verbruggen, H., Leliaert, F., Zechman, F., Sabbe, K., and De Clerck, O. (2009). Gain and loss of elongation factor genes in green algae. *BMC Evol. Biol.* 9:39. doi: 10.1186/1471-2148-9-39

Delsuc, F., Brinkmann, H., and Philippe, H. (2005). Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* 6, 361–375. doi: 10.1038/nrg1603

Deng, Y., Zhan, Z., Tang, X., Ding, L., and Duan, D. (2014). Molecular cloning and expression analysis of *rbc*L cDNA from the bloom-forming green alga *Chaetomorpha valida* (Cladophorales, Chlorophyta). *J. Appl. Phycol.* 26, 1853–1861. doi: 10.1007/s10811-013-0208-z

Felsenstein, J. (1985). Confidence-limits on phylogenies: an approach using the bootstrap. *Evolution* 39, 783–791. doi: 10.2307/2408678

Fučíková, K., Lewis, P. O., and Lewis, L. A. (2014a). Putting incertae sedis taxa in their place: a proposal for ten new families and three new genera in Sphaeropleales (Chlorophyceae, Chlorophyta). *J. Phycol.* 50, 14–25. doi: 10.1111/jpy.12118

Fučíková, K., Lewis, P. O., and Lewis, L. A. (2014b). Widespread desert affiliations of trebouxiophycean algae (Trebouxiophyceae, Chlorophyta) including discovery of three new desert genera. *Phycol. Res.* doi: 10.1111/pre.12062. [Epub ahead of print].

Gaysina, L., Němcová, Y., Škaloud, P., Ševčíková, T., and Eliáš, M. (2013). *Chloropyrula uraliensis* gen. et sp nov (Trebouxiophyceae, Chlorophyta), a new green coccoid alga with a unique ultrastructure, isolated from soil in south urals. *J. Syst. Evol.* 51, 476–484. doi: 10.1111/jse.12014

Glenn, T. C. (2011). Field guide to next-generation DNA sequencers. *Mol. Ecol. Resour.* 11, 759–769. doi: 10.1111/j.1755-0998.2011.03024.x

Huelsenbeck, J. P., and Ronquist, F. (2001). MRBAYES: bayesian inference of phylogenetic trees. *Bioinformatics* 17, 754–755. doi: 10.1093/bioinformatics/17.8.754

Kosakovsky Pond, S. L., Frost, S. D. W., and Muse, S. V. (2005). HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21, 676–679. doi: 10.1093/bioinformatics/bti079

La Claire, J. W., Loudenslager, C. M., and Zuccarello, G. C. (1998). Characterization of novel extrachromosomal DNA from giant celled marine green algae. *Curr. Genet.* 34, 204–211. doi: 10.1007/s002940050387

La Claire, J. W., and Wang, J. (2000). Localization of plasmidlike DNA in giant-celled marine green algae. *Protoplasma* 213, 157–164. doi: 10.1007/BF01282153

Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., Mcgettigan, P. A., McWilliam, H., et al. (2007). Clustal W and Clustal X version 2.0. *Bioinformatics* 23, 2947–2948. doi: 10.1093/bioinformatics/btm404

Leliaert, F., Smith, D. R., Moreau, H., Herron, M. D., Verbruggen, H., Delwiche, C. F., et al. (2012). Phylogeny and molecular evolution of the green algae. *Crit. Rev. Plant Sci.* 31, 1–46. doi: 10.1080/07352689.2011.615705

Leliaert, F., Verbruggen, H., and Zechman, F. W. (2011). Into the deep: new discoveries at the base of the green plant phylogeny. *Bioessays* 33, 683–692. doi: 10.1002/bies.201100035

Lemieux, C., Otis, C., and Turmel, M. (2007). A clade uniting the green algae *Mesostigma viride* and *Chlorokybus atmophyticus* represents the deepest branch of the Streptophyta in chloroplast genome-based phylogenies. *BMC Biol.* 5:2. doi: 10.1186/1741-7007-5-2

Lewis, P. O., Holder, M. T., and Swofford, D. L. (2011). *Phycas ver. 1.2.* Avaliable online at: http://www.phycas.org

Lokhorst, G. M., and Star, W. (1993). The flagellar apparatus in the marine flagellate algal genus *Oltmannsiellopsis* (Dunaliellales, Chlorophyceae). *Arch. Protistenk.* 143, 13–32. doi: 10.1016/S0003-9365(11)80268-X

Lopez-Giraldez, F., and Townsend, J. (2011). PhyDesign: an online application for profiling phylogenetic informativeness. *BMC Evol. Biol.* 11:152. doi: 10.1186/1471-2148-11-152

Lü, F., Xü, W., Tian, C., Wang, G., Niu, J., Pan, G., et al. (2011). The *Bryopsis hypnoides* plastid genome: multimeric forms and complete nucleotide sequence. *PLoS ONE* 6:e14663. doi: 10.1371/journal.pone.0014663

Marin, B. (2012). 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. doi: 10.1016/j.protis.2011.11.004

Marin, B., and Melkonian, M. (2010). Molecular phylogeny and classification of the Mamiellophyceae class. nov. (Chlorophyta) based on sequence comparisons of the nuclear- and plastid-encoded rRNA operons. *Protist* 161, 304–336. doi: 10.1016/j.protis.2009.10.002

Massjuk, N. P. (2006). Chlorodendrophyceae class. nov. (Chlorophyta, Viridiplantae) in the Ukrainian flora: I. the volume, phylogenetic relations and taxonomical status. *Ukr. Bot. J.* 63, 601–614 (in Ukrainian).

Matsumoto, T., Shinozaki, F., Chikuni, T., Yabuki, A., Takishita, K., Kawachi, M., et al. (2011). Green-colored plastids in the dinoflagellate genus *Lepidodinium* are of core chlorophyte origin. *Protist* 162, 268–276. doi: 10.1016/j.protis.2010.07.001

Mattox, K. R., and Stewart, K. D. (1984). "Classification of the green algae: a concept based on comparative cytology," in *Systematics of the Green Algae,* eds D. E. G. Irvine and D. M. John. (London: Academic Press), 29–72.

Moestrup, Ø. (1991). Further studies of presumedly primitive green algae, including the description of Pedinophyceae class. nov. and Resultor gen. nov. *J. Phycol.* 27, 119–133. doi: 10.1111/j.0022-3646.1991.00119.x

Nakayama, T., Marin, B., Kranz, H. D., Surek, B., Huss, V. A. R., and Melkonian, M. (1998). The basal position of scaly green flagellates among the green algae (Chlorophyta) is revealed by analyses of nuclear-encoded SSU rRNA sequences. *Protist* 149, 367–380. doi: 10.1016/S1434-4610(98)70043-4

Němcová, Y., Eliáš, M., Škaloud, P., Hodač, L., and Neustupa, J. (2011). Jenufa gen. nov.: a new genus of coccoid green algae (Chlorophyceae, incertae sedis) previously recorded by environmental sequencing. *J. Phycol.* 47, 928–938. doi: 10.1111/j.1529-8817.2011.01009.x

Neustupa, J., Eliáš, M., Škaloud, P., Němcová, Y., and Šejnohová, L. (2011). *Xylochloris irregularis,* gen. et sp. nov. (Trebouxiophyceae, Chlorophyta), a novel subaerial coccoid green alga. *Phycologia* 50, 57–66. doi: 10.2216/08-64.1

Neustupa, J., Němcová, Y., Veselá, J., Steinová, J., and Škaloud, P. (2013a). *Leptochlorella corticola* gen. et sp. nov. and Kalinella apyrenoidosa sp. nov.: two novel Chlorella-like green microalgae (Trebouxiophyceae, Chlorophyta) from subaerial habitats. *Int. J. Syst. Evol. Microbiol.* 63, 377–387. doi: 10.1099/ijs.0.047944-0

Neustupa, J., Němcová, Y., Veselá, J., Steinová, J., and Škaloud, P. (2013b). *Parachloroidium* gen. nov. (Trebouxiophyceae, Chlorophyta), a novel genus of

coccoid green algae from subaerial corticolous biofilms. *Phycologia* 52, 411–421. doi: 10.2216/13-142.2

Novis, P. M., Smissen, R., Buckley, T. R., Gopalakrishnan, K., and Visnovsky, G. (2013). Inclusion of chloroplast genes that have undergone expansion misleads phylogenetic reconstruction in the Chlorophyta. *Am. J. Bot.* 100, 2194–2209. doi: 10.3732/ajb.1200584

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20, 289–290. doi: 10.1093/bioinformatics/btg412

Pažoutová, M., Škaloud, P., and Nemjová, K. (2010). Phylogenetic position of *Ooplanctella planoconvexa*, gen. et comb. nova and *Echinocoleum elegans* (Oocystaceae, Trebouxiophyceae, Chlorophyta). *Fottea* 10, 75–82.

Pombert, J.-F., and Keeling, P. J. (2010). The mitochondrial genome of the entomoparasitic green alga *Helicosporidium*. *PLoS ONE* 5:e8954. doi: 10.1371/journal.pone.0008954

Pombert, J. F., Lemieux, C., and Turmel, M. (2006). The complete chloroplast DNA sequence of the green alga *Oltmannsiellopsis viridis* reveals a distinctive quadripartite architecture in the chloroplast genome of early diverging ulvophytes. *BMC Biol.* 4:3. doi: 10.1186/1741-7007-4-3

Rambaut, A., and Drummond, A. J. (2007). *Tracer v1.4*. Available online at: http://beast.bio.ed.ac.uk/Tracer

Rodríguez-Ezpeleta, N., Brinkmann, H., Roure, B., Lartillot, N., Lang, B. F., and Philippe, H. (2007a). Detecting and overcoming systematic errors in genome-scale phylogenies. *Syst. Biol.* 56, 389–399. doi: 10.1080/10635150701397643

Rodríguez-Ezpeleta, N., Philippe, H., Brinkmann, H., Becker, B., and Melkonian, M. (2007b). Phylogenetic analyses of nuclear, mitochondrial, and plastid multigene data sets support the placement of *Mesostigma* in the Streptophyta. *Mol. Biol. Evol.* 24, 723–731. doi: 10.1093/molbev/msl200

Ronquist, F., and Huelsenbeck, J. P. (2003). MrBayes 3: bayesian phylogenetic inference under mixed models. *Bioinformatics* 19, 1572–1574. doi: 10.1093/bioinformatics/btg180

Ruhfel, B., Gitzendanner, M., Soltis, P., Soltis, D., and Burleigh, J. (2014). From algae to angiosperms-inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evol. Biol.* 14:23. doi: 10.1186/1471-2148-14-23

Salisbury, J. L., Swanson, J. A., Floyd, G. L., Hall, R., and Maihle, N. J. (1981). Ultrastructure of the flagellar apparatus of the green alga *Tetraselmis subcordiformis* with special consideration given to the function of the rhizoplast and rhizanchora. *Protoplasma* 107, 1–11. doi: 10.1007/bf01275602

Sanderson, M. J. (2002). Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol. Biol. Evol.* 19, 101–109. doi: 10.1093/oxfordjournals.molbev.a003974

Servín-Garcidueñas, L. E., and Martínez-Romero, E. (2012). Complete mitochondrial and plastid genomes of the green microalga Trebouxiophyceae sp. strain MX-AZ01 isolated from a highly acidic geothermal lake. *Eukaryot. Cell.* 11, 1417–1418. doi: 10.1128/ec.00244-12

Shimodaira, H. (2004). Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. *Ann. Stat.* 32, 2616–2641. doi: 10.1214/009053604000000823

Škaloud, P., Kalina, T., Nemjová, K., De Clerck, O., and Leliaert, F. (2013). Morphology and phylogenetic position of the freshwater green microalgae *Chlorochytrium* (Chlorophyceae) and *Scotinosphaera* (Scotinosphaerales, ord. nov., Ulvophyceae). *J. Phycol.* 49, 115–129. doi: 10.1111/jpy.12021

Sluiman, H. J. (1989). The green algal class Ulvophyceae. an ultrastructural survey and classification. *Crypt. Bot.* 1, 83–94.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* 22, 2688–2690. doi: 10.1093/bioinformatics/btl446

Steinkötter, J., Bhattacharya, D., Semmelroth, I., Bibeau, C., and Melkonian, M. (1994). Prasinophytes form independent lineages within the Chlorophyta: evidence from ribosomal RNA sequence comparison. *J. Phycol.* 30, 340–345. doi: 10.1111/j.0022-3646.1994.00340.x

Sun, S., Chen, J., Li, W., Altintas, I., Lin, A., Peltier, S., et al. (2011). Community cyberinfrastructure for advanced microbial ecology research and analysis: the CAMERA resource. *Nucleic Acids Res.* 39, D546–D551. doi: 10.1093/nar/gkq1102

Tippery, N. P., Fučíková, K., Lewis, P. O., and Lewis, L. A. (2012). Probing the monophyly of the Sphaeropleales (Chlorophyceae) using data from five genes. *J. Phycol.* 48, 1482–1493. doi: 10.1111/jpy.12003

Turmel, M., Brouard, J.-S., Gagnon, C., Otis, C., and Lemieux, C. (2008). Deep division in the Chlorophyceae (Chlorophyta) revealed by chloroplast phylogenomic analyses. *J. Phycol.* 44, 739–750. doi: 10.1111/j.1529-8817.2008.00510.x

Turmel, M., Gagnon, M.-C., O'Kelly, C. J., Otis, C., and Lemieux, C. (2009a). The chloroplast genomes of the green algae *Pyramimonas*, *Monomastix*, and *Pycnococcus* shed new light on the evolutionary history of prasinophytes and the origin of the secondary chloroplasts of euglenids. *Mol. Biol. Evol.* 26, 631–648. doi: 10.1093/molbev/msn285

Turmel, M., Otis, C., and Lemieux, C. (2009b). The chloroplast genomes of the green algae *Pedinomonas minor*, *Parachlorella kessleri*, and *Oocystis solitaria* reveal a shared ancestry between the Pedinomonadales and Chlorellales. *Mol. Biol. Evol.* 26, 2317–2331. doi: 10.1093/molbev/msp138

Van Den Hoek, C., Mann, D. G., and Jahns, H. M. (1995). *Algae: An Introduction to Phycology*. Cambridge, UK: Cambridge University Press.

Verbruggen, H. (2012). *SiteStripper, v. 1.01*. Avaliable online at: http://phycoweb.net/software/SiteStripper

Waddell, P. J., Cao, Y., Hauf, J., and Hasegawa, M. (1999). Using novel phylogenetic methods to evaluate mammalian mtDNA, including amino acid invariant sites LogDet plus site stripping, to detect internal conflicts in the data, with special reference to the positions of hedgehog, armadillo, and elephant. *Syst. Biol.* 48, 31–53. doi: 10.1080/106351599260427

Watanabe, S., and Nakayama, T. (2007). Ultrastructure and phylogenetic relationships of the unicellular green algae *Ignatius tetrasporus* and *Pseudocharacium americanum* (Chlorophyta). *Phycol. Res.* 55, 1–16. doi: 10.1111/j.1440-1835.2006.00439.x

Wolf, M., Buchheim, M., Hegewald, E., Krienitz, L., and Hepperle, D. (2002). Phylogenetic position of the Sphaeropleaceae (Chlorophyta). *Plant Syst. Evol.* 230, 161–171. doi: 10.1007/s006060200002

Zechman, F. W., Theriot, E. C., Zimmer, E. A., and Chapman, R. L. (1990). Phylogeny of the Ulvophyceae (Chlorophyta): cladistic analysis of nuclear-encoded rRNA sequence data. *J. Phycol.* 26, 700–710. doi: 10.1111/j.0022-3646.1990.00700.x

Zuccarello, G. C., Price, N., Verbruggen, H., and Leliaert, F. (2009). Analysis of a plastid multigene data set and the phylogenetic position of the marine macroalga *Caulerpa filiformis* (Chlorophyta). *J. Phycol.* 45, 1206–1212. doi: 10.1111/j.1529-8817.2009.00731.x