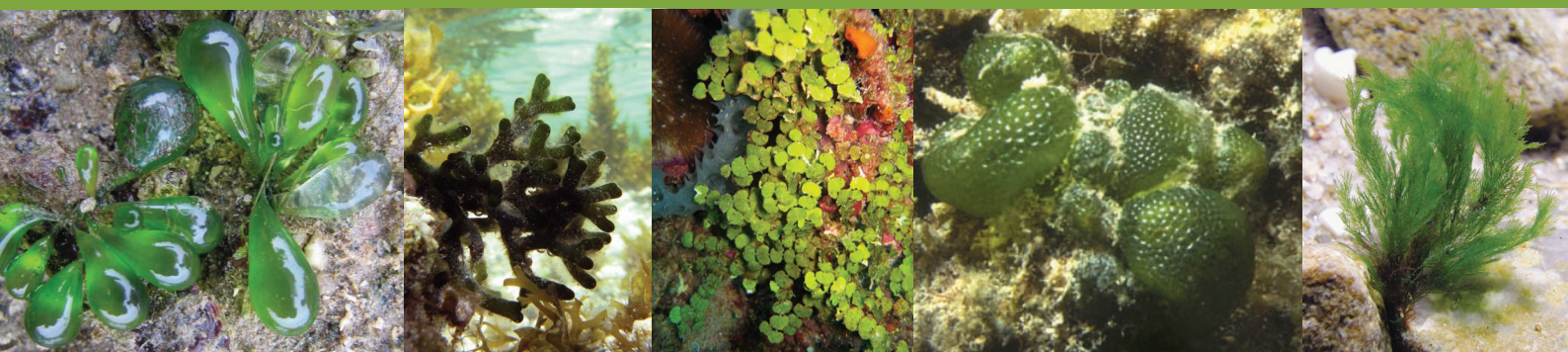# Phylogeny and molecular evolution of green algae

Ellen Cocquyt

Universiteit Gent
Faculteit Wetenschappen, Vakgroep Biologie
Onderzoeksgroep Algologie

Phylogeny and molecular evolution of green algae

Fylogenie en moleculaire evolutie van groenwieren

Ellen Cocquyt

Proefschrift voorgelegd tot het behalen van de graad van
Doctor in de Wetenschappen: Biologie
Academiejaar 2008-2009

UNIVERSITEIT
GENT

Research Group

Phycology

CeMoFE
UGent

Promotor: Prof. Dr. O. De Clerck (Universiteit Gent)

Co-promotor: Dr. H. Verbruggen (Universiteit Gent)

Leden van de leescommissie:

Prof. Dr. K. Hoef-Emden (Universität zu Köln, Germany)

Dr. P. Rouzé (Universiteit Gent)

Prof. Dr. A. Vanderpoorten (Université de Liège)

Overige leden van de examencommissie:

Prof. Dr. K. Sabbe (Universiteit Gent)

Prof. Dr. Erik Smets (Katholieke Universiteit Leuven and National Herbarium of the Nederlands)

Prof. Dr. W. Vyverman (Universiteit Gent)

Dankwoord

Vooreerst zou ik mijn promotor Olivier willen bedanken. Bijna negen jaar geleden stapte ik op het vliegtuig richting Zuid-Afrika. Ik mocht gedurende 2 maanden wieren inzamelen langsheen de Zuid-Afrikaanse kust en zou leren 'kijken' naar roodwieren, samen met Olivier die daar toen gedurende een jaar aan de Universiteit van Kaapstad werkte. Op de luchthaven van Kaapstad aangekomen vroeg Olivier verwondert: "Is dat alles wat je mee hebt?", wijzend naar mijn klein rugzakje. Tja, er zat een rugby team op het vliegtuig en niet alle bagage was meegeraakt. De mijne stond nog in Londen. 't Is gelukkig allemaal goed gekomen en ik heb daar een fantastisch tijd gehad!

Eenmaal ik mijn licentiaatdiploma behaalde, ging ik nog even langs het labo om goedendag te zeggen. Olivier stelde toen voor om met een Marie Curie beurs een tijdje bij Christine Maggs aan de Queen's University of Belfast te gaan werken. Hm, ik zou eigenlijk net gaan samenwonen met Toon. Uiteindelijk ben ik toch vertrokken voor een half jaartje. Daar heb ik voor het eerst DNA geëxtraheerd en PCR's gedaan, en eveneens een fantastisch tijd beleeft. Terug in België mocht ik onder voorwaarde dat ik een IWT beurs zou aanvragen, beginnen als laborante bij onze onderzoeksgroep. Die IWT beurs werd niets, maar na anderhalf jaar kon ik dan toch beginnen aan een doctoraat met een BOF beurs. Het resultaat daarvan is dit doctoraat!

Olivier, bedankt om me te begeleiden doorheen al die jaren.

Ten tweede, zou ik mijn co-promotor Heroen willen bedanken. Als laborante heb ik voor zijn doctoraat veel praktisch werk gedaan, maar het laatste anderhalf jaar heb ik ontzettend veel hulp van hem gekregen. Heroen, bedankt voor de hulp bij het analyseren van mijn gegevens en het snel en grondig nalezen en verbeteren van de teksten.

Ten derde, zou ik Olivier, Heroen en Frederik willen bedanken om me te steunen. Zonder de talloze brainstormmomenten en de hulp van jullie alle drie bij het verwerken en uitschrijven van de resultaten, was ik nooit tot dit resultaat gekomen.

Caroline, het was leuk om gedurende drie jaar bureau en labo met je te delen. Eveneens bedankt voor de hulp bij het praktisch werk. Andy en Renata, bedankt voor al het sequentiewerk.

Kadriye thanks to help me with PCR's and cloning of some of the nuclear genes. It was often a frustrating job, with a lot of trial and error.

Koen Sabbe, Ann Willems en Paul De Vos, bedankt voor de tijd die jullie hebben vrij gemaakt om, vooral in het begin, te luisteren naar mijn vorderingen en me met jullie suggesties telkens een stapje vooruit te helpen. Ook Steven Robbens en Yves van de Peer hielpen me door me in het begin de kans te geven over het nog niet gepubliceerde *Ostreococcus* genoom te beschikken.

Klaus Valentin, thanks for the cDNA service. The generation of this cDNA library was a big step forward during this PhD study. The people from VERTIS Biotechnologie AG (Freising , Germany) also helped a lot to solve the problems I encountered during the screening of the cDNA library.

Aan de mensen van de plantkunde in de Ledeganckstraat, het was altijd leuk en gezellig tijdens de middag of aan de koffietafel. Liesbeth, op het bankje aan het kleine vijvertje was het ook steeds gezellig vertoeven. Ik heb daar goeie herinneringen aan!

Het laatste anderhalf jaar was het met de mensen van op de Sterre minstens even gezellig tijdens de middagen in de Resto.

Katrien en Elke, bedankt voor het nalezen van een stukje Nederlandstalige tekst.

Eric, al wist je nooit goed waar ik nu precies mee bezig was, toch zou ik je willen bedanken om me warm te maken voor de algologie, en om je vlucht naar Sri Lanka te verzetten zodat je aanwezig kunt zijn op mijn publieke verdediging.

Ook mijn ouders en vrienden zou ik willen bedanken om me steeds te blijven steunen doorheen de jaren.

En tenslotte, Toon die nu al bijna negen jaar lang mijn vriend is... nog vele fijne jaren voor ons!


Ellen

juni 2009

# Contents

1

Introduction

Algae

Algae are a large and diverse group of eukaryotic photosynthetic organisms occurring in almost every habitat. They exhibit a huge morphological diversity, ranging from tiny unicells to huge kelps over 50 m long. The first algal groups arose between 1 and 1.5 billion years ago (Douzery et al. 2004, Yoon et al. 2004) after the symbiogenesis of a heterotrophic eukaryotic organism with a photosynthetic cyanobacterium. This event gave rise to the primary plastids which are still present in the Glaucophyta, red algae and green lineages including land plants (Reyes-Prieto et al. 2007). These three lineages are collectively called Plantae or Archaeplastida (Cavalier-Smith 1981, Adl et al. 2005). The other photosynthetic protists arose through secondary endosymbiosis of either a green or a red alga. The euglenids and chlorarachniophytes are thought to have acquired their plastids from a green alga in two separate secondary endosymbiotic events, while molecular evidence suggests that the red algal plastid of cryptomonads, heterokonts, haptophytes, apicomplexans and dinoflagellates was acquired by a single secondary endosymbiosis in their common ancestor (Archibald 2005, Archibald 2008). This process of serial cell capture and subsequent enslavement explains the diversity of photosynthetic eukaryotes. Endosymbiosis forms the landmark evolutionary event, responsible for the spread of photosynthesis through the Eukaryotic tree of life. Photosynthesis occurs in four of the six supergroups: Archaeplastida (Glaucophyta, red algae, green plants), Chromalveolata (cryptophytes, Stramenopila or heterokonts including diatoms and brown algae, haptophytes and dinoflagallates), Rhizaria (Chlorarachniophyta) and Excavata (euglenoids) (Fig. 1).
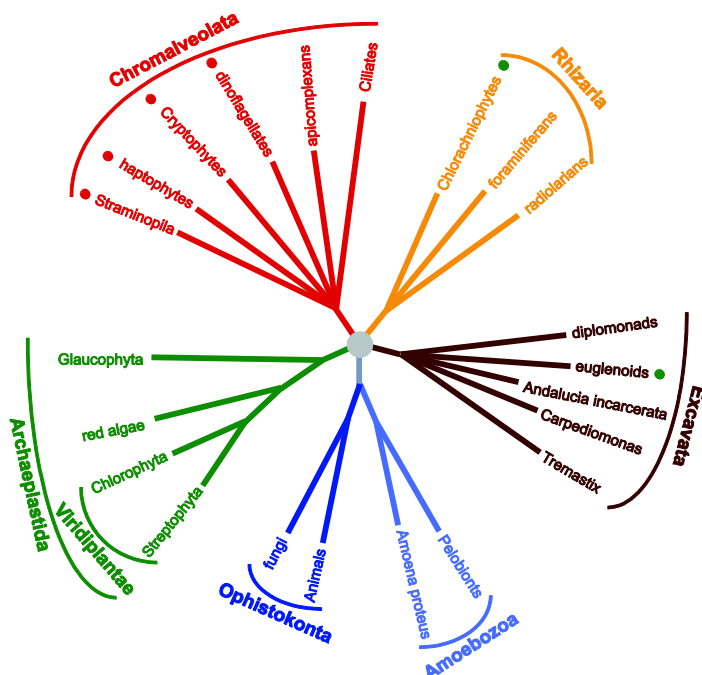
Figure 1. Eukaryotic tree of life. The first algae arose after the symbiogenesis of a heterotrophic eukaryotic organism with a photosynthetic cyanobacterium, giving rise to the Archaeplastida. The other photosynthetic protists arose through secondary endosymbiosis of either a green or a red alga and occur in four of six supergroups (marked with respectively green and red circles). The monophyly of the Archaeplastida is well-supported and most recent evidence favours the Glaucophyta as earliest diverging lineage within the Archaeplastida (modified after Baldauf 2008, Lane and Archibald 2008).

Archaeplastida

The monophyly of primary plastids has long been suggested by several features, such as a similar gene content of plastid genomes, the presence of plastid-specific gene clusters that are distinct from those found in Cyanobacteria, the conservation of the plastid-protein import machinery and protein-targeting signals, and phylogenies based on plastid and cyanobacterial gene sequences (Palmer 2003). Nevertheless, several single-gene phylogenies and a few multigene phylogenies have challenged this hypothesis (e.g., Stiller et al. 2001, Nozaki et al. 2003a, Nozaki et al. 2003b, Stiller and Harrell 2005). Conclusive evidence for the monophyly of the Glaucophyta, red algae and green plants was provided only relatively recently by Rodriguez-Ezpeleta et al. (2005) based on: (1) chloroplast gene phylogenies showing the monophyly of primary plastid and (2) a phylogenomic dataset containing 143 nuclear genes, ca. 30,000 amino acid positions which show the monophyly of all organisms with a primary plastid (Fig. 1). The latter study, however, could not reveal the relation among the three major lineages. Several nuclear genes suggest that red algae are the earliest diverging Archaeplastida, but such results are inconsistent with many plastid gene trees that identify glaucophytes as the earliest divergence. Most recent evidence favours the early divergence of glaucophytes, as demonstrated by Reyes-Prieto et al. (2007) using a concatenated dataset of conserved nuclear-encoded plastid targeted proteins of cyanobacterial origin. The latter evolutionary scenario corroborates with two important putatively ancestral characters shared by glaucophyte plastids and the cyanobacterial endosymbiont that gave rise to this organelle: the presence of carboxysomes and a peptidoglycan deposition between the two organelle membranes. Both traits were apparently lost in the common ancestor of red and green algae after the divergence of glaucophytes.
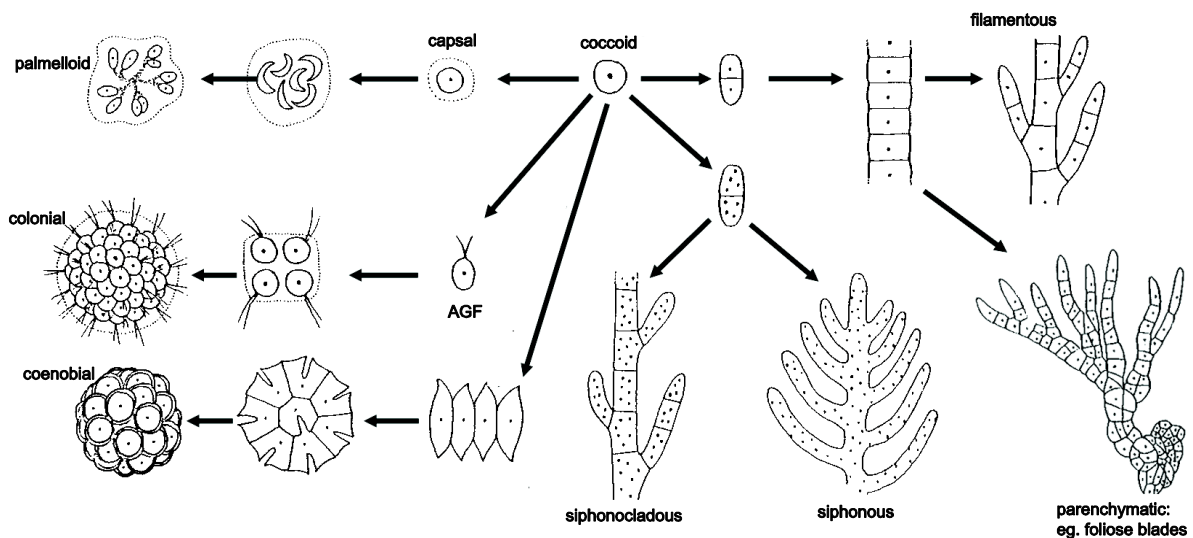
Figure 2. The green algae exhibit a remarkable cytological diversity ranging from unicellar organisms (coccoid or flagellates), over multicellular filaments and foliose blades, to coenocytic and siphonous life forms that are essentially composed of a single giant cell containing countless nuclei (after Coppejans 1998). Arrows indicate trends in morphological complexity rather than evolutionary hypotheses. For example, green algae are thought to have evolved from a unicellular flagellate (the Ancestral Green Flagellate, AGF) rather than a coccoid life form.

Green lineage or Viridiplantae

Green algae are distributed worldwide and can be found in almost every habit ranging from polar to tropical marine, freshwater and terrestrial environments and as symbionts (Pröschold and Leliaert 2007). They exhibit a remarkable cytological diversity ranging from the world's smallest free-living eukaryote known to date *Ostreococcus taurii* (Derelle et al. 2006), over multicellular filaments and foliose blades, to siphonous life forms that are essentially composed of a single giant cell containing countless nuclei (Fig. 2). Together with land plants, green algae form the green lineage or Viridiplantae (also written as Virideaplantae or known as green plants, Chlorobionta, Chloroplastida or Chlorophycophyta). Morphological and molecular studies have identified a major split within the Viridiplantae giving rise to two monophyletic lineages, the Chlorophyta and the Streptophyta (Pickett-Heaps and Marchant 1972, Lewis and McCourt 2004) (Fig. 3). The streptophytes comprise several lineages of predominantly freshwater green algae (often called charophytes or charophyte green algae) and the land plants (Embryophyta) which evolved roughly 470 million years ago from a charophyte ancestor. The majority of green algae, however, belong to the Chlorophyta.

Streptophyta

When motile cells are present, the Streptophyta are characterized by biflagellate cells with asymmetrically flagellar roots including a multilayered structure or MLS (a distinct parallel arrangement of microtubules) and a smaller root. In all representatives the nuclear envelope breaks down before the chromosomes separate (open mitosis) and the mitotic spindle is persistent which helps to keep the daughter nuclei separate until cytokinesis has been accomplished. Biochemical characters such as photorespiratory enzymes are different from those found in most chlorophyte green algae (Figs. 4 and 5; Table 1).

Several phylogenetic studies have tried to determine the origins of the land plants, focussing on the green algal progenitors of the Streptophyta. Recent studies have indicated the scaly green flagellate *Mesostigma* as the earliest diverging streptophyte. Initially, the scaly flagellate was placed within the prasinophytes (Mattox and Stewart 1984), later ultrastructural investigations (e.g. a flagellum which is anchored in the cell by means of an asymmetric root) revealed the association with the Streptophyta (Melkonian 1989). Molecular phylogenies also showed conflicting results regarding the phylogenetic relationships of this enigmatic species: *Mesostigma* either diverges before the Chlorophyta/Streptophyta split (Lemieux et al. 2000, Turmel et al. 2002a, Turmel et al. 2002b) or as an early diverging flagellate within the Streptophyta (Bhattacharya et al. 1998, Marin and Melkonian 1999, Karol et al. 2001). Increasing taxon and gene sampling and the use of more realistic models of evolution provide evidence that *Mesostigma* is an early diverging lineage within the Streptophyta (Petersen et al. 2006, Lemieux et al. 2007, Rodriguez-Ezpeleta et al. 2007). The colonial soil alga *Chlorokybus* diverges after *Mesostigma* in most phylogenies (Karol et al. 2001), although more recent studies united *Mesostigma* and *Chlorokybus* as the earliest diverging branch of the Streptophyta (Lemieux et al. 2007). Unbranched filaments that form the class Klebsormidiophyceae diverge next, followed by the Zygnematophyceae clade, which includes unicells and unbranched filaments with isogamous sexual reproduction. The more complex charophytes are Coleochaetophyceae and Charophyceae both consisting of branched filaments with oogamous sexual reproduction. It remains

inconclusive from which group of algae Embryophytes emerged. Karol et al. (2001) resolved the stoneworts (Charophyceae) as the closest relatives of the Embryophyta, while evidence from plastid genomes point toward the conjugating algae or Zygnematophyceae (Lemieux et al. 2007).
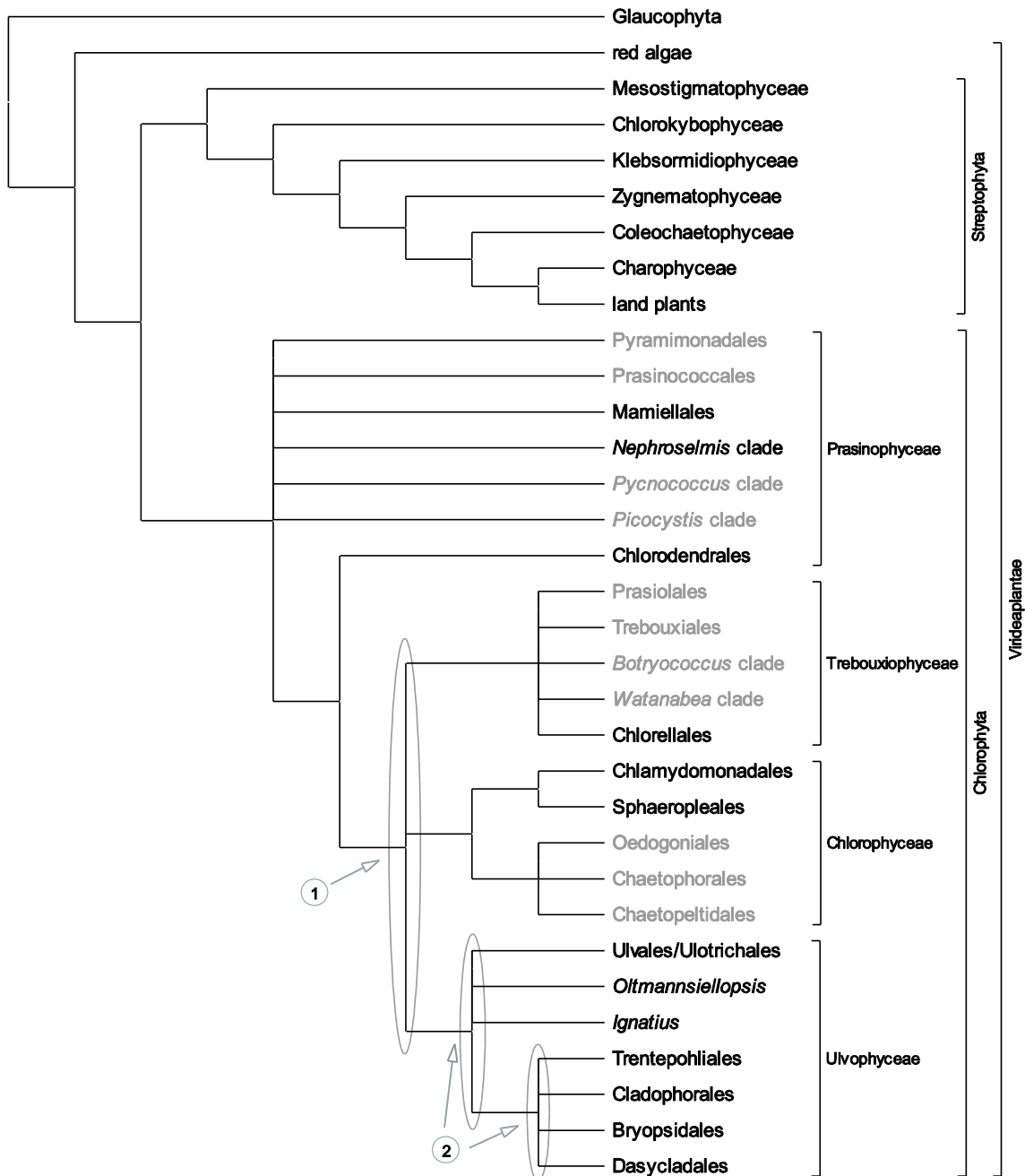
Figure 3. Phylogenetic relationships between green algae inferred from several phylogenetic studies: Streptophyta (Karol et al. 2001: SSU nrDNA, mitochondrial nad5 gene and plastid rbcL and atpB genes), prasinophytes (Guillou et al. 2004: SSU nrDNA), Trebouxiophyceae (Karsten et al. 2005: SSU nrDNA), Chlorophyceae (Turmel et al. 2008: chloroplast genes), Ulvophyceae (Lopez-Bautista and Chapman 2003: SSU nrDNA) and (Watanabe and Nakayama 2007: SSU nrDNA). The main interest of chapter 2 is to resolve relations between Ulvophyceae, Trebouxiophyceae and Chlorophyceae (1) and among Ulvophyceae (2). Representatives from each clade are studied (except those marked in gray).

counter-clockwise	direct oppposite	clockwise	asymmetric
A	**B**	**C**	**D**
Trebouxiophyceae Ulvophyceae	Chlorophyceae	Chlorophyceae	Streptophyta Trentepohliales (Ulvophyceae)
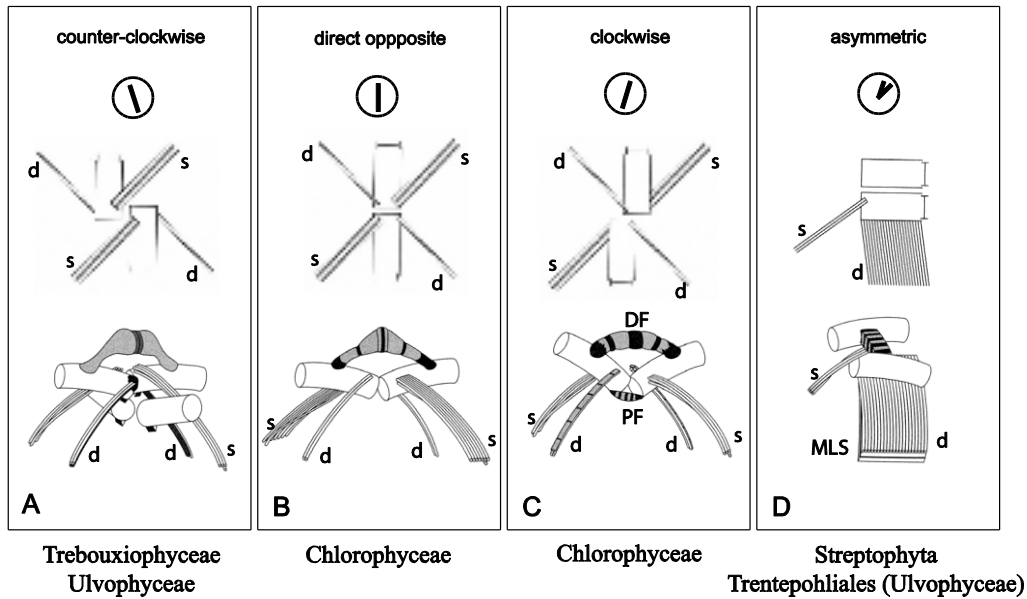
Figure 4. Variation in flagellar apparatuses found among green algae, viewed from the top (upper figure) and from the side (lower figure), modified after Graham et al. (2009) and Pröshold and Leliaert (2007). The flagellar apparatus generally include two or four basal bodies (shown here as rectangles or cylinders), microtubular roots (s or d), and distal (DF) or proximal (PF) connecting fibers. A. Flagellar apparatus with cruciate roots and basal bodies displaced in counter-clockwise direction. B. Flagellar apparatus with cruciate roots showing directly opposed placement of flagellar basal bodies. C. Flagellar apparatus with clockwise displaced flagellar basal bodies. D. Flagellar apparatus with parallel basal bodies and asymmetrical distribution of the flagellar roots, showing the characteristic multilayered structure (MLS).
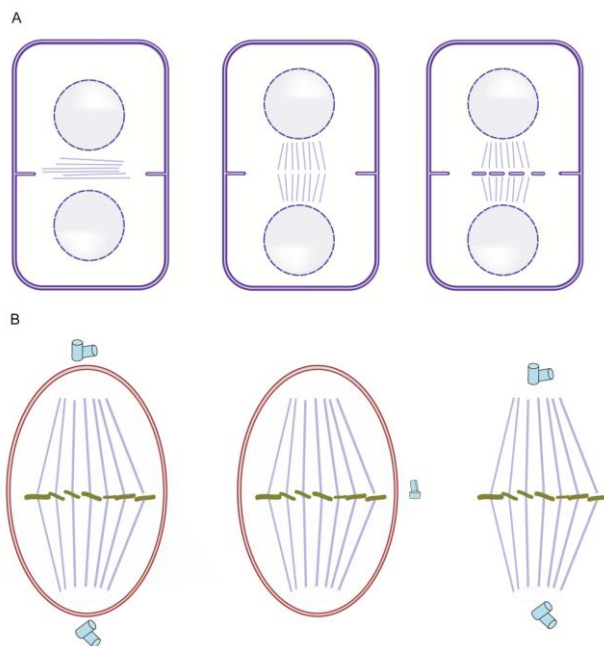
Figure 5. Ultrastructural features of green algae. **1.** Comparison of cytokinesis among green algae (after Graham et al. 2009). A. Phycoplasts, arrays of microtubules which lie parallel to the developing cleavage furrow, are often present in the Chlorophyceae and Trebouxiophyceae. B. Furrowing, sometimes with involvement of microtubules, is observed in the early charophytes and some Ulvophyceae. C. Phragmoplasts very similar, if not identical, to those of land plants are found in later diverging charophytes and the Trentepohliales (Ulvophyceae). Little furrowing is involved, the cell plates develop from the center toward the cell periphery. Microtubules arranged perpendicular to the developing cell plate. **2.** Different types of mitosis among green algae during the metaphase indicating the fate of the nuclear envelope and position of the centrioles (after Graham et al. 2009). A. Closed mitosis. B. Metacentric mitosis C. Open mitosis.
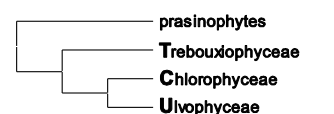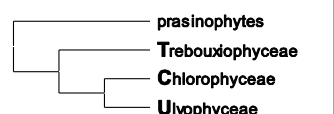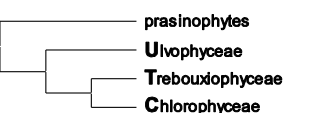
Chlorophyta

The current classification of the Chlorophyta, which relies on a combination of morphology, ultrastructural features of the flagellar root system, and characters relating to the mitotic spindle during cell division and cytokinesis, has been largely confirmed by phylogenetic analysis (Figs. 4 and 5, Table 1) (Mattox and Stewart 1984, Pröschold and Leliaert 2007). Four major groups, commonly regarded as classes, are recognized by consensus: "Prasinophyceae", Chlorophyceae, Trebouxiophyceae and Ulvophyceae (Fig. 3) (reviewed in Lewis and McCourt 2004). The prasinophytes form a paraphyletic group of unicellular flagellates or coccoid cells at the base of the Chlorophyta (Steinkötter et al. 1994, Fawley et al. 2000, Lopez-Bautista and Chapman 2003, Guillou et al. 2004). The Ulvophyceae, Trebouxiophyceae and Chlorophyceae are resolved as a well-supported clade (UTC clade) in most studies (Mishler et al. 1994), but the relationships among these lineages form the basis of a longstanding debate. Furthermore, the monophyly of the three classes remains to be demonstrated unequivocally (O'Kelly and Floyd 1984a, Zechman et al. 1990, Krienitz et al. 2003).

Certain ultrastructural characteristic are shared between the Ulvophyceae, Trebouxiophyceae and Chlorophyceae. In all representatives the nuclear envelope remains intact until the chromosomes finally separate (closed mitosis). The flagella are anchored in the cell by means of cruciate flagellar roots with mostly an X-2-X-2 configuration of the microtubules (Moestrup 1978, Lewis and McCourt 2004). Other ultrastructural observations are useful diagnostic characters to separate the three classes. The orientation of the basal bodies, short cylindrical arrays of microtubules at the base of a flagellum, is one of those discriminative characters. The Ulvophyceae and Trebouxiophyceae have a counter-clockwise orientation of the basal bodies, while the Chlorophyceae have a direct opposite or clockwise orientation of the basal bodies (Fig. 4) (Lewis and McCourt 2004). The Ulvophyceae have a persistent mitotic spindle which helps to keep the daughter nuclei separate until cytokinesis has been accomplished. The Trebouxiophyceae and Chlorophyceae both have a non-persistent mitotic spindle and a phycoplast composed of a set of microtubules which lie parallel to the plane of cytokinesis (Fig. 5) (Friedl 1995, Lewis and McCourt 2004). Based on these ultrastructural observations Mattox and Stewart (1984) suggest that the Ulvophyceae diverged first, followed by the Trebouxiophyceae and Chlorophyceae (Fig. 6). While Mattox and Stewart (1984) based their classification on the orientation of the basal bodies in the flagellar apparatus and differences of the mitotic spindle during cell division and cytokinesis, Sluiman (1989) only used the orientation of basal bodies in the flagellar apparatus as diagnostic character and merged the Trebouxiophyceae with the Ulvophyceae based on the counter-clockwise orientation of the basal bodies in the flagellar apparatus (Fig. 6).

Molecular phylogenetic studies have been highly inconclusive about the relationships between UTC classes (Fig. 6). The first molecular phylogenies based on small subunit nuclear ribosomal DNA (SSU or 18S nrDNA) sequences all observed that Ulvophyceae branch first, leaving Trebouxiophyceae and Chlorophyceae as sisters (Friedl 1995, Bhattacharya et al. 1996, Krienitz et al. 2001, Lopez-Bautista and Chapman 2003), while more recent SSU nrDNA phylogenitic studies using expanded taxon sampling and likelihood-based methods with more realistic models of sequence evolution revealed a sister relation between Chlorophyceae and Ulvophyceae (Friedl and O'Kelly 2002, Lewis and Lewis 2005, Watanabe and Nakayama 2007). Chloroplast gene order data and genomic structural features

(shared gene losses and rearrangements within conserved gene clusters), along with a phylogenetic analysis of seven mitochondrial genes supported this sister relation between Ulvophyceae and Chlorophyceae, while phylogenetic analysis of 58 concatenated chloroplast genes supported a sister relation between Ulvophyceae and Trebouxiophyceae (Pombert et al. 2004, Pombert et al. 2005). In this way, all possible relations between the UTC algae have been proposed the latest decennia.

Topology	prasinophytes **T**rebouxiophyceae **C**hlorophyceae **U**lvophyceae	prasinophytes **T**rebouxiophyceae **C**hlorophyceae **U**lvophyceae	prasinophytes **U**lvophyceae **T**rebouxiophyceae **C**hlorophyceae
Phylogenies	recent SSU nrDNA studies; mitochondrial genes	Chloroplast genome phylogenies	Some early SSU phylogenies
Ultrastructure and other evidence	Structural evidence in plastid genomes	- **U** and **T** share a counter-clockwise orientation of the basal bodies - **C** has a direct opposite or clockwise orientation of the basal bodies	- **T** and **C** share a phycoplast and a non-persistent mitotic spindle - **U** has a persistent mitotic spindle

Figure 6. The three alternative topologies for Ulvophyceae, Trebouxiophyceae and Chlorophyceae (UTC) have been supported by ultrastructural characteristics, molecular phylogenies and/or genomic structural features.

"Prasinophyceae" – The prasinophytes are planktonic and predominantly marine unicellular flagellates (with one to eight flagella) or coccoid organisms exhibiting a wide variety of ultrastructural features and photosynthetic pigment signatures (see O'Kelly 2007 and references therein). The cells and flagella of many members are covered by up to seven distinct types of organic scales, which are formed in the Golgi apparatus. The absence of clear synapomorphies suggest that the prasinophytes are not monophyletic but rather form a cluster of several independent lineages (Chlorodendrales, *Picocystis* clade, *Pycnococcus* clade, *Nephroselmis* clade, Mamiellales, Prasinococcales, Pyramimonadales) at the base of the Chlorophyta, which have been formally described and confirmed by SSU nrDNA phylogenies (Fig. 3 + Table 1) (Steinkötter et al. 1994, Fawley et al. 2000, Guillou et al. 2004, Turmel et al. 2009). It should be noted that our knowledge of these diverse, mainly picoplanktonic algae is far from comprehensive. The application of new technologies such environmental sequencing results in the discovery of new lineages, the majority of which remains hitherto uncultured and undescribed.

The Chlorodendrales (*Tetraselmis* and *Scherffelia*) is the only prasinophyte lineage which is robustly placed at the base of the UTC clade in molecular phylogenies and which is characterised by the presence of a counter-clockwise orientation of the basal bodies, a non-persistent mitotic spindle and the occurrence of a phycoplast. All the rest of the prasinophytes contain ancestral morphological treats, which makes it difficult to assign them either to the Chlorophyta or Streptophyta. In this view, the Pyramimonadales are crucial in our understanding of the early history of green plants. The combination of a well-represented fossil record, a flagellar apparatus configuration from which all other patterns in the green plants can be plausibly derived by reduction, and the only documented instances of phagotrophic mixotrophy suggests that the Pyramimonadales are the modern representatives of the earliest green algae (O'Kelly 2007). The prasinophytes further contain *Ostreococcus tauri* (Mamiellales), which is the smallest free-living eukaryote and which has the

smallest nuclear genome of all photosynthetic eukaryotes (Derelle et al. 2006) and *Nephroselmis* (Nephroselmidales) a small unicellular flagellate.

Trebouxiophyceae - The Trebouxiophyceae mainly consist of freshwater (e.g. *Chlorella*) and terrestrial algae (e.g. the phycobiont *Trebouxia* in lichens), some members (e.g. *Prasiola*) occur in marginally marine habitats. In contrast to most trebouxiophytes, the genus *Prototheca* is colorless and obligately heterotrophic and mainly lives in soil, but there are also some disease causing species. The enigmatic parasitic alga *Helicosporidium* is closely related to *Prototheca* (Tartar et al. 2002). Trebouxiophyceae occur as non-flagellate unicells or colonies, unbranched or branched filaments, or small blades (e.g. *Prasiola*) similar to those found among Ulvophyceae. Trebouxiophyceae commonly produce asexual, non-motile autospores. Sexual reproduction, involving flagellate sperm and non-motile eggs, is only known for some representatives (e.g. *Prasiola*).

The Trebouxiophyceae are characterized by a combination of ultrastructural characteristics: counter-clockwise orientation of the basal bodies (Fig. 4), non-persistent metacentric mitotic spindles (Fig. 5B) and the presence of a phycoplast (Fig. 5A), none of which is unique to the class. The basal body orientation is shared with the Ulvophyceae, metacentric spindles with the prasinophytes and non-persistent spindles and phycoplasts with the Chlorophyceae. The monophyly of the Trebouxiophyceae has still to be proven due to the lack of unique structural or reproductive features and the failure of molecular phylogenetic studies to support monophyly (Krienitz et al. 2003, Lewis and McCourt 2004).

Chlorophyceae - The Chlorophyceae mainly contain freshwater and terrestrial green algae. The class exhibits all major body plans found among green algae: unicells (flagellates and non-flagellates), sarcinoid organisms which are composed of packets of non-motile cells, colonies, unbranched or branched filaments and some multinucleate organisms. Asexual reproduction often occurs by means of flagellated zoospores but can also occur with non-motile aplanospores or autospores. Sexual reproduction may be isogamous, involving gametes that are morphological identical; anisogamous, in which case flagellate gametes are structurally distinguishable; or oogamous, with a large non-motile egg and smaller flagellate sperm. When sexual reproduction is present, Chlorophyceae always have a haploid vegetative phase and a single celled, often dormant, zygote as diploid stage (zygotic meiosis). Cell division is fairly uniform among Chlorophyceae and includes closed mitosis and a non-persitent mitotic spindle which collapses before cytokinesis. Cleavage is mostly mediated by a phycoplast but some Chlorophyceae divide by simple centripetal furrowing of the cell wall. The production of a cell plate that develops centrifugally by fusion of Golgi-derived vesicles in which case plasmodesmata or channels through the cell wall will allow intercellular communication after division have been reported for *Cylindrocapsa* and *Uronema* (see chapter 7 and van den Hoek et al. 1995, Graham et al. 2009).

The monophyly of the Chlorophyceae is corroborated by the presence of clockwise or direct opposite flagellar root supports (Fig. 4). The two major clades are formed by the Chlamydomonadales (clockwise basal body orientation) which include genera like *Chlamydomonas* and *Volvox* and the

Sphaeropleales (direct opposite basal body orientation) which contain for instance *Scenedesmus*. Other orders are Chaetophorales, Chaetopeltidales and Oedogoniales (Turmel et al. 2008).

Ulvophyceae - The Ulvophyceae mainly include marine green macroalgae among which some well-known green algae such as the sea lettuce *Ulva*, the model organism *Acetabularia* and the weedy *Codium* and *Bryopsis*. Morphological diversity ranges from flagellate and non-flagellate unicells and colonies to branched and unbranched filaments, foliose blades and multinucleate life forms. The sexual life cycle of most Ulvophyceae involves an alternation between a diploid sporophyte and haploid, free-living gametophytes (sporic meiosis), but zygotic meiosis with a haploid vegetative phase and single celled zygote as diploid stage also occurs. In some representatives (e.g. some Ulotrichales), flagellate reproductive cells are covered by a layer of small scales similar to those found in some prasinophytes (Sluiman 1989).

Since the class lacks clear synapomorphies, the monophyly of the Ulvophyceae has been doubted since its establishment (O'Kelly and Floyd 1984b). Therefore, the Ulvophyceae are defined by a suite of characters such as a counter clockwise orientation of the basal bodies (Fig. 4), a persistent mitotic spindle (Fig. 5B) and cytokinesis mediated by centripetal furrowing of the plasma membrane, none of which is exclusive for the Ulvophyceae (Mattox and Stewart 1984, O'Kelly and Floyd 1984a). The Trentepohliales for which molecular phylogenies revealed their relationship to ulvophycean seaweeds, have some enigmatic streptophyte-like characteristics such as an asymmetric flagellar root anchored in the cell with a multilayered structures and the formation of a phragmoplast during cytokinesis.

Molecular phylogenies based on SSU nrDNA failed to provide solid support for the monophyly of the Ulvophyceae and revealed the presence of two distinct groups within the Ulvophyceae. One group is represented by the orders Ulvales and Ulotrichales and contains unicellular, filamentous and blade-like organisms. The other group is composed of the branched filamentous order Trentepohliales, the siphonocladous seaweed order Cladophorales (synonymous with Siphonocladales), and the siphonous seaweed orders Dasycladales and Bryopsidales (Zechman et al. 1990, Watanabe et al. 2001, Lopez-Bautista and Chapman 2003, Watanabe and Nakayama 2007). In these trees, long evolutionary distances separate the green seaweed orders from the rest of the Ulvophyceae which consequently obscures the relationships among them. Based on these long branches and the apparent differences in thallus architecture, cellular organization, chloroplast morphology, cell wall composition, and life histories, some authors suggested to elevate the ulvophycean orders Cladophorales, Bryopsidales, Dasycladales and Trentepohliales to the class level (van den Hoek et al. 1995). In addition to the uncertainty regarding the affinities between the main orders, the position of some enigmatic Ulvophyceae such as *Ignatius, Oltmannsiellopsis* and *Blastophysa* is not well-known.

Green algal phylogenies

The estimated divergence time between Streptophyta and Chlorophyta varies considerably among studies. The split between these lineages is generally situated in the Neoproterozoic to late Mesoproterozoic (Fig. 7) (between 1200 and 700 million years ago: Douzery et al. 2004, Hedges et al. 2004, Yoon et al. 2004, Berney and Pawlowski 2006, Roger and Hug 2006, Zimmer et al. 2007, Herron et al. 2009). The ancient age in combination with the short time span over which the major lineages of green algae diverged likely caused the unstable relationships among these classes. Determining relationships among the different orders of the Ulvophyceae poses a similar problem. Up to now, green algal phylogenies using a broad taxon sampling were always based on SSU nrDNA (Zechman et al. 1990), while studies using whole plastid genome sequences contend with a small taxon sampling (Pombert et al. 2004, Pombert et al. 2005). It is well-known that phylogenetic information contained in a single gene is often insufficient to obtain firm statistical support for deep nodes (Philippe et al. 2005). Hence, improved character sampling by the inclusion of a larger number genes often leads to more accurate estimates of phylogenetic relationships. Unfortunately there is a limited availability of genomic data for green algae. Whole genome sequences were only available for the chlorophycean *Chlamydomonas reinhardtii* and the prasinophycean *Ostreococcus tauri* at the start of this project. Some other whole genome sequences have been recently released or are on the way for the chlorophyceae *Volvox* and the prasinophytes *Micromonas* and *Bathycoccus*. However, up till present no genome has been sequenced for a representative of the Ulvophyceae or Trebouxiophyceae. In addition to whole genome sequences, there are small to moderately sized EST libraries and complete organelle genomes available for some green algae (table 2). However, the ulvophycean seaweed orders remain underrepresented: no sequenced plastid genome and only 2 smaller EST libraries are available.

This limited availability of genomic information for the Chlorophyta inevitably poses a restriction on taxon sampling. Most studies which have addressed deep relationships of Chlorophyta are seriously constrained in taxon sampling (e.g., Pombert et al. 2005, Rodriguez-Ezpeleta et al. 2007). Furthermore, sparse and uneven taxon sampling, with most species belonging to either the prasinophytes or Chlorophyceae, increases the risk of systematic error due to long branch attraction and other biases (Verbruggen and Theriot 2008).

Resolving deep green algal relationships will inevitably request the combined analysis of many markers. Since genomic data are rare, it will be necessary to design primers for new markers. Low-copy nuclear markers proved to be good candidates due to a higher evolutionary rate than organellar genes, the potential to analyze multiple unlinked genes and the bi-parentally inheritance (Small et al. 2004). However, the greater difficulty of isolating and characterizing low-copy nuclear markers in comparison with SSU nrDNA and chloroplast genes which can be amplified with universal primers and the difficulty of assessing orthology are the major challenge when using low-copy nuclear markers (Small et al. 2004).
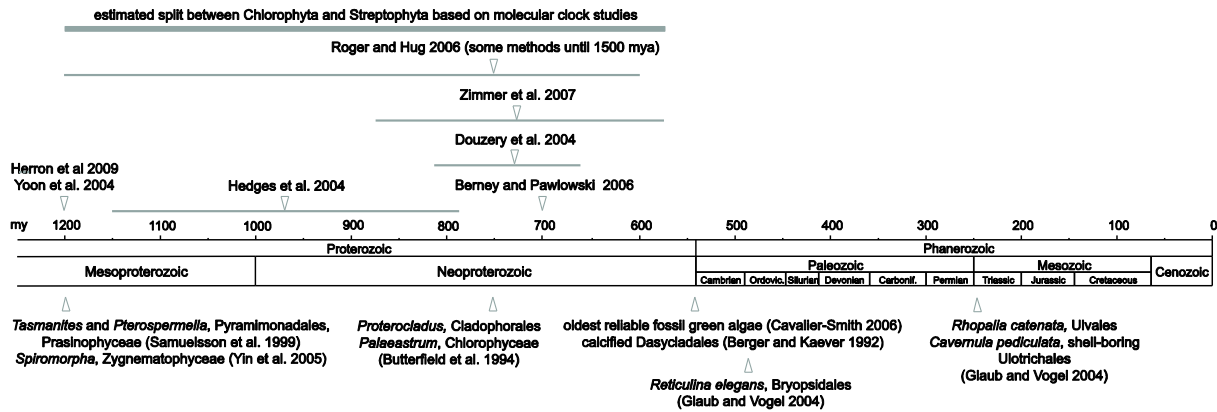
Figure 7. The estimated divergence time between Streptophyta and Chlorophyta, which are based on molecular clock studies, vary considerably among studies. The split between these lineages is generally situated in the Neoproterozoic to late Mesoproterozoic (between 1200 and 700 million years ago). The best characterized Precambrian fossils which most likely are green algae are *Proterocladus* and *Palaeastrum* at Svanbergfjellet, *Tasmanites* and *Pterospermella* at Thule and *Spiromorpha* at Ruyang. The oldest reliable fossil green algae are 540 million years old (Dasycladales). Other fossil Ulvophyceae are found more recently.

Evolution of nuclear markers

One of the major challenges when using nuclear genes for phylogenetic purposes is to assess orthology. Gene duplications lead to the formation of gene families (e.g. actin gene family, and GapA/B gene duplication in Streptophytes and the prasinophyte *Ostreococcus*), in which case extreme caution should be taken to ensure that the same orthologues gene is used in each species (An et al. 1999, Petersen et al. 2006, Robbens et al. 2007). Apart from difficulties in assessing orthology, there are indications of the occurrence of lateral gene transfer between eukayotes (Andersson 2005, Keeling and Palmer 2008). Two key genes of the translational apparatus, elongation factor-1 alpha (EF-1α) and elongation factor-like (EFL) have an almost mutually exclusive distribution in eukaryotes. In the green plant lineage, the Chlorophyta encode EFL except *Acetabularia* (Dasycladales, Ulvophyceae) where EF-1α is found, and the Streptophyta possess EF-1α except *Mesostigma*, which has EFL (Noble et al. 2007). The punctuated distribution of these two genes make it worth to further explore their distribution and evolutionary patterns of gain and loss within the green plant lineage. In this context it is interesting that *Acetabularia* not only has a different elongation factor gene compared to the rest of the Chlorophyta but also uses a different genetic code. In *Acetabularia,* and its close relative *Batophora,* stop codons TAG and TAA code for the amino acid glutamine (Schneider et al. 1989, Schneider and de Groot 1991). It is interesting to see if other green algae, especially close relatives of *Acetabularia*, also have the EF-1α gene or use the same alternative (non-canonical) code.

Designing primers for low-copy nuclear markers

For phylogenetic analyses we will concentrate on the development of low-copy nuclear markers. There are several alternative approaches to isolate and characterize novel low-copy nuclear markers: 1) the design of new primers from information in sequence databases (e.g., GenBank), 2) obtaining novel sequences via cDNA cloning 3) isolation of homologous DNA using a gene probe from another organism, and 4) characterization of sequence markers from DNA fingerprints (Schluter et al. 2005). In order to find suitable nuclear markers we first searched the literature for potentially useful low-copy nuclear markers and then applied the first two approaches (Fig. 8). When designing new primers from information already available in GenBank, we used genomic sequences of two unicellular green algae (*Chlamydomonas reinhartii* and *Ostreococcus tauri*) and aligned them with Genbank sequences from land plants and other green algae if available. For the second approach we made an EST library for the siphonocladous ulvophyte *Cladophora coelothrix*. All *Cladophora coelothrix* EST sequences were blasted against the GenBank protein database (blastx) for annotating the genes. Only genes with a certain annotation (hits to Swiss-Prot database) were retained and aligned with GenBank sequences from other green algae and land plant. If those genes were long enough and contained conserved regions, primers were designed and tested on a variety of green algae.
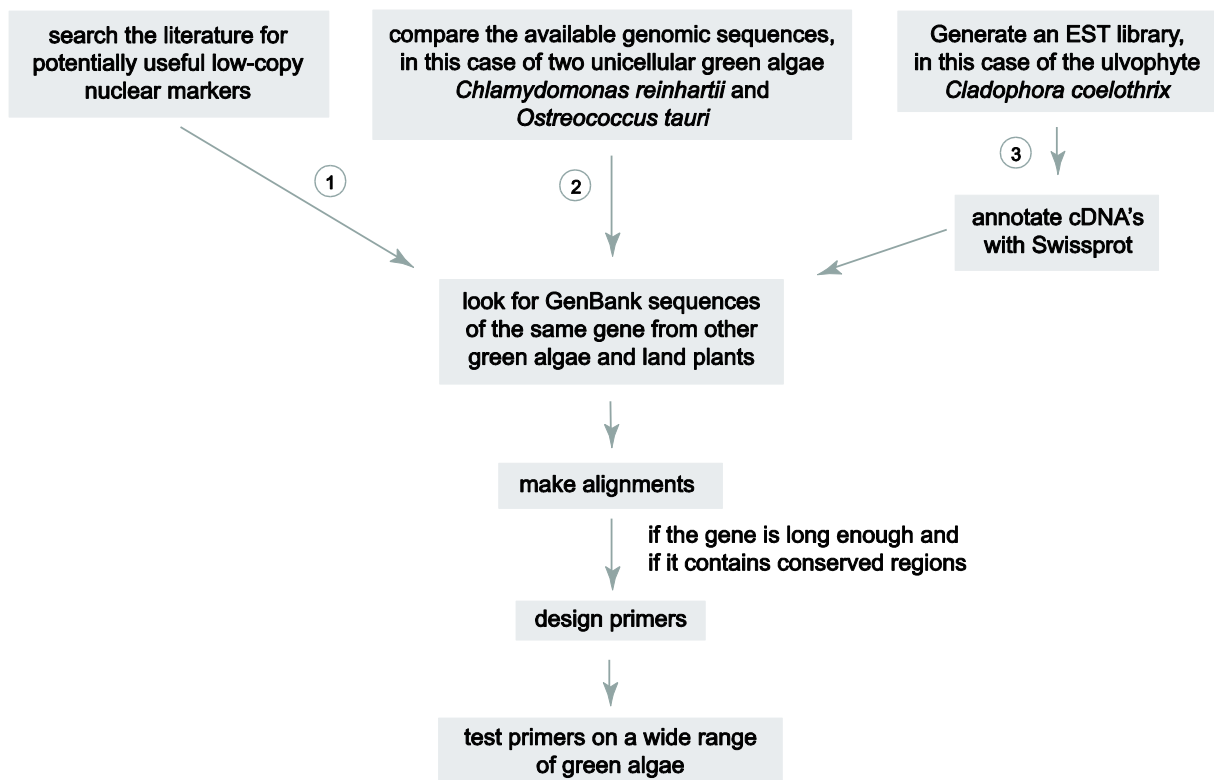
Figure 8. Flow chart showing three different methods to select suitable nuclear markers in green algae for which primers can be designed and tested on a wide range of green algae.

Tree building methods

Maximum likelihood and Bayesian Inference

Using the most accurate tree building methods and evolutionary models available is a basic necessity to obtain accurate phylogenies (Fig. 9) (Delsuc et al. 2005). Likelihood-based methods (maximum likelihood and Bayesian Inference) generally outperform methods based on distance or parsimony criteria because they allow the explicit incorporation of the processes of character evolution into probabilistic models to calculate the likelihood of the data given the model and tree. Maximum likelihood (ML) selects the tree that maximizes the probability of observing the data under a given model of sequence evolution. Bayesian methods derive the distribution of trees according to their posterior probability, using Bayes' mathematical formula to combine the likelihood function (including tree and model parameters) with prior probabilities on trees. Since prior knowledge is mostly lacking or bias towards one or the other tree is not generally desirable, flat priors are usually chosen, i.e. giving the same prior probability to all trees. Consequently, posterior tree probabilities depend primarily on the tree likelihood. Unlike ML, which optimizes model parameters to find the highest peak in parameter space and where confidence is obtained by non-parametric bootstrapping, Bayesian approaches integrate the model parameters by measuring the volume under a posterior probability surface rather than finding its maximum height and simultaneously estimates trees and measurements of uncertainty for every branch (Holder and Lewis 2003, Delsuc et al. 2005, Verbruggen and Theriot 2008). During Bayesian analysis, Markov chain Monte Carlo (MCMC) simulation is used to approximate the posterior probability distribution because the complexity of the phylogenetic likelihood functions prevents its analytical calculation. During each generation, a parameter change is proposed (topology, branch lengths and model parameters) and accepted if it increases the posterior probability. If the posterior probability decreases, the parameter change is either accepted or rejected depending on the amount of change in posterior probability. Whereas small changes are often accepted, large decreases are usually rejected. Because parameters are usually not near their optimal values during initial generations these first generations, called burn-in, need to be removed before a consensus tree of all post-burn-in samples can be made. In order to search tree space even more thoroughly, Metropolis-coupled MCMC, in which several chains are run in parallel can be applied. Metropolis-coupled MCMC is implemented in the commonly used BI program MrBayes (Ronquist and Huelsenbeck 2003). The first chain is the called the cold chain and only propose small parameter changes. The other chains are incrementally heated and propose larger parameter changes in order to find distant regions with high posterior probabilities. After each generation, chains can be swapped, i.e. a heated chain in a higher posterior probability region than the current cold chain can become the cold chain in order to find the local optimum. Only the output from the cold chain is used to summarize the posterior distribution and, due to chain swapping, this chain will contain a more complete image of the high posterior probability regions of tree space compared with a BI analysis based on a single MCMC chain. The downside of Metropolis-coupled MCMC is a considerably higher computational cost because several chains have to be run in parallel.
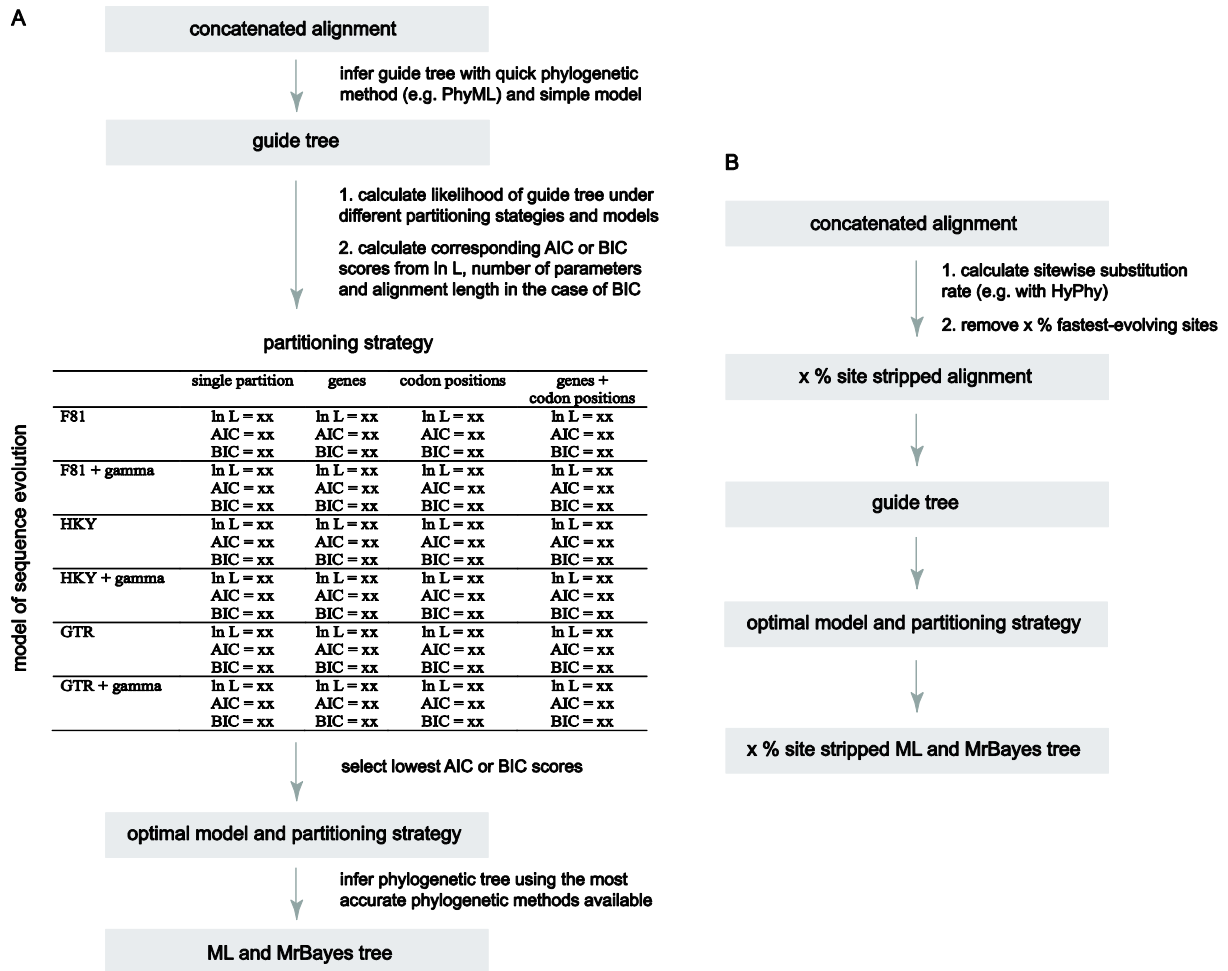
Missing data

Deep phylogenies require the simultaneous analysis of many characters and many taxa (Delsuc et al. 2005). Individual, orthologous genes can be combined into a supermatrix which inevitably involves a certain amount of missing data. Many studies have studied the effects of missing data on phylogenetic reconstruction. A simulation study suggests that the placement of individual taxa in a tree is robust to large amounts of missing data in the sequences of the taxa in question (up to 50% under the simulated conditions) and that model-based methods can deal with even greater amounts of missing data (Wiens 2005). Another simulations study demonstrates that Bayesian analyses are even more robust to missing data, i.e. the phylogenetic position of taxa with 95% of missing data in their sequence is still accurate, as long as the total number of characters in the dataset is large (Wiens and Moen 2008). Studies of empirical datasets have shown that datasets with up to 92% of missing data are still able to provide insights into various parts of the tree of life (Driskell et al. 2004, Philippe et al. 2004, Delsuc et al. 2005).

Models of sequence evolution

The General Time Reversible (GTR) model and its simpler variants include one or more parameters to describe the substitution rate between the different bases. The GTR model uses a set of parameters to describe the relative **substitution rate** between all combinations of bases (AC, AG, AT, CG, CT, and GT). The simpler models only consider transitions versus transversions or attribute an equal substitution rate to all possible changes. A second important component of a model are the **base frequencies**. They can be calculated directly from the dataset ('empirical' base frequencies) or optimized along with the other parameters of the model. A third common element of the model allows for variations of evolutionary rate across site (e.g. different codon positions in protein coding genes, loops and stems in ribosomal DNA). Such among site rate variation is commonly accounted for by assuming that the site rates follow a **gamma distribution** and/or by incorporating a proportion of **invariable sites.**

Partitioning strategies

A supermatrix, a dataset composed of different genes, often demands data partitioning to account for across site heterogeneity in evolutionary rate (Delsuc et al. 2005). Therefore, careful attention has to be paid to the selection of suitable partitioning strategies (Brown and Lemmon 2007, Li et al. 2008, Verbruggen and Theriot 2008). Protein coding genes usually benefit from partitioning into codon position. Empirical studies showed that codon position models perform better than models which do not take codon position into account (Shapiro et al. 2006). In order to accommodate differences in evolutionary rate among partitions **rate multipliers** can be used.

A

concatenated alignment

infer guide tree with quick phylogenetic
method (e.g. PhyML) and simple model

guide tree

1. calculate likelihood of guide tree under
different partitioning stategies and models

2. calculate corresponding AIC or BIC
scores from ln L, number of parameters
and alignment length in the case of BIC

partitioning strategy

model of sequence evolution	single partition	genes	codon positions	genes + codon positions
F81	ln L = xx AIC = xx BIC = xx	ln L = xx AIC = xx BIC = xx	ln L = xx AIC = xx BIC = xx	ln L = xx AIC = xx BIC = xx
F81 + gamma	ln L = xx AIC = xx BIC = xx	ln L = xx AIC = xx BIC = xx	ln L = xx AIC = xx BIC = xx	ln L = xx AIC = xx BIC = xx
HKY	ln L = xx AIC = xx BIC = xx	ln L = xx AIC = xx BIC = xx	ln L = xx AIC = xx BIC = xx	ln L = xx AIC = xx BIC = xx
HKY + gamma	ln L = xx AIC = xx BIC = xx	ln L = xx AIC = xx BIC = xx	ln L = xx AIC = xx BIC = xx	ln L = xx AIC = xx BIC = xx
GTR	ln L = xx AIC = xx BIC = xx	ln L = xx AIC = xx BIC = xx	ln L = xx AIC = xx BIC = xx	ln L = xx AIC = xx BIC = xx
GTR + gamma	ln L = xx AIC = xx BIC = xx	ln L = xx AIC = xx BIC = xx	ln L = xx AIC = xx BIC = xx	ln L = xx AIC = xx BIC = xx

select lowest AIC or BIC scores

optimal model and partitioning strategy

infer phylogenetic tree using the most
accurate phylogenetic methods available

ML and MrBayes tree

B

concatenated alignment

1. calculate sitewise substitution
rate (e.g. with HyPhy)

2. remove x % fastest-evolving sites

x % site stripped alignment

guide tree

optimal model and partitioning strategy

x % site stripped ML and MrBayes tree

Figure 9. Flow chart for accurate phylogenetic reconstruction. A. Phylogenetic reconstruction. B. Removal of fast-evolving sites.

Selection of the optimal partitioning strategy and model

A great number of models of sequence evolution have been described, ranging from simple models to complex models incorporating a lot of parameters. To reconstruct an accurate phylogenetic tree it is important to select a model of sequence evolution that approximates the evolutionary history of genes under study. A number of criteria have been developed to evaluate the fit of the different models to the data. The Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are two of these criteria that can be used for selection of the optimal partitioning strategy and evolutionary model (Figure 9A). Whereas AIC only penalizes for the number of model parameters, BIC also incorporates alignment length and thus penalizes a situation in which many parameters have to be estimated from a small dataset. In other words for the same dataset, AIC will prefer a more complex model than BIC. Because Bayesian analysis appears to be more sensitive to model under-specifications than ML, some authors have suggested that AIC scores can be used to choose a complex model for Bayesian analysis and BIC to choose a less complex model for ML analysis (Verbruggen and Theriot 2008). AIC and BIC calculation starts from a guide tree which is inferred with a fast distance based method (e.g. NJ) or a fast ML search under a simple model (e.g. PhyML,

Treefinder). In the second step log likelihoods of the guide tree under different partitioning strategies and models are calculated. Subsequently, the corresponding AIC or BIC scores are calculated and compared. The condition with the lowest AIC and/or BIC score is chosen for phylogenetic analysis.

Alternatively, Bayes factors (Nylander et al. 2004) can be used to compare different partitioning strategies and models. For each tested condition a separate Bayesian analyses has to be run which implies high computational times. This makes it unrealistic to compare many partitioning strategies and models in a Bayesian framework.

Complex models of sequence evolution

The secondary structure of ribosomal RNA consists of loops and stems. The nucleotides in the stem regions form base pairs and are interdependent because a change on one side of the stem has to be compensated in the other side of stem to avoid malfunction of the molecule. Since models of sequence evolution have to approach real evolution as close by as possible, it is recommended to incorporate this site interdependence in the model. This can be done by partitioning the ribosomal RNA into loops and stems and using a **doublet model** for the stem regions (Schöniger and Von Haeseler 1994). However, the use of a doublet model is computational demanding.

Instead of partitioning protein coding genes into codon positions, a **codon substitution model** can be applied. In this model, nucleotide triplets are considered as a single character and changes from one triplet to another one are considered taking into account that some changes are more likely than others (e.g. synonymous versus non-synonymous substitution). Although codon substitution models are a more realistic approximation of protein sequence evolution than codon position models, they come with a very high computational cost, hindering their use for large datasets (Shapiro et al. 2006).

Mixture models

Mixture models offer an attractive alternative to data partitioning and applying different models to the partitions. Whereas a partitioned analysis assumes that all sites within a partition arose from a single evolutionary process, mixture models relax this assumption by not expecting any prior partitioning and applying a set of different models to each site in the alignment. The log likelihood of each site is calculated as a weighted sum of the log likelihoods of each model for that site. The model weights correspond to the probability that the site has evolved under the model in question. Mixture models can thus apply different rate matrices to different parts of the dataset without explicitly partitioning it (Pagel and Meade 2004, Venditti et al. 2008). This is an elegant way to incorporate across site heterogeneity in the evolutionary process because it does not require prior knowledge about differences of evolutionary processes between different parts of the dataset and it avoids problems associated with differences of the evolutionary process within partitions that are defined a priori. Although analyses using mixture models outperform analyses based on partitioned datasets, they are restrictively time-consuming for large datasets.

Removal of fast-evolving sites

It has been shown that even the most accurate phylogenetic method and evolutionary models are unable to account for differences in the evolutionary process between species causing the well-known long-branch attraction artifact (Delsuc et al. 2005). Long-branch attraction artifacts can be reduced by improving taxon sampling and removing fast-evolving sites (Brinkmann et al. 2005, Delsuc et al. 2005). Fast-evolving sites are particularly challenging because they mask the true phylogenetic signal, resulting in loss of resolution and decrease of accuracy (Rodriguez-Ezpeleta et al. 2007). In chapter 2, we will illustrate that removal of fast-evolving sites (site stripping: Waddell et al. 1999) improves phylogenetic signal in the desired epoch (Fig. 9B).

Aims and outline of this thesis

First, we want to improve the phylogeny of the green lineage. To avoid biases caused by taxon or character sampling we examine deep relations of Chlorophyta by analyzing a combined dataset of several single copy nuclear markers, together with SSU nrDNA and plastid *rbc*L and *atp*B genes, including taxa from all major green algal lineages. Broad taxon and character sampling in combination with careful selection of models of sequence evolution and partitioning strategies, and thorough phylogenetic analyses, including methods to reduce data saturation will yield optimal resolution in the desired epoch (relation between UTC classes and relations within Ulvophyceae).

Guided by this improved green algal phylogenetic tree, we address various topics relating to molecular evolution of the Chlorophyta. The patterns of gain and loss of EF-1α and EFL genes, key enzymes in the translational apparatus, are interpreted in this phylogenetic framework. In addition we will screen green algal nuclear genes for the presence of non-canonical genetic codes, with emphasis on the ulvophycean relatives of the Dasycladales, where a non-canonical code had been detected earlier. Their phylogenetic distribution pattern will be studied and we will look for biological factors influencing the origin of a non-canonical code.

In **Chapter 2** deep relations of Chlorophyta will be examined and interpreted with respect to evolution of multicellularity and multinucleate cells.

In **Chapter 3** a 74-taxon phylogeny of the green lineage based on SSU nrDNA and two plastid genes (*rbc*L and *atp*B) is inferred and used to study the distribution and gain-loss patterns of elongation factor genes in green algae.

In **Chapter 4** the presence of a non-canonical code in some green algae is evaluated and discussed. Codon usage of canonical and non-canonical glutamine codons are calculated and their evolution is reconstructed.

Chapter 5 explores synonymous codon usage bias and GC content among the green plant lineage.

Chapter 6 calibrates a multi-locus phylogeny of the siphonous green algae in a geological timeframe.

In **Chapter 7** the phylogenetic positions of the enigmatic microfilamentous species *Uronema curvatum* and *Urospora microscopica* are assessed by molecular phylogenetic analysis of nuclear-encoded small and large subunit rDNA sequences

Finally, the main conclusions are summarized and discussed in **Chapter 8** and perspectives for future research are provided.

Authors' contribution

Chapter 2: Ellen Cocquyt and Olivier De Clerck designed the study. Ellen Cocquyt carried out lab work. Ellen Cocquyt and Frederik Leliaert maintained algal cultures and performed sequence alignment. Ellen Cocquyt and Heroen Verbruggen analyzed data. All authors wrote, revised and approved the final manuscript.

Chapter 3: Ellen Cocquyt, Olivier De Clerck, Heroen Verbruggen and Koen Sabbe designed the study. Ellen Cocquyt carried out lab work. Ellen Cocquyt and Frederik Leliaert maintained algal cultures and performed sequence alignment. Ellen Cocquyt and Heroen Verbruggen analyzed data and wrote the manuscript. Fredirick W. Zechman provided *atp*B sequences. All authors revised and approved the final manuscript.

C**hapter 4**: Ellen Cocquyt and Olivier De Clerck designed the study. Ellen Cocquyt and Gillian H. Gile carried out labwork. Ellen Cocquyt, Frederik Leliaert and Gillian H. Gile maintained algal cultures. Ellen Cocquyt and Heroen Verbruggen analyzed data. Ellen Cocquyt, Heroen Verbruggen and Olivier De Clerck wrote the manuscript. All authors revised and approved the final manuscript.

Chapter 5: Ellen Cocquyt and Olivier De Clerck designed the study. Ellen Cocquyt carried out lab work. Ellen Cocquyt and Heroen Verbruggen analyzed data and wrote the manuscript. All authors revised and approved the final manuscript.

Chapter 6: Heroen Verbruggen and Steven T. LoDuca designed and wrote the manuscript. Frederick W. Zechman, Diane S. Littler, Mark M. Littler, Frederik Leliaert, and Olivier De Clerck collected samples. Ellen Cocquyt, Matt Ashworth, Caroline Vlaeminck and Thomas Sauvage carried out lab work. Heroen Verbruggen analyzed data. All authors revised and approved the final manuscript.

Chapter 7: Frederik Leliaert, Jan Rueness and Christine A. Maggs designed the study. Jan Rueness provided the culture strains of *Uronema curvatum* and *Urospora microscopica*. Christian Boedeker provided sequences of *Chaetomorpha* and *Wittrockiella*. Frederik Leliaert and Ellen Cocquyt carried out lab work and analyzed the data. All authors revised and approved the final manuscript.

2

Ancient relationships among green algae inferred from nuclear and chloroplast genes [1]

Ellen Cocquyt, Heroen Verbruggen, Frederik Leliaert and Olivier De Clerck

Phycology Research Group and Center for Molecular Phylogenetics and Evolution, Ghent University, Krijgslaan 281 S8, 9000 Ghent, Belgium

Abstract

The Chlorophyta is one of the two divisions of green plants and harbors a wide range of green algae. The ancient relationships among three classes of this division, the Ulvophyceae, Trebouxiophyceae and Chlorophyceae (UTC) have been at the center of a long-standing debate. Our phylogenetic analyses (ML and BI) of seven nuclear genes, SSU nrDNA and two plastid markers with carefully chosen partitioning strategies and models of sequence evolution result in high support across the topology of the Chlorophyta, show the monophyly of the UTC classes and resolve the branching order among them. Even though topology tests (AU) do not exclude an alternative branching order of UTC classes, we show that moderate removal of fast-evolving sites improves the phylogenetic signal in the desired epoch. We also infer the relationships among the orders of the Ulvophyceae, providing novel insights into the evolution of multicellularity and multinucleate cells in the green tree of life.

Keywords

single-copy nuclear genes, Chlorophyta, green algae, molecular phylogenetics, Ulvophyceae, Chlorophyceae, Trebouxiophyceae

---

[1] submitted article

Introduction

The green plant lineage or Viridiplantae represents one of three groups of photosynthetic eukaryotes that diverged after enslavement of a cyanobacterium to make a primary chloroplast (Rodriguez-Ezpeleta et al. 2005). Ultrastructural and molecular studies have identified a major split within the Viridiplantae giving rise to two lineages, the Chlorophyta and the Streptophyta (Pickett-Heaps and Marchant 1972, Lewis and McCourt 2004). The Streptophyta consists of several lineages of freshwater green algae from which land plants evolved approximately 470 million years ago [2] (Karol et al. 2001, McCourt et al. 2004, Hall and Delwiche 2007). Whereas considerable progress has been made during the past decade in clarifying the relationships among the streptophyte green algae and land plants (Parkinson et al. 1999, Turmel et al. 2003, Turmel et al. 2006, Lemieux et al. 2007, Moore et al. 2007, Rodriguez-Ezpeleta et al. 2007, Saarela et al. 2007), the phylogeny and evolutionary history of the Chlorophyta has been more difficult to elucidate.

Members of the Chlorophyta are common inhabitants of aquatic environments and exhibit a remarkable morphological and cytological diversity ranging from the world's smallest free-living eukaryote *Ostreococcus*, over multicellular filaments and foliose blades, to highly complex siphonous life forms. Four classes are recognized within the Chlorophyta: Prasinophyceae, Ulvophyceae, Trebouxiophyceae and Chlorophyceae. The predominantly marine planktonic Prasinophyceae form a paraphyletic assemblage of unicellular organisms from which the Ulvophyceae, Trebouxiophyceae and Chlorophyceae (UTC) are derived (Steinkötter et al. 1994, Fawley et al. 2000, Lopez-Bautista and Chapman 2003, Guillou et al. 2004). The latter three classes form a monophyletic group termed UTC clade but the relationships among them form the basis of a longstanding debate (Pröschold and Leliaert 2007). All possible relationships between the UTC classes have been hypothesized over the years, depending on which ultrastructural characters were interpreted or which DNA locus or phylogenetic analysis method was used (Fig. 1). The unstable relationships exhibited among these three classes are likely due to a combination of their ancient age and the short time span over which they diverged from one another (O'Kelly 2007). The fossil record indicates the presence of the classes in the mid-Neoproterozoic (Butterfield et al. 1994) and molecular clock estimates situate the UTC divergence in the early Neoproterozoic (Douzery et al. 2004, Zimmer et al. 2007, Herron et al. 2009).

The UTC lineages occupy distinct and widely divergent ecological niches. The Chlorophyceae and Trebouxiophyceae are ubiquitous in freshwater and soil ecosystems. Furthermore, trebouxiophycean algae are the favored partner in symbiotic relationships with fungi to form lichens. Members of the Ulvophyceae, on the other hand, are best known as marine macroalgae in coastal ecosystems, but some of them live in damp subaerial habitats such as humid soil, rocks, tree bark and leaves (Lopez-Bautista and Chapman 2003, Watanabe and Nakayama 2007). Having evolved from a marine prasinophyte ancestor, the UTC lineages have experienced one to several transitions to a freshwater

---

[2] This estimate is based on a plant phylogeny derived from 27 plastid protein-coding genes calibrated with plant fossils (Sanderson et al. 2003). Estimates based on phylogenies of eukaryotes derived from 75 nuclear protein coding genes calibrated with vertebrate fossil suggest that vascular plant already diverged from mosses about 700 mya, which suggest that land plants are at least 700 my old (Heckman et al. 2001; Hedges 2004). A possible explanations for this huge difference might be the use of other calibration points (i.e. plant versus vertebrate fossils).
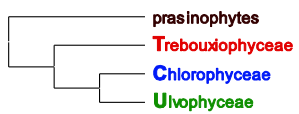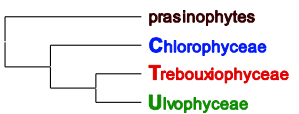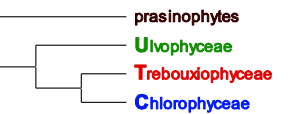
environment, which involves profound physiological adaptation (Mann 1996, Becker and Marin 2009). Furthermore, marine versus freshwater lifestyles coincide with differentiations in life-style: whereas the marine Ulvophyceae have a haplodiplontic life cycle with alternating generations of free-living gametophyte and sporophyte phases and sporic meiosis, freshwater green algae predominantly have a haploid vegetative phase and a single-celled, often dormant zygote as the diploid stage (haplontic life cycle with zygotic meiosis) (Lewis and McCourt 2004). Resolving the branching order among the three classes can help to understand the diversification of ecological, physiological and life history features of the UTC clade.

Whereas Prasinophyceae are unicellular marine plankton and Trebouxiophyceae and Chlorophyceae are unicells, colonies or in a few cases simple filaments, the predominantly marine class Ulvophyceae has evolved an unrivalled diversity of morphologies and cytological types. Morphologies in the class range from simple unicells to multicellular organisms of various levels of complexity (the green seaweeds). Cytological diversity is equally diverse and ranges from uni- to multinucleate cells and, in the extreme, giant tubular cells with millions of nuclei. Five major groups are recognized in the Ulvophyceae. The first has fairly typical cells and ranges in morphology from unicells to multicellular filaments and sheets, the multicellular types being more derived (orders Oltmannsiellopsidales, Ulvales and Ulotrichales) (Cocquyt et al. 2009). The second group has similar cells in a filamentous organization (order Trentepohliales). In contrast to these two groups, the other groups possess multinucleate cells. Representatives of the third group (Cladophorales) are branched filamentous seaweeds consisting of multinucleate cells with a few to thousands of nuclei arranged in non-motile cytoplasmic domains (siphonocladous organization). Members of the fourth group (Bryopsidales) are essentially unicellular: they consist of a single, giant tubular cell with thousands to millions of nuclei and complex patterns of cytoplasmic flow (siphonous organization). This siphonous cell can branch and coalesce to form highly complex seaweeds that dominate the flora of tropical coastal ecosystems (Vroom and Smith 2003). The seaweeds belonging to the fifth group (order Dasycladales) also feature a siphonous organization. *Acetabularia* and *Batophora* (Dasycladales) form an exception, because they have a single, giant nucleus situated in the rhizoid with which the alga is attached to its rocky substrate. Before sexual reproduction, this nucleus divides meiotically followed by several mitotic divisions, leading to the formation of numerous nuclei that populate the gametes. In addition to these five major multicellular or multinucleate groups, the Ulvophyceae also include the genus *Ignatius*, a soil alga forming clusters of a few cells (Watanabe and Nakayama 2007), and *Blastophysa*, a microscopic endophytic green alga. Several problems surround the classification of the Ulvophyceae. First, the monophyly of the Ulvophyceae has been questioned since its establishment because it lacks unique ultrastructural synapomorphies (Mattox and Stewart 1984, O'Kelly and Floyd 1984, Lewis and McCourt 2004). Second, molecular phylogenetic studies have not fully resolved the relationships among the orders. Most studies did reveal the presence of two major groups: Oltmannsiellopsidales–Ulvales–Ulotrichales, and a clade consisting of Trentepohliales and the siphonocladous and siphonous seaweed orders (Zechman et al. 1990, Watanabe et al. 2001, Lopez-Bautista and Chapman 2003, Watanabe and Nakayama 2007). However, the relationships within the latter clade, and the positions of the enigmatic genera *Ignatius* and *Blastophysa* have not been resolved satisfactorily.

The Chlorophyta thus offer a unique opportunity to study the proliferation of cytological types, morphological complexity, ecophysiological adaptations and life cycle evolution. A first step towards

understanding this diversification is to resolve the phylogenetic relationships among the UTC classes and the ulvophycean orders. The fact that previous molecular phylogenetic analyses have not yielded a satisfactory resolution of these taxa can be attributed at least in part to the difficulty in obtaining a good combination of taxon and gene sampling. Single-gene analyses with good taxon sampling have suffered from poor resolving power due to the low number of characters (Watanabe and Nakayama 2007). Conversely, studies based on longer alignments from complete organelle genomes have suffered from sparse and uneven taxon sampling (Pombert et al. 2004, Pombert et al. 2005), which can lead to systematic error in phylogenetic analysis (Brinkmann et al. 2005, Philippe et al. 2005).

Our goal in this study is to infer ancient relationships within the Chlorophyta, more specifically the divergence of UTC classes and ulvophycean orders. Our approach consists of phylogenetic analysis of a dataset that offers a good balance between taxon and gene sampling. Our alignment consists of seven single-copy nuclear markers, SSU nrDNA and two plastid genes for 43 taxa representing the major lineages of the Viridiplantae. Phylogenetic analyses are carried out with model-based techniques, paying careful attention to the selection of suitable partitioning strategies and models of sequence evolution. Noise-reduction techniques targeted at improving signal in the epoch of interest are applied.

Topology	prasinophytes, Trebouxiophyceae, Chlorophyceae, Ulvophyceae	prasinophytes, Chlorophyceae, Trebouxiophyceae, Ulvophyceae	prasinophytes, Ulvophyceae, Trebouxiophyceae, Chlorophyceae
Phylogenies	recent SSU nrDNA studies; mitochondrial genes	Chloroplast genome phylogenies	Some early SSU phylogenies
Ultrastructure and other evidence	Structural evidence in plastid genomes	- U and T share counter-clockwise orientation of the basal bodies - C has a direct opposite or clockwise orientation of the basal bodies (Sluiman 1989)	- T and C share a phycoplast and a non-persistent mitotic spindle - U has a persistent mitotic spindle (Mattox and Stewart, 1994)
AU test	ln L = -123822.64; p = 0.910	ln L = -123831.41; p = 0.114	ln L = -123834.31; p = 0.019
Bayes Factor	ln marginal L = -54020.39	ln marginal L = -54021.72 $2 \ln B_{T(UC), C(UT)} = 13.62$	ln marginal L = -54026.78 $2 \ln B_{T(UC), U(TC)} = 32.62$

Figure 1. The three alternative topologies for Ulvophyceae, Trebouxiophyceae and Chlorophyceae (UTC) have been supported by ultrastructural characteristics, molecular phylogenies and/or genomic structural features. Topology hypothesis testing using an AU tests based on our dataset containing seven nuclear genes, SSU nrDNA and two plastid genes suggested that an alternative branching order for the UTC classses (second column) was not significantly worse than the ML solution (first column), which supports a sister relationship between the Ulvophyceae and Chlorophyceae. Topology hypothesis testing using Bayesian factors provided very strong support in favor of the T(UC) topology.

Results

We generated EST data from a siphonocladous ulvophyte, *Cladophora coelothrix*, to facilitate the development of PCR primers for single-copy nuclear genes across the green algae. We selected 43 taxa representing all major lineages of the green algae and obtained DNA sequences for seven nuclear genes, the nuclear SSU rDNA and two plastid genes (Fig. S1, Table S1). The concatenated matrix is 10209 bp long and is 63% filled at the gene level and 61% at the nucleotide level (Fig. S1).

Our phylogenetic tree shows high resolution of the backbone of the green algal tree of life, including the branching order among the UTC classes and most ulvophycean orders (Figs. 2, S2 and S3). It reveals a sister relationship between the Ulvophyceae and Chlorophyceae, with the branch joining these two classes being very short, indicative of a rapid diversification. Although Bayes factors strongly favor the T(UC) configuration over the C(UT) and U(TC) topologies, AU tests suggested that an alternative branching order for the UTC classes was not significantly worse than the ML solution (Fig. 1).

Because the former result may in part be due to the application of tests that rely on character resampling to a dataset with a suboptimal signal to noise ratio, we used a site removal approach to counteract the erosion of ancient phylogenetic signal caused by fast-evolving sites. Our approach consisted of maximizing the phylogenetic signal in the epoch corresponding to the radiation of UTC classes and ulvophycean orders (Fig. 3). Incremental removal of fast-evolving sites did not change the topology but yielded a positive trend in phylogenetic signal in the relevant epoch (Fig. 3D, G_E = gain in relevant epoch), with an optimum at moderate amounts of site removal (25%). Fig. 3C clearly illustrates that statistical support for phylogenetic relationships within the relevant epoch is substantially improved by removing the 25% fastest sites, at the expense of signal in more recent epochs that are outside the scope of our study. The tree in Fig. 2 was inferred from the dataset with optimal signal in the relevant epoch (25% sites removed); the result of the phylogenetic analysis on the full data matrix is given in Fig. S2.

Discussion

Ancient, rapid radiations are difficult to resolve even with large datasets because there has been little time for the accumulation of substitutions, the underlying genes may have discordant phylogenetic histories, and inference methods can suffer from systematic error in some circumstances, e.g. the long branch attraction artifact (Delsuc et al. 2005, Wiens et al. 2008). We have taken some precautions to avoid systematic error: the use of multiple genes, broad taxon sampling, and the incremental removal of fast-evolving sites. The use of broad taxon sampling has been proven to be one of the most important design criteria leading to improved accuracy in phylogenetic studies (Zwickl and Hillis 2002, Geuten et al. 2007). Additionally, because fast-evolving sites can erode the overall phylogenetic signal of a dataset by masking the more reliable signal of slow sites, removing them can reveal a more accurate signal (Delsuc et al. 2005).

Radiation of the UTC classes

Our phylogenetic tree indicates an early divergence of the freshwater-terrestrial clades Trebouxiophyceae and Chlorophyceae from marine prasinophycean progenitors, closely followed by the diversification of the marine Ulvophyceae. A consistent timeframe of chlorophytan evolution is still lacking, mainly due to ambiguities in the interpretation of the scarce fossils, the estimated age of the clade being highly dependent on the taxonomic assignment of key fossils like *Proterocladus* and *Palaeastrum* (Butterfield et al. 1994, O'Kelly 2007). Yet, our data clearly shows that the radiation of

the UTC clades and the coinciding ecological diversification was rapid. Even with a liberal estimate of the Streptophyta-Chlorophyta split set at 1200 My (Herron et al. 2009), the UTC radiation would have occurred in ca. 20 to 30 million years (unpublished results), complicating interpretation of the ecological diversification.

The relationships between the Ulvophyceae, Trebouxiophyceae and Chlorophyceae have not been unambiguously resolved up until now. Phylogenies based on SSU nrDNA sequences have been inconclusive, with the branching order of the UTC classes depending on taxon sampling, alignment methods and phylogenetic techniques (Fig. 1). Some early SSU phylogenies showed a sister relationship between Chlorophyceae and Trebouxiophyceae (Friedl 1995, Bhattacharya et al. 1996, Krienitz et al. 2001, Lopez-Bautista and Chapman 2003) while others revealed a sister relation between the Chlorophyceae and Ulvophyceae (Friedl and O'Kelly 2002, Lewis and Lewis 2005, Watanabe and Nakayama 2007). Although chloroplast phylogenomics has been valuable to resolve problematic relationships among green algae (Qiu et al. 2006, Jansen et al. 2007, Lemieux et al. 2007, Turmel et al. 2008, Turmel et al. 2009), it has not been able to resolve the branching order of the UTC classes conclusively. Whereas phylogenetic analysis of 58 concatenated chloroplast genes supported a sister relation between Ulvophyceae and Trebouxiophyceae, chloroplast gene order data and genome structure (gene losses and rearrangements within conserved gene clusters), as well as a phylogenetic analysis of seven mitochondrial genes supported a sister relationship between Ulvophyceae and Chlorophyceae (Pombert et al. 2004, Pombert et al. 2005).

Ultrastructural evidence has been interpreted as either providing support for a sister relation between Chlorophyceae and Trebouxiophyceae (Mattox and Stewart 1984) or between Trebouxiophyceae and Ulvophyceae (Sluiman 1989) based on the shared presence of a phycoplast and a non-persistent mitotic spindle in the first case and a counterclockwise orientation of the flagellar basal bodies in the second case (Fig 1). Our phylogeny suggests an alternative branching order that necessitates reinterpretation of the evolution of ultrastructural features. It now appears more likely 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 Chlorodendrales (e.g. *Tetraselmis*), the closest prasinophyte relatives of the UTC clade (Mattox and Stewart 1984). 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 *Tetraselmis*. The modern UTC clade appeared after that a new mode of cell division, which is mediated by a phycoplast and enables complex multicellular growth, was introduced in the marine ancestor of the UTC clade. The following ecophysiological adaptations lead to the success of the Chlorophyceae and Trebouxiophyceae in freshwater habitats (Becker and Marin 2009).
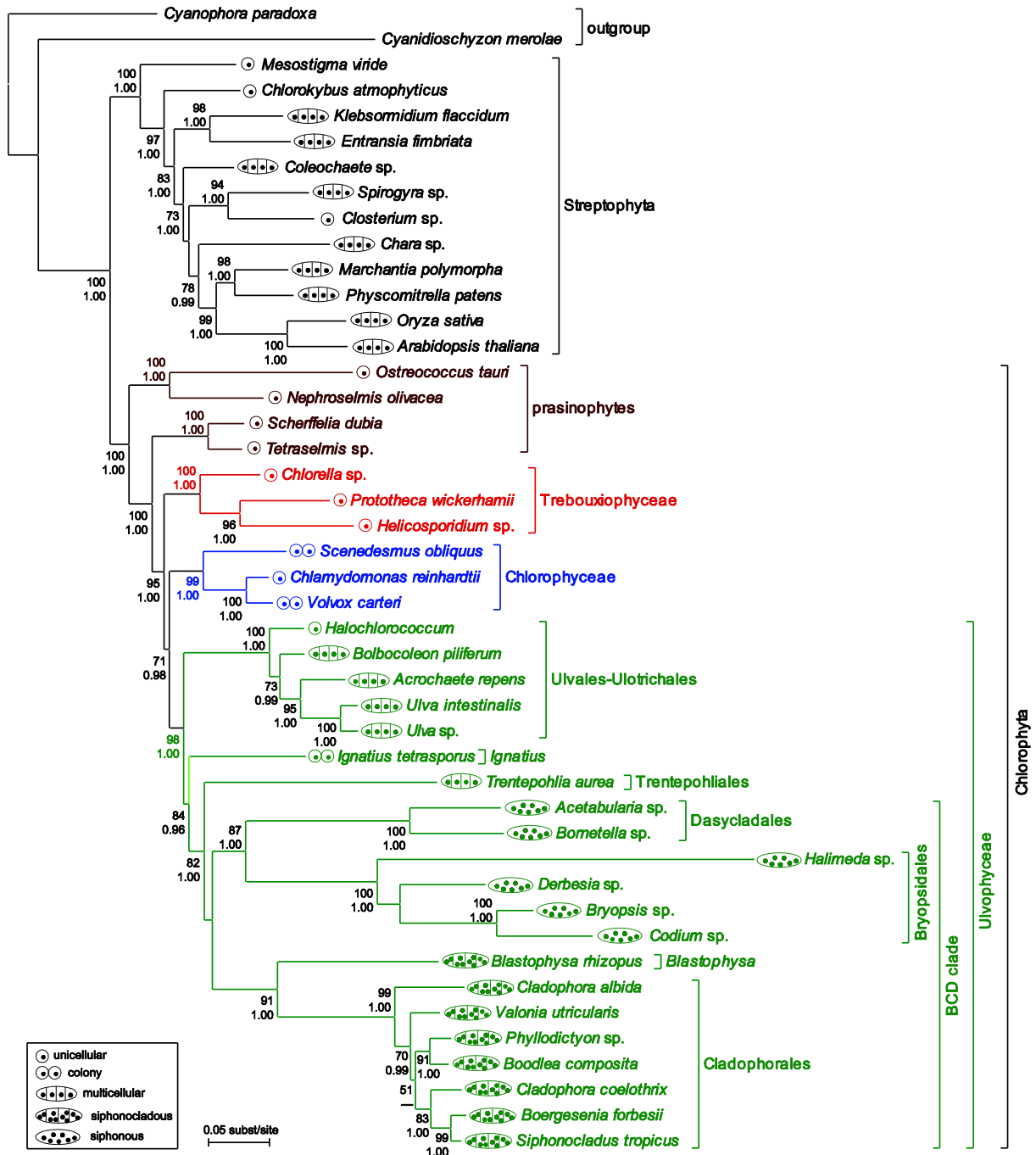
Figure 2. Phylogeny of the green plant lineage obtained by maximum likelihood inference of the 25% site stripped dataset containing 7 nuclear genes, nuclear SSU nrDNA and plastid genes *rbc*L and *atp*B. Numbers at nodes indicate ML bootstrap values (top) and posterior probabilities (bottom); values below respectively 50 and 0.9 are not shown. BCD clade stands for the orders Bryopsidales, Cladophorales and Dasycladales.

Diversification and specialization of the Ulvophyceae

For the first time, the monophyly of the Ulvophyceae is inferred with high statistical support (BV = 98; PP = 1.00). In addition, our phylogeny provides a solid hypothesis to explain the cytological and ecophysiological differentiation of this predominantly marine class of green algae, by resolving the

relationships among the major clades. Our phylogeny reveals several ancient lineages which diverged early in the evolution of the Ulvophyceae. In contrast to previous studies, the basal divergences in the present tree are well-supported. Congruent with earlier studies, the clade comprising Ulvales and Ulotrichales branches first. *Ignatius* and *Trentepohliales*, two relatively inconspicuous and somewhat neglected taxa both confined to subaerial habitats, form separate, ancient lineages. They are associated with the clade consisting of Bryopsidales, Cladophorales and Dasycladales (BCD clade) (BV 84; PP 0.96). Contrary to expectations based on ultrastructure (O'Kelly and Floyd 1984), the Dasycladales group with the Bryopsidales instead of the Cladophorales [3].

The position of *Ignatius* has never been fully resolved, being either embedded in the Ulvales— Ulotrichales clade or clustering with the Trentepohliales and the BCD clade (Watanabe and Nakayama 2007). The relationship of *Ignatius* with the Trentepohliales and the BCD clade revealed in this study is corroborated by features of the flagellar apparatus (Watanabe and Nakayama 2007) and the shared presence of the elongation factor-1 alpha gene instead of the elongation factor-like gene used in all earlier diverging chlorophytan lineages (Cocquyt et al. 2009). The enigmatic, endophytic green alga *Blastophysa*, which has often been allied with the Bryopsidales (Burrows 1991, Brodie et al. 2007, Kraft 2007), is here revealed to be sister to the Cladophorales (BV 91 and PP 1.00). This relationship is supported by morphological, ultrastructural, cytological, and biochemical features [4] (O'Kelly and Floyd 1984, Chappell et al. 1991).

The observed branching pattern yields insight in the morphological and cytological diversification of the Ulvophyceae. Multicellularity evolved independently in the Ulvales—Ulotrichales and the Trentepohliales + BCD clade. After multinucleate cells evolved in a common ancestor of the BCD clade, the siphonous thallus structure evolved from a multicellular state in the Bryopsidales and Dasycladales. This interpretation is further supported by the occurrence of cross walls at the base of reproductive structures in one of the two major bryopsidalean lineages (Bryopsidineae). The evolution of siphonocladalean and siphonous architectures coincides with several cytological and cytoskeletal specializations. These giant-celled green algae have evolved a unique mechanism of wounding response, in which injured cells contract their protoplasm into numerous spheres that regenerate into new cells and thalli (O'Neil and La Claire 1984, Menzel 1988, Kim et al. 2001). This mechanism has also given rise to segregative cell division, a specialized type of cell division found in some Cladophorales (Leliaert et al. 2007). Unlike the Cladophorales, the siphonous seaweeds Bryopsidales and Dasycladales are characterized by vigorous cytoplasmic streaming in which organelles are transported in a thin parietal layer of cytoplasm. This cytoplasmic transport has been shown to have certain selective advantages, such as transport of nutrients from the rhizoidal holdfast

---

[3] Common features of the Dasycladales and Cladophorales relate to certain aspects of the flagellar apparatus: its flattened aspect, structure of the striated distal fibers and absence of terminal caps. Based on these observations O'Kelly and Floyd O'Kelly C. J. and G. L. Floyd. 1984. Correlations among patterns of sporangial structure and development, life histories, and ultrastructural features in the Ulvophyceae. Pp. 121-156 *in* D. Irvine, and D. John, eds. Systematics of the green algae. The Systematics Association Special Volume 27, Academic Press, London and Orlando. suggested a closer relationship between Dasycladales and Cladophorales than to any other group within the Ulvophyceae.

[4] See Chapter 8 for a more detailed discussion

to the upright photosynthetic thallus parts (Littler et al. 1988, Chisholm et al. 1996), or chloroplast migration for optimal photosynthesis (Drew and Abel 1992). Despite the fact that siphonous thalli are essentially composed of a single giant cell, many bryopsidalean and dasycladalean representatives form large, differentiated thalli with a complex organization of interwoven siphons (Verbruggen et al. 2009).

The Ulvophyceae are an essentially marine clade, but a few representatives have adapted to freshwater environments (a number of Ulvales, Ulotrichales and Cladophorales) and subaerial habitats (*Ignatius*, Trentepohliales and two Cladophorales species). Our phylogeny thus implies that the transition to freshwater and terrestrial habitats must not only have occurred in the Chlorophyceae and Trebouxiophyceae (Becker and Marin 2009), but also several times independently within the Ulvophyceae.
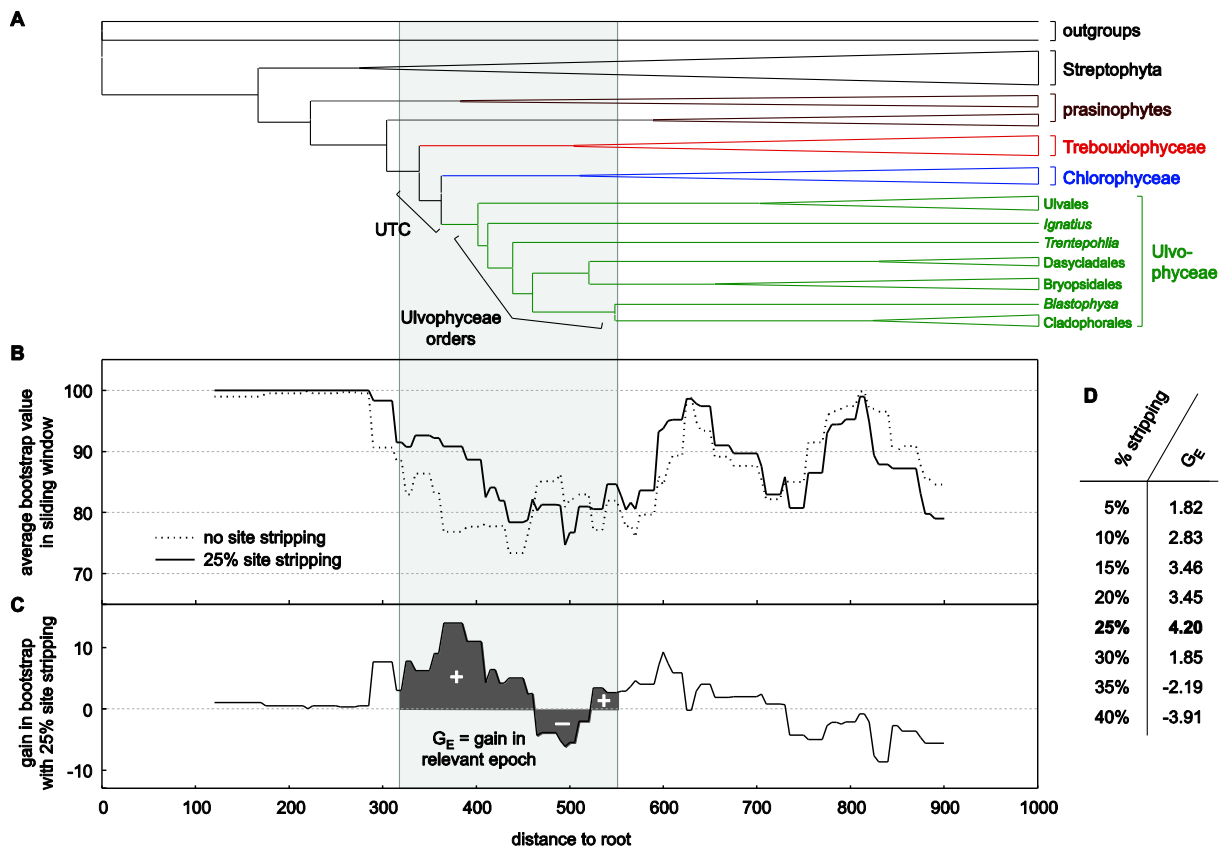
Figure 3. Illustration of the site stripping approach.

(A) Based on a rate-smoothed tree, the epoch in which the relationships of interest are situated is selected (gray band across figure).

(B) The strength of the phylogenetic signal in the data (measured as average bootstrap values [BV] in a sliding window across the tree) is plotted for an analysis from which the 25% fastest-evolving characters had been stripped and compared to the condition without site stripping.

(C) The gain in bootstrap values, calculated as the BV of site stripped condition minus the BV of the condition without site stripping is plotted. This graph shows net gain in older epochs and net loss in younger epochs. In this example, there is net gain in the epoch of interest (dark gray).

(D) The gain in the epoch of interest is calculated for all site stripping conditions. This table shows that moderate site stripping yields optimal phylogenetic signal in the relevant epoch.

Progress and perspectives

The slow progress in green algal phylogenetics can in part be attributed to the limited availability of genomic data and the difficulties in consistently amplifying single-copy nuclear markers over the entire spectrum of a group of organisms that dates back well into the Proterozoic. Our results indicate that advanced phylogenetic analyses of multiple markers from different genomic compartments show great promise in resolving ancient divergences within the green algae. The present phylogeny has provided us with a framework to advance our understanding of the nature of the morphological, cytological and ecological diversification of the Chlorophyta. Additional insight would be gained from a dated phylogeny, which would allow an assessment of the timing of these evolutionary events and, by consequence, the global ecological circumstances under which this radiation took place.

Material and methods

Algal strains

Algal strain information is provided in Table S2. Cultures were grown under cool white fluorescent lights at 18°C, with a 12/12h light/dark cycle. Dasycladales, Cladophorales and *Derbesia* were grown at 23°C. Bold's basal medium and f/2 medium was used for freshwater and marine species, respectively (Andersen 2005).

RNA isolation

Total RNA was extracted with a RNeasy Plant Mini Kit (Qiagen) or a NucleoSpin® RNA Plant kit (Machery-Nagel) according to the manufacturer's instructions, including a DNAse step to eliminate genomic DNA contamination. RNA quality was checked on a 1% agarose gel and with spectrophotometric analysis (Sambrook et al. 1989).

cDNA library construction and primer design

A standard cDNA library was constructed from 30 μg of total RNA of *Cladophora coelothrix* by VERTIS Biotechnologie AG (Germany) and screened for nuclear genes potentially useful for green algal phylogeny. After the cDNA library had been plated, 200 clones were randomly picked and sequenced with M13 reverse primer (Table S3). Clones with a positive blastx hit to the Swiss-Prot database were also sequenced with the M13 forward primer (Table S3).

For each recovered gene, an alignment including GenBank sequences from other green plants was made. Primers were designed for sufficiently long genes that contained conserved regions and tested on a variety of green algae, leading to successful amplification of 4 nuclear genes across a wide range of green algae: 40S ribosomal protein S9, 60S ribosomal proteins L3 and L17 and oxygen-evolving enhancer protein (OEE1). Actin and some glyceraldehyde-3-phosphate dehydrogenase (GapA)

primers were found in literature (Table S3). Glucose-6-phosphate isomerase (GPI) and other GapA primers were based upon aligned GenBank sequences from green algae (e.g., available genome sequences of *Ostreococcus* and *Chlamydomonas*) and land plants (Table S3). Primers of *rbc*L (Hanyuda et al. 2000), *atp*B (Wolf 1997, Karol et al. 2001) and SSU nrDNA (Zechman et al. 1990, Lewis and Lewis 2005) were taken from the literature.

Reverse Transcriptase and Polymerase Chain Reaction

For amplification of the nuclear genes, cDNA was constructed from total RNA with the Omniscript RT kit (Qiagen) according to the manufacturer's instructions. The reaction was incubated for several hours at 37°C. PCR amplification was performed with the following reaction mixture: 1 µl of cDNA, 2.5 µl of 10x Buffer (Qiagen), 0.5 µl dNTP's (10 mM), 0.5 µl MgCl (25 mM, Qiagen), 0.5 µl of each primer (10 µM), 0.25 µl BSA (10 µg/µl), 18.125 µl sterilized MilliQ water and 0.125 µl Taq polymerase (5 U/µl, Qiagen). The amplification profile consisted of an initial denaturation of 2 min at 94 °C, followed by 35 cycles of 30 s at 94 °C, 30 s at 50°C or 55 °C and 45 s at 72 °C and a final extension of 10 min at 72 °C. Products of expected size (Table S4) were either sequenced directly or cloned first.

Cloning and sequencing

PCR products were first sequenced with the forward primer with an Applied Biosystems 3130xl. Sequences were blasted against the GenBank protein database (blastx) to check for potential bacterial contaminants. Sequences without ambiguous base calls yielding a significant hit for Viridiplantae were further sequenced with the reverse primer. When ambiguous base calls were present in sequences, samples were cloned if the rough sequence gave a significant blastx hit for Viridiplantae. Cloning was performed with the pGEM®-T Vector System (Promega) according to the manufacturer's instructions. After ligation, transformation and incubation, the white colonies were transferred to 15 µl distilled water, boiled for 10 minutes to lyse cells and centrifuged to pellet cells walls and harvest the DNA in the liquid phase. Three to five clones were PCR amplified and sequenced with the vector specific primers T7 and SP6 following the protocol described above.

Alignment

Sequence reads were assembled with AutoAssembler 1.4.0 (ABI Prism, Perkin Elmer). Additional sequences were retrieved from GenBank and aligned with our own sequences (Table S1). Nuclear and plastid genes were aligned by eye, taking their corresponding amino acids into account. SSU nrDNA sequences were aligned based on their RNA secondary structure (Cocquyt et al. 2009). The ten loci were concatenated, yielding an alignment of 10209 bases.

Model selection procedure

A suitable partitioning strategy and suitable models of sequence evolution were selected with a three-step procedure based on the Bayesian Information Criterion (BIC). The guide tree used during the entire procedure was obtained by ML analysis of the unpartitioned concatenated alignment with PhyML, using a JC+Γ_8 model (Guindon and Gascuel 2003). All subsequent likelihood optimizations and BIC calculations were carried out with Treefinder (Jobb et al. 2004). The first step consisted of optimizing the likelihood for eight potential partitioning strategies, assuming a HKY+Γ_8 model for each partition. The three partitioning strategies with the best fit to the data (lowest BIC scores) were retained for further evaluation. The second step involved model selection for individual partitions. The likelihood of each partition present in the three retained partitioning strategies was optimized for three variants of the general time reversible model (F81, HKY and GTR), with and without among-site rate heterogeneity (+Γ_8). Because not all genes were sampled for all taxa, the guide tree was pruned to the taxa present in the partition in question before each optimization. The partition-specific models obtaining the lowest BIC score were passed on to the third step, which consisted of a re-evaluation of the three partitioning strategies retained from the first step using the models selected for these partitions in the second step. The partitioning strategy plus model combination that received the lowest BIC score in the third step was used in the phylogenetic analyses. The model selection procedure proposed 8 partitions: SSU nrDNA was partitioned into loops and stems (2 partitions), nuclear and plastid genes were partitioned according to codon position (2×3 partitions). GTR+Γ_8 was the optimal model for all partitions.

Phylogenetic analysis

Maximum likelihood analysis was performed with TreeFinder, which allows likelihood searches under partitioned models (Jobb et al. 2004). Due to the relatively low tree space coverage in TreeFinder compared to other ML programs, an analysis pipeline was created to increase tree space coverage by running analyses from many starting trees. A first set of starting trees was created by randomly modifying the PhyML guide tree by 100 and 200 nearest neighbor interchange (NNI) steps (50 replicates each). ML searches were run from these 100 starting trees and the three trees yielding the highest likelihood were used as the starting point for another set of NNI modifications of 20 and 50 steps (50 replicates each). A second set of ML searches was run from each of the resulting 300 starting trees. The tree with the highest likelihood resulting from this set of analyses was retained as the global ML solution. The bootstrap resampling method was used to assess statistical support (1000 pseudo-replicates).

Bayesian phylogenetic inference was carried out with MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003). Two parallel runs, each consisting of four incrementally heated chains were run for 25 million generations, sampling every thousand generations. Convergence of log-likelihoods and parameter values was assessed in Tracer v1.4 (Rambaut and Drummond 2007). A burnin sample of 2.5 million trees was removed before constructing the majority rule consensus tree.

Topological hypothesis testing

Approximately unbiased tests (AU test, Shimodaira 2002) were used to test the alternative relationships between UTC classes. Site likelihoods were calculated by maximum likelihood optimization in Treefinder using the same model specifications as for phylogenetic inference. AU tests were performed with CONSEL V0.1i (Shimodaira and Hasegawa 2001).

In addition Bayes Factors were used to evaluate the alternative branching order between UTC classes (Kass and Raftery 1995). Two additional Bayesian analyses using identical conditions as the original analysis that yielded the T(UC) phylogeny were run while constraining the topology to conform to either C(UT) or U(TC). The harmonic mean estimates of the marginal likelihoods of the competing hypotheses were obtained with the sump command in MrBayes and used to calculate the Bayes Factors (BF) as twice the difference in marginal likelihood between the competing hypotheses (Nylander et al. 2004).
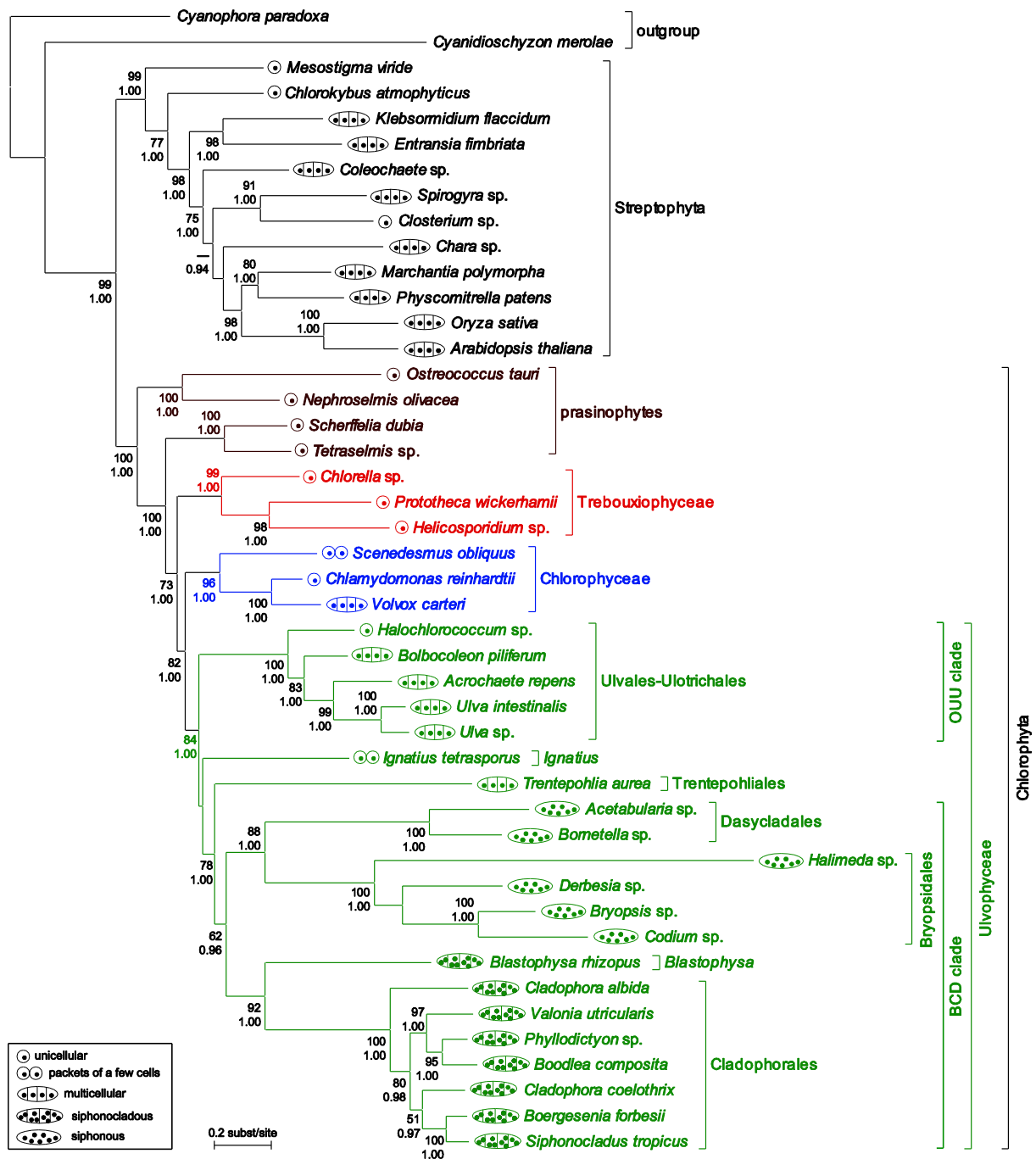
Removal of fast-evolving sites: site stripping

We applied a modified site stripping approach (Waddell et al. 1999) to improve the signal to noise ratio in the epoch of interest. Site-specific rates were calculated in HyPhy (Pond et al. 2005). Progressive removal of the fastest sites, 5% each time, resulted in a set of alignments that were subjected to ML analysis as described above. We used a novel modus operandi to reveal the trend of signal change with increasing amounts of site stripping and select an optimal site stripping condition (Fig. 3). First, we performed rate smoothing on the ML tree obtained from the complete alignment to make branch lengths roughly proportional to evolutionary time (Fig. 3A), using the nonparametric rate smoothing (NPRS) method with the Powell optimization algorithm implemented in r8s (Sanderson 2003). Second, the epoch of interest (containing the UTC and ulvophycean diversification) was located in the tree. Third, the strength of phylogenetic signal, measured as average bootstrap values (BV) in a sliding window across the rate-smoothed tree, was plotted for all site-stripping conditions (e.g. Fig. 3B). Fourth, we calculated signal gain in the epoch of interest (G_E) by integrating the function representing the difference between the BV of the site-stripped condition and the BV of the analysis on the complete alignment across the epoch of interest (Fig. 3C). Finally, these gain values were compared between site stripping conditions, leading to the conclusion that moderate site stripping (25% of characters removed) yielded maximum signal in the epoch of interest (Fig. 3D). The rationale of using BV as a proxy of signal strength is that these values reflect the clarity of the signal in the data, i.e., are inversely related to the amount of conflict in the data (Felsenstein 1985, Soltis and Soltis 2003). It may seem as circular reasoning to use BV to decide on the most suitable site stripping condition and subsequently use BV to come to conclusions about the strength of relationships. However, this is not entirely true because G_E is calculated over a wide temporal range and across all lineages in the relevant epoch, so it does not rely exclusively on the BV of the branches we wish to resolve.

Acknowledgements

Additional files

Figure S2. Phylogeny of the green plant lineage obtained by maximum likelihood inference of the complete dataset containing 7 nuclear genes, nuclear SSU rDNA and plastid genes *rbc*L and *atp*B. Numbers at nodes indicate ML bootstrap values (top) and posterior probabilities (bottom); values below respectively 50 and 0.9 are not shown.

Fig. S1. Data availability matrix. Graphical representation of our concatenated alignment, showing the availability of sequence data. The color of column and row headers indicate the amount of data available for that column or row. Green indicates high data availability, red indicates low data availability and yellow/orange represents intermediate data availability.
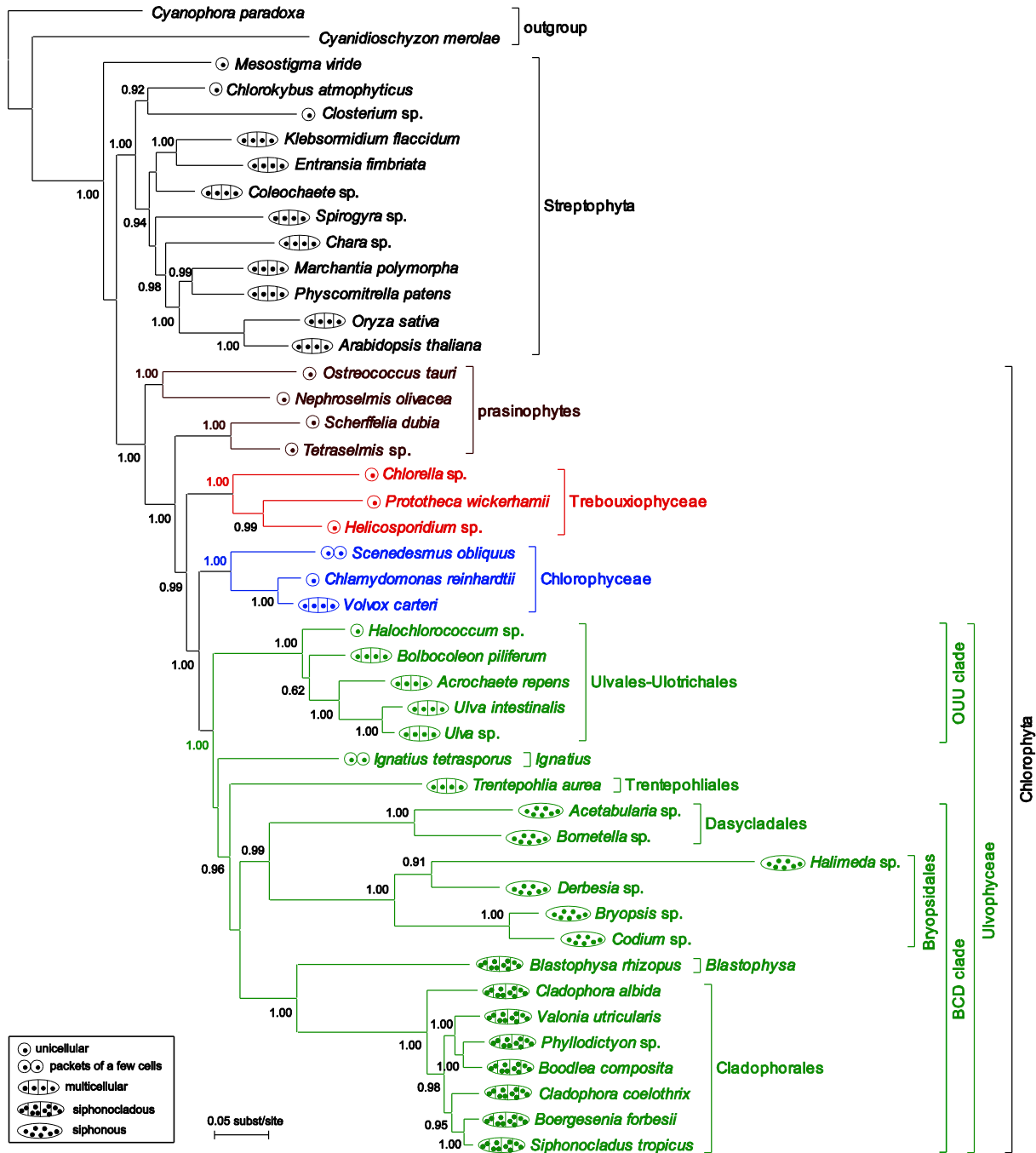
Figure S3. Phylogeny of the green plant lineage inferred from amino-acid sequences of 7 nuclear genes and plastid genes *rbc*L and *atp*B and nucleotide sequences of nuclear SSU rDNA with Bayesian techniques. Numbers at nodes indicate posterior probabilities; values below 0.9 are not shown.

Because TreeFinder cannot simultaneously analyze nucleotide and amino acid sequences only MrBayes was used to infer this tree. Amino acid sequences used a WAG model with among-site rate heterogeneity (+Γ8), SSU nrDNA was partitioned into loops and stems and used a GTR+Γ8 model. Two parallel runs, each consisting of four incrementally heated chains were run for 3 million generations, sampling every thousand generations. Convergence of log-likelihoods and parameter values was assessed in Tracer v1.4 (Rambaut and Drummond 2007). A burnin sample of 300.000 trees was removed before constructing the majority rule consensus tree.

Table S1. Genbank accession numbers for nucleotide sequences of actin, glucose-6-phosphate isomerase, glyceraldehydes-3-phosphate dehydrogenase, histone, oxygen evolving protein, 40S ribosomal protein S9 and 60S ribosomal protein L3 and L17.

	actin	G6PI	GapA	histone	OEE1	40S_S9	60S_L3	60S_L17
Chlorophyta								
Ulvophyceae								
Ulotrichales								
Halochlorococcum moorei	R102	R102		R73_2				
Ulvales								
Acrochaete repens	R94	R94	R94_Gap2	R94_2	R94	R94_2		
Bolbocoleon piliferum		R95	R95_GapM	R60_2				
Enteromorpha sp.	R109	R106	R109_GapM	R109_1	R109_1_Oxy1	R109_3UTR		
Ulva sp.	AB106563 (*U. pertusa*)	R110_10	R49_2	R110_1	R110_Oxy2			
Ignatius-clade								
Ignatius tetrasporus			R66_GapM	R66_4	R66			R66
Bryopsidales								
Blastophysa rhizopus		R104	R104_GapM	R104	R104			
Bryopsis sp.			R105	R105_1	R105/ AB293980 (*B. plumosa*)		R105	
Codium sp.			R63	R63	R63_1_Oxy1			R63
Derbesia sp.	R98		R98_2	R98	R98			
Halimeda spp	KZN2K4_20 (H. cuneata2)			R65_1				
Dasycladales								
Acetabularia acetabulum	R7	AAL00001487 (TBestDB)		R103	R103_1_Oxy2		EC096877	CF259099
Bornetella sphaerica			R80	R80_1	R80			
Siphonocladales								
Boergesenia forbesii	R21_1	R21_2		R78		R78_1		R78
Boodlea composita	R8_1	R38_6	R79_GapM		R79	R79_1		R79
Cladophora albida		R88_2	R88_19	R88				R88
Cladophora coelothrix	R35_3	R35_12		cDNA_lib117	cDNA_lib122	cDNA_lib128	cDNA_lib44	cDNA_lib80
Phyllodictyon spp	R18	R85		R85	R85	R85_2	R85	R85
Siphonocladus tropicus	R29_3	R84		R84_5	R84	R84_1		R84
Valonia utricularis		R86		R86_5	R86			R86
Trentepohliales								
Trentepohlia aurea				R107	R107_Oxy2			
Chlorophyceae								
Chlamydomonas reinhardtii	D50838	R44_2	L27668	R96_5	X13826	R96_2 = RPS9	R96 = RPL3	XM_001693402
Scenedesmus obliquus	R69	SOL00005809 (TBestDB)		R69	R69	EC189050	EC189501	R69
Volvox carteri	M33963			X06963	AF110780	FD918100	FD837306	
Trebouxiophyceae								
Chlorella kessleri	R97	R97_5	R97_5	R97	R97			R97_1
Helicosporidium sp.	AF317896			CX128748		CX128902		CX128917
Prototheca wickerhamii	EC181529					EC181191	EC183246	EC180670

	actin	G6PI	GapA	histone	OEE1	40S_S9	60S_L3	60S_L17
Prasinophyceae								
Chlorodendrales								
Scheffelia dubia	AF061018		DQ270259		AJ919716	AJ919712	AL132935	AJ919390
Tetraselmis spp	R76 (*T. striata*)	R76 (*T. striata*)		R76 (*T. striata*)	R76_Oxy2/ AB293977 (*T. cordiformis*)			R76 (*T. striata*)
Mamiellales								
Ostreococcus tauri	CR954216	Steven	DQ649076	R68_2	R68		R68	CR954206
Pseudoscourfieldiales								
Nephroselmis olivacea	EC732532	R112_clon	R112_2_GapM	R112	R112/ AB293978			
Streptophyta								
Mesostigmatales								
Mesostigma viride	AF061020	R111	R111_11	R111	R111/ DN255652		R111	R111
Chlorokybales								
Chlorokybus atmophyticus		R99	DQ270263	R99				
Klebsormidiales								
Entransia fimbriata				R101	R101_1_Oxy1		R101	
Klebsormidium flaccidum			R46_7	R67	R67_1_Oxy2	R67_2	R67	R67
Zygnematales								
Closterium sp.	R4	R4_3			AB293981			
Spirogyra sp.	AF061021		AJ246030					
Coleochaetales								
Coleochaete spp	AF061019 (*C. scutata*)		DQ270264 (*C. scutata*)	R113_3 (*C. orbicularis*)				
Charales								
Chara spp	DQ846905 (*C. contraria*)		DQ270262 (*C. vulgaris*)		AB293979 (*C. braunii*)			
Embryophyta (land plants)								
Marchantiophyta (Liverworts)								
Marchantia polymorpha	AB100427	BJ863864	AJ246026	AB185062	BJ844290	BJ861755	BJ852563	BJ853031
Bryophyta (Mosses)								
Physcomitrella patens	XM_001782636	XM_001760154	DQ270266	BJ975811	XM_001763206	XM_001754317	XM_001782386	XM_001767807
Spermatophyta (Seed plants)								
Arabidopsis thaliana	M20016	AB044951	NM_101161	X60429	AJ145957	AB010077	M32655	AC004393
Oryza sativa	X16280	D45217	NM_001059519	NM_001056811	NM_001049669	NM_001055539	D12630	NM_001069236
other eukaryotes								
Archaeplastida								
Glaucocystophyta								
Cyanophora paradoxa	CPU90325	DQ812897	DQ270258	ES235644	AJ784854	EC666063	EC661414	EC666069
Rhodophyta								
Cyanidiophyceae								
Cyanidioschyzon merolae	D32140	CMT497C_Steven	AP006492	AP006496	AB159597		AP006495	

Table S1 continued. Genbank accession numbers for nucleotide sequences of *atp*B, *rbc*L and SSU rDNA.

	*atp*B	*rbc*L	SSU rDNA
Chlorophyta			
Ulvophyceae			
Ulotrichales			
Halochlorococcum moorei			AY198122
Ulvales			
Acrochaete repens		FJ715715	FJ715684
Bolbocoleon piliferum		FJ715716	AY205330
Enteromorpha sp.		AY422552	AJ000040
Ulva sp.		AF499668 (*U. fenestrate*)	
Ignatius-clade			
Ignatius tetrasporus			AB110439
Bryopsidales			
Bryopsis sp.	FJ480417 (*B. plumosa*)	FJ715718	FJ715685 (*B. plumosa*)
Codium sp.		M67453 (*C. fragile*)	FJ715686 (*C. platylobium*)
Derbesia sp.		AF212142 (*D. marina*)	
Halimeda spp	FJ480416 (*H. discoidea*)	FJ715719 (*H. incrassate*)	AY786526 (*H. gracilis*)
Dasycladales			
Acetabularia spp	FJ480413 (*A. dentate*)	FJ715714 (*A. acetabulum*)	Z33461 (*A. acetabulum*)
Bornetella spp	FJ480414 (*B. nitida*)	FJ715717 (*B. sphearica*)	Z33464 (*B. nitida*)
Siphonocladales			
Boergesenia forbesii			AM498746
Boodlea composita			AF510157
Cladophora albida			Z35317
Cladophora coelothrix			Z35315
Phyllodictyon spp			AF510163 (*P. pulcherrimum*)
Siphonocladus tropicus			AM498761
Valonia utricularis			Z35323
Trentepohliales			
Trentepohlia aurea		FJ534608/**FJ715722**	AB110783
Chlorophyceae			
Chlamydomonas reinhardtii	NC 005353	NC 005353	M32703
Scenedesmus obliquus	NC008101	NC008101	X56103
Volvox carteri	AB013999	D63446	
Trebouxiophyceae			
Chlorella spp	NC 001865 (*C. vulgaris*)	NC 001865 (*C. vulgaris*)	X13688 (*C. vulgaris*)
Prototheca wickerhamii	AJ245645		X74003
Prasinophyceae			
Chlorodendrales			
Tetraselmis spp	DQ173248 (*T. chuii suecica*)	DQ173247 (*T. chuii suecica*)	X70802 (*T. striata*)
Mamiellales			
Ostreococcus tauri	NC 008289	NC 008289	AY329635
Pseudoscourfieldiales			
Nephroselmis olivacea	NC 00927	NC 00927	X74754
Streptophyta			
Mesostigmatales			
Mesostigma viride	NC 002186	NC 002186	AJ250109
Chlorokybales			
Chlorokybus atmophyticus	DQ422812	DQ422812	M95612
Klebsormidiales			
Entransia fimbriata	AY823688	E88 slordig	LB2793
Klebsormidium flaccidum	AF408801	E87 rbcL2	X75520
Zygnematales			
Closterium spp		AJ553936 (*C. lunula*)	AF352237 (*C. gracile*)

	*atp*B	*rbc*L	SSU rDNA
Spirogyra spp	AF408797 (*S. maxima*)	FJ715721 (*S.* sp.)	
Coleochaetales			
Coleochaete scutata	AY082303	AY082313	X68825
Charales			
Chara spp	AF408782 (*C. connivens*)	L13476 (*C. connivens*)	U18493 (*C. connivens*)
Embryophyta (land plants)			
Marchantiophyta (Liverworts)			
Marchantia polymorpha	NC 001319	NC 001319	AB021684
Bryophyta (Mosses)			
Physcomitrella patens	AP005672	AB066207	X80986
Spermatophyta (Seed plants)			
Arabidopsis thaliana	NC 000932	NC 000932	AC006837
Oryza sativa	AC092750	AJ746297	X00755
other eukaryotes			
Archaeplastida			
Glaucocystophyta			
Cyanophora paradoxa	NC 001675	NC 001675	AY823716
Rhodophyta			
Cyanidiophyceae			
Cyanidioschyzon merolae	AB002583		AB158485

Table S2. Algal strain information.

	number	Culture collection
Chlorophyta		
Ulvophyceae		
Ulvales		
Acrochaete repens	E093db	Barbara Rinkel (Natural History Museum, London)
Bolbocoleon piliferum	E344pc	Barbara Rinkel (Natural History Museum, London)
Ulva intestinalis	EE3	Field collection at Goese Sas (Netherlands)
Ulva sp.	EE2 and EE6	Field collection at Goese Sas (Netherlands)
Ulotrichales		
Halochlorococcum tenue	19.92	Sammlung von Algenkulturen (University of Göttingen, Germany)
Ignatius-clade		
Ignatius tetrasporus	2012	UTEX culture collection of algae (University of Texas at Austin, USA)
Bryopsidales		
Blastophysa rhizopus	LB 1029	UTEX culture collection of algae (University of Texas at Austin, USA)
Bryopsis sp.	EE4	Field collection at Goese Sas (Netherlands)
Codium sp	HEC 15711	Field collection in Madeira
Derbesia sp.	2773-1	Tatiana Klotchkova (Kongju National University, Korea)
Ostreobium quekettii	6.99	Sammlung von Algenkulturen (University of Göttingen, Germany)
Trentepohliales		
Trentepohlia aurea	483-1	Sammlung von Algenkulturen (University of Göttingen, Germany)
Dasycladales		
Acetabularia acetabulum	LB 2694	UTEX culture collection of algae (University of Texas at Austin, USA)
Bornetella sphaerica	LB 2690	UTEX culture collection of algae (University of Texas at Austin, USA)
Siphonocladales		
Boergesenia forbesii	Boerg2 and 3	Phycology Research Group, Ghent University*
Boodlea composita	Bcomp2, 4 and 6	Phycology Research Group, Ghent University*
Cladophora albida	Calb3	Phycology Research Group, Ghent University*
Cladophora coelothrix	Ccoel1 and 2	Phycology Research Group, Ghent University*
Phyllodictyon orientale	Struv1, West 1631	John West (University of Melbourne, Australia)
Siphonocladus tropicus	Siph2 and 3	Phycology Research Group, Ghent University*
Valonia utricularis	Vutric2	Phycology Research Group, Ghent University*
Chlorophyceae		
Chlamydomonas reinhardtii	CC1690	Chlamydomonas Center (Duke University, USA)
Scenedesmus obliquus	1450	UTEX culture collection of algae (University of Texas at Austin, USA)
Trebouxiophyceae		
Chlorella kessleri	211-11g	Sammlung von Algenkulturen (University of Göttingen, Germany)
Prasinophyceae		
Tetraselmis striata	41.85	Sammlung von Algenkulturen (University of Göttingen, Germany)
Ostreococcus tauri	OTH95	Hervé Moreau (Observatoire Océanologique de Banyuls)
Nephroselmis olivacea	40.89	Sammlung von Algenkulturen (University of Göttingen, Germany)
Streptophyta		
Chlorokybus atmophyticus	48.80	Sammlung von Algenkulturen (University of Göttingen, Germany)
Entransia fimbriata	LB 2793	UTEX culture collection of algae (University of Texas at Austin, USA)
Klebsormidium flaccidum	KL1	Hans Sluiman (Royal Bot Garden Edinburgh, Scotland); described in G.M. Lokhorst (1996)
Mesostigma viride	50-1	Sammlung von Algenkulturen (University of Göttingen, Germany)
Nephroselmis olivacea	40.89	Sammlung von Algenkulturen (University of Göttingen, Germany)
Spirogyra sp.	169.80	Sammlung von Algenkulturen (University of Göttingen, Germany)

*donated by Jeanine Olsen and Wytze Stam (University of Groningen, Netherlands)

Table S3. Primer sequences used for PCR amplification and sequencing. Primers were found in literature, adapted from existing primers or based on aligned Genbank sequence of green algae and land plants (Viridiplantae), often complemented with a *Cladophora coelothrix* cDNA sequence.

Primer name	Primer sequence (5'-3')	Primers based on
Actin		
Actin_F	GAC ATG GAR AAR ATN TGG CAC CAC AC	244F; An et al. (1999)
Actin_R	ATC CAC ATC TGY TGA AAR GTR G	Ac3-R; Bhattacharya, Stickes, and Sogin (1993)
Ac3-R	GAA GCA YTT GCG RTG SAC RAT	Bhattacharya, Stickes, and Sogin (1993)
Fern5-F	CTT GTY TGY GAC AAT GGA TCW GGA ATG GT	An et al. (1999)
Glucose-6-phosphate isomerase		
G6PI_F	TTY GCR TTY TGG GAC TGG G	*Ostreococcus* and land plants
G6PI_R	CCC CAC TGG TCR AAI GAR TT	*Ostreococcus* and land plants
Glyceraldehyde-3-phosphate dehydrogenase		
GapA_Fb	GSN ATH AAY GGN TTY GG	Petersen et al. (2006)
GapA_Re	CCA YTC RTT RTC RTA CCA	Petersen et al. (2006)
GapA_Fm	GCI YSI TGC ACI ACY AAC TG	green algae and land plants
Oxygen-evolving enhancer protein, chloroplastic (OEE1)		
Oxy_F1	TGA CCT AYI TGC ARR TYA AGG G	Viridiplantae and cDNA library *Cladophora*
Oxy_F2	CIG GCI TKG CCA ACA CIT GCC C	Viridiplantae and cDNA library *Cladophora*
Oxy_R	GAR CCA CCR CGG CCC TTG GGG TC	Viridiplantae and cDNA library *Cladophora*
40S ribosomal protein S9 (RPS9)		
40S_S9_F	ATG CCG AAG ATC GSC YAK TAY CGC	Viridiplantae and cDNA library *Cladophora*
40S_S9_R	CAC KCK AGC GTG GTG GAT GGA CTT	Viridiplantae and cDNA library *Cladophora*
40S_S9_3'UTR	GTC TAT CCT GTC CAG CCT GAG CAG C	Viridiplantae and cDNA library *Cladophora*
60S ribosomal protein L3 (RPL3)		
60S_L3_F	ATG TCI CAC AGI AAG TTT GAA CAC CC	Viridiplantae and cDNA library *Cladophora*
60S_L3_R	GTC TTG CGI GGM AGI CGG GTG AC	Viridiplantae and cDNA library *Cladophora*
60S ribosomal protein L17 (RPL17)		
60S_L17_F	CAA GGC GCG CGG GTC GGA YCT	Viridiplantae and cDNA library *Cladophora*
60S_L17_R	CAT GTA CGG GTT GAT GCG RCC RTG CGC	Viridiplantae and cDNA library *Cladophora*
cDNA library pExCell primers		
M13_F	GTA AAA CGA CGG CCA GT	
M13_R	GGA AAC AGC TAT GAC CAT G	

Table S4. PCR conditions, expected length of PCR products and total length of the alignment used for phylogenetic inferences.

primer combination	Annealing temperature	Length of PCR product	Total length alignment	Comment
Actin			1122 bp	
Actin_F + Actin_R	55°C	≈ 850 bp		
Ac3-R + Fern5-F	55°C	≈ 1000 bp		*Closterium*
Glucose-6-phosphate isomerase			753 bp	
G6PI_F + G6PI_R	55°C	≈ 760 bp		
Glyceraldehyde-3-phosphate dehydrogenase			966 bp	
GapA_Fb + GapA_Re	50°C	≈ 1000 bp		
GapA_Fm +GapA_Re	50°C	≈ 550 bp		
Oxygen-evolving enhancer protein, chloroplastic			762 bp	
Oxy_F1 + Oxy_R	50°C	≈ 470 bp		
Oxy_F2 + Oxy_R	50°C	≈ 500 bp		
40S ribosomal protein S9 (RPS9)			537 bp	
40S_S9_F + 40S_S9_R	55°C	≈ 400 bp		
40S_S9_F + 40S_S9_3'UTR	55°C	≈ 550 bp		*Ulva intestinalis*
60S ribosomal protein L3 (RPL3)			1185 bp	
60S_L3_F + 60S_L3_R	55°C	≈ 730 bp		
60S ribosomal protein L17 (RPL17)			477 bp	
60S_L17_F + 60S_L17_R	55°C	≈ 380 bp		

3

Gain and loss of elongation factor genes in green algae [1]

Abstract

Background

Two key genes of the translational apparatus, elongation factor-1 alpha (EF-1α) and elongation factor-like (EFL) have an almost mutually exclusive distribution in eukaryotes. In the green plant lineage, the Chlorophyta encode EFL except *Acetabularia* where EF-1α is found, and the Streptophyta possess EF-1α except *Mesostigma*, which has EFL. These results raise questions about evolutionary patterns of gain and loss of EF-1α and EFL. A previous study launched the hypothesis that EF-1α was the primitive state and that EFL was gained once in the ancestor of the green plants, followed by differential loss of EF-1α or EFL in the principal clades of the Viridiplantae. In order to gain more insight in the distribution of EF-1α and EFL in green plants and test this hypothesis we screened the presence of the genes in a large sample of green algae and analyzed their gain-loss dynamics in a maximum likelihood framework using continuous-time Markov models.

Results

Within the Chlorophyta, EF-1α is shown to be present in three ulvophycean orders (i.e., Dasycladales, Bryopsidales, Siphonocladales) and the genus *Ignatius*. Models describing gene gain-loss dynamics revealed that the presence of EF-1α, EFL or both genes along the backbone of the green plant phylogeny is highly uncertain due to sensitivity to branch lengths and lack of prior knowledge about ancestral states or rates of gene gain and loss. Model refinements based on insights gained from the EF-1α phylogeny reduce uncertainty but still imply several equally likely possibilities: a primitive EF-1α state with multiple independent EFL gains or coexistence of both genes in the ancestor of the Viridiplantae or Chlorophyta followed by differential loss of one or the other gene in the various lineages.

Conclusions

EF-1α is much more common among green algae than previously thought. The mutually exclusive distribution of EF-1α and EFL is confirmed in a large sample of green plants. Hypotheses about the gain-loss dynamics of elongation factor genes are hard to test analytically due to a relatively flat likelihood surface, even if prior knowledge is incorporated. Phylogenetic analysis of EFL genes indicates misinterpretations in the recent literature due to uncertainty regarding the root position.

Background

Elongation factor-1 alpha (EF-1α) is a core element of the translation apparatus and member of the GTPase protein family. The gene has been widely used as a phylogenetic marker in eukaryotes; either to resolve their early evolution (e.g., Baldauf et al. 1996, Roger et al. 1999) or more recent phylogenetic patterns (e.g., Hashimoto et al. 1994, Baldauf and Doolittle 1997, Beltran et al. 2007, Sung et al. 2007, Zhang and Qiao 2007). The evolutionary history of genes used for such inferences should closely match that of the organisms and not be affected by ancient paralogy or lateral gene transfer (Keeling and Inagaki 2004). A gene related to but clearly distinguishable from EF-1α, called elongation factor-like (EFL), appears to substitute EF-1α in a scattered pattern: several unrelated eukaryote lineages have representatives that encode EFL and others that possess EF-1α. The EFL and EF-1α genes are mutually exclusive in all but two organisms: the zygomycete fungus *Basidiobolus* and the diatom *Thalassiosira* (James et al. 2006, Kamikawa et al. 2008). Although EFL is found in several eukaryotic lineages, EF-1α is thought to be the most abundant of both (Rogers et al. 2007). So far, EFL has been reported in chromalveolates (*Perkinsus*, dinoflagellates, diatoms, haptophytes, cryptophytes), the plant lineage (green and red algae), rhizarians (cercozoans, foraminifera), unikonts (some Fungi and choanozoans) and centrohelids (Keeling and Inagaki 2004, Gile et al. 2006, Noble et al. 2007, Kamikawa et al. 2008, Sakaguchi et al. 2008).

The mutually exclusive distribution of EF-1α and EFL suggests similar functionality. The main function of EF-1α is translation initiation and termination, by delivering aminoacyl tRNAs to the ribosomes (Negrutskii and El'skaya 1998). Other functions include interactions with cytoskeletal proteins: transfer, immobilization and translation of mRNA and involvement in the ubiquitine-dependent proteolytic system, as such forming an intriguing link between protein synthesis and degradation (Negrutskii and El'skaya 1998). In contrast, the function of EFL is barely known. It is assumed to have a translational function because the putative EF-1β, aa-tRNA, and GTP/GDP binding sites do not differ between EF-1α and EFL (Keeling and Inagaki 2004). Based on a reverse transcriptase quantitative PCR assay in the diatom *Thalassiosira*, which possesses both genes, it was proposed that EFL had a translation function while EF-1α performed the auxiliary functions (Kamikawa et al. 2008).

The apparently scattered distribution of EFL across eukaryotes raises questions about the gain-loss patterns of genes with an important role in the cell. This mutually exclusive and seemingly scattered distribution can be explained by two different mechanisms: ancient paralogy and lateral gene transfer. Ancient paralogy was considered unlikely because this would imply that both genes were present in ancestral eukaryotic genomes during extended periods of evolutionary history while the genes rarely coexist in extant species (Keeling and Inagaki 2004). Furthermore, a prolonged coexistence of both genes in early eukaryotes would have likely resulted in either functional divergence or pseudogene formation of one or the other copy (Van de Peer et al. 2001), as is suggested for EFL and EF-1α coexisting in the diatom *Thalassiosira* (Kamikawa et al. 2008). Keeling and Inagaki (Keeling and Inagaki 2004) proposed lateral gene transfer of the EFL gene between eukaryotic lineages as the most likely explanation for the scattered distribution of both genes.

In the green plants (Viridiplantae), EF-1α and EFL seem to show a mutually exclusive distribution. Of the two major green plant lineages, the Chlorophyta were shown to have EFL with the exception of *Acetabularia* where EF-1α is found, and the Streptophyta were shown to possess EF-1α with the

exception of *Mesostigma*, which has EFL (Noble et al. 2007). Noble et al. (2007) proposed the hypothesis that EFL was introduced once in the ancestor of the green lineage, followed by differential loss of EF-1α or EFL in the principal clades of the Viridiplantae (i.e., Streptophyta and Chlorophyta).

The goals of the present study are to extend our knowledge of the distribution pattern of EF-1α and EFL in the green algae and investigate patterns of gain and loss of these key genes of the translational apparatus. We applied a RT-PCR and sequencing-based screening approach across a broad spectrum of green algae, with emphasis on the ulvophycean relatives of *Acetabularia*. To test the hypothesis of Noble et al. (2007), we modeled patterns of gene gain and loss. To this goal, a reference phylogeny based on three commonly used loci was inferred, and gain-loss dynamics of EFL and EF-1α were optimized along this phylogeny using continuous-time Markov models.

## *Results and Discussion*

### Distribution of elongation factors in the green algae

EF-1α sequences were retrieved from streptophytes *Entransia* (Klebsormidiophyceae) and *Chlorokybus* (Chlorokybophyceae), confirming previous observations that all Streptophyta except *Mesostigma* have EF-1α. We found EFL sequences in *Chlorella* (Trebouxiophyceae), *Acrochaete* and *Bolbocoleon* (Ulvophyceae), *Nephroselmis* and *Tetraselmis striata* (prasinophytes), further confirming the formerly established distribution pattern within the Chlorophyta. We reaffirmed the presence of EFL in *Chlamydomonas* and *Scenedesmus* (Chlorophyceae), *Ulva intestinalis* and *U. fenestra* (Ulvophyceae) and *Ostreococcus* (prasinophytes), previously shown by Noble et al. (2007). In addition to *Acetabularia*, EF-1α was discovered in representatives of the ulvophycean orders Dasycladales (*Bornetella*), Bryopsidales (*Blastophysa*, *Bryopsis*, *Codium*, *Derbesia*, *Ostreobium*), Siphonocladales (*Boodlea*, *Cladophora*, *Dictyosphaeria*, *Ernodesmis*, *Phyllodictyon*) and in *Ignatius* (see Figures 1 and 2). The RT-PCR approach did not reveal the presence of both genes in any of the screened species despite the fact that our primers could amplify the target genes across the Viridiplantae. Our RT-PCR experiments on two species whose genomes have been sequenced (*Chlamydomonas* and *Ostreococcus*) yielded a single gene for each species, a result in compliance with the knowledge derived from their genome sequences (DOE Joint Genome Institute).

The reference phylogeny, inferred from a DNA matrix consisting of 72 taxa representing all major plant lineages and three loci (SSU rDNA, *rbc*L and *atp*B), is in accordance with recent phylogenetic studies, including the position of *Mesostigma* within the Streptophyta (Lemieux et al. 2007, Rodriguez-Ezpeleta et al. 2007). Figure 1 shows the phylogenetic relationships among the taxa for which we have information on elongation factors; the full 72-taxon phylogeny can be found as an online supplement [see Additional file 1]. Even though the tree shows improved resolution from previous studies, large parts of the backbone remained poorly resolved. In order to obtain a solid hypothesis of green algal evolution, much additional sequence data may have to be gathered. The occurrence of EF-1α and EFL in terminal taxa was plotted on the reference phylogeny in Figure 1. *Mesostigma* is the only streptophyte which encodes EFL. Within the chlorophytan class Ulvophyceae,

the order Ulvales possesses EFL whereas the other orders encode EF-1α (Dasycladales, Siphonocladales, Bryopsidales and *Ignatius*).
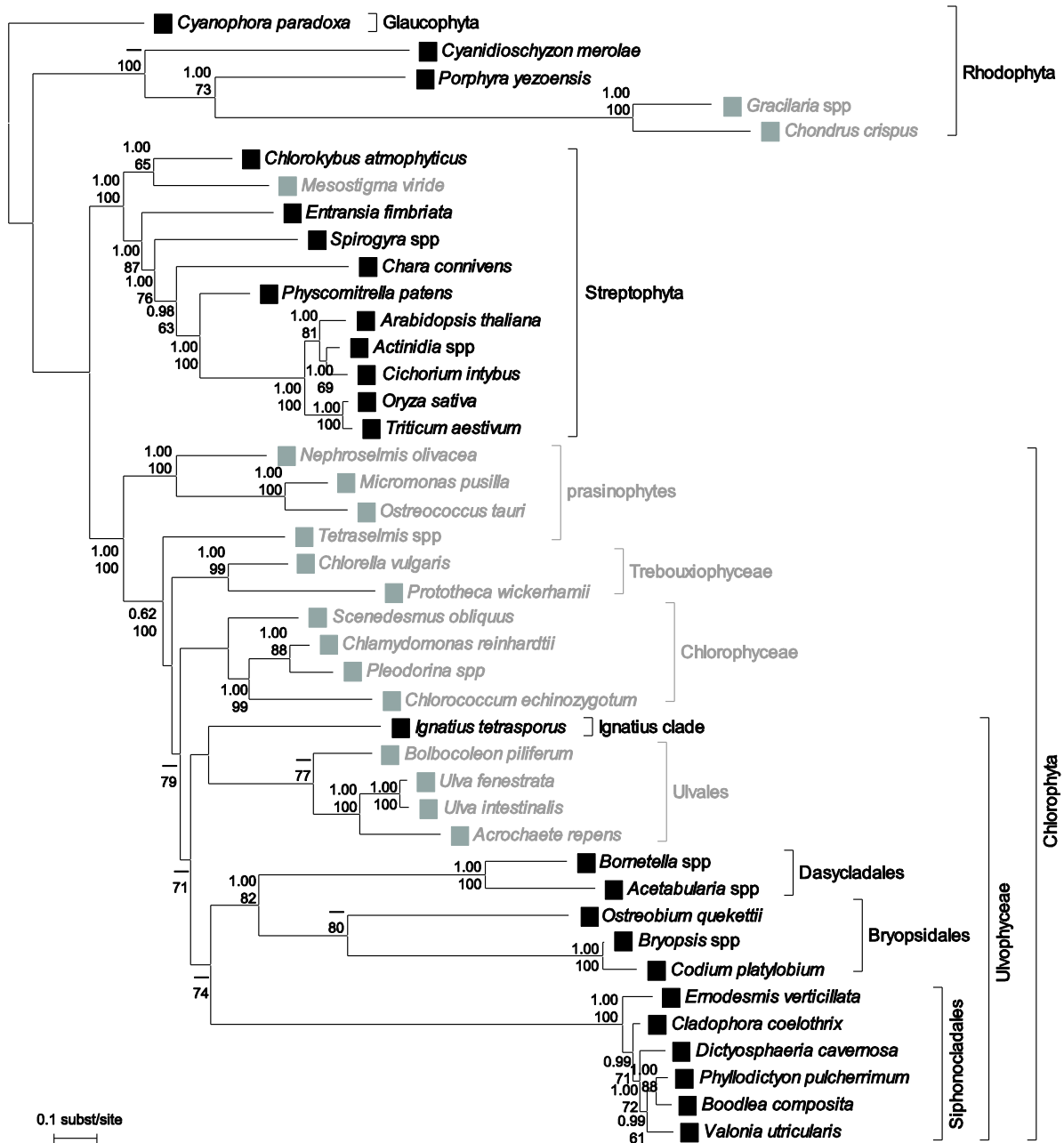
Figure 1. Distribution of EF-1α and EFL in the green plants. The type of elongation factor is indicated with black (EF-1α) or gray (EFL) squares. The reference phylogeny was obtained by Bayesian phylogenetic inference of nuclear SSU rDNA and the plastid genes *rbc*L and *atp*B. Numbers at nodes indicate posterior probabilities (top) and ML bootstrap values (bottom); values below respectively 0.9 and 50 are not shown.
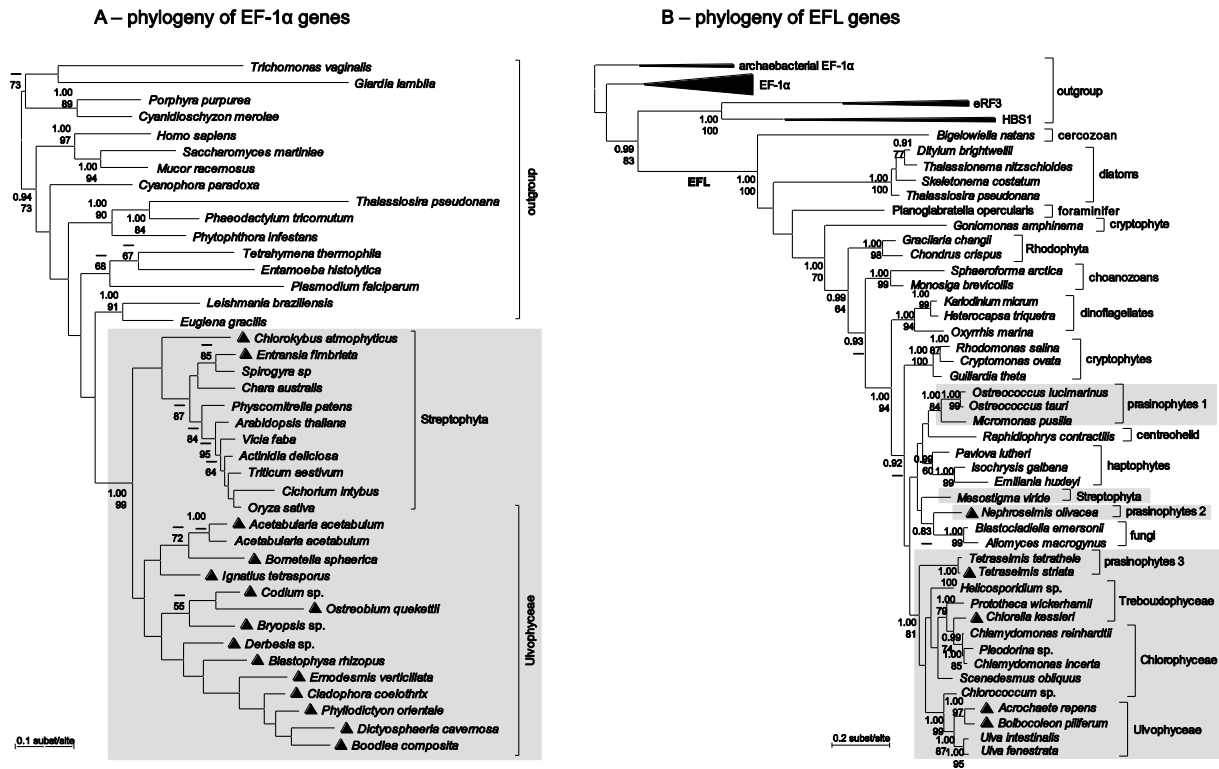
Figure 2. Phylogenies inferred from EF-1α and EFL amino-acid sequences with Bayesian techniques. Sequences belonging to the green plant lineage are in gray boxes. Whereas all green plant EF-1α sequences group in a single clade, the green plant EFL sequences seem to form separate lineages. Sequences generated for this study are indicated with triangles. Numbers at nodes indicate posterior probabilities (top) and ML bootstrap values (bottom); values below respectively 0.9 and 50 are not shown.

Phylogenies of EF-1α and EFL

All green plant EF-1α sequences form a monophyletic group clearly differentiated from EF-1α sequences of a variety of other eukaryotes (Figure 2A). Even though the Viridiplantae form a strongly supported group, resolution among and within Streptophyta and Chlorophyta is generally low, which could in part be due to some short EF-1α sequences included in the analysis.

In contrast, green plant EFL genes do not form a monophyletic lineage (Figure 2B). Although the backbone of the phylogeny is moderately resolved, monophyly of green plant EFL genes is unlikely because it is not observed in the MCMC output (zero posterior probability). EFL sequences of the Viridiplantae can be found in several clades. The chlorophytes, trebouxiophytes, ulvophytes and prasinophyte *Tetraselmis* form a single monophyletic group. The other prasinophyte EFL sequences form two separate groups. The last clade consists of the streptophyte *Mesostigma*.

To obtain an accurate root position for our EFL tree, we included related subfamilies of the GTPase translation factor superfamily: EF-1α, eukaryotic release factor 3 (eRF3), heat shock protein 70 subfamily B suppressor (HBS1) and archaebacterial EF-1α sequences in our analyses. In accordance with Keeling and Inagaki (Keeling and Inagaki 2004), the tree is rooted with archaebacterial EF-1α sequences. All analyses (Bayesian and ML) resulted in a phylogeny very similar to the one shown in

Figure 2B, the complete phylogeny with all related subfamilies can be found as an online supplement [see Additional file 2]. This phylogeny shows seven EFL clades, with the following branching order: *Bigelowiella,* the diatoms, *Planoglabratella*, the cryptophyte *Goniomonas*, red algae, choanozoans, and a large clade containing the green plant lineage, chromalveolates (dinoflagellates, haptophytes, cryptophytes), fungi and *Rhaphidiophrys* (Figures 2B). Deep branches generally received low statistical support, preventing strong conclusions about the relationship between the seven clades.
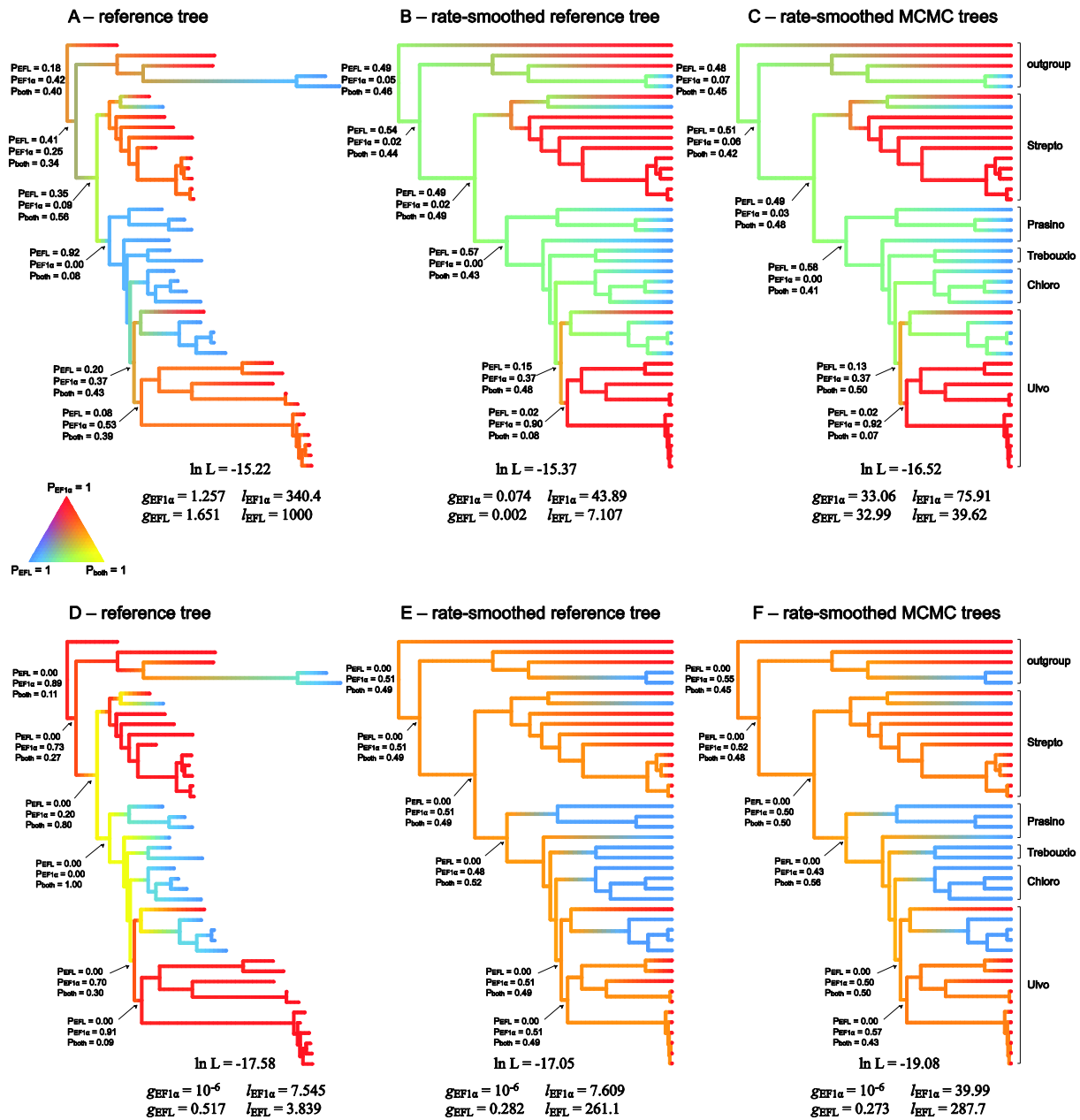
Gain-loss dynamics

The scattered distribution of EF-1α and EFL in the green plant lineage is a remarkable phenomenon that raises questions about evolutionary patterns of gain and loss of both genes. Noble et al. (2007) proposed the hypothesis that EF-1α was present in the common ancestor of the plant lineage, followed by a single gain of EFL early in evolution of the green lineage and subsequent differential loss of one or the other gene in the various lineages. Our aim was to test this hypothesis by modeling gain-loss dynamics and inferring ancestral presence-absence patterns of both genes in a maximum likelihood framework. Gene gain and loss rates were estimated by maximum likelihood (ML) optimization, using a dataset of presence-absence patterns of EF-1α and EFL and a reference phylogeny derived from the Bayesian analysis of three commonly used loci (SSU nrDNA, *rbc*L and *atp*B).

A first analysis, based on the reference tree, shows uncertain character state probabilities along the backbone of the Viridiplantae and suggests a loss of EF-1α in early Chlorophyta evolution and regain in some Ulvophyceae (Figure 3A). Because branch lengths play a crucial role in model optimization, the analysis was repeated on an alternative version of the reference tree in which branch lengths were transformed using a rate smoothing approach. Since our tree deviates from the molecular clock, we performed rate smoothing to obtain branch lengths roughly proportional to time. Rate smoothing techniques relax the assumption of constant rates of evolution throughout the tree: differences in rates of molecular evolution are smoothed out by assuming that evolutionary rates change gradually throughout the phylogeny. The result is an ultrametric tree in which branch lengths are roughly proportional to evolutionary time instead of amounts of molecular evolution. Modeling gain-loss dynamics of elongation factor genes along the rate-smoothed tree yields results that strongly deviate from those obtained with the original reference tree: probabilities of the character states along the major part of backbone are now around 50 % for EFL and around 50 % for the presence of both genes (Figure 3B). Subsequently, an additional level of realism was introduced by taking phylogenetic uncertainty into account because several nodes in the reference tree are poorly supported. To this goal, all post-burnin MCMC trees were rate-smoothed and analyzed individually. The results were summarized on the rate-smoothed reference tree. Taking phylogenetic uncertainty into consideration had a minor influence on the probabilities of the characters states (Figure 3C).

Although the exact numbers differ between analyses, gene gain rates were always lower than gene loss rates, reinforcing the notion that gene transfers are rare events in comparison to losses of redundant genes (Barker et al. 2007). Whereas the analysis based on the original reference tree returned faster gain and loss rates for EFL than for EF-1α, analyses based on rate-smoothed trees

(including MCMC trees) suggested the inverse, marking the sensitivity of Markov models to the unit of operational time.

Figure 3. Gain-loss dynamics of green algal elongation factor genes and their inferred presence in ancestral genomes. Gain and loss rates, as well as the estimated probabilities for presence of the genes in ancestral genomes are given for a variety of analysis conditions. Panels A-C show the outcome of models in which EF-1α and EFL gain and loss rates were not constrained. In panels D-F, the gain rate of EF-1α was constrained to be 10^{-6}. Colors were used to visualize estimated probabilities for presence of genes along the tree. Red indicates a high probability for EF-1α, blue marks a high probability of EFL and yellow stands for a high probability of the presence of both genes. Intermediate colors indicate uncertainty.

From these results, it seems fair to conclude that there is major uncertainty about the ancestral states for a variety of reasons, including sensitivity to branch lengths and lack of prior knowledge about ancestral states or rates of gene gain and loss. Considering that the ancestors must have had either EF-1α, EFL or both genes opens perspectives for hypothesis comparison in a likelihood framework. Additionally, information about rates of gene gain and loss could be gleaned from the EF-1α and EFL phylogenies.

Analyses constrained with various hypotheses about ancestral gene content resulted in a confidence set of 8 trees that differ extensively [see Additional file 3]. The fact that strongly different hypotheses are also present in the confidence set denotes that the likelihood surface is too flat to draw firm conclusions in favor of one or another hypothesis.

The last option to reduce uncertainty is to inform the Markov models with information on gains and losses gleaned from the EF-1α and EFL trees (cf. Barker et al. 2007). Because green plant EF-1α sequences form a monophyletic and strongly supported lineage, it seems fair to assume vertical descent of EF-1α throughout the Viridiplantae. This knowledge can be introduced in our Markov model by setting a very low gain rate of EF-1α. If the analysis is constrained in this way, both EFL and EF-1α were inferred to be present along the backbone of the Viridiplantae in the original reference tree (Figure 3D) and a 50/50 probability for the presence of EF-1α or both genes was obtained in the rate-smoothed trees (Figures 3E-F). Comparison of hypotheses about ancestral gene content constrained with a very low EF-1α gain rate reduced the confidence set to 3 trees in which either EF-1α or both genes are present along the backbone [see Additional file 4]. The ML solution (hypothesis 122) assumes that only EF-1α was present along the backbone of the tree and consequently shows independent gains of EFL in *Mesostigma*, prasinophytes, Chlorophyceae, Trebouxiophyceae and Ulvales. An alternative scenario (hypothesis 123) in the confidence set has EF-1α at the base of the Viridiplantae, a gain of EFL in the ancestor of the Chlorophyta, and subsequent differential loss of one or the other gene in the various lineages. Information from the EFL phylogeny may provide clues for further distinction between either multiple transfers or ancient paralogy with subsequent losses.

The green EFL sequences form a highly supported clade together with dinoflagellates, cryptophytes, haptophytes, fungi and *Rhaphidiophrys*, suggesting lateral gene transfer of the EFL gene between these distant eukaryotic lineages (Andersson 2005, Keeling and Palmer 2008). Considering the ability of chromalveolates (i.e., dinoflagellates, cryptophytes and haptophytes) and *Raphidiophrys* to feed through phagocytosis (Sakaguchi et al. 2002) and the absence of this behavior in green algae, it would be tempting to assume that lateral gene transfer occurred from green algae to the dinoflagellates, cryptophytes, haptophytes and *Raphidiophrys* instead of the other way around. Phagotrophic eukaryotes have been shown to have elevated rates of lateral gene transfer (Gogarten 2003, Andersson 2005) because this feeding mechanism enables them to continually recruit genes from engulfed prey (Nosenko et al. 2006). Lateral gene transfers to fungi, although known to exist (Andersson et al. 2003), would require a different explanation because neither phagotrophy nor endosymbiosis occur in fungi. However, in the light of this peripheral information, it would be tempting to conclude that both EF-1α and EFL essentially show vertical descent in green plants and that the observed mutually exclusive pattern of EFL and EF-1α sequences results from differential loss. In this scenario, lateral gene transfer must have occurred from green algal cells to other eukaryotic lineages.

In previous studies of functionally similar eukaryotic genes with mutually exclusive distributions, distinction between ancient paralogy with subsequent losses and multiple transfers was made based on two main criteria (Rogers et al. 2007). The first criterion states that if one gene dominates the tree and the other occurs in only a few lineages, multiple independent transfers should be regarded as the most probable explanation whereas equal representation would suggest common ancestry with subsequent differential loss. The second criterion is about the age of the taxa involved. If the mutually exclusive pattern occurs between closely related species, one can conclude common ancestry with subsequent losses. If the pattern is more ancient, multiple lateral transfers are a more probable explanation. It is obvious that such criteria are very difficult to apply in real situations. These difficulties can be overcome by taking a probabilistic angle on the problem and modeling gain-loss dynamics with continuous-time Markov models. This approach brings statistical rigor to the analysis of gene presence-absence patterns and has the potential to discriminate between the alternative scenarios of ancient paralogy with differential losses and multiple independent lateral transfers. Application of this technique to our dataset of green algal elongation factors revealed the difficulty of arriving at firm conclusions about ancient gene transfer events because of a relatively flat likelihood surface and, consequently, ambiguous probabilities for gene content at ancestral nodes. When informed with external information, the analyses allow somewhat more definitive conclusions.
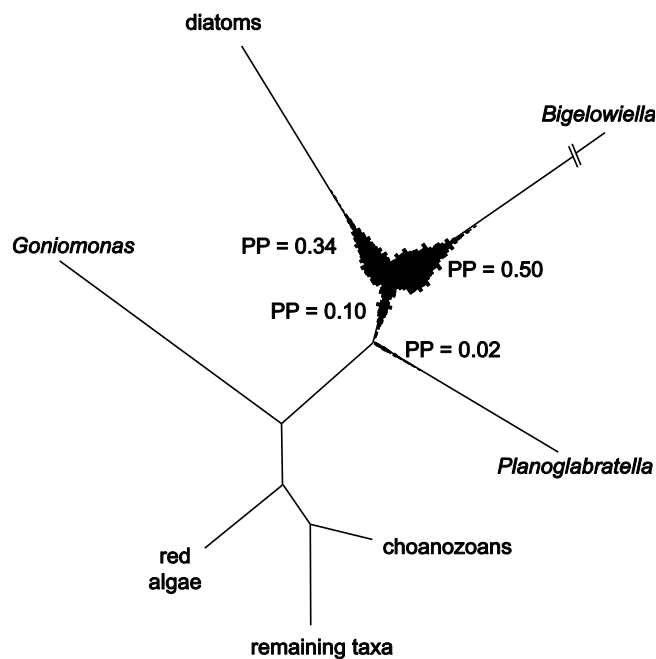
Figure 4. Visualization of the posterior probability of rooting of the EFL tree. The topology represents the unrooted topology of EFL genes. Branch width is proportional to the posterior probability that the outgroup, consisting of archaebacterial EF-1α, EF-1α, eRF3 and HBS1 sequences, attaches to the ingroup tree at that point. Numbers at branches represent the total posterior probability that the root is situated along the branch in question.

The broader eukaryotic picture

In addition to the information gained about elongation factor evolution in green algae, our results also highlight misinterpretations in recent literature on EFL evolution across the eukaryotes. Previous studies have not been explicit about whether or how their phylogenetic trees were rooted, but have drawn conclusions that require directionality in the tree. Kamikawa et al. (2008) concluded that lateral gene transfer from a foraminifer (*Planoglabratella*) to the ancestor of the diatoms must have

occurred because the diatom sequences were nested within the Rhizaria (foraminifera and cercozoans). In case their tree was unrooted, this conclusion is flawed due to a lack of directionality in the tree. In their presentation of the tree, choanozoans are used as one of the basal clades, probably because they were the earliest-branching lineage in the tree presented by Keeling and Inagaki (2004). Our EFL tree, which includes EF-1α, eRF3 and HBS1 sequences and is rooted with archaebacterial EF-1α sequences, indicates that the directionality inferred by Kamikawa et al. (2008) is likely to be wrong. Our phylogram (Figures 2B) suggest that the root position of EFL lies on the branch leading towards the cercozoan *Bigelowiella*, but support is lacking for the basal relationships. A plot of the posterior distribution of root placements (Figure 4) illustrates the uncertainty about the root placement more clearly. It is evident from this plot that the choanozoans are not the oldest diverging lineage. This finding overturns the conclusion from Kamikawa et al. (2008) because the nested position of the diatom EFL genes within the Rhizaria sequences can no longer be maintained. Our EFL phylogeny supports the presence of lateral gene transfer between eukaryotic lineages, however, the direction of lateral gene transfer is difficult to evaluate.

Conclusions

The mutually exclusive nature of EF-1α and EFL is confirmed in a large sample of green algae. The Streptophyta possess EF-1α with the exception of *Mesostigma*, which has EFL. The Chlorophyta encode EFL with the exception of Dasycladales, Bryopsidales, Siphonocladales and *Ignatius*, where EF-1α is found. This result establishes EF-1α as a widespread gene among green algae.

Gain-loss models revealed that the probabilities of the presence of EF-1α, EFL or both genes along the backbone of the plant phylogeny are highly uncertain, and that a previously published hypothesis (Noble et al. 2007) is as likely as several other hypotheses. Model refinements based on insights gained from the EF-1α phylogeny were unable to distinguish between three possibilities: (1) multiple, independent gains of EFL throughout the plant lineage, (2) a single gain of EFL early in evolution of the plant lineage followed by differential loss, or (3) independent gains of EFL in *Mesostigma* and the ancestor of the Chlorophyta followed by differential loss of one or the other gene in the various lineages (Figure 3 D-F and Additional file 4).

Further research into the gain-loss dynamics of elongation factors of green plants and eukaryotes in general is needed to come to more definitive conclusions about their evolution. First, the EFL phylogeny should be refined by obtaining full-length sequences for a set of relevant taxa to confirm or reject the presence of multiple independent green lineages in this tree. The use of codon models may help to achieve this (Seo and Kishino 2008). An alternative approach would be to learn about the processes responsible for lateral transfer of elongation factors by studying their flanking regions for signature sequences of mobile elements (Zhang et al. 2006, Dybvig et al. 2007). Finally, studying gain-loss dynamics across a wider spectrum of eukaryotic supergroups should lead to more stable conclusions. In addition to yielding more precise parameter estimates for gene gain and loss rates, a eukaryote-wide study would allow the use of more specific models for lateral gene transfer because both donor and recipient lineages would be present in the analysis (Than et al. 2006, Galtier 2007, Lake 2008). It remains an enigma that the evolution of elongation factors, genes crucial for cell functioning, is marked by such complex gain-loss patterns.

Methods

Algal strains

Algal strain information is provided as additional material online [see Additional file 5]. All cultures were grown at 18°C, except Dasycladales, Siphonocladales and *Derbesia* (23°C). Cool white fluorescent lamps were used for a 12/12 h light/dark cycle. Marine cultures were maintained in f/2 medium and freshwater cultures in Bold's Basal Medium (Andersen 2005).

RNA isolation and cDNA library construction of Cladophora coelothrix

Total RNA was extracted with a RNeasy Plant Mini Kit (Qiagen Benelux b.v., Venlo, the Netherlands) or a NucleoSpin® RNA Plant kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany) according to the manufacturer's instructions, including a DNase step to eliminate genomic DNA contamination. RNA quality was checked on a 1 % agarose gel (made with 1x TAE diluted in 0,1 % DEPC water). RNA concentration and purity were measured in a spectrophotometer at 260 and 280 nm according to standard methods (Sambrook et al. 1989).

Approximately 30 μg of total RNA of *Cladophora coelothrix* was extracted as described above. A standard cDNA library was constructed by VERTIS Biotechnologie AG (Freising , Germany). An EF-1α sequence of 624 bp was obtained by sequencing randomly picked clones.

Reverse Transcriptase and Polymerase Chain Reaction

cDNA construction was performed with the Omniscript RT kit (Qiagen) and oligodT primers according to the manufacturer's instructions; the reaction was incubated for several hours at 37°C.

Primers were designed to fit the most conserved regions of EF-1α and EFL sequences across Viridiplantae. Primers for EF-1α were based upon aligned GenBank sequences from green algae (*Acetabularia* and *Chara*) and land plants, completed with our *Cladophora coelothrix* cDNA sequence (EF-1α-F: 5'-GGC CAT CTT ATC TAC AAG CTT GGC GG-3' and EF-1α-R: 5'-CCA GGA GCA TCA ATC ACG GTG CAG-3'). EFL primers were adapted from Noble et al. (Noble et al. 2007) (EFL-F: 5'-TCC ATY GTS ATY TGC GGN CAY GTC GA-3' and EFL-R: 5'-CTT GAT GTT CAT RCC RAC RTT GTC RCC-3'). PCR amplification was performed with the following reaction mixture: 1 μl of cDNA, 2.5 μl of 10x Buffer (Qiagen), 0.5 μl dNTP's (10 mM), 0.5 μl MgCl (25 mM, Qiagen), 0.5 μl of each primer (10 μm), 0.25 μl BSA (10 μg/μl), 18.125 μl sterilized MilliQ water and 0.125 μl Taq polymerase (5 U/μl, Qiagen). The amplification profile consisted of an initial denaturation of 2 min at 94 °C, followed by 35 cycles of 30 s at 94 °C, 30 s at 55 °C and 45 s at 72 °C and a final extension of 10 min at 72 °C. Products of expected size (300 bp for EF-1α and 900 bp for EFL) were either sequenced directly or cloned and sequenced.

Cloning and sequencing

PCR products were first sequenced with the forward primer with an Applied Biosystems 3130xl. Sequences were blasted against the GenBank protein database (blastx), to check for potential bacterial contaminants. Sequences without ambiguous base calls yielding a significant hit for Viridiplantae were further sequenced with the reverse primer. When ambiguous base calls were present in sequences, samples were cloned if the rough sequence gave a significant blastx hit for Viridiplantae. Cloning was performed with the pGEM®-T Vector System (Promega Benelux b.v., Leiden, the Netherlands) according to the manufacturer's instructions. After ligation, transformation and incubation, the white colonies were transferred to 15 µl double distilled water, boiled for 10 minutes to lyse cells and subsequently centrifuged to pellet the cells walls and allow harvest of the DNA in the liquid phase. Between three and five clones were PCR amplified and sequenced with the vector specific primers T7 and SP6 following the protocol described above. Cloning showed minor polymorphisms that most likely represent different alleles.

Alignments and phylogenetic analysis of EF-1α and EFL

Sequences [see Additional file 6] were assembled with AutoAssembler 1.4.0 (ABI Prism, Perkin Elmer, Foster City, CA, USA) and aligned manually for both genes separately, resulting in EF-1α and EFL alignments of 1374 and 1653 bp, respectively [see Additional file 7]. Sequences generated with our primers begin in the N-terminal part the of the gene and are 900 bp for EFL and 150-300 bp for EF-1α. We included eukaryotic EF-1α, eRF3 and HBS1 sequences as well as archeabacterial EF-1α sequences to serve as outgroups for the EFL phylogeny (Keeling and Inagaki 2004). Due to the large divergences between EFL and the other genes, Gblocks was run to remove ambiguously aligned regions (Castresana 2000). We ran Gblocks v.0.91b, allowing smaller final blocks, gap positions within the final blocks and less strict flanking positions, resulting in an alignment of 358 amino acids [see Additional file 7]. The resulting EFL and EF-1α alignments were subjected to Bayesian phylogenetic inference with MrBayes 3.1.2 (Ronquist and Huelsenbeck 2003) using the model suggested by ProtTest 1.4 (Abascal et al. 2005) (WAG with among site rate heterogeneity: gamma distribution with 8 categories). Two parallel runs, each consisting of four incrementally heated chains were run for 1,000,000 generations, sampling every 1,000 generations. Convergence of log-likelihoods was assessed in Tracer v1.4 (Rambaut and Drummond 2007). A burnin sample of 100 trees was removed before constructing the majority rule consensus tree for each of the genes. Maximum likelihood phylogenies were inferred for EF-1α and EFL with Treefinder (Jobb 2008). The analyses were based on amino acid sequences and used a WAG model with among site rate heterogeneity (gamma distribution with 8 categories). One thousand non-parametric bootstrap trees were inferred. Bootstrap values were summarized with consense from the Phylip package (Felsenstein 2005) and plotted onto the Bayesian consensus tree.

Phylogeny of the green plants: SSU rDNA, *rbc*L and *atp*B

A reference phylogeny of green plants for which the presence of EF-1α or EFL is known was constructed using three commonly used phylogenetic markers: nuclear SSU rDNA and plastid *atp*B

and *rbc*L genes and rooted with red algae and a glaucophyte. To obtain an even species distribution and consequently a better phylogenetic tree (Zwickl and Hillis 2002), many additional species were included in the phylogenetic analysis [see Additional file 7]. Sequences were retrieved from GenBank and aligned with our own sequences [see Additional file 6]. DNA was extracted using a standard CTAB method. PCR conditions followed standard protocol. Primers were based on other publications: SSU rDNA (Zechman et al. 1990, Lewis and Lewis 2005), *rbc*L (Hanyuda et al. 2000) and *atp*B (Wolf 1997, Karol et al. 2001). The *rbc*L and *atp*B sequences were aligned by eye. The SSU rDNA sequences were aligned based on their RNA secondary structure with DCSE [see Additional file 8].

The model selection procedure [see Additional file 8] proposed eight partitions: *atp*B and *rbc*L genes were partitioned into codon positions (6 partitions) and the SSU rDNA was partitioned into RNA loops and stems (2 partitions). Bayesian phylogenetic inference was carried out using a GTR model with gamma distribution and 8 rate categories per partition (all parameters unlinked) and rate multipliers to accommodate rate differences among partitions. Two parallel runs, each consisting of four incrementally heated chains were run for 5,000,000 generations, sampling every 1,000 generations. Convergence of log-likelihoods was assessed in Tracer v1.4 (Rambaut and Drummond 2007). A burnin sample of 3,000 trees was removed before constructing the majority rule consensus tree. A maximum likelihood phylogeny was inferred with Treefinder (Jobb 2008). The analysis used a GTR model with among site rate heterogeneity (gamma distribution with 8 categories). One thousand non-parametric bootstrap trees were inferred. Bootstrap values were summarized with consense from the Phylip package (Felsenstein 2005) and plotted onto the Bayesian consensus tree.

To obtain trees suitable for modeling gene gain and loss, the Bayesian consensus tree and the complete post-burnin set of trees were pruned to the set of species for which the type of elongation factor is known using the APE package (Paradis et al. 2004). Because our data deviate from the molecular clock, we performed rate smoothing to obtain branch lengths that are roughly proportional to time. We used the penalized likelihood method (Sanderson 2002) implemented in the r8s program (Sanderson 2003), with a log-additive penalty and a smoothing value of 2, which was the optimal value in cross-validation (Sanderson 2002). PL rate smoothing was applied to the Bayesian consensus tree as well as the post-burnin set of MCMC trees.

Modeling gene gain and loss

If the presence of EF-1α and EFL are coded as two binary characters, their gain-loss dynamics can be modeled along a reference phylogeny using a continuous-time Markov model. Given the likely dependency of gain and loss between EF-1α and EFL, a model designed to study interdependent trait evolution was used (Pagel 1994). The rate matrix of this model is given by:

$$
Q_D = \begin{array}{c} \\ 0,0 \\ 0,1 \\ 1,0 \\ 1,1 \end{array}
\begin{array}{c} 0,0 \quad 0,1 \quad 1,0 \quad 1,1 \end{array}
\begin{pmatrix}
\cdot & q_{12} & q_{13} & 0 \\
q_{21} & \cdot & 0 & q_{24} \\
q_{31} & 0 & \cdot & q_{34} \\
0 & q_{42} & q_{43} & \cdot
\end{pmatrix}
\tag{1}
$$

where (0,0) indicates the absence of both genes from the genome, (0,1) and (1,0) denote the presence of EFL and EF-1α, respectively, and (1,1) is the state where both genes are present in the genome. Different q's indicate relative rates of the respective changes in gene content. Transitions that require more than one event (e.g. 1,0 → 0,1) are not allowed to occur as a single step in this model, the logic being that the probability of two traits changing at exactly the same time is negligible. This is consistent with the fact that transitions from EF-1α to EFL and vice versa should pass through a stage where both genes are present in the genome. The elements of the diagonal are determined by the requirement that each row sums to zero. Because the absence of both genes is likely to be lethal, the matrix was constrained by introducing a series of very low rates as follows:

$$Q_D = \begin{pmatrix} \cdot & 10^{-6} & 10^{-6} & 0 \\ 10^{-6} & \cdot & 0 & g_{EF1\alpha} \\ 10^{-6} & 0 & \cdot & g_{EFL} \\ 0 & l_{EF1\alpha} & l_{EFL} & \cdot \end{pmatrix} \quad (2)$$

In this matrix, $g_{EF1\alpha}$ and g_{EFL} denote gain rates and $l_{EF1\alpha}$ and l_{EFL} loss rates. It must be noted that the model does not take gene duplications into account because our data provided no indications for the presence of such events.

The rate matrix (2) was specified as a special case of the "discrete dependent" model in BayesTraits (Pagel and Meade 2006). The model parameters were estimated by maximum likelihood (ML) optimization, using a dataset of presence-absence patterns of EF-1α and EFL. One hundred optimization attempts were carried out to find the ML solution. Ancestral state probabilities were calculated using the addNode command. The reference phylogeny used for inferring patterns of gain and loss was derived from the Bayesian analysis of SSU nrDNA, *rbc*L and *atp*B, and was varied as follows. First, the majority rule consensus tree provided by MrBayes was used. Second, a rate-smoothed version of this consensus tree was used to have branch lengths roughly proportional to evolutionary time. Third, topological uncertainty was introduced in the analysis by repeating analyses on the entire post-burnin set of MCMC trees after they had been rate-smoothed. For the analysis on MCMC trees, ancestral state probabilities were calculated with the addMRCA instead of the addNode command. Rate estimates and ancestral state probabilities were averaged across the MCMC trees. We opted not to use BayesTraits' Bayesian inference because we found its output to be strongly influenced by prior settings.

In addition to these analyses, several specific hypotheses about ancestral genome content (EFL, EF-1α or both) were compared using ML optimization on the rate-smoothed reference tree. Constraints on ancestral genome content were placed on 5 ancestral nodes with the fossil command in BayesTraits, resulting in 3^5 = 243 hypotheses for which the log-likelihoods could be compared. Only hypotheses within two log-likelihood units from the ML solution were retained for interpretation. This set of hypotheses can be seen as a confidence set because two log-likelihood units is considered a significance threshold for such analyses (Pagel 1999).

The BayesTraits output was mapped onto the trees with TreeGradients v1.02 (Verbruggen 2008). This program plots ancestral state probabilities on a phylogenetic tree as colors along a color gradient.

Authors' contributions

EC, ODC, HV and KS designed the study. EC carried out lab work. EC and FL maintained algal cultures and performed sequence alignment. EC and HV analyzed data and drafted the manuscript. FWZ provided *atp*B sequences. All authors revised and approved the final manuscript.

## *Acknowledgements*

Additional files

Additional file 1

Figure S1. Reference phylogeny of the green plant lineage obtained by Bayesian inference of nuclear SSU rDNA and the plastid genes rbcL and atpB. Numbers at nodes indicate posterior probabilities (top) and ML bootstrap values (bottom); values below respectively 0.9 and 50 are not shown.

Additional file 2

Figure S2. Phylogeny inferred from amino-acid sequences of EFL and related subfamilies of the GTPase translation factor superfamily with Bayesian techniques. Sequences belonging to the green plant lineage are in gray boxes. The green plant EFL sequences seem to form separate lineages. Numbers at nodes indicate posterior probabilities (top) and ML bootstrap values (bottom); values below respectively 0.9 and 50 are not shown.
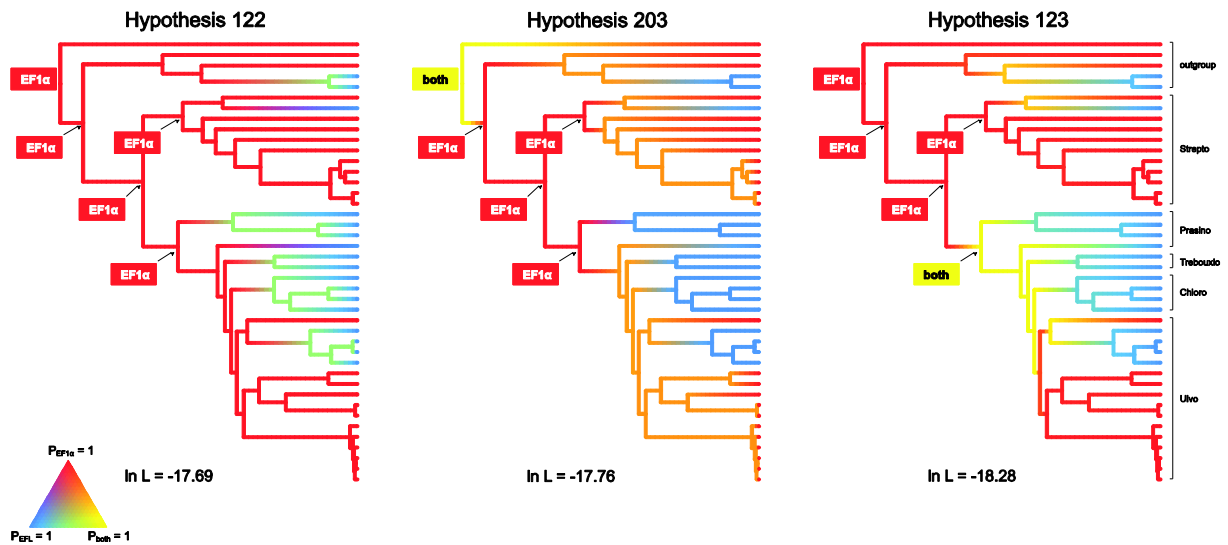
Additional file 3
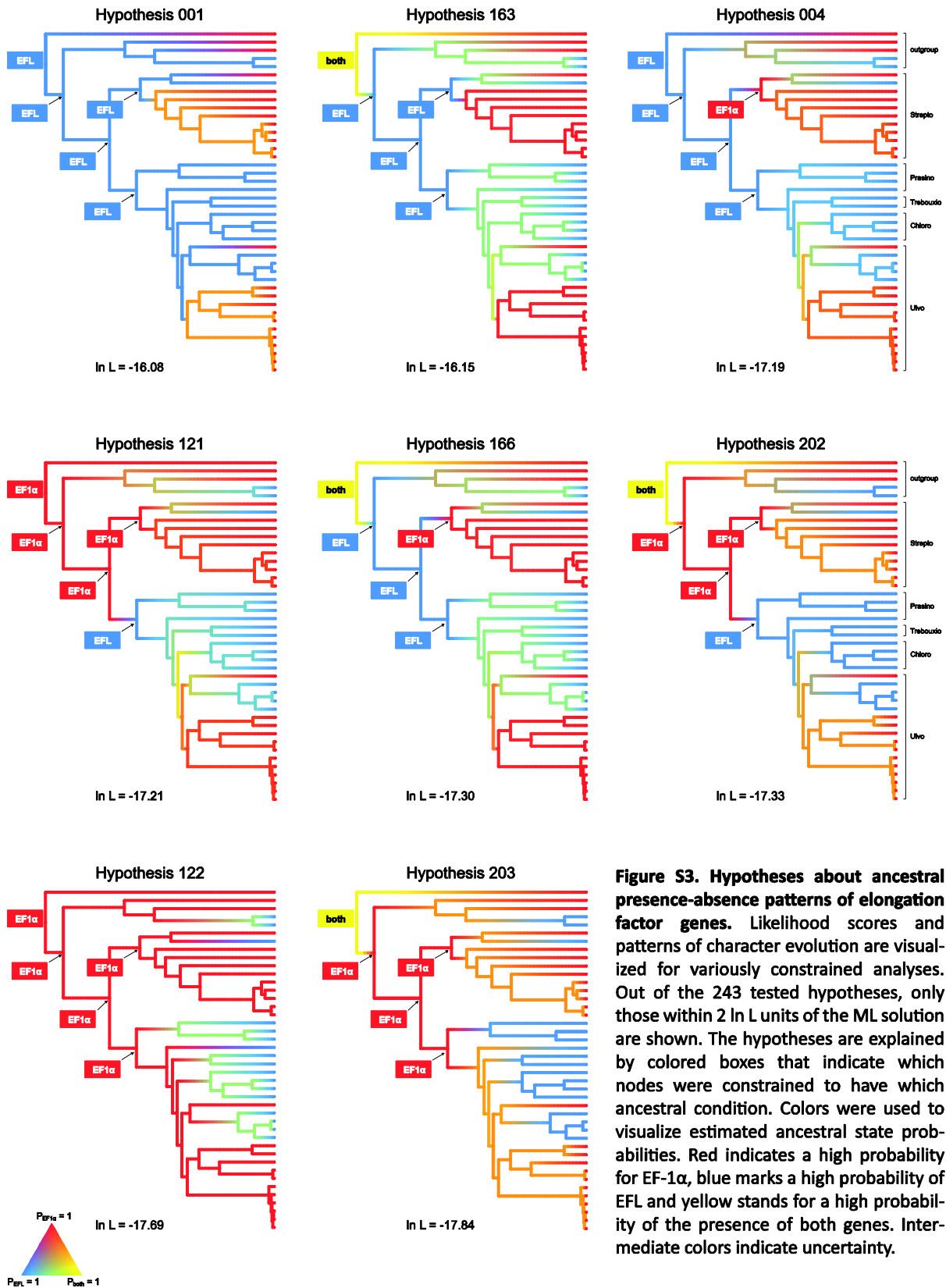
Figure S3. Hypotheses about ancestral presence-absence patterns of elongation factor genes. Likelihood scores and patterns of character evolution are visualized for variously constrained analyses. Out of the 243 tested hypotheses, only those within 2 ln L units of the ML solution are shown. The hypotheses are explained by colored boxes that indicate which nodes were constrained to have which ancestral condition. Colors were used to visualize estimated ancestral state probabilities. Red indicates a high probability for EF-1α, blue marks a high probability of EFL and yellow stands for a high probability of the presence of both genes. Intermediate colors indicate uncertainty.

Additional file 4

Figure S4. Hypotheses about ancestral presence-absence patterns of elongation factor genes, assuming that EF-1 α is not acquired by lateral gene transfer. Likelihood scores and patterns of character evolution are visualized for variously constrained analyses. Out of the 243 tested hypotheses, only those within 2 ln L units of the ML solution are shown. The hypotheses are explained by colored boxes that indicate which nodes were constrained to have which ancestral condition. Colors were used to visualize estimated ancestral state probabilities. Red indicates a high probability for EF-1α, blue marks a high probability of EFL and yellow stands for a high probability of the presence of both genes. Intermediate colors indicate uncertainty.

Additional file 5

Table S1. Algal strain information.

	number	Culture collection
Chlorophyta		
Ulvophyceae		
Ulvales		
Acrochaete repens	E093db	Barbara Rinkel (Natural History Museum, London)
Bolbocoleon piliferum	E344pc	Barbara Rinkel (Natural History Museum, London)
Ulva fenestrata	EE2 and EE6	Field collection at Goese Sas (Netherlands)
Ulva intestinalis	EE3	Field collection at Goese Sas (Netherlands)
Ignatius-clade		
Ignatius tetrasporus	B 2012	UTEX culture collection of algae (University of Texas at Austin, USA)
Trentepohliales		
Trentepohlia aurea	483-1	Sammlung von Algenkulturen (University of Göttingen, Germany)
Bryopsidales		
Blastophysa rhizopus	LB 1029	UTEX culture collection of algae (University of Texas at Austin, USA)
Bryopsis sp.	EE4	Field collection at Goese Sas (Netherlands)
Codium sp	HEC 15711	Field collection in Madeira
Derbesia sp.	2773-1	Tatiana Klotchkova (Kongju National University, Korea)
Ostreobium quekettii	6.99	Sammlung von Algenkulturen (University of Göttingen, Germany)
Dasycladales		
Acetabularia acetabulum	LB 2694	UTEX culture collection of algae (University of Texas at Austin, USA)
Bornetella sphaerica	LB 2690	UTEX culture collection of algae (University of Texas at Austin, USA)
Siphonocladales		
Boodlea composita	Bcomp4, BoTTd75	Jeanine Olsen and Wytze Stam (University of Groningen, Netherlands)*
Cladophora coelothrix	Ccoel2, C83.14	Jeanine Olsen and Wytze Stam (University of Groningen, Netherlands)*
Dictyosphaeria cavernosa	Dcav3, D.cavSJ25b	Jeanine Olsen and Wytze Stam (University of Groningen, Netherlands)*
Ernodesmis verticillata	Erno4, EvVGa88	Jeanine Olsen and Wytze Stam (University of Groningen, Netherlands)*
Phyllodictyon orientale	Struv1, West 1631	John West (University of Melbourne, Australia)
Valonia utricularis	Vutric2, VU1546	Jeanine Olsen and Wytze Stam (University of Groningen, Netherlands)*
Chlorophyceae		
Chlamydomonas reinhardtii	CC1690	Chlamydomonas Center (Duke University, USA)
Scenedesmus obliquus	1450	UTEX culture collection of algae (University of Texas at Austin, USA)
Trebouxiophyceae		
Chlorella kessleri	211-11g	Sammlung von Algenkulturen (University of Göttingen, Germany)
Prasinophyceae		
Tetraselmis striata	41.85	Sammlung von Algenkulturen (University of Göttingen, Germany)
Ostreococcus tauri	OTH95	Hervé Moreau (Observatoire Océanologique de Banyuls)
Nephroselmis olivacea	40.89	Sammlung von Algenkulturen (University of Göttingen, Germany)
Streptophyta		
Chlorokybus atmophyticus	48.80	Sammlung von Algenkulturen (University of Göttingen, Germany)
Entransia fimbriata	LB 2793	UTEX culture collection of algae (University of Texas at Austin, USA)
Spirogyra sp.	169.80	Sammlung von Algenkulturen (University of Göttingen, Germany)

*now maintained in the Phycology Research Group, Ghent University

Additional file 6

Table S2. GenBank accession numbers or DOE Joint Genome Institute (JGI) transcript identifiers for nucleotide sequences of *atpB*, *rbc*L, SSU rDNA, EF-1α and EFL. Newly generated sequences are in boldface. GenBank accession numbers for protein sequences of eRF3 (*Candida maltosa* BAB12681, *Homo sapiens* NP_002085, *Nicotiana tabacum* AAA79032, *Saccharomyces cerevisiae* EDN60510), HBS1 (*Arabidopsis thaliana* NP_196625, *Homo sapiens* NP_006611, *Saccharomyces cerevisiae* NP_01301) and archaebacterial EF-1α (*Sulfolobus sulfataricus* CAA50033, *Termoplasma acidophilum* P19486) are given here between brackets.

	atpB	rbcL	SSU rDNA	EF-1α	EFL
Chlorophyta					
Ulvophyceae					
Ulotrichales					
Halochlorococcum moorei			AY198122		
Ulothrix zonata			Z47999		
Ulvales					
Acrochaete repens		**FJ715715**	**FJ715684**		**FJ715704**
Bolbocoleon piliferum		**FJ715716**	AY205330		**FJ715705**
Pseudendoclonium akinetum	NC 008114	NC 008114	DQ011230		
Ulva fenestrata		AF499668			EF551331/ **FJ715706**
Ulva intestinalis		AY422552	AJ000040		EF551324/ **FJ715707**
Ignatius-clade					
Ignatius tetrasporus			AB110439	**FJ715687**	
Oltmannsiellopsidales					
Oltmanssiellopsis viridis	NC 008099	NC 008099	D86495		
Bryopsidales					
Blastophysa rhizopus				**FJ715688**	
Bryopsis spp	FJ480417 (*B. plumosa*)	**FJ715718**	**FJ715685** (*B. plumosa*)	**FJ715689**	
Caulerpa spp	FJ480415 (*C. prolifera*)		AF479703 (*C. sertularioides*)		
Codium spp			**FJ715686** (*C. platylobium*)	**FJ715690**	
Derbesia sp.				**FJ715691**	
Halimeda spp	FJ480416 (*H. discoidea*)	**FJ715719** (*H. incrassate*)	AY786526 (*H. gracilis*)		
Ostreobium quekettii		**FJ715720**		**FJ715692**	
Dasycladales					
Acetabularia spp	FJ480413 (*A. dentata*)	**FJ715714** (*A. acetabulum*)	Z33461 (*A. acetabulum*)	EF551321/ **FJ715693** (*A. acetabulum*)	
Bornetella spp	FJ480414 (*B. nitida*)	**FJ715717** (*B. sphearica*)	Z33464 (*B. nitida*)	**FJ715694** (*B. sphearica*)	
Siphonocladales					
Boergesenia forbesii			AM498746		
Boodlea composita			AF510157	**FJ715695**	
Cladophora albida			Z35317		
Cladophora coelothrix			Z35315	**FJ715696**	
Dictyosphaeria cavernosa			AM498756	**FJ715697**	

	atpB	rbcL	SSU rDNA	EF-1α	EFL
Ernodesmis verticillata			AM498757	**FJ715698**	
Phyllodictyon spp			AF510163 (*P. pulcherrimum*)	**FJ715699**	
Siphonocladus tropicus			AM498761		
Struvea plumosa			AF510161		
Valonia utricularis			Z35323	**FJ715700**	
Trentepohliales					
Trentepohlia aurea		**FJ715722**	AB110783		
Chlorophyceae					
Chlamydomonas incerta					DQ122889
Chlamydomonas reinhardtii	NC 005353	NC 005353	M32703		DS496185/ **FJ715708**
Chlorococcum echinozygotum	EF113500	EF113430	U57698		EF551323
Cylindrocapsa geminella	EF119849	EF113434	AF387159		
Oedogonium cardiacum	EF113523	EF113458	U83133		
Paulschulzia pseudovolvox	AB014040	D86837	U83120		
Pleodorina spp	AB214424 (*P. starrii*)	D86834 (*P.indica*)			AB095178 (*P.* sp.)
Scenedesmus obliquus	NC008101	NC008101	X56103		SOL0000077/ **FJ715709**
Sphaeroplea robusta	EF113536	EF113472	U73472		
Stigeoclonium helveticum	NC 008372	NC 008372	U83131		
Tetraspora sp.	EF113540	EF113477	U83121		
Uronema belkae	EF113544	EF113481	AF182821		
Trebouxiophyceae					
Chlorella spp	NC 001865 (*C. vulgaris*)	NC 001865 (*C. vulgaris*)	X13688 (*C. vulgaris*)		**FJ715710** (*C. kessleri*)
Closteriopsis acicularis	EF113502	EF113433	Y17470		AY729488
Helicosporidium sp.					
Oocystis spp	EF113524 (*O. apiculata*)	EF113549 (*O. apiculata*)	AF228686 (*O. solitaria*)		
Prototheca wickerhamii	AJ245645	AJ969630	Z21552		PWL00000297
Trebouxia magna	EF113541				
Prasinophyceae					
Chlorodendrales					
Tetraselmis spp	DQ173248 (*T. suecica*)	DQ173247 (*T. suecica*)	X70802 (*T. striata*)		EF551330 (*T. tetrathele*) **FJ715711** (*T. striata*)
Mamiellales					
Micromonas pusilla		AY955031	AJ010408		EF551325
Ostreococcus lucimarinus					XM 001420948
Ostreococcus tauri	NC 008289	NC 008289	AY329635		CR954213/ **FJ715713**
Pseudoscourfieldiales					
Nephroselmis olivacea	NC 00927	NC 00927	X74754		**FJ715712**
Streptophyta					
Mesostigmatales					
Mesostigma viride	NC 002186	NC 002186	AJ250109		DQ394295

	atpB	rbcL	SSU rDNA	EF-1α	EFL
Chlorokybales					
Chlorokybus atmophyticus	DQ422812	DQ422812	M95612	**FJ715701**	
Klebsormidiales					
Entransia fimbriata	AY823688	E88 slordig	LB2793	**FJ715702**	
Klebsormidium flaccidum	AF408801	E87 rbcL2	X75520		
Zygnematales					
Gonatozygon spp	AF408796 (G. monotaenium)	U71438 (G. monotaenium)	X91346 (G. aculeatum)		
Spirogyra spp	AF408797 S. maxima	**FJ715721** (S. sp.)		EF551328/ **FJ715703**	
Staurastrum punctulatum	NC 008116	NC 008116	AF115442		
Zygnema spp	NC 008117 (Z. circumcarin)	NC 008117 (Z. circumcarin)	AJ853450 (Z. pseudogedeanum)		
Coleochaetales					
Chaetosphaeridium globosum	NC 004115	NC 004115	AF113506		
Coleochaete scutata	AY082303	AY082313	X68825		
Charales					
Chara spp	AF408782 (C. connivens)	L13476 (C. connivens)	U18493 (C. connivens)	EF551322 (C. australis)	
Nitella flexilis	AB110837	AB076056	U05261		
Embryophyta (land plants)					
Anthoceratophyta (Hornworts)					
Anthoceros spp	NC 004543 (A. formosa)	NC 004543 (A. formosa)	X80984 (A. agrestis)		
Marchantiophyta (Liverworts)					
Marchantia polymorpha	NC 001319	NC 001319	AB021684		
Bryophyta (Mosses)					
Physcomitrella patens	AP005672	AB066207		XM 001753007	
Spermatophyta (Seed plants)					
Actinidia spp	AJ235382 (A. chinensis)	L01882 (A. chinensis)	U42495 (A. sp.)	AY946009 (A. deliciosa)	
Arabidopsis thaliana	NC 000932	NC 000932	AC006837	AK230352	
Cichorium intybus		L13652		AY378166	
Oryza sativa	AC092750	AJ746297	X00755	AF030517	
Triticum aestivum	NC 002762	NC 002762	AY049040	M90077	
Vicia faba				AJ222579	
other eukaryotes					
Archaeplastida					
Glaucocystophyta					
Cyanophora paradoxa	NC 001675	NC 001675	AY823716	AF092951	
Rhodophyta					
Bangiophyceae					
Porphyra yezoensis	AP006715		D79976	U08844	
Porphyridium aerugineum			AJ421145		
Cyanidiophyceae					
Cyanidioschyzon merolae	AB002583		AB158485	AB095182	

	atpB	*rbcL*	SSU rDNA	EF-1α	EFL
Cyanidium caldarium	X66698		AB091232		
Florideophyceae					
Chondrus crispus		CCU02984	Z14140		CO652990 - CO652099- CO652788
Gracilaria spp	AY673996 (*G. tenuistipitata*)	AY673996 (*G. tenuistipitata*)	L26210 (*G. verrucosa*)		DV963090 - DV963156 - DV964877 (*G. changii*)
Chromalveolates					
Apicomplexans					
Plasmodium falciparum				X60488	
Ciliates					
Tetrahymena pyriformis				D11083	
Tetrahymena thermophila				XM 001032213	
Cryptophytes					
Goniomonas amphinema					AB332031
Guillardia theta					AM183813
Rhodomonas salina					DQ659244
diatoms					
Ditylum brightwellii					AB368772
Phaeodactylum tricornutum				18475 (JGI, v2.0)	
Skeletonema costatum					AB368773
Thalassionema nitzschioides					AB368774
Thalassiosira pseudonana				3858 (JGI, v3.0)	41829 (JGI, v3.0)
Dinoflagellates					
Heterocapsa triquetra					AY729485
Karlodinium micrum					EF134135
Oxyrrhis marina					DQ659243
Haptophytes					
Emiliania huxleyi					CV068986
Isochrysis galbana					AY729486
Pavlova lutheri					AY729487
Heterokonts					
Oomycetes					
Phytophthora infestans				AJ249839	
Opisthokonts					
Fungi					
Allomyces macrogynus					EC637105
Blastocladiella emersonii					EF064246
Mucor racemosus				MRATEF1A	
Saccharomyces martiniae				AF402021	
Metazoa (Animals)					
Homo sapiens				NM 001958	

	atpB	rbcL	SSU rDNA	EF-1α	EFL
choanoflagellates					
Monosiga brevicollis					XM_001745603
Sphaeroforma arctica					DQ403164
Excavates					
Diplomonads					
Giardia intestinalis				D14342	
Giardia lamblia				XM 76292	
Euglenids					
Euglena gracilis				X16890	
Parabasalids					
Trichomonas tenax				D78479	
Trichomonas vaginalis				XM 001325448	
Kinoplastids					
Leishmania braziliensis				XM 001563727	
Amoebozoa					
Entamoeba histolytica				ENHEF1ALPH	
Rhizaria					
Cercozoa					
Bigelowiella natans					AY729489
Foraminifera					
Planoglabratrella opecularis					AB334123
Centrohelids					
Raphidiophrys contractilis					AB332032

Additional file 7

Nexus file of EF-1α, EFL, GBlocks stripped EFL and SSU-*rbc*L-*atp*B alignments

Additional file 8

SSU rDNA alignment and partitioning

The SSU rDNA sequences were aligned on the basis of their rRNA secondary structure information using the following procedure. The SSU rDNA sequences of several green plant representatives incorporated in the European Ribosomal RNA Database (Wuyts et al. 2004, http://www.psb.ugent.be/rRNA/) were used as an initial model for building the alignment. The alignment editor DCSE v2.6 (De Rijk and De Wachter 1993) was used to annotate and check the secondary structure, and manually align the sequences. For the newly generated sequences, the alignment of the highly variable helices 43 and 49 (see De Rijk et al. 1999 for secondary structure nomenclature of the SSU gene) was refined and aided by folding the RNA sequences using mfold v.3.2 with default temperature and RNA parameters (Zuker 2003, http://mfold.bioinfo.rpi.edu/). Finally, we used the Xstem software (Telford et al. 2005) to extract the RNA secondary structure information from DCSE to a nexus format with stem/loop partitions.

Model selection procedure

Selection of a suitable partitioning strategy and suitable models for the partitions followed a three-step procedure and uses the Akaike Information Criterion (AIC) as a selection criterion. The guide tree used during the entire procedure was obtained by MP analysis of the concatenated data using PAUP* 4.0b10 (Swofford 2002). All subsequent likelihood optimizations and AIC calculations were carried out with Treefinder (Jobb et al. 2004). The first step consisted of optimizing the likelihood for nine potential partitioning strategies, assuming a HKY+G8 model for each partition. The three partitioning strategies with the lowest AIC scores (i.e., those providing the best fit to the data) were retained for further evaluation. The second step involved model selection for individual partitions. The likelihood of each partition present in the three retained partitioning strategies was optimized for three variants of the general time reversible model (F81, HKY and GTR), with and without inclusion of a discrete gamma distribution (eight categories) to model among-site rate heterogeneity. Because not all genes were sampled for all taxa, the guide tree was pruned to the taxa present in the partition in question before each optimization. The partition-specific models obtaining the lowest AIC score were passed on to the third step, which consisted of re-evaluation of the three partitioning strategies retained from the first step using the models selected for these partitions in the second step. The partitioning strategy + model combination that received the lowest AIC score in the third step was used in the phylogenetic analyses documented in the main text.

4

Complex phylogenetic distribution of a non-canonical genetic code in green algae

Ellen Cocquyt[1], Gillian H. Gile[2], Frederik Leliaert[1], Heroen Verbruggen[1], Patrick Keeling[2] and Olivier De Clerck[1]

[1]Phycology Research Group and Center for Molecular Phylogenetics and Evolution, Ghent University, Krijgslaan 281 S8, 9000 Ghent, Belgium

[2]Canadian Institute for Advanced Research, Department of Botany, University of British Columbia, Vancouver, V6 T 1Z4 Canada

Abstract

A non-canonical code, in which TAG and TAA have been reassigned from stop codons to glutamine, has previously been reported for the green algal orders Dasycladales. This study demonstrates that the Dasycladales share an identical alternative genetic code with the related clade Trentepohliales and the genus *Blastophysa*, but not with its sister clade the Bryopsidales. A single transition to the non-canonical code followed by a reversal to the canonical code in the Bryopsidales is highly improbable due to the profound genetic changes that coincide with codon reassignment. Multiple independent gains of the non-canonical code are not very plausible because a single type of non-canonical code evolved in some closely related ulvophyte lineages. Instead we favor a stepwise acquisition model, congruent with the ambiguous intermediate theory, whereby the alternative codes observed in these green algal orders share a single origin. The transition from a canonical to a non-canonical code has been completed only in the Trentepohliales, Dasycladales, Cladophorales and *Blastophysa*. This transition process, however, has not been completed in the Bryopsidales and hence no codon reassignment has taken place.

Keywords

non-canonical genetic code, glutamine, Chlorophyta, Ulvophyceae, stop codon, genome evolution

Introduction

The genetic code, which translates nucleotide triplets (codons) into amino acids, is one of the most highly conserved features in living organisms. The genetic code is universal in nearly all genetic systems, including viruses, bacteria, archaebacteria, eukaryotic nuclei and organelles. However, a small number of eubacterial, eukaryotic nuclear, plastid, and mitochondrial genomes have evolved slight variations on this standard or canonical genetic code (Jukes and Osawa 1993, Knight et al. 2001). Various models have been proposed to explain the evolutionary changes of the genetic code, including codon capture and ambiguous intermediate models (reviewed in Knight et al. 2001, Santos et al. 2004). Only five lineages of eukaryotes are known to have evolved non-canonical nuclear genetic codes, including ciliates, hexamitid diplomonads, fungi in the genus *Candida* and many Ascomycetes, polymastigid oxymonads and dasycladalean green algae (reviewed in Knight et al. 2001, Keeling and Leander 2003, de Koning et al. 2008).

By far the most common change to the genetic code in nuclear genomes is the reassignment of stop codons TAG and TAA to glutamine. This has happened independently in at least four eukaryotic lineages, including hexamitid diplomonads (Keeling and Doolittle 1996, Kolisko et al. 2008), some ciliates (Horowitz and Gorovsky 1985, Hanyu et al. 1986), polymastigid oxymonads (Keeling and Leander 2003, de Koning et al. 2008) and dasycladalean green algae (Schneider et al. 1989, Schneider and de Groot 1991). Interestingly, this particular non-canonical code has never evolved in prokaryotes or organelles. This bias towards reassignment of the stop codons TAG and TAA to encode glutamine in eukaryotes has been attributed to differences in the translation termination apparatus, tRNAs and tRNA synthetases, and mutation frequencies (Knight et al. 2001, Keeling and Leander 2003). Knowledge about the frequency and distribution of non-canonical codes across the branches of the tree of life will enable a better understanding of the evolution of genetic codes (Keeling and Leander 2003).

The aim of the present study is to extend our knowledge of the phylogenetic distribution of the non-canonical genetic code in green algae, with emphasis on the ulvophycean relatives of the Dasycladales. Our approach consists of screening green algal nuclear genes for the presence of non-canonical glutamine codons and interpreting the evolution of the genetic code and glutamine codon usage in a phylogenetic framework.

Methods

Polymerase Chain Reaction and sequencing

Total RNA was extracted from 43 taxa representing the major lineages of the Viridiplantae as described previously (Cocquyt et al. submitted). Portions of seven nuclear genes (actin, GPI, GapA, OEE1, 40S ribosomal protein S9 and 60S ribosomal proteins L3 and L17) were amplified, cloned when necessary and sequenced as described in Cocquyt et al. (submitted). A histone gene was amplified using the same PCR conditions with an annealing temperature of 55°C. The primers were based on a *Cladophora coelothrix* cDNA sequence aligned with GenBank sequences from green algae and land plants (His-F: 5'-ATG GCI CGT ACI AAG CAR AC-3' and His-R: 5'-GGC ATG ATG GTS ACS CGC TT-3'). In

addition, total RNA was extracted from *Ignatius tetrasporus* and the bryopsidalean species *Caulerpa cf. racemosa* as described previously (Gile et al. 2009). Portions of actin, β-tubulin, and HSP90 genes including the stop codon were amplified from these taxa by 3' RACE using the First Choice RLM-RACE kit (Ambion) using nested degenerate primers (actin-F: 5'-TAC GAA GGA TAC GCA CTN CCN-3' C and actin-R: 5'- GAG ATC GTG CGN GAY ATH AAR GA-3'; β-tubulin-F: 5'-GAT AAC GAG GCT CTN TAY GAY ATH TG-3' and β-tubulin-R: 5'-CCT TTC CGA CGG CTN CAY TTY TT-3'; HSP90-F: 5'-ATG GTC GAT CCN ATH GAY GAR TA-3' and 5'-GCT AAG ATG GAG MGN ATH ATG AA-3').

Genetic codes

The presence of a non-canonical code in green algal taxa was detected by screening alignments of nuclear genes for apparent stop codons at positions coding for glutamine in other green plant taxa and by the presence of only TGA as a functional stop codon at the predicted 3' end of genes. The presence of the standard code is confirmed by the presence of only canonical glutamine codons and the use of all three stop codons at the predicted end of genes.

Molecular phylogenetics

We constructed a reference phylogeny of the Viridiplantae based on the analysis of multiple genes as described in Cocquyt et al. (submitted) to study the phylogenetic occurrence of the standard and non-canonical genetic code. The phylogenetic analysis was carried out on an alignment consisting of the nuclear genes mentioned above, together with SSU nrDNA and the plastid genes *rbc*L and *atp*B. Histone genes were excluded from the analysis because they are known to be duplicated across genomes (Nei and Rooney 2005, Wahlberg and Wheat 2008). Phylogenetic analyses were carried out with model-based techniques (ML and BI), paying careful attention to the selection of suitable partitioning strategies and models of sequence evolution. The model selection procedure proposed 8 partitions: SSU nrDNA was partitioned into loops and stems (2 partitions) and nuclear and plastid genes were partitioned into codon positions (2×3 partitions). GTR+Γ_8 was the preferred model for all partitions. Noise-reduction techniques were applied to counteract the erosion of ancient phylogenetic signal caused by fast-evolving sites. The phylogenetic tree presented here is based on the 75% slowest-evolving sites (Cocquyt et al. submitted).

Evolution of glutamine codon usage

The evolution of glutamine codon usage was estimated using ancestral state estimation techniques. The frequency of the two canonical and two non-canonical glutamine codons was calculated for each species in the phylogenetic tree. Codon frequencies were mapped along the reference tree using the ace function of the APE package (Paradis et al. 2004). This function estimates ancestral character states by maximum likelihood optimization (Schluter et al. 1997). The branch lengths were based on ML estimates because we consider them to be a more relevant approximation of the amount of codon usage evolution that can be expected to take place than absolute time (cf. Cocquyt et al.

2009). The output from APE was mapped onto the reference tree with TreeGradients v1.03 (Verbruggen 2009) to plot ancestral states of continuous characters on a phylogenetic tree as colors along a color gradient.

Topological hypothesis testing

Approximately unbiased tests (AU test, Shimodaira 2002) were used to test an alternative relationship between ulvophycean orders as suggested by the distribution of the canonical genetic code (see results). Site-specific likelihoods were calculated by maximum likelihood optimization in Treefinder using the same model specifications as for phylogenetic inference (Cocquyt et al. submitted). AU tests were performed with CONSEL V0.1i (Shimodaira and Hasegawa 2001).

Results

The dasycladaceans *Acetabularia* and *Batophora* have been shown previously to use the non-canonical code (Schneider et al. 1989, Schneider and de Groot 1991). Looking more broadly, we find that the non-canonical glutamine codons appear in coding sequences of additional members of the Dasycladales (genus *Bornetella*), as well as Cladophorales (genera *Boergesenia, Boodlea*, *Cladophora*, *Dictyosphaeria*, *Phyllodictyon, Siphonocladus, Valonia*), Trentepohliales (genus *Trentepohlia*) and the genus *Blastophysa*, which is currently not assigned to an order. Most non-canonical codons were found at highly conserved positions that encoded glutamine codons in the rest of the green plant lineage. Only TGA is used as stop codon in ribosomal protein 40S S9 and OEE1 of *Cladophora*. Moreover, other sequences from both *Cladophora* and *Acetabularia* deposited in GenBank only have TGA as stop codon: GapA gene (DQ270261) of *Cladophora*, and EF-1α (EF551321) and PsbS genes (BK006014) of *Acetabularia*. The presence of the standard code is confirmed for the genus *Ignatius* and the order Bryopsidales (genera *Caulerpa*, *Bryopsis*) by the presence of only canonical glutamine codons and the use of all three stop codons at the ends of various genes. The *Ignatius* actin gene has TAG as stop codon, while the β tubulin and HPS90 genes have TAA as stop codons. β tubulin and HPS90 genes of *Caulerpa* have TAA as stop codons. Additional *Bryopsis* sequences deposited in GenBank confirmed the presence of all three stop codons in the Bryopsidales: TAA in ribonuclease Bm2 gene (AB164318), TAG in lectin precursor and oxygen evolving protein of photosystem II genes (EU410470 and AB293980), and TGA in the bryohealin precursor gene (EU769118).

The occurrence of the standard and non-canonical code are plotted onto the reference phylogeny in Fig. 1. The Streptophyta, prasinophytes, Trebouxiophyceae and Chlorophyceae possess the standard genetic code. Within the class Ulvophyceae, the standard code is found in the orders Ulvales, Ulotrichales, Bryopsidales and the genus *Ignatius* whereas the orders Dasycladales, Siphonocladales, Trentepohliales and the genus *Blastophysa* have a non-canonical code. However, the taxa with a non-canonical code do not form a monophyletic group (Fig. 1).
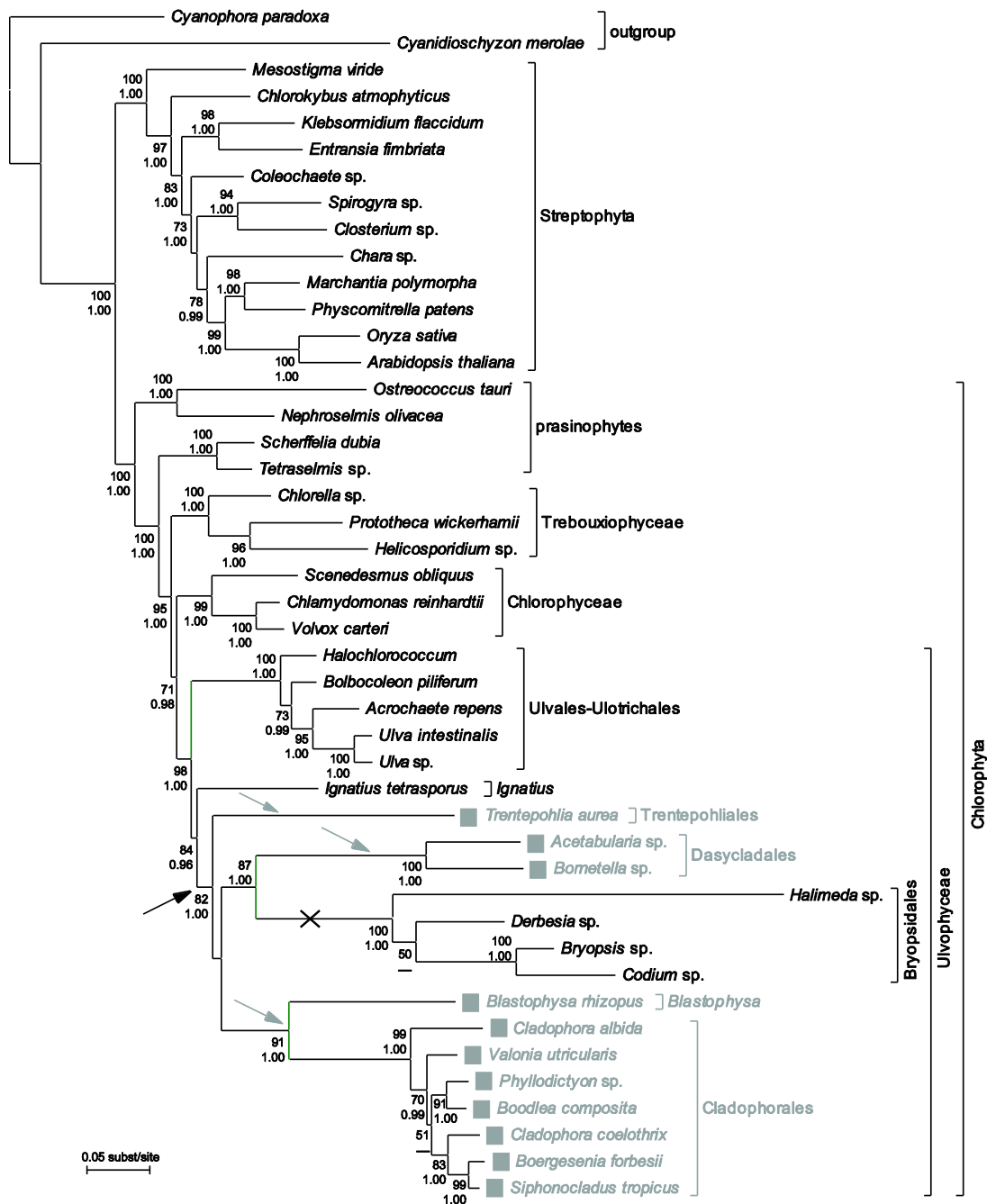
Figure 1. The occurrence of a non-canonical genetic code in green algae is indicated with gray squares. Taxa with a non-canonical genetic code are not monophyletic. Three alternative scenarios can explain its distribution on the tree:

(1) a single origin of the non-canonical code along the branch leading to the orders Trentepohliales, Dasycladales, Bryopsidales, Cladophorales and the genus *Blastophysa* and a reversal to the universal code in the Bryopsidales (indicated with black arrow and cross)

(2) three independent gains of the non-canonical code in the Trentepohliales, the Dasycladales and the Cladophorales + *Blastophysa* (indicated with gray arrows)

(3) a single initiation of the formation of the non-canonical code along the branch leading to the orders Trentepohliales, Dasycladales, Bryopsidales, Cladophorales and the genus *Blastophysa* that has been completed in all lineages except the Bryopsidales (black arrow combined with gray arrows)

The reference phylogeny of the green plant lineage was obtained by maximum likelihood inference of the 25% site stripped dataset containing 7 nuclear genes, SSU nrDNA and plastid genes *rbc*L and *atp*B (Cocquyt et al. submitted). Numbers at nodes indicate ML bootstrap values (top) and posterior probabilities (bottom); values below respectively 50 and 0.9 are not shown.

If both the phylogeny and the distribution of the genetic codes shown in Fig. 1 are correct, then more than one gain and/or loss event of the non-canonical code must be postulated. To examine the validity of the phylogeny, we performed topology tests to evaluate whether the data rejected a topology in which all taxa with the non-canonical code formed a monophyletic group. Specifically, we found that the topology where Bryopsidales is sister to all Ulvophyceae with a non-canonical code is rejected with high significance (ΔlnL = 27.9; p < 0.0001).

The estimated evolution of glutamine codon usage frequencies is shown in Fig. 2. The canonical codon CAG is most commonly used, followed by the canonical codon CAA. Among the non-canonical codons, TAG is used more commonly then TAA. This bias is likely a product of the overall bias of these species and/or genes for GC residues, which favors a G in the third position. This observation is also congruent with canonical glutamine codons CAG and CAA which mutate to non-canonical glutamine codons TAG and TAA by a single transition (C $\rightarrow$ T) at the first codon position (Keeling and Leander 2003).

Discussion

Our results reveal the distribution of a non-canonical genetic code in the Ulvophyceae, where glutamine is encoded by both canonical CAG and CAA codons as well as non-canonical TAG and TAA codons. Unexpectedly, we find the taxa with this non-canonical code do not form a monophyletic group according to the seemingly robust phylogeny of the organisms in which it is found (Fig. 1). If the inferred phylogeny is indeed correct, three alternative scenarios can explain the distribution of the code on that tree: (1) a single origin of the non-canonical code along the branch leading to the orders Trentepohliales, Dasycladales, Bryopsidales, Cladophorales and the genus *Blastophysa* and a reversal to the universal code in the Bryopsidales (Fig. 1: indicated with black arrow and cross); (2) three independent gains of the non-canonical code in the Trentepohliales, the Dasycladales and the Cladophorales + *Blastophysa* (Fig. 1: indicated with gray arrows), and (3) a single initiation of a stepwise formation of the non-canonical code along the branch leading to the orders Trentepohliales, Dasycladales, Bryopsidales, Cladophorales and the genus *Blastophysa,* but where the process has gone to completion in all lineages except the Bryopsidales (Fig. 1: black arrow combined with gray arrows). Alternatively, because changes in the genetic code are so rare, the possibility that the reference phylogeny is wrong should not be passed over too easily. More specifically, if one assumes that the current position of the Bryopsidales is wrong and that in reality this group is sister to all the taxa with a non-canonical code, only a single transition from the genetic code would have to be invoked. In what follows, we will discuss each of these possibilities in more detail and report on some cytological correlates of the non-canonical genetic code.
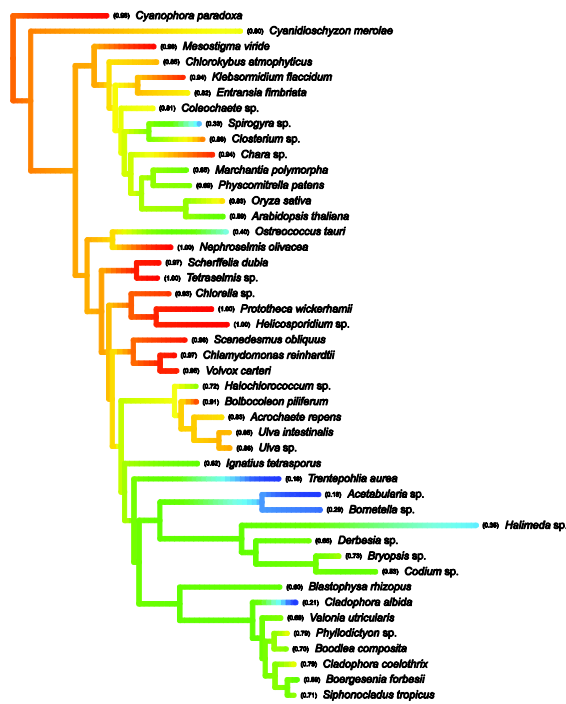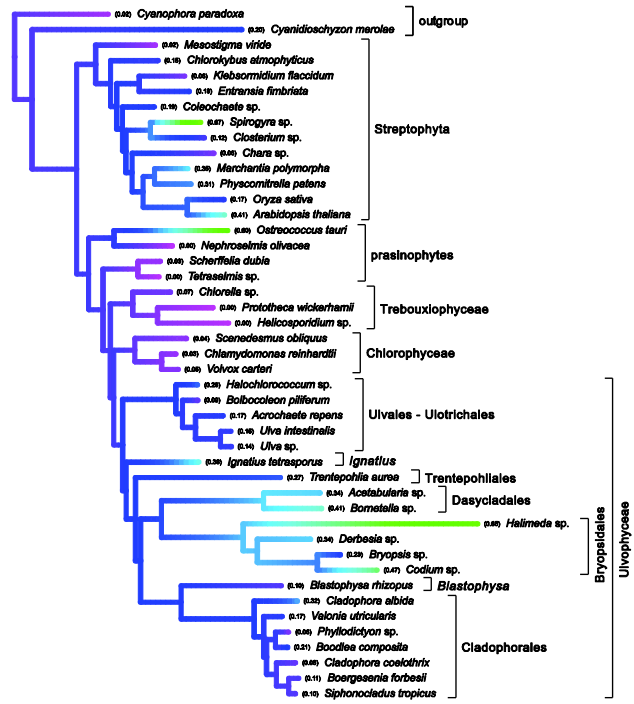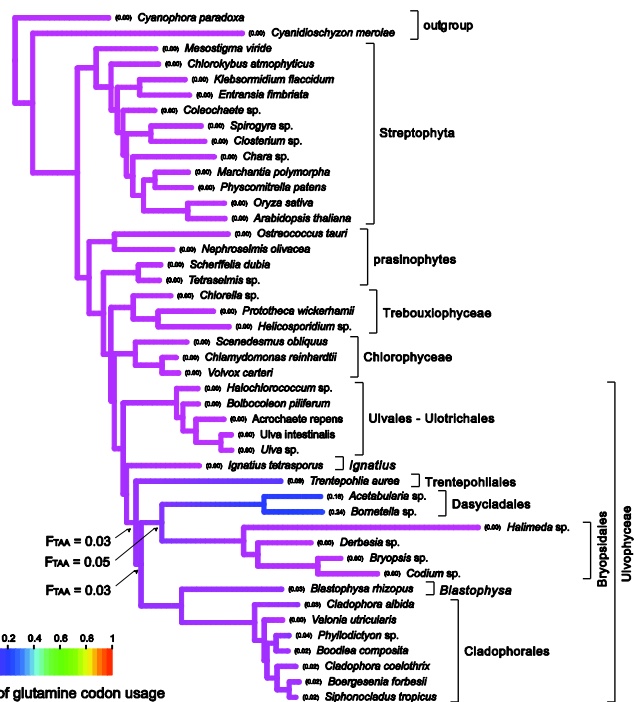
Figure 2. The estimated ancestral frequencies of glutamine codon usage. Canonical codon CAG is most commonly used, followed by CAA, the second canonical codon. Among the non-canonical codons, TAG is used more commonly then TAA. The estimate ancestral frequencies of non-canonical codon usage along the nodes of interest are indicated with an arrow.

Phylogenetic uncertainty

Our phylogenetic tree is based on the most comprehensive dataset currently available for the Chlorophyta. It is inferred from a concatenated dataset including seven nuclear genes, SSU nrDNA and two plastid genes using model-based techniques with carefully chosen partitioning strategies and models of sequence evolution and application of a site removal approach to optimize signal for the relevant level of divergence (Cocquyt et al. submitted). Our tree shows a sister relationship between Dasycladales and Bryopsidales with moderately high statistical support (BV 87 and PP 1.00). This relationship is concordant with a recently published 74-taxon phylogeny of the green lineage based on SSU nrDNA and two plastid genes (Cocquyt et al. 2009). Both phylogenies show major improvements in taxon and gene sampling within the Ulvophyceae compared to previously published phylogenies, which were either based on a single marker, did not include the Bryopsidales, or could not resolve the relationships among the Bryopsidales, Dasycladales and Cladophorales (Zechman et al. 1990, Lopez-Bautista and Chapman 2003, Watanabe and Nakayama 2007). Based on the dataset used to infer our reference tree, the alternative topology in which the Bryopsidales are sister to all taxa with a non-canonical code is significantly less likely than the ML tree as shown by AU tests. As a consequence, unless the phylogenetic signal contained in the molecular data is fundamentally wrong, the ulvophycean taxa with the non-canonical code form a paraphyletic group and one of the more complex evolutionary scenarios for the gain of the non-canonical code has to be invoked.

Gain–reversal hypothesis

A reversal from the non-canonical to the standard genetic code is unlikely for several reasons. First, following a transition to the non-canonical code, TAG and TAA codons would be present in many coding sequences. In order to revert to the standard or canonical code, these codons would all have to revert to canonical codons or they would terminate translation, with obvious detrimental effects. Our ancestral state estimates indicate a non-negligible usage frequency of both non-canonical codons along the ancestral nodes of interest (Fig. 2 C,D, indicated with arrows). These results must be considered with caution because of the intrinsic limitations of ancestral state estimation (Martins 1999) and the fact that the non-independence of the evolution of the genetic code and glutamine codon usage cannot be captured by the Brownian motion model. Nevertheless, it seems likely that glutamine-encoding TAA and TAG codons in a genome using the non-canonical code would be more common than TAA and TAG were in the ancestral genome where the standard code was still in use, because the frequency of stop codons is at most once per gene whereas even disfavored glutamine codons can appear several times in a single gene. The major argument against a reversal from the non-canonical to the canonical code is that the original change would have been accompanied by changes in the translation apparatus, e.g. alterations in tRNAs, aminoacyl-tRNA-synthethases, eukaryotic release factor (eRF1) and also potentially a nonsense-mediated mRNA decay (NMD) mechanism, all of which would have to reverse in the Bryopsidales in order to support this scenario (Knight et al. 2001, Keeling and Leander 2003). Taken together, the reversal of a non-canonical code to the standard code appears highly unlikely.

Multiple independent gains

Several independent acquisitions of non-canonical codes have been reported for ciliates (Tourancheau et al. 1995, Knight et al. 2001, Lozupone et al. 2001). Indeed, stop codon reassignments are surprisingly frequent in this group of organisms: the same non-canonical code than the one present in the Ulvophyceae has evolved several times, another non-canonical code in which TGA codes for tryptophan evolved twice, and a third non-canonical code that translates TGA to cysteine evolved once. In the present study, the distribution of the non-canonical code in the phylogenetic tree would require three gains: in the Trentepohliales, the Dasycladales and the *Cladophorales + Blastophysa*. Contrary to the situation in ciliates, however, Ulvophyceae only evolved a single type of non-canonical code and they did so in closely related lineages. The combination of these two arguments weakens the plausibility of multiple independent gains although the possibility cannot be excluded altogether.

Stepwise acquisition model

Several studies have revealed that stop codon reassignment is a gradual process requiring changes to tRNA and eRF1 genes (reviewed in Knight et al. 2001, Santos et al. 2004, Sengupta and Higgs 2005). In several eukaryotes, the two glutamine tRNAs, recognizing CAG and CAA, are also able to read TAG and TAA codons by a G·U wobble base pairing at the third anticodon position (Beier and Grimm 2001). In normal circumstances (canonical code), eRF1 out-competes the glutamine tRNAs when it comes to binding TAG and TAA. Mutations in glutamine tRNA copies that allow them to bind TAG and TAA may increase the capacity to translate TAG and TAA to glutamine. This leads to an intermediate step in which both eRF1 and the mutated tRNAs can easily bind to TAG and TAA. The ciliate *Tetrahymena thermophila* has three major glutamine tRNAs, one that recognizes the normal glutamine codons CAG and CAA, and two supplementary tRNAs that recognize the non-canonical codons TAG and TAA. These two supplementary tRNAs were shown to have evolved from the normal glutamine tRNA (Hanyu et al. 1986). A similar situation likely exists in diplomonads (Keeling and Doolittle 1996). In order to alter the genetic code, mutations are required not only in glutamine tRNAs but also in eukaryotic release factor 1 (eRF1) so it no longer recognises TAG and TAA codons. In ciliates it has been shown that eRF1 sequences are highly divergent in domain 1 between species with a canonical and non-canonical code, which could indicate that eRF1 can no longer recognize TAG and TAA codons in the species with a non-canonical code (Lozupone et al. 2001, Inagaki et al. 2002). An additional mechanism that constrains the genetic code and therefore represents yet another potential step in the process of codon assignment is nonsense-mediated mRNA decay (NMD) (Baker and Parker 2004, Maquat 2004). NMD reduces errors in gene expression by eliminating mRNAs that encode for an incomplete polypeptide due to the presence of stop codons. In the case of a non-canonical code where stop codons TAG and TAA encode glutamine, as observed here, this NMD mechanism would have to be altered also in order to prevent degradation of mRNAs containing TAG and TAA codons.

A stepwise acquisition model can reconcile the opposed and problematic hypotheses of multiple gains versus a single gain with subsequent loss. For example, the ambiguous intermediate theory (Schultz and Yarus 1994, Santos et al. 2004) would explain the distribution of the non-canonical code

in Ulvophyceae as follows: mutations in the anticodons of canonical glutamine tRNAs occurred once along the branch leading to the orders Trentepohliales, Dasycladales, Bryopsidales, Cladophorales and the genus *Blastophysa* (Fig. 1, black arrow). The presence of these mutated tRNAs allowed TAG and TAA codons to be translated to glutamine instead of terminating translation. At this step, the mutated tRNAs compete with eRF1 for the TAA and TAG codons. To complete the transition to the non-canonical code, a subsequent mutation of eRF1 that prevents binding of eRF1 with TAG and TAA is required. If one assumes that this step occurred three times independently in the Trentepohliales, Dasycladales and Cladophorales + *Blastophysa* (Fig. 1, gray arrows), whereas the mutated tRNAs decreased in importance or went extinct through selection or drift along the branch leading to the Bryopsidales, the distribution of the non-canonical code in the Ulvophyceae would be explained. However, a detailed comparison of eukaryotic release factors (eRF1) and glutamine tRNAs in the respective clades of the Ulvophyceae is needed to test this evolutionary scenario.

Cytological correlates of non-canonical code

In the ciliates, the multiple appearances of alternative codes have been attributed to their nuclear characteristics. Ciliates have two nuclei: a small, diploid micronucleus which is not expressed and is the germ line for DNA exchanges during the sexual process, and a large, polyploid macronucleus which is actively transcribed and ensures vegetative cell growth, but is not passed on to progeny after sexual conjugation and is replaced by a newly formed macronucleus after a number of rounds of mitotic division. There is therefore a time lag between the occurrence of mutations in the micronucleus and the expression of these mutations in the macronucleus, and this has been postulated to be a contributing factor to why ciliates have evolved alternative genetic codes more frequently (Cohen and Adoutte 1995). In this context it is worth mentioning that hexamitid diplomonads, for which a single origin of a non-canonical code together with the uninucleate enteromonads has been shown, have two similar nuclei per cell (Keeling and Leander 2003, Kolisko et al. 2008) and that several ulvophycean groups are characterized by multinucleate cells namely the Dasycladales, Bryopsidales, Cladophorales and *Blastophysa*. The Cladophorales and *Blastophysa* are branched filamentous seaweeds consisting of multinucleate cells with a few to thousands of nuclei arranged in non-motile cytoplasmic domains (siphonocladous organization). Members of the Bryopsidales and Dasycladales have a siphonous organization: they consist of a single, giant tubular cell with numerous nuclei and complex patterns of cytoplasmic flow. *Acetabularia* (Dasycladales) forms an exception in the sense that the number of nuclei has been reduced to one in this genus: its members possess a single, giant nucleus situated in the rhizoid. It thus appears that the ancestors of these taxa in which changes of the genetic code have occurred were very likely all multinucleate. Despite the fact that some eukaryotes with a non-canonical code do not feature multinucleate cells and that there are plenty of examples of groups with multinucleate cells that have not evolved alternative codes, this observation suggests that a multinucleate cytology can facilitate codon reassignments. To our knowledge, no explanatory mechanisms have been described.

Acknowledgements

5

Codon usage bias and GC content in green algae

Ellen Cocquyt, Heroen Verbruggen and Olivier De Clerck

[1]Phycology Research Group and Center for Molecular Phylogenetics and Evolution, Ghent University, Krijgslaan 281 S8, 9000 Ghent, Belgium

Abstract

Large differences in synonymous codon usage bias and GC content have been observed within and among genomes and a plethora of hypotheses have been put forward to explain them. In this study, we characterize patterns of codon usage bias and GC content in eight nuclear housekeeping genes from 43 green algae and land plants. We analyze the evolution of these biases in a phylogenetic framework. We observe stronger codon usage bias in the ancestral streptophytes *Mesostigma* and *Chlorokybus* than in the remainder of the Streptophyta. Within the Chlorophyta, the prasinophytes, Trebouxiophyceae and Chlorophyceae have markedly stronger codon usage bias than the Ulvophyceae. One exception is *Ignatius*, which is a member of the Ulvophyceae yet has a markedly stronger codon usage bias than other members of this class. GC content patterns show congruent trends, species with strong codon usage bias having a high GC content. We interpret these results along with the biology of the organisms in the framework of two models: the mutation-selection-drift model and the co-evolutionary model of genome composition and resource allocation. It is remarkable that unicellular organisms and colony-forming species have much more pronounced GC and codon usage biases as compared to multicellular and macroscopic species. This may follow from unicells having large population sizes, which leads to more codon usage bias due to stronger selection as compared to species with smaller population sizes where drift can more rapidly fix mutation. We also interpret the correlation of the observed biases with growth rates, generation time and rates of molecular evolution of the organisms under study.

Keywords

codon usage bias, GC content, unicellular, multicellular, effective population size, green plants

Introduction

The genetic code, which allows translating codons (nucleotide triplets) to amino acids, is redundant because all amino acids except methionine and tryptophan are encoded by two to six synonymous codons. Even though the fact that synonymous codons translate to the same amino acid would lead to the expectation that synonymous codons are used in equal proportions, DNA sequence data from a wide range of organisms have shown that this is not the case: some synonymous codons are clearly preferred over others, and different organisms often have different preferences. This phenomenon is called codon usage bias and has been associated with a wide range of biological factors (Salim and Cavalcanti 2008). Within a single organism, codon usage bias is more pronounced in highly expressed and long genes, and near the 3'end of genes. Codon usage bias is stronger when the population of tRNAs is less diverse and correlates with GC content. Differences in codon usage bias between organisms appear to correlate with their population size; species with large effective population sizes having more biased codon usage than species with smaller population sizes (Cutter et al. 2006, Ingvarsson 2008, Subramanian 2008).

Conceptual models of various levels of complexity have been proposed to explain the presence and degree of codon usage bias in and between organisms. The mutation-selection-drift balance model assumes that codon usage ensues from the delicate balance between mutation, selection and genetic drift in populations (Bulmer 1991, Hershberg and Petrov 2008). Whereas natural selection can be expected to favor the organism's preferred (major) codons over minor codons, mutational pressure and genetic drift allow minor codons to persist in the population (Fig 1. A external factors influencing mutation bias and selection, reviewed in Hershberg and Petrov 2008). A more complex model, which we will refer to as the co-evolutionary model, assumes that genome biases (e.g. GC and codon usage bias) result from the co-evolution between genome composition and resource concentration, both of which act on genome bias via mutation bias and selection (Fig 1. B internal mechanism influencing mutation bias and selection, Vetsigian and Goldenfeld 2009). More specifically, for a certain genome composition the nucleotide and tRNA concentrations are altered to increase speed and accuracy, which in turn leads to mutation and selection pressures favoring the preferred nucleotides and codons. The combination of all these factors leads to differences in genome biases between and within species. To date, most studies have concentrated on genome bias within a single species and relatively few studies examined genome biases between organisms. As far as we know, only one study examined codon usage bias between more distantly related nematode species which inhabit different ecological niches being either free-living or parasitic (Cutter et al. 2006).

The green algae form a promising model system to study the evolution of codon usage bias and GC content over longer time-scales for several reasons. First, some ulvophycean algae were shown to have a non-canonical genetic code in which the stop codons TAG and TAA have been reassigned to glutamine (Cocquyt et al. in prep.). This non-canonical code involves changes in glutamine codon usage, the translation termination apparatus, tRNAs and tRNA synthetases, and mutations converting canonical glutamine codons (CAG and CAA) to non-canonical glutamine codons (TAG and TAA) by a single transition at the first codon position (C $\rightarrow$ T). Second, green algae and land plants occupy a wide variety of ecological niches, including marine, freshwater and terrestrial habitats. Finally, the green algae are characterized by markedly different rates of molecular evolution as

indicated by root to tip path length differences in phylogenetic trees, indicating differences in molecular evolution between different groups of green algae. Finally, green algae exhibit remarkable morphological and cytological diversity ranging from unicells and colonies, over multicellular filaments and foliose blades, to highly complex multicellular and multinucleate life forms. Two types of multinucleate organisms are present: (1) a siphonocladous organization with cells that contain a few to thousands of nuclei arranged in non-motile cytoplasmic domains (e.g. order Cladophorales and genus *Blastophysa*), and (2) a siphonous organization with a single, giant tubular cell that contains numerous nuclei and complex patterns of cytoplasmic flow (e.g. orders Bryopsidales and most Dasycladales). These morphological and cytological differences may correlate with differences in population size and as such have an impact on genome biases.
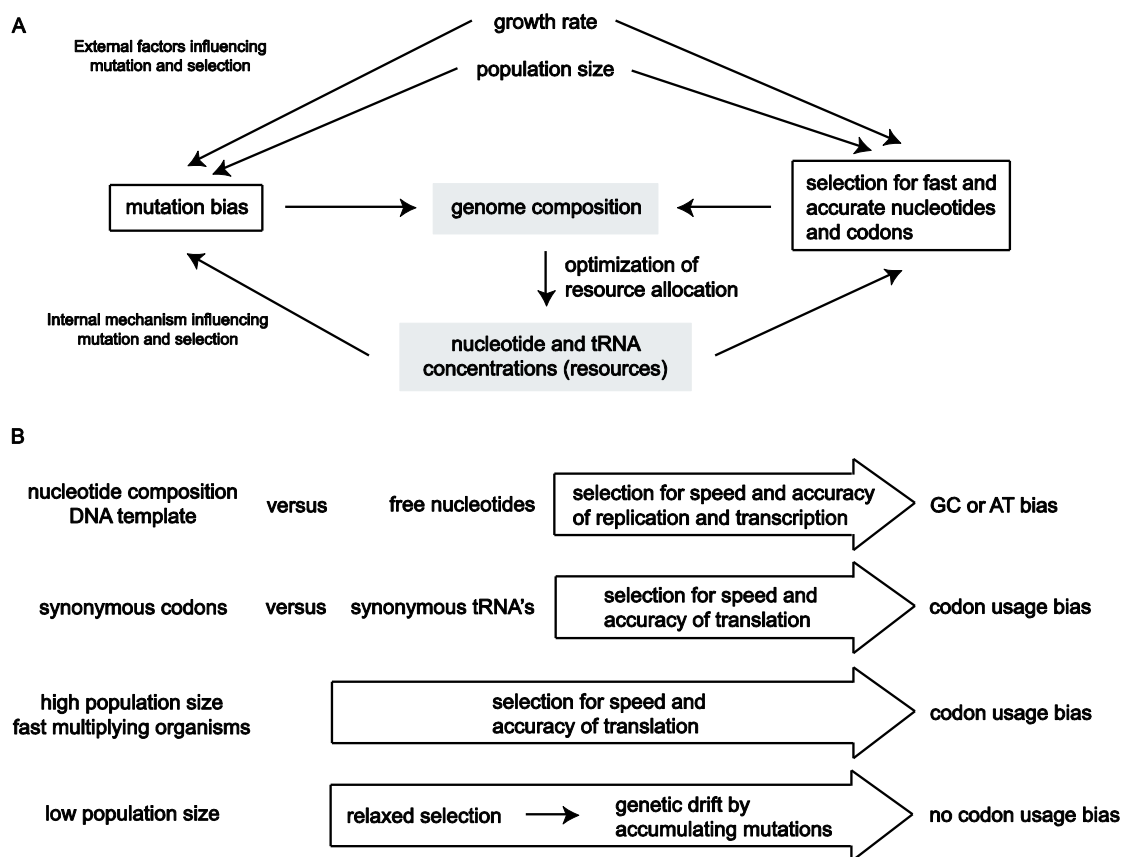
Figure 1. Origin of genome biases such as GC or AT bias and codon usage bias.

A. Genome composition is influenced by a delicate balance between mutation bias and selection for fast and accurate transcription and translation. Selection and mutation are in turn influenced by several external and internal factors.

(1) species with large population sizes have more codon usage bias due to stronger selection compared to species with smaller population sizes where drift can more rapidly fix mutations.

(2) In rapidly multiplying organisms, the time spent in translation is a limiting factor in cell division, increased selection leads to an optimal translation.

(3) Selection for optimal speed and accuracy of replication and translation involves a co-evolution between genome composition and resource concentration, which in turn influence the direction of mutation bias.

B. Selection and mutation are influenced by the co-evolution between genome composition and resource concentration, effective population size and growth rate, each causing a specific genome bias.

The goal of this study is to characterize patterns of codon usage bias and GC content in green algal nuclear genes and analyze their evolution in an explicit phylogenetic framework. Our approach consists of calculating codon usage bias and GC content in 43 representatives of green algae and land plants based on eight nuclear housekeeping genes. The evolutionary patterns of codon usage bias and GC content are subsequently approximated with ancestral state estimation techniques and mapped along the reference tree.

Methods

DNA amplification and sequencing

Seven nuclear genes (actin, GPI, GapA, OEE1, 40S ribosomal protein S9 and 60S ribosomal proteins L3 and L17) were amplified and sequenced for 43 taxa representing the major lineages of the Viridiplantae as described in Cocquyt et al. (submitted). A histone gene was amplified using the same PCR conditions with an annealing temperature of 55°C. The primers were based on a *Cladophora coelothrix* cDNA sequence aligned with GenBank sequences from green algae and land plants (His-F: 5'-ATG GCI CGT ACI AAG CAR AC-3' and His-R: 5'-GGC ATG ATG GTS ACS CGC TT-3').

Molecular phylogenetics

We constructed a reference phylogeny of the Viridiplantae as described in Cocquyt et al. (submitted) to study patterns of codon usage bias and GC content. The phylogenetic analysis was carried out on an alignment consisting of the nuclear genes mentioned above, SSU nrDNA and the plastid genes *rbc*L and *atp*B. Histone genes were excluded from the analysis because they are known to be duplicated across genomes (Nei and Rooney 2005, Wahlberg and Wheat 2008). The analysis was based on the 75% slowest-evolving sites to improve signal for the reconstruction of ancient relationships (Cocquyt et al. submitted).

Synonymous codon usage bias and GC content

Synonymous codon usage order (SCUO) was used to measure synonymous codon usage bias in each species and was calculated with CodonO based on concatenated sequences of the eight nuclear genes mentioned above (Wan et al. 2006, Angellotti et al. 2007). SCUO represents a value between 0 to 1, larger values indicating stronger codon usage bias. GC content was calculated with TreeFinder (Jobb et al. 2004) based on the same eight nuclear genes. The evolution of SCUO and GC content was approximated using ancestral state estimation techniques and mapped along the reference tree. Ancestral state estimation was carried out with the ace function of the APE package (Paradis et al. 2004), using maximum likelihood optimization for continuous characters (Schluter et al. 1997). The output from APE was mapped onto the reference tree with TreeGradients v1.03 (Verbruggen 2009). This program plots the estimated ancestral character states on a phylogenetic tree as colors along a color gradient.

Results and discussion

The estimated evolution of synonymous codon usage bias (SCUO) and GC content is shown in Fig. 2. *Mesostigma*, closely followed by the genera *Chlorokybus* and *Entransia*, have stronger codon usage bias than the remainder of the Streptophyta. Within the Chlorophyta, the prasinophytes, Trebouxiophyceae, Chlorophyceae have markedly stronger codon usage bias than the Ulvophyceae. One exception is *Ignatius*, which is a member of the Ulvophyceae yet has a markedly stronger codon usage bias than other members of this class. GC content patterns show congruent trends: within the Streptophyta, GC bias is highest for *Mesostigma* and *Chlorokybus*. Within the Chlorophyta, GC bias is much more pronounced in the prasinophytes, Trebouxiophyceae and Chlorophyceae as compared to the Ulvophyceae. Within the Ulvophyceae, the orders Trentepohliales, Dasycladales and Bryopsidales have the weakest GC bias.

It follows from Fig. 2 that there have been only a few transitions in GC and codon usage bias during the evolution of the green algae, more specifically two independent reductions of the biases in the Streptophyta and Ulvophyceae. This apparent conservatism of genome biases over long evolutionary periods stands in strong contrast with the notion that variation of codon usage bias and GC content has been shown between closely related species and even within individual genomes (e.g. Duret and Mouchiroud 1999, Derelle et al. 2006, Ingvarsson 2008). In the light of these indications for fast evolution of genome biases, it seems unlikely that their conservatism over long periods of time in the green algae is due to slow evolution but rather that it is a consequence of the association of genome biases with biological traits that have evolved slowly. In the following paragraphs, we will highlight several correlations between the observed patterns of genome bias and organismal traits that are associated with them and potentially responsible for them.

Body and population size

As illustrated in Fig. 2, the genes under study have more pronounced GC and codon usage bias in unicellular organisms and colony-forming species as compared to multicellular and/or multinucleate species. As such, the trends towards multicellularity and macroscopic algal bodies in both the Streptophyta and Chlorophyta are associated with decreasing amounts of GC and codon usage bias. It has been documented that codon usage bias is stronger for species with large populations due to stronger selection compared to species with smaller population sizes where drift can more rapidly fix mutations (Cutter et al. 2006, Ingvarsson 2008). Because the selective advantages associated with the usage of alternative codons are subtle, selection can only offset the stochastic effects of genetic drift in species with large effective population sizes (Akashi 1995, Cutter et al. 2006, dos Reis and Wernisch 2009). Because body size and population size appear to be inversely related for aquatic organisms (Finlay 2002, Fenchel and Finlay 2004), it seems likely that the strong codon usage bias observed in unicellular and colonial organisms may be a result of their larger effective population sizes.

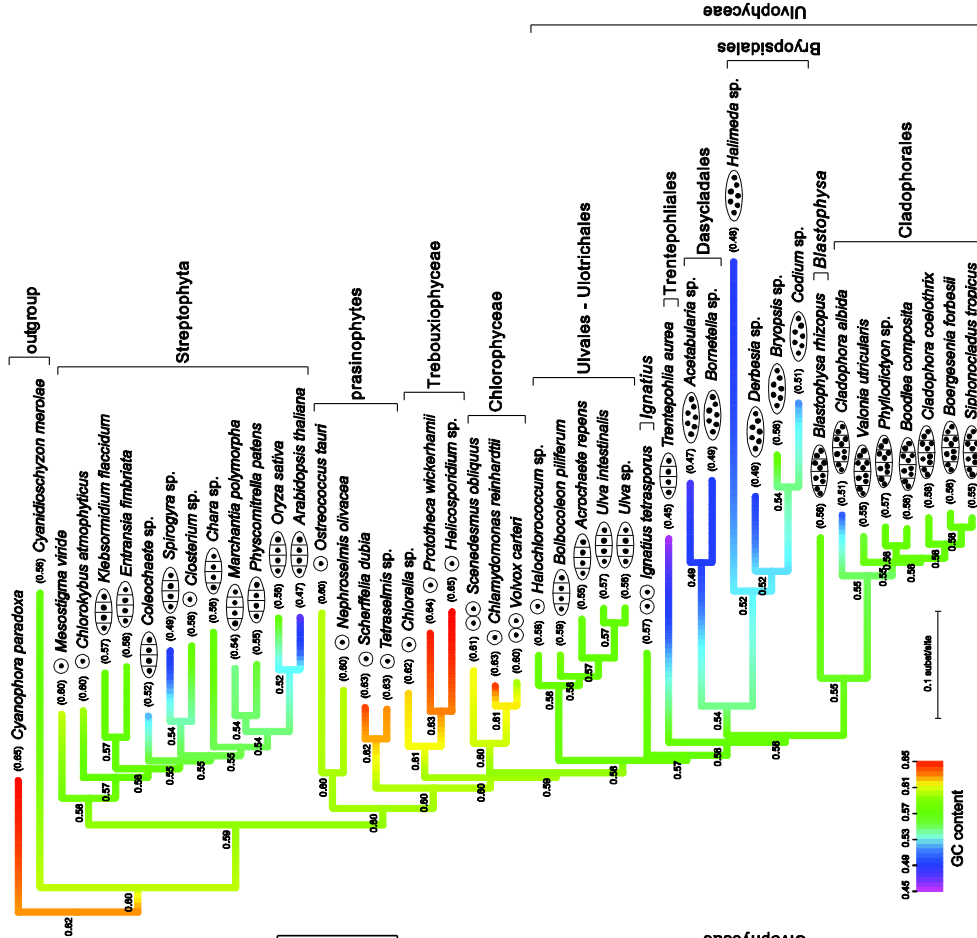# A. Synonymous codon usage bias (SCUO)
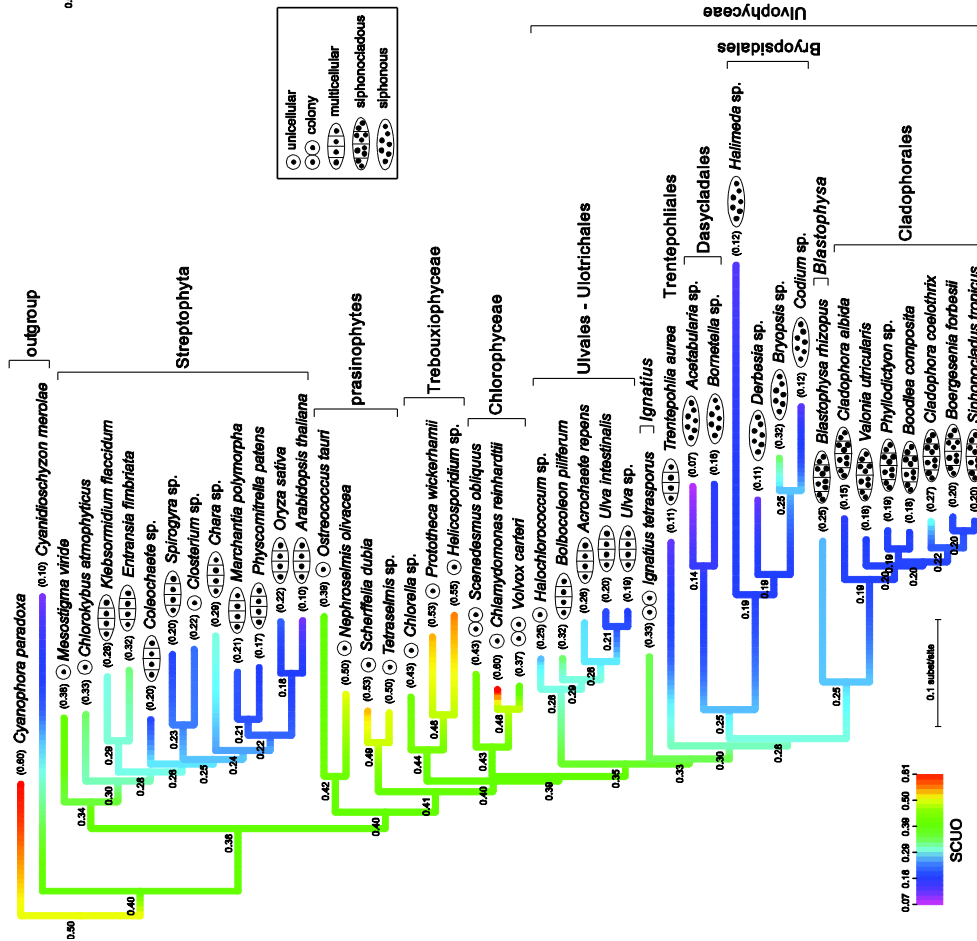
B. GC content

Figure 2. Estimated evolution of synonymous codon usage bias and GC content. Unicells and organisms forming simple and complex colonies (e.g. Ignatius and Volvox, respectively) tend to have more codon usage bias and a higher GC content than macroscopic and multicellular species.

Growth rate and generation time

In rapidly multiplying organisms, the time spent in translation is among the time-limiting factors in cell division (Higgs and Ran 2008). The speed of translation can be improved through the use of preferred or optimal codons accompanied with a higher concentration of the corresponding tRNAs most likely mediated by tRNA gene duplications (Higgs and Ran 2008). In other words, stronger codon usage bias leads to more efficient translation and as such allows for higher growth rates.

Under optimal environmental conditions, unicellular organisms have very high growth rates and short generation time. Since in small, unicellular organisms every division leads to a doubling of the number of individuals, large population sizes are quickly reached. On the other hand, multicellular organisms and macroscopic unicellular algae (e.g. siphonous ulvophytes) make a larger investment in structural components and have lower population sizes. For example, under optimal conditions the unicellular flagellate *Chlamydomonas* divides every 6.7 h (Badour 1981), whereas the siphonous ulvophyte *Acetabularia* needs 115 days to complete its life cycle (Berger and Liddle 2003). These characteristic growth rates and generation times can be compared to r/K strategies in ecology. r-selected species are opportunists characterized by high fecundity, small body sizes, early onset of maturity, short generation times and the ability to disperse offspring widely. K-selected species live in more stable environments and are characterized by large body sizes, long life expectation, and the production of fewer offspring. Generally speaking, unicellular algae fit the r-selection theory, whereas macroscopic algae can be indicated as K-selected species. The strong codon usage bias observed in small unicellular algae can likely be explained by ecological traits such as fast growth rates, short generation time, small body size and large population size all being characteristic for r-selected species.

Rate of molecular evolution

Some Ulvophyceae, i.e. Trentepohliales, Dasycladales, Bryopsidales, Cladophorales and *Blastophysa*, have markedly elevated rates of molecular evolution as indicated by long root to tip path lengths in phylogenetic trees (branch lengths). In other words, these taxa have accumulated mutations at a faster pace than the rest of the green algae, potentially suggesting higher levels of genetic drift in those algae. The amount of codon usage bias depends on the delicate balance between selection for optimal translation and mutations fixed by genetic drift. Selection for speed and accuracy of translation is expected to favor optimal codons over minor codons, leading to codon usage bias. On the other hand, when selection for optimal translation is less stringent, mutation events that are fixed through genetic drift will result in a uniform codon usage. The observed uniform codon usage in some Ulvophyceae is likely due to elevated rates of molecular evolution leading to the fixation of many mutations and pronounced genetic drift. Studies on various other groups of organisms (both eukaryotes and prokaryotes) have also revealed a negative correlation between codon usage bias and rate of synonymous nucleotide substitution (Powell and Moriyama 1997, Kawahara and Imanishi 2007).

Codon usage bias and GC content

In our data, GC content and codon usage bias show congruent trends, species with strong codon usage bias having a high GC content (Fig. 2). Similar correlations have been recovered by other workers (Knight et al. 2001, Chen et al. 2004). The causal direction of this correlation has mostly been interpreted from GC content to codon usage bias. More specifically, it has been suggested that base composition drives codon usage through mutational bias in favor of the common nucleotides (Knight et al. 2001). Knight et al. (2001) conclude that 71-87% of the observed differences in codon usage within and between genomes can be explained by GC content influencing codon usage, the remaining 29-23% being a consequence of selective pressures for optimal translation.

The co-evolutionary model offers a more mechanistic perspective on the correlation between GC content and codon usage bias. More specifically, this model predicts that base composition and codon usage are linked through internal feedback loops. If one assumes that translation efficiency improves with increased usage of G and C at synonymous codon positions, a GC-rich genome will result if selection for translation efficiency is sufficiently strong. Although many authors have reported a positive relationship between GC content and codon usage bias, we are not aware of studies showing empirical evidence for intrinsically more efficient translation of codons that have G and C at synonymous positions. It is possible that the use of G and C improves codon–anticodon recognition because of the extra hydrogen bond that is formed between them, with positive effects on the speed and efficiency of translation. However, this is in contradiction with result from the recently published genomes of two strains of the marine prasinophyte *Micromonas* (Worden et al. 2009). This study revealed that the average GC content is about 65 %, and that both strains contained a region with a GC content 14 % lower than average (comprising about 7% of the total genome). These low GC regions have a higher transcriptional activity and a different codon usage compared to the rest of the genome.

Conclusion and perspectives

The mutation-selection-drift model (Bulmer 1991, Hershberg and Petrov 2008) and the co-evolutionary model of genome composition and resource allocation (Vetsigian and Goldenfeld 2009) form theoretical frameworks that allow predicting how a variety of biological factors can be expected to influence genome biases (Fig. 1). In the green plants studied here, we have been able to show that several biological factors that have previously been associated with genome bias are effectively associated with genome bias. All of these factors may have some explanatory power in the observed evolutionary patterns of codon usage bias and GC content but at present it is difficult to estimate their relative contributions. Incorporation of these and other biological factors (e.g. population size, growth rate, life style) in more detailed models would allow for a more precise estimation about their contribution.

The strong correlation observed between ecological characteristics and genome biases invites studying the mechanisms responsible for this apparent association with sets of more closely related

taxa that allow addressing the questions within the timeframe they act. The green plant lineage offers some useful cases as the members of this lineage occupy many distinct and widely divergent ecological niches, including freshwater, terrestrial and marine ecosystems, and exhibit a remarkable morphological and cytological diversity, ranging from unicells to complex macroscopic and multicellular organisms. Studying closely related organism exhibiting less biological variation but that differ in terms of r/K selection would allow testing the impact of these ecological affinities on genome biases.

Comparison of genome biases among different species is usually performed on EST data (e.g. Cutter et al. 2006, Ingvarsson 2008). Such data are not available for the broad sample of taxa we have used here. Our calculations are based on eight housekeeping genes, implying that the codon usage biases we are observing may be higher than average due to the relatively high expression level one can expect of these genes. However, because we use the same eight genes for comparisons across the phylogeny, we expect that the observed patterns of genome bias will generally be preserved when EST data are compared. Obviously, studying larger datasets would allow for more definitive conclusions but this awaits the generation of EST or whole-genome data for a broader range of green algae, particularly siphonocladous and siphonous Ulvophyceae.

Acknowledgments

6

A multi-locus time-calibrated phylogeny of the siphonous green algae [1]

Abstract

The siphonous green algae are an assemblage of seaweeds that consist of a single giant cell. They comprise two sister orders, the Bryopsidales and Dasycladales. We infer the phylogenetic relationships among the siphonous green algae based on a five-locus data matrix and analyze temporal aspects of their diversification using relaxed molecular clock methods calibrated with the fossil record. The multi-locus approach resolves much of the previous phylogenetic uncertainty, but the radiation of families belonging to the core Halimedineae remains unresolved. In the Bryopsidales, three main clades were inferred, two of which correspond to previously described suborders (Bryopsidineae and Halimedineae) and a third lineage that contains only the limestone-boring genus *Ostreobium*. Relaxed molecular clock models indicate a Neoproterozoic origin of the siphonous green algae and a Paleozoic diversification of the orders into their families. The inferred node ages are used to resolve conflicting hypotheses about species ages in the tropical marine alga *Halimeda*.

Keywords

molecular phylogenetics; relaxed molecular clock; fossil algae; Bryopsidales; Dasycladales; Ulvophyceae; Chlorophyta

---

Introduction

The siphonous green algae form a morphologically diverse group of marine macroalgae. They are readily distinguished from other green algae by their ability to form large, differentiated thalli comprised of a single, giant tubular cell. This tubular cell branches and fuses in various patterns to form a broad range of forms (Fig. 1). Individual branches of the tubular cell are called siphons. The siphonous green algae include two sister orders (Bryopsidales and Dasycladales) and belong to the chlorophytan class Ulvophyceae. The Cladophorales, an order closely related to the Bryopsidales and Dasycladales (Zechman et al. 1990), is sometimes included in the siphonous algae but its members are not truly siphonous because they are generally multicellular. The siphonous green algae are unique among their chlorophytan relatives in having a relatively rich fossil record because many of them biomineralize.
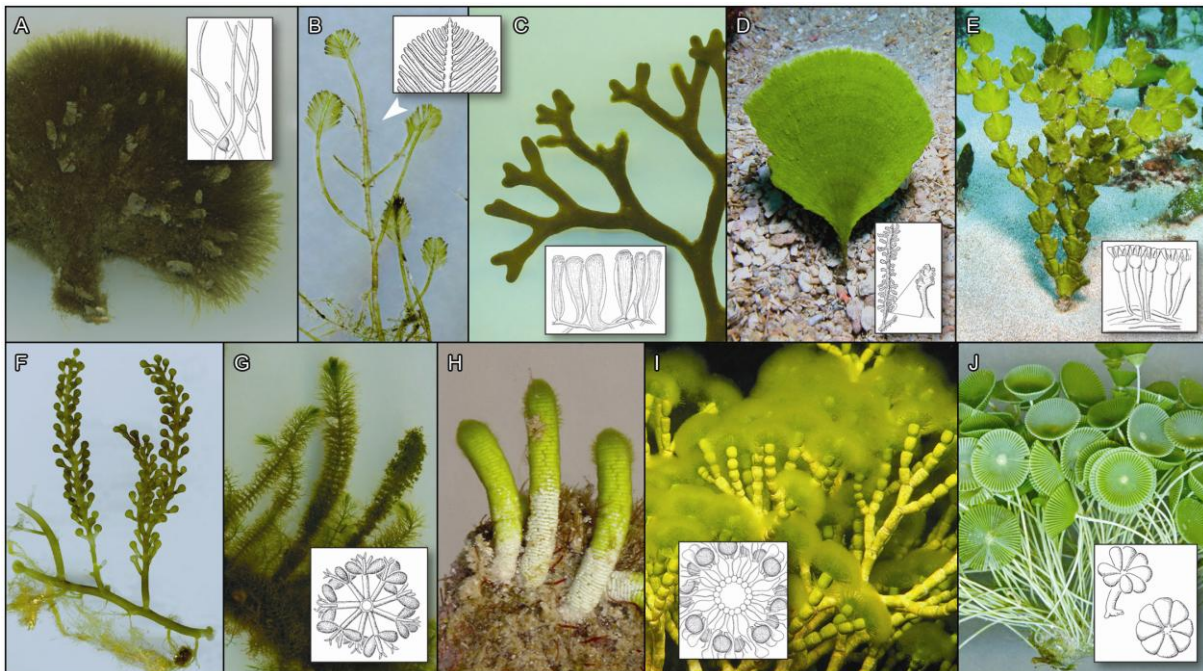
Fig. 1. Morphology and anatomy of the siphonous green algae comprising the orders Bryopsidales (A–F) and Dasycladales (G–J). A. *Derbesia*, B. *Bryopsis*, C. *Codium*, D. *Udotea*, E. *Halimeda*, F. *Caulerpa*, G. *Batophora*, H. *Neomeris*, I. *Cymopolia*, J. *Acetabularia*.

The Bryopsidales range in morphology from simple, branched thalli (Fig. 1A,B) to more complex organizations with interwoven siphons, differentiated thalli and various morpho-ecological specializations (Fig. 1C-F). They have multiple nuclei scattered throughout the siphons. These algae form an important component of the marine flora, particularly in tropical marine environments, where they are among the major primary producers on coral reefs, in lagoons and seagrass beds. They comprise roughly 500 recognized species (Guiry and Guiry 2008). The calcified representatives are major contributors of limestone to coral reef systems and are well-represented in the fossil record (Hillis-Colinvaux 1986; Mu 1990). The Bryopsidales are also common in warm-temperate

marine habitats where they form a significant component of the flora (e.g., *Codium*). Several bryopsidalean taxa are vigorous invasive species (e.g., *Codium fragile*, *Caulerpa taxifolia* and *Caulerpa racemosa* var. *cylindracea*) that are known to have profoundly affected the ecology and native biota in areas of introduction. Molecular phylogenetic studies have shown two principal bryopsidalean lineages, both comprising simple as well as complex forms (Lam and Zechman, 2006; Verbruggen et al., 2009; Woolcott et al., 2000).

Extant Dasycladales are characterized by a central axis surrounded by whorls of lateral branches (Fig. 1G-J). Members of this group contain a single giant nucleus situated in the rhizoid during the vegetative phase of their life cycle. Albeit relatively inconspicuous, they are common algae of shallow tropical and subtropical marine habitats. Both calcified and non-calcified representatives have left a rich fossil record dating back to the Cambrian Period (540–488 my) (Berger and Kaever 1992; LoDuca, Kluessendorf, and Mikulic 2003). Fossil remains suggest that non-calcified dasyclads were most diverse during the Ordovician and Silurian periods and declined in favor of calcified representatives after the Early Devonian (± 400 my), perhaps as a result of selection for resistance to herbivory (LoDuca, Kluessendorf, and Mikulic 2003). In all, over 700 species are known from the fossil record, and fossil dasyclads are important tools for both biostratigraphic and paleoenvironmental studies (Berger and Kaever 1992; Bucur 1999). The present dasycladalean diversity consists of 37 species belonging to 10 genera and the two families Dasycladaceae and Polyphysaceae (Berger 2006). Molecular phylogenetic studies have shown that the Polyphysaceae arose from the Dasycladaceae, leaving the latter paraphyletic (Berger et al., 2003; Olsen et al., 1994; Zechman, 2003).

Currently available phylogenetic studies of the siphonous green algae suffer from a few shortcomings. First, the studies have been based on single loci, yielding partially resolved trees with some unresolved taxa. Second, several important groups within the Bryopsidales have not been included. Finally, the temporal aspects of siphonous green algal diversification have not been explored in sufficient detail. Several recent studies point to the necessity of a time-calibrated phylogenetic framework. For example, the large discrepancy between species ages resulting from interpretations of molecular data and the fossil record creates confusion (Dragastan, Littler, and Littler 2002; Kooistra, Coppejans, and Payri 2002). Furthermore, biogeographic interpretations are difficult without reference to a temporal framework (Verbruggen et al., 2005; Verbruggen et al., 2007). So far, the fossil record has been used on one occasion to calibrate a dasycladalean molecular phylogeny in geological time (Olsen et al. 1994). In the years that have passed since the publication of this study, however, several important discoveries have been made in dasyclad paleontology (e.g. Kenrick and Li, 1998; LoDuca et al., 2003) and more advanced calibration methods have been developed (reviewed by e.g., Verbruggen and Theriot 2008).

The present study sets out to achieve two goals. First, it aims to resolve the phylogeny of the siphonous green algae more fully and include previously omitted taxonomic groups. Second, it aspires to provide a temporal framework of siphonous green algal diversification by calibrating the phylogeny in geological time using information from the fossil record. Our approach consists of model-based phylogenetic analyses of a five-locus alignment spanning both of the orders and uses a composite (partitioned) model of sequence evolution. Calibration of the phylogeny in geological time is achieved with Bayesian implementations of relaxed molecular clock models, using a selection of fossil reference points.

Materials and methods

Data generation

In preparation for this study, the *rbc*L gene of a wide spectrum of taxa was sequenced as described below and additional *rbc*L sequences were downloaded from GenBank. From a neighbor joining guide tree produced from these sequences, 56 taxa representing all major clades were selected. For these taxa, we attempted to amplify four additional loci (plastid encoding *atp*B, *tuf*A and 16S rDNA, and nuclear 18S rDNA) or downloaded this information from GenBank.

DNA extraction followed a CTAB protocol modified from Doyle and Doyle (1987). DNA amplification followed previously published protocols (Famà et al., 2002; Hanyuda et al., 2000; Karol et al., 2001; Kooistra, 2002; Lam and Zechman, 2006; Olson et al., 2005; Wolf, 1997), with additional primers for amplification of partial sequences of the *rbc*L gene and the 16S rDNA. A complete overview of the primers used is given in Fig. 2. Newly generated sequences were submitted to GenBank. A complete list of Genbank accession numbers is provided in Table 1. Our specimens are vouchered in the Ghent University Herbarium or the US National Herbarium (see GenBank records for voucher information). The ulvophycean algae *Trentepohlia aurea*, *Oltmannsiellopsis viridis*, *Ulva intestinalis*, *Ulothrix zonata* and *Pseudendoclonium akinetum* were used as outgroup taxa (Lopez-Bautista and Chapman 2003). Despite the fact that the order Cladophorales is the closest relative of Bryopsidales and Dasycladales (Zechman et al. 1990), we did not use it as an outgroup because it has proven impossible to amplify chloroplast DNA in its members using standard primer combinations.

Sequence alignment

After removal of introns in the *rbc*L gene (Hanyuda, Arai, and Ueda 2000), all sequences were of equal length and their alignment was unambiguous. The *atp*B sequences were of equal length and their alignment was straightforward. The *tuf*A gene was aligned by eye on the basis of the corresponding amino-acid sequences and ambiguous regions were removed. After introns had been removed, the 16S rDNA sequences were aligned using Muscle v. 3.6 using standard parameters (Edgar 2004). Ambiguous alignment regions were localized by eye and removed. Alignment of 18S rDNA sequences followed an analogous procedure. The insert near the 3' terminus of the 18S gene of certain Bryopsidales (Durand et al., 2002; Hillis et al., 1998; Kooistra et al., 1999) was removed because it was not present in all taxa and virtually impossible to align among genera. After removal of this region, sequences were aligned with Muscle v. 3.6 and stripped of their ambiguous regions. The concatenated alignment used for analysis is available through TreeBase and www.phycoweb.net.

Partitioning and model selection

Selection of a suitable partitioning strategy and suitable models for the various partitions used the Bayesian Information Criterion (BIC) as a selection criterion (e.g., Sullivan and Joyce 2005). The guide tree used during the entire procedure was obtained by maximum likelihood (ML) analysis of the concatenated data with a JC69 model with rate variation among sites following a discrete gamma

distribution with 8 categories (JC69+Γ_8) inferred with PhyML (Guindon and Gascuel 2003). All subsequent likelihood optimizations and BIC calculations were carried out with TreeFinder (Jobb, von Haeseler, and Strimmer 2004). Six alternative partitioning strategies were considered (see results). Per partitioning strategy, six substitution models were optimized with unlinked parameters between partitions and a partition rate multiplier (see results). The partitioning strategy + model combination receiving the lowest BIC score was used in the phylogenetic analyses documented below.

*rbc*L: ribulose-1,5-bisphosphate carboxylase/oxygenase - large subunit

numbering corresponds to CDS of *Codium fragile* sequence M67453

*tuf*A: elongation factor Tu

numbering corresponds to CDS of *Ostreococcus tauri* sequence CR954199

*atp*B: ATP synthase beta subunit

numbering corresponds to CDS of *Mesostigma viride* sequence AF166114

16S: plastid 16S (small subunit) ribosomal DNA

numbering corresponds to rRNA-coding parts of *Caulerpa sertularioides* sequence AY389514

18S: nuclear 18S (small subunit) ribosomal DNA

rbcL	
7F	5' CCA MAA ACW GAA ACW AAA GC 3'
436F	5' AAA ACW TTY CAA GGI CCI CC 3'
530R	5' TTW GGT TTA ATW GTA CAR CC 3'
712F	5' CAT TAY TTA AAT GCW ACW GC 3'
791R	5' GGN AYA CCN AAW TCT TTI GC 3'
808F	5' TTA ACT GGW GGT TKK ACI GC 3'
893R	5' TGC ATK GCA CGR TGI ATR TG 3'
904R	5' CAA TAA CMG CRT GCA TAG C 3'
1391R	5' TCT TTC AAA ACT TCA CAA GC 3'

tufA	
1F	5' ATG GYW CGW GAA AAR TTT 3'
142F	5' ATT GAY TCW GCA CCT GAA 3'
F	5' TGA AAC AGA AMA WCG TCA TTA TGC 3'
R	5' CCT TCN CGA ATM GCR AAW CGC 3'

atpB	
147F	5' TGT TAC TTG TGA AGT TCA ACA 3'
242F	5' TTG ATA CGG GAG CYC CTC TWA GTG T 3'
1162R	5' TGT AAC GYT GYA AAG TTT GCT TAA 3'
1396R	5' CTA AAT AAA ATG CTT GTT CAG G 3'

16S	
F16S-A	5' GGA GAG TTT GAT CCT GGC TC 3'
F16S-B	5' CCA GTG TAG CGG TGA AAT GC 3'
730F	5' GCG AAA GCG CTC TRC TGG G 3'
830R	5' AAT CCT ATT GGC TCC CYT AGC TTT CGT 3'
R16S-C	5' CGT GCA ACA TCT AGT ATC 3'
R16S-E	5' GGA GGT GAT CCA GCC GCA CC 3'

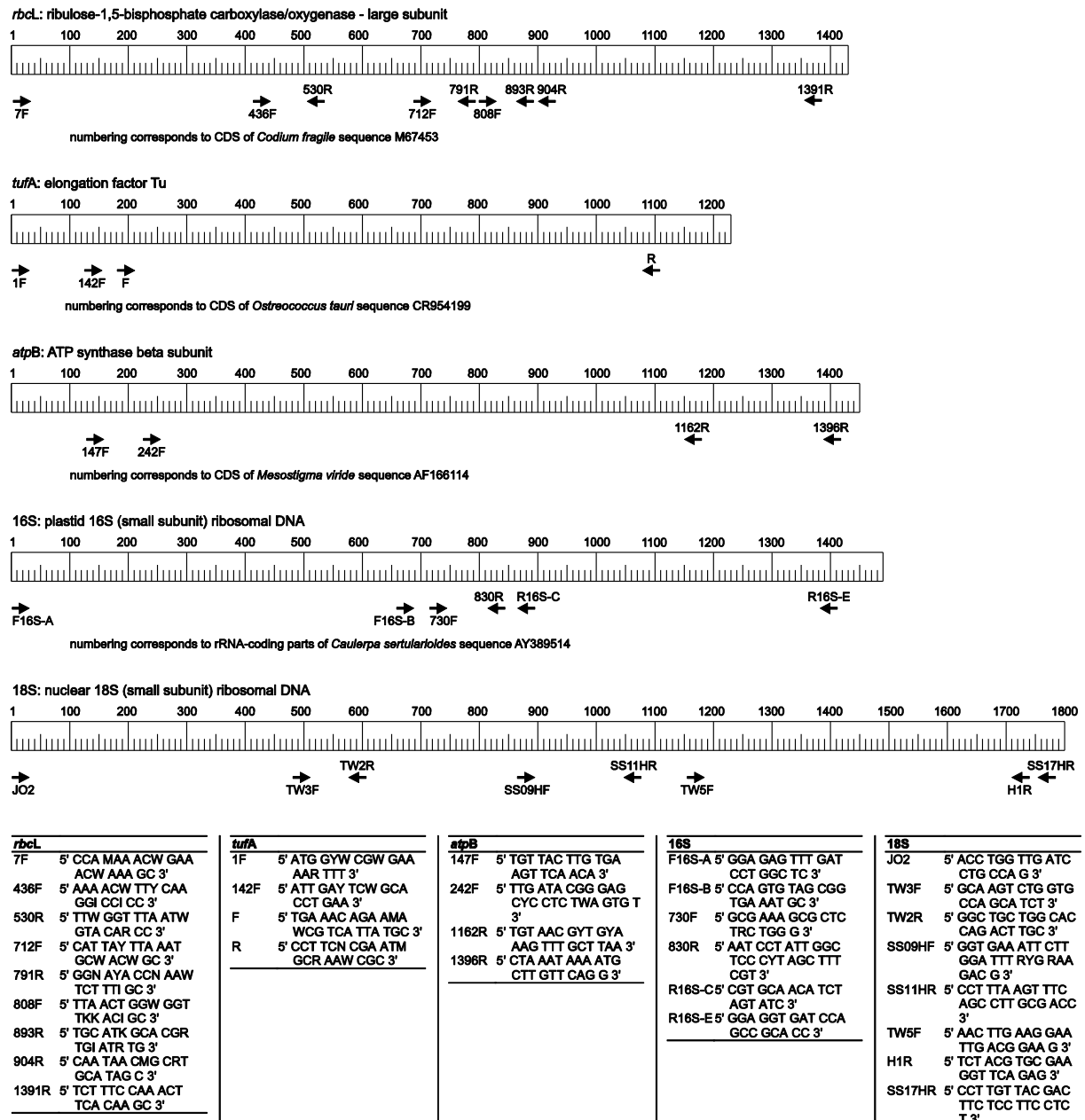| **18S**	
JO2	5' ACC TGG TTG ATC CTG CCA G 3'
TW3F	5' GCA AGT CTG GTG CCA GCA TCT 3'
TW2R	5' GGC TGC TGG CAC CAG ACT TGC 3'
SS09HF	5' GGT GAA ATT CTT GGA TTT RYG RAA GAC G 3'
SS11HR	5' CCT TTA AGT TTC AGC CTT GCG ACC 3'
TW5F	5' AAC TTG AAG GAA TTG ACG GAA G 3'
H1R	5' TCT ACG TGC GAA GGT TCA GAG 3'
SS17HR	5' CCT TGT TAC GAC TTC TCC TTC CTC T 3'

Fig. 2. Primers used for amplification and/or sequencing of the five loci used in this study.

Phylogenetic analyses

Phylogenetic analyses consisted of Bayesian inference (BI) and maximum likelihood tree searches using the unrooted partitioned model selected using the BIC procedure (section 2.3). For Bayesian inference, three independent runs, each consisting of four incrementally heated, Metropolis-coupled chains were run for ten million generations using MrBayes v. 3.1.2 (Ronquist and Huelsenbeck 2003). The default heating factor, priors, proposal mechanisms and other settings were used. Rate heterogeneity among partitions was modeled by using a variable rate prior. Parameter values and trees were saved every thousand generations. Convergence and stationarity of the runs was assessed using Tracer v. 1.4 (Rambaut and Drummond 2007). An appropriate burn-in value was determined with the automated method proposed by Beiko et al. (2006). Their method was applied to each run individually, with a sliding window of 100 samples. The post-burnin trees from different runs were concatenated and summarized using MrBayes' sumt command.

Maximum likelihood tree searches were carried out with TreeFinder, which allows likelihood tree searches under partitioned models (Jobb, von Haeseler, and Strimmer 2004). Tree space coverage in the TreeFinder program is low compared to other ML programs. Therefore, an analysis pipeline was created to increase tree space coverage by running analyses from many start trees. First, a set of start trees was created by randomly modifying the guide tree used for model selection by 100 and 200 nearest neighbor interchange (NNI) steps (50 replicates each). Analyses were run from these 100 start trees. The 3 trees yielding the highest likelihood were used as the starting point for another set of NNI modifications of 20 and 50 steps (50 replicates each). A second set of ML searches was run from each of the resulting 300 start trees. The ML tree resulting from this set of analyses was retained as the overall ML solution. The same partitions and models as in the BI analysis were used, with the second-level tree search and proportional partition rates. Branch support was calculated using the bootstrap resampling method (1000 pseudo-replicates). Bootstrap analyses used the same settings but started from the ML tree.

Relaxed molecular clock

Likelihood ratio tests significantly rejected a strict (uniform) molecular clock for the alignment. Node age estimates were therefore obtained by fitting a relaxed clock model to our molecular data. The assumption of the molecular clock was relaxed by allowing rates of molecular evolution to vary along the tree according to a Brownian motion model (Kishino et al., 2001; Thorne et al., 1998). First, an F84 model with rate variation across sites following a discrete gamma distribution with 20 rate categories was optimized in PAML v.4 (Yang 2007), using the topology obtained with Bayesian inference from which all but two outgroup taxa were removed (*Trentepohlia* and *Oltmannsiellopsis*). Second, a variance-covariance matrix was built with estbNew from the T3 package (Thorne et al. 1998; http://abacus.gene.ucl.ac.uk/software.html). Finally, node age estimates were obtained by running three independent MCMC chains with the PhyloBayes program (Lartillot, Blanquart, and Lepage 2007) on the variance-covariance matrix, using the lognormal auto-correlated clock model. Sensitivity analyses were carried out to evaluate the impact of some potentially erroneous fossil assignments with chains of 50,000 generations that were sampled every 100th generation. The final analysis consisted of three independent runs of 1,000,000 generations, sampled every 100th

generation. Convergence and stationarity of the chains was evaluated by plotting trace files in Tracer v. 1.4 (Rambaut and Drummond 2007). The fossils used as calibration points are documented in the results section.

Results

Data exploration

The data compiled for our analyses consisted of 155 sequences. The resulting five-locus data matrix was 51% filled. The *rbc*L gene was best represented (90% filled), followed by 18S rDNA (62%), *tuf*A (49%), 16S rDNA (33%) and *atp*B (20%). After ambiguously aligned parts had been removed, the matrix measured 61 taxa by 5588 characters. The 18S rDNA provided most characters (1555 bp), followed by *rbc*L (1386 bp), 16S rDNA (1327 bp), *tuf*A (804 bp) and *atp*B (516 bp).

The BIC-based model selection procedure selected a model with four partitions and GTR+Γ_8 substitution models for each partition (Table 2). The four partitions were: (1) ribosomal loci: 18S and 16S, (2) first codon positions of *rbc*L, *tuf*A and *atp*B combined, (3) second codon positions of the three genes, and (4) third codon positions of the three genes.

Phylogeny of the siphonous algae

Different runs of the Bayesian phylogenetic analysis of the five-locus dataset under an unrooted partitioned model rapidly converged onto highly similar posterior distributions. The chains reached convergence after at most 384,000 generations. The analyses resulted in a well-resolved phylogenetic hypothesis of the siphonous green algae (Fig. 3A). The backbone of the dasycladalean clade was determined with high confidence except for the branching order between the *Bornetella* lineage, the *Cymopolia-Neomeris* clade and the Polyphysaceae. Relationships among *Acetabularia* species are not resolved with our dataset. In the Bryopsidales, three main clades were inferred, two of which correspond to previously described suborders (Bryopsidineae and Halimedineae). A third lineage contained only the limestone-boring genus *Ostreobium*. Relationships among the three families of the Bryopsidineae were inferred with high confidence (Fig. 3B) whereas within the Halimedineae, relationships among what we will refer to as the "core Halimedineae" (families Caulerpaceae, Rhipiliaceae, Halimedaceae, Pseudocodiaceae and Udoteaceae) remained poorly resolved (Fig. 3C).
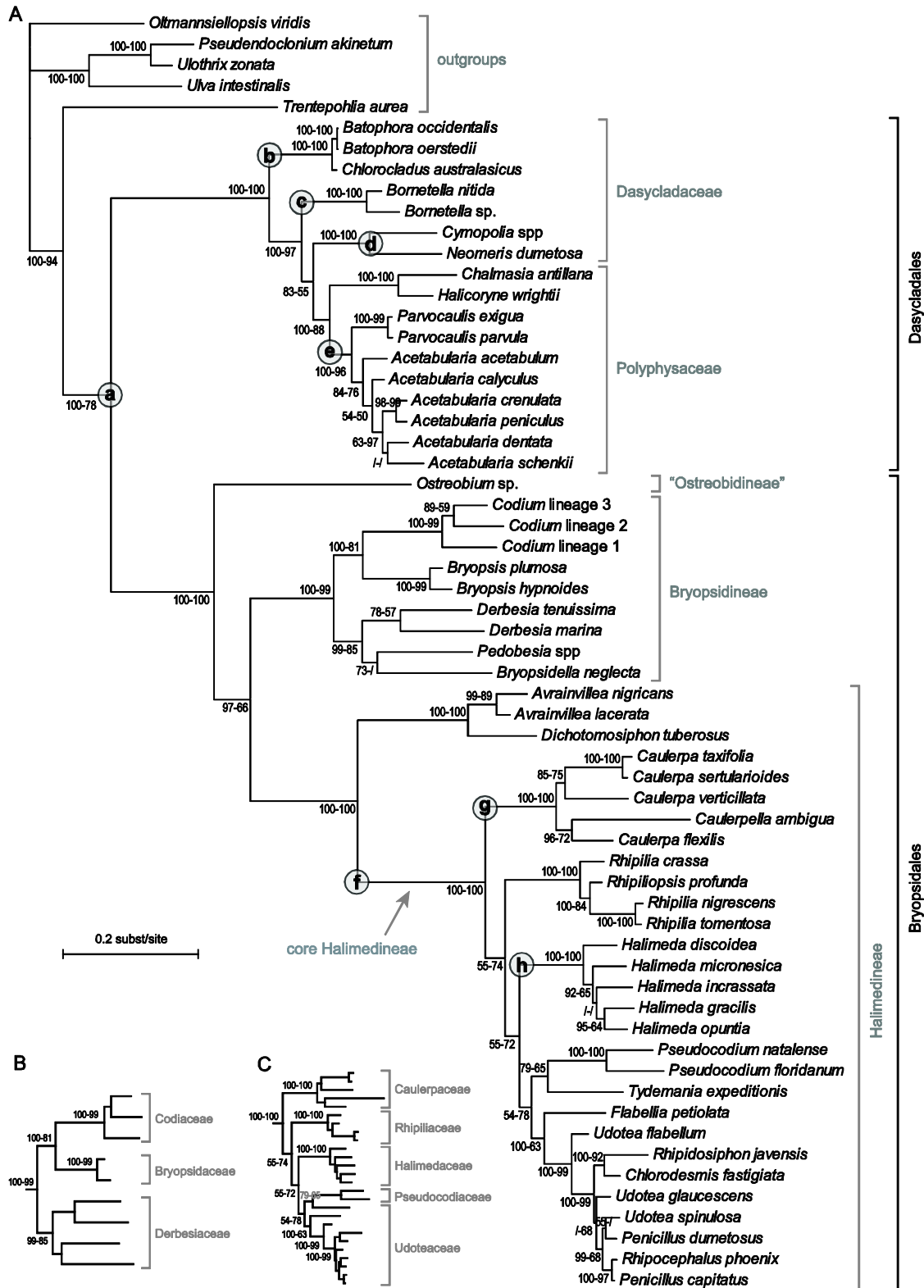
Fig 3. Phylogenetic relationships among the siphonous green algae inferred from a five-locus DNA alignment using Bayesian analysis under a partitioned, unrooted model. A. Majority-rule consensus phylogram of post-burnin trees. B. Relationships among families of the Bryopsidineae. C. Relationships among families of the core Halimedineae. Numbers at nodes indicate statistical support, posterior probabilities before the dash and ML bootstrap proportions after (both given as percentages). Encircled letters indicate calibration points (Table 3). The scale bar only applies to panel A.

Time-calibrated phylogeny

We compiled a list of fossils that are thought to represent early records of lineages observed in our phylogenetic tree (Table 3). Prior to carrying out the main calibration analysis, we ran several analyses to test the sensitivity of results to the choice of certain fossils as calibration points and of the maximum age constraint imposed on the root of the siphonous algae. The exact combination of constraints used in these trial runs can be found in the online appendix.

Analyses with *Dimorphosiphon* (min. 460.9 my) constrained at node f resulted in considerably older node age estimates throughout the tree than analyses in which this constraint was not imposed. *Dimorphosiphon* has been regarded as an ancestral taxon of the Halimedineae (Dragastan, Littler, and Littler 2002) but its combination of characters did not allow unambiguous placement on any node of the tree. From the incompatibility of this calibration point with the others, we concluded that our admittedly speculative assignment of this taxon to node f was unjustified. With *Dimorphosiphon* excluded from analyses, various combinations of the remaining calibration points led to very similar results.

The analysis was sensitive to the maximum age constraint imposed on the phylogeny. By default, we used a maximum age constraint of 635 my on the root of the siphonous algae (node a) because there is no clear evidence for either Bryopsidales or Dasycladales in Ediacaran Konservat-Lagerstätten, in which macroalgae are abundant and well preserved (Xiao et al., 2002; Zhang et al., 1998). These deposits do, however, contain simple cylindrical and spherical forms (e.g., *Sinospongia*, *Beltanelliformis*) that may be ancestral to these extant lineages (Xiao et al. 2002). Changing the maximum age constraint to 800 my as a sensitivity experiment increased average node ages and widened their 95% confidence intervals. Analogously, lowering the maximum age constraint to 500 my decreased the average ages and narrowed the confidence intervals. The chronogram presented in Fig. 4 resulted from the analysis without *Dimorphosiphon* and with a maximum age constraint of 635 my; more specifically, the analysis was carried out with calibration points a1, a2, b1, c, d2, e1 and h1 (Table 3). Repeating this analysis without age constraint on node b resulted in a chronogram that was very similar.

As an alternative strategy to using a maximum age constraint on the root of the siphonous algae (node a), we constrained the age of node b to be Early Devonian (416.0–397.5 my) based on the first occurrence of fossils with thalli comparable to those of Dasycladaceae, in terms of both general thallus form and reproductive structures, in strata of this age (*Uncatoella*: Kenrick and Li, 1998). This analysis, which is illustrated in the online appendix, should be regarded as a conservative alternative to the analysis presented in Fig. 4 and node ages in this chronogram should be interpreted as minimum values rather than realistic estimates. We will use the chronogram in Fig. 4 as the preferred reference time frame for the remainder of the paper because the maximum age constraint used for this analysis has an empirical basis (although it is based on absence of evidence rather than evidence of absence) and because we consider the intermediate-sized confidence intervals on the node ages of this tree to yield a fairly realistic image of the true uncertainty surrounding the node ages. In what follows, we report node age estimates as the mean node age followed by the 95% confidence interval in square brackets.

The origin of the orders Dasycladales and Bryopsidales was inferred to be in the Neoproterozoic (571 my [628–510]). The earliest divergence between extant Dasycladaceae lineages occurred during the Ordovician or Silurian (458 my [517–407]) and the family diversified into its main extant lineages during the Devonian. The Polyphysaceae originated from the Dasycladaceae during the second half of the Paleozoic (367 my [435–306]). In the Bryopsidales, the suborders Bryopsidineae and Halimedineae diverged from each other in the Early Paleozoic (456 my [511–405]) and diversified into their component families during the second half of the Paleozoic. The families belonging to the core Halimedineae appear to have diverged from one another during the Permian.

Discussion

Taxonomic implications

The increased sampling of taxa and loci compared to previous studies has produced some results that deserve mention from a taxonomic viewpoint. First, the lineage leading to the limestone-boring genus *Ostreobium* seems to deserve recognition at the suborder level, hence the tentative clade name Ostreobidineae. The bryopsidalean family Udoteaceae disintegrates. Besides the transfer of *Avrainvillea* and *Cladocephalus* to the Dichotomosiphonaceae, which was previously indicated by Curtis et al. (2008), a number of *Rhipilia* and *Rhipiliopsis* species form a lineage of their own and the remainder of the Udoteaceae receives little statistical phylogenetic support. We have applied the name Rhipiliaceae to the clade of *Rhipilia* and *Rhipiliopsis* species. This family name was proposed earlier but its phylogenetic relevance had not yet been demonstrated (Dragastan et al. 1997; Dragastan and Richter 1999). The same authors proposed the family Avrainvilleaceae to harbor the extant genera *Avrainvillea* and *Cladocephalus* and some fossil genera. However, the ultrastructural affinities between *Dichotomosiphon*, *Avrainvillea* and *Cladocephalus* (Curtis, Dawes, and Pierce 2008) and the limited DNA sequence divergence between *Dichotomosiphon* and *Avrainvillea* shown here suggest that transferring *Avrainvillea* and *Cladocephalus* to the Dichotomosiphonaceae would be more appropriate. Although the fossil taxa in the Avrainvilleaceae are more difficult to evaluate because no ultrastructural evidence is available to link them unequivocally to the Dichotomosiphonaceae, we nevertheless propose their transfer to this family for taxonomic convenience. We refrain from formally describing a suborder Ostreobidineae for *Ostreobium* because we feel that the description of such high-level taxa should be accompanied by detailed ultrastructural observations.

Our results for the phylogeny of the Dasycladales are consistent with previous studies and do not add much insight into their pattern of diversification. One thing worth mentioning is that the close relationship between *Batophora* and *Chlorocladus*, which was suggested by previous molecular phylogenetic studies (Olsen et al. 1994; Zechman 2003), is confirmed by our multi-locus dataset. As reported in the aforementioned papers, this finding contradicts the traditional tribe-level classification by Berger and Kaever (1992).

Fig. 4. Chronogram of the siphonous green algae. Node ages were inferred using Bayesian inference assuming a relaxed molecular clock and a set of node age constraints derived from the fossil record. Values at nodes indicate average node ages and bars represent 95% confidence intervals. The calibration points used for this analysis are a1, a2, b1, c, d2, e1 and h1 (Table 3).

Dasycladalean diversification

Of the five dasycladalean families, only Dasycladaceae and Polyphysaceae have recent representatives and Seletonellaceae, Triploporellaceae and Diploporellaceae are entirely extinct (Berger and Kaever 1992). The Seletonellaceae includes the oldest fossils assigned to Dasycladales, *Yakutina aciculata* from the Middle Cambrian (Kordé 1973; Riding 1994; Riding 2001) and *Seletonella mira* from the Upper Cambrian (Kordé 1950a; Riding 1994; Riding 2001). *Cambroporella* from the Lower Cambrian of Tuva, was initially described as a dasyclad (Kordé 1950b) but was reinterpreted as a probable bryozoan (Elias 1954). Unlike living dasyclads, Seletonellaceae contained gametes within the main axis (endospore reproduction) as opposed to within gametophores (sensu Dumais and Harrison 2000), and developed laterals irregularly along the length of the main axis (aspondyl structure), rather than in whorls as is the case in living representatives (euspondyl structure). Nonetheless, two lines of evidence support a dasyclad affinity for the Seletonellaceae. First, in terms of reproduction, development of reproductive cysts within the main axis is known as a teratology among living dasyclads (Valet 1968) and all living dasyclads pass through an "endospore stage" as swarms of haploid secondary nuclei generated by meiosis of the large primary nucleus in the rhizoid migrate upward through the main axis towards the gametophores (Elliott 1989; Berger and Kaever 1992). Second, an aspondyl structure is a predicted outcome of a detailed reaction-diffusion model of dasyclad whorl morphogenesis (Dumais and Harrison 2000).

Our chronogram suggests a late Neoproterozoic origin for the Dasycladales, with the 95% confidence interval ranging into the Cambrian: 571 my [628–510] (Fig. 4). The upper boundary on this node age (628 my) should not be overinterpreted because it is largely determined by the upper age constraint on this node (635 my). The Middle Cambrian lower boundary on this node age (510 my), however, can be taken as a fairly safe minimum age estimate for the Dasycladales as well as the Bryopsidales. Unlike *Yakutina* and *Seletonella*, most early Paleozoic dasyclad taxa did not biomineralize (LoDuca, Kluessendorf, and Mikulic 2003). This fact, when taken together with the results of the present analysis, suggests that noncalcified dasyclads may yet be discovered within Konservat-Lagerstätten of latest Neoproterozoic or Early Cambrian age.

Taxa assigned to Dasycladales are both abundant and diverse in Ordovician strata, and include euspondyl (e.g., *Chaetocladus*) as well as aspondyl (e.g., *Dasyporella*) forms (Berger and Kaever, 1992; LoDuca, 1997; LoDuca et al., 2003). The earliest euspondyl forms, like aspondyl taxa, were endospore. Later, however, at least one lineage of euspondyl forms emerged in which gametogenesis occurred within the laterals (cladospore reproduction). Collectively, endospore and cladospore taxa with a euspondyl structure comprise the family Triploporellaceae (Berger and Kaever 1992; LoDuca 1997). Triploporellaceae are though to have originated from the Seletonellaceae, indicating that the latter is a paraphyletic group.

The family Dasycladaceae is characterized by euspondyl thalli with spherical gametophores along the sides or tips of laterals (choristospore reproduction). It is though to have originated from the Triploporellaceae, in which case the latter group, like Seletonellaceae, is paraphyletic. Studies indicate that the dasycladacean gametophore, rather than being a modified lateral, is instead a separate structure with its own morphogenetic identity (Dumais and Harrison 2000). The oldest taxon with gametophores and associated reproductive cysts comparable to those of living Dasycladaceae is *Uncatoella verticillata* from the Lower Devonian (416–397 my) of China (Kenrick

and Li 1998). This taxon is somewhat problematic, however, in that the asymmetrical manner of lateral branching and instances of pseudodichotomies of the main axis are otherwise unknown among dasyclads (Kenrick and Li 1998). *Uncatoella* is the only taxon known from Paleozoic strata with gametophores and associated reproductive cysts comparable to those of living Dasycladaceae.

Our molecular results suggest that the earliest divergence between the extant Dasycladaceae lineages occurred during the Ordovician or Silurian (458 my [517-407]). Notably, this result emerges regardless of whether *Uncatoella* is used to constrain the minimum age of node b. *Archaeobatophora*, from the Upper Ordovician of Michigan (Nitecki 1976), is of interest in this regard, as it bears a striking resemblance to the vegetative thallus of the extant dasycladacean *Batophora*. However, gametophores are not present among the few known specimens (all on a single small slab), and thus the status of this form as an early dasycladacean remains uncertain. Our results suggest that the main extant lineages belonging to the Dasycladaceae originated during the Devonian. This is somewhat older than expected because, with the possible exception of the Batophoreae (*Archaeobatophora*), the oldest fossils assigned to extant tribes within the Dasycladaceae are from the Mesozoic (Berger and Kaever 1992). Similarly, the age estimate for the split between *Cymopolia* and *Neomeris*, two calcified genera known from numerous occurrences in the fossil record, is somewhat older than anticipated: 211 my [300–152] vs. a Barremian age for the earliest *Neomeris* fossil (130–125 my) (Sotak and Misik 1993). Two factors could explain these discrepancies. First, it is possible that many early Dasycladaceae did not biomineralize. This is certainly true for most early Triploporellaceae and Selenonellaceae (LoDuca, Kluessendorf, and Mikulic 2003), and applies to *Uncatoella* and several living Dasycladaceae as well (e.g., *Batophora*). The lack of biomineralization would have severely limited the preservation potential of these taxa. Second, it may be that some Paleozoic and early Mesozoic calcified taxa previously inferred on the basis of lateral morphology to have been endospore or cladospore were in fact choristospore. A choristospore condition for such forms could be masked if calcification was restricted to the outermost part of the thallus, beyond the level of the gametophores, as is known for the living taxon *Bornetella*.

The family Polyphysaceae is characterized by the development of clavate gametophores arranged in whorls (Berger and Kaever 1992; Berger et al. 2003). The distinctive gametophores of this group appear to have originated through modification of the spherical gametophores of Dasycladaceae (Dumais and Harrison 2000). Our results suggest that the Polyphysaceae originated and began their diversification in the second half of the Paleozoic. An Early Carboniferous origin of the family had been suggested based on the fossil taxon *Masloviporella* (Deloffre 1988; Berger and Kaever 1992). *Eoclypeina* and *Clypeina* are taxa classified as polyphysaceans from the Permian and Triassic, respectively (Berger and Kaever 1992). The extant polyphysacean *Halicoryne* has been reported from the Triassic (Misik, 1987; Tomasovych, 2004). However, because these reports are based solely on isolated reproductive elements, both the genus- and family-level assignment of this material must be regarded as tentative (Barattolo and Romano 2005).

According to the molecular clock results, the polyphysacean taxon *Acetabularia* originated as early as the early Mesozoic. The oldest *Acetabularia* fossils, however, are from the Oligocene (Berger and Kaever 1992). Here, too, this discrepancy may reflect poor representation of the group in the fossil record, as living members of the genus are only very weakly calcified. Material from Mesozoic-age

strata has been assigned to the closely related extant taxon *Acicularia* (Iannace, Radiocic, and Zamparelli 1998). As is the case for reports of *Halicoryne* from the Triassic, however, such an assignment must be regarded as equivocal because the material at hand comprises only isolated reproductive elements (Barattolo and Romano 2005).

Overall, the results of the relaxed molecular clock analysis line up rather well against major aspects of dasyclad evolution inferred from the dasyclad fossil record. Nonetheless, our results indicate that the fossil record of certain aspects of the evolutionary history of the Dasycladaceae may be more incomplete than previously suspected (see Kenrick and Li 1998). The node age estimates obtained in our relaxed molecular clock analysis are older than those recovered by Olsen et al. (1994). Causes for this discrepancy are difficult to pinpoint because differences between the studies are manifold: the molecular clock method, calibration technique, molecular dataset and selection of fossils all differ. Whereas the previous study relied exclusively on 18S rDNA sequences, our dataset includes several additional loci, which should yield more reliable results (Magallón and Sanderson 2005). Furthermore, at the time the previous study was published, only uniform (strict) molecular clock methods were available. Based on our alignment, 18S rDNA violates the uniform molecular clock (LRT: $-2 \Delta \ln L = 399.403$, $p < 0.0001$), a statement also true for the other loci in our dataset. As a consequence, node age estimates from our analysis with relaxed molecular clock models should match the true divergence times more closely. The calibration method differs in that fossils are used as minimum age estimates for stem groups in our analyses, whereas they were used as point estimates to determine the rate of 18S rDNA evolution in Olsen et al. (1994). Finally, we use a more extensive list of fossil calibration points, including some recently described fossils that represent older occurrences of some dasyclad clades. Several of the aspects mentioned above would predict our age estimates to be older than those of Olsen et al. (1994), which is in agreement with the empirical results.

Bryopsidalean diversification

The fossil record of the Bryopsidales is not as well characterized as that of the Dasycladales. Many fossil taxa that are generally considered to belong to the order have been described (reviewed by Bassoullet et al., 1983; Dragastan et al., 1997; Dragastan and Schlagintweit, 2005), but the taxonomic placement of these taxa is often ambiguous (Mu 1990). Dragastan and Schlagintweit (2005) presented an evolutionary scenario of bryopsidalean diversification based on their interpretation of calcified fossil taxa. They proposed a temporal succession of three main lineages. The primitive Dimorphosiphonaceae, characterized by a single medullar siphon and a simple cortex, originated in the Neoproterozoic or early Paleozoic and persisted through the Devonian (± 360 my). The Protohalimedaceae, characterized by multiple medullar siphons and a cortex of variable complexity, diverged from the Dimorphosiphonaceae in the early Silurian (± 440 my) and thrived up to the PT-boundary (± 250 my), when it suffered major losses, and continued into the Mesozoic to go extinct during the Cretaceous. The extant family Halimedaceae, which features multiple medullar siphons and a complex cortex, was thought to have diverged from the Protohalimedaceae in the second half of the Permian (270–250 my) and diversified through the Mesozoic and Cenozoic.

In our opinion, relationships between fossil taxa and lineages in the phylogeny of extant taxa are not evident. This uncertainty is reflected in our study by the lower number of fossil calibration points used within the Bryopsidales. Our time-calibrated phylogeny indicates that after the initial diversification of the order into its suborders during the early Paleozoic, current families originated in the second half of the Paleozoic. It also suggests that calcification is a relatively recent phenomenon in the extant lineages of the Bryopsidales because the Halimedaceae and Udoteaceae, the only extant families with calcified, corticated representatives, originated during the Permian (± 300–250 my) and diversified during the Mesozoic (Fig. 4). All other lineages of the Halimedineae, including some recently discovered lineages (Verbruggen et al. 2009), are not calcified. As a consequence, the presence of the calcified, corticated families Dimorphosiphonaceae and Protohalimedaceae in older deposits is difficult to interpret in the context of our phylogeny. Based on our results, the classical paleontological interpretation that the Dimorphosiphonaceae and Protohalimedaceae are direct ancestral forms of the serial-segmented Halimedaceae seems doubtful. First, our phylogenetic results show no indication of the presence of calcification in Bryopsidales prior to the Permian. Second, it follows from our phylogeny that the internal architecture of thalli was relatively simple up until the late Paleozoic. Thalli consisting of a medulla and a cortex appear to have evolved from simple, siphonous thalli several times independently during the Permian–Triassic period. They evolved once in the Bryopsidineae (*Codium*), a second time in the Dichotomosiphonaceae (*Avrainvillea*) and a third time in the core Halimedineae (Halimedaceae, Udoteaceae, Pseudocodiaceae).

Some alternative hypotheses may be posited to explain this disparity of results, all of which should be regarded as speculative. Assuming that the Dimorphosiphonaceae and Protohalimedaceae are genuine Bryopsidales, these two families could very well represent an early-diverging bryopsidalean lineage that went extinct. Alternatively, they could represent a collection of taxa that branched off at various places along the lineage leading from the origin of the Halimedineae to the Halimedaceae and Udoteaceae.

Perspectives

The genera *Codium* and *Halimeda* have been proposed as model systems for studying marine algal speciation, biogeography and macroevolution (Kooistra et al., 2002; Verbruggen et al., 2007). Both these genera are species-rich, ecologically diverse and geographically widespread, making them ideal case studies for a spectrum of evolutionary questions. Furthermore, these genera contrast in their climatic preferences, *Halimeda* being mostly tropical and *Codium* being more diverse in temperate seas. Further development of these case studies will greatly benefit from the temporal framework provided here. Both genera diverged from their respective sister lineages in the late Paleozoic (*Halimeda*: 273 my [327–223]; *Codium*: 307 my [370–245]). The most recent common ancestor of the extant species in both genera are highly comparable and can be situated in the Late Jurassic – Early Cretaceous (*Halimeda*: 147 my [198–103]; *Codium*: 153 my [217–97]). To get an initial idea of the time-frame of evolutionary diversification we have included representatives of each of the five sections of *Halimeda* (Verbruggen and Kooistra 2004) and each of the three major lineages of *Codium* (Verbruggen et al. 2007). It follows from our results that both genera spawned their major extant lineages during the Cretaceous. This close match between the timeframes of evolutionary diversification of both genera is convenient for comparative studies between them.

One advantage of using the calcified genus *Halimeda* as a model for marine evolution is that phylogenetic results can be contrasted with its extensive fossil record. An initial comparison of our results to interpretations of the fossil record marks a different timeframe of diversification: whereas the fossil record suggests diversification of the genus into its main lineages (taxonomic sections) in the Eocene (roughly 56–34 my) (Dragastan and Herbig 2007), our results indicate a considerably older, Cretaceous divergence (between approximately 147 and 97 my). The contradiction of these results cannot be explained at present, but three hypotheses are suggested. First, a deviation of the true molecular evolutionary process from the relaxed molecular clock model used here could cause the contradiction. Second, misinterpretation of the fossil record used to calibrate our tree may be at the basis. Third, a significant gap could be present in the fossil record. To distinguish between these alternative scenarios, additional paleontological as well as molecular phylogenetic research is needed. Paleontologists should document additional time-series in the fossil record, especially through the Cretaceous and Paleogene systems (145–23 my). Molecular systematists should evaluate the fit of different clock relaxation models to the molecular data (e.g., Lepage et al. 2007) and take different calibration approaches, either top-down (as was done here) or bottom-up (from calibration points within the genus).

The literature on the genus *Halimeda* is marked by a contradiction of species ages inferred from paleontological and molecular phylogenetic data. In the paleontological literature, extant species are commonly reported from Miocene (> 5 my) and Paleogene deposits (> 23 my), and some are even thought to date back to the Triassic (> 200 my) (Dragastan, Littler, and Littler 2002; Dragastan and Herbig 2007). In contrast, interpretation of vicariance patterns in molecular phylogenetic trees implied that extant species were younger than 3 my, and reports of extant species in the fossil record were interpreted as the result of iterative convergent evolution (Kooistra, Coppejans, and Payri 2002). Because of this possibility, we have not used any calibration points within the genus; only its earliest appearance was used to calibrate the clock. Even though the present study was not designed to answer questions about species ages in *Halimeda*, we can glean some information from its results. These indicate that the five sections of *Halimeda* diverged from one another during the first half of the Cretaceous (between approximately 147 and 97 my). Because each of the sections diversified relatively soon after they originated, we conclude that species ages must be considerably older than implied by Kooistra et al. (2002). This result also implies that the iterative morphological convergence hypothesis should be re-evaluated. It should be clear that we do not imply that the hypothesis is false, and records of extant species in Miocene and older sediments should certainly not be accepted without scrutiny. For example, Dragastan et al. (2002) synonymized several Mesozoic taxa, including some Triassic ones (> 200 my) with the extant taxon *H. cylindracea* because of similar segment shape. It follows directly from our results that no extant species can be of Triassic age and we are of the opinion that using extant species as form taxa for fossils is undesirable. Comparative studies of extant taxa in a molecular phylogenetic framework have presented unambiguous evidence that morphological convergence, especially of segment shape, has occurred during *Halimeda* evolution (Kooistra et al., 2002; Verbruggen et al., 2005). Consequently, interpretation of fossils as extant species should follow only from statistically sound morphometric analyses, preferably using time-series in various parts of the world.

Acknowledgments

Table 1. Taxon list with Genbank accessions and voucher numbers (in parentheses). Vouchers are deposited in the Ghent University Herbarium, the US National Herbarium or the Zechman lab herbarium (CSU Fresno).

taxon	rbcL	tufA	atpB	16S	18S
Acetabularia acetabulum	AY177738	FJ535854 (HV1287)			Z33461
Acetabularia calyculus		FJ535855 (HV389)			
Acetabularia crenulata	AY177737		FJ539159		Z33460
Acetabularia dentata	AY177739		FJ480413		Z33468
Acetabularia peniculus	AY177743		FJ539163		Z33472
Acetabularia schenkii	AY177744				Z33470
Avrainvillea lacerata	FJ432635 (HV599)	FJ432651 (HV599)		FJ535833 (HV599)	
Avrainvillea nigricans	FJ432636 (HV891)	FJ432652 (HV891)		FJ535834 (HV891)	
Batophora occidentalis	AY177747		FJ539160		Z33465
Batophora oerstedii	AY177748				Z33463
Bornetella nitida	AY177746		FJ480414		Z33464
Bornetella sp.	FJ535850 (LB1029)				
Bryopsidella neglecta	AY004766				
Bryopsis hypnoides	AY942169			AY221722	
Bryopsis plumosa	FJ432637 (HV880)	FJ432653 (HV880)	FJ480417	FJ535835 (HV880)	FJ432630 (HV880)
Caulerpa flexilis	AJ512485	DQ652532			
Caulerpa sertularioides	AY942170	FJ432654 (HV989)	FJ539164	AY389514	AF479703
Caulerpa taxifolia	AJ316279	AJ417939			
Caulerpa verticillata		AJ417967			
Caulerpella ambigua	FJ432638 (TS78)	FJ432655 (TS24)	FJ539161	FJ535836 (TS78)	
Chalmasia antillana					AY165785
Chlorocladus australasicus	AY177750				Z33466
Chlorodesmis fastigiata	FJ432639 (HV102)	FJ432662 (H0882)		FJ535837 (HV102)	AF416396
Codium lineage 1	AB038481			FJ535838 (SD0509370)	
Codium lineage 2	M67453	U09427		U08345	FJ535848 (KZN2K4.1)
Codium lineage 3	EF108086 (DHO2.178)	FJ535856 (KZN2K4.10)		FJ535839 (H0773)	FJ535849 (KZN2K4.10)
Cymopolia spp.	FJ535851 (WP011)				Z33467
Derbesia marina	AF212142				
Derbesia tenuissima	FJ535852 (H0755)	FJ535857 (H0755)			

Taxon					
Dichotomosiphon tuberosus	AB038487				AF416389
Flabellia petiolata	FJ432640 (HV1202)	FJ535858 (HV565)		FJ535847 (HV1202)	AY165786
Halicoryne wrightii	AY177745				AF407254 (SOC299)
Halimeda discoidea	AB038488	AY826360 (SOC299)			AF407257 (HEC11839)
Halimeda gracilis	AY942167	AM049965 (HV317)	FJ480416		AF407233 (H0179)
Halimeda incrassata		AM049958 (H0179)			AF407243
Halimeda micronesica		AM049964 (WLS420-02)			AF407267 (H0484)
Halimeda opuntia	AB038489	AM049967 (HV61)			Z33469
Neomeris dumetosa					D86495
Oltmannsiellopsis viridis	DQ291132	DQ291132	DQ291132	DQ291133	
Ostreobium sp.	FJ535853 (H0753)	FJ535859 (H0753)		FJ535840 (H0753)	
Parvocaulis exigua	AY177740		FJ539162		Z33471
Parvocaulis parvula	AY177741				
Pedobesia spp.	AY004768			FJ535841 (HV1201)	
Penicillus capitatus	FJ432641 (HV338)				AF416404
Penicillus dumetosus	AY942175				AF416406
Pseudendoclonium akinetum	AY835431	AY835432	AY835433	AY835434	DQ011230
Pseudocodium floridanum	AM909692 (NSF.I23)	AM909697 (NSF.I23)			FJ432631 (NSF.I23)
Pseudocodium natalense	AM909693 (KZNb2241)	AM049969 (KZNb2241)		FJ535842 (KZNb2241)	FJ432632 (KZNb2242)
Rhipidosiphon javensis	FJ432644 (DML40134)			FJ535843 (DML40134)	
Rhipilia crassa	FJ432645 (H0748)	FJ432657 (HV738)		FJ535844 (H0748)	
Rhipilia nigrescens	FJ432646 (H0847)	FJ432658 (HV788)			FJ432633 (HV788)
Rhipilia tomentosa	AY942164				
Rhipiliopsis profunda	FJ432647 (DML51973)	FJ432659 (DML51973)		FJ535845 (DML51973)	
Rhipocephalus phoenix	FJ432648 (HV404)			FJ535846 (HV404)	AF416402
Trentepohlia aurea	FJ534608				DQ399590
Tydemania expeditionis	AY942161	FJ432661 (HV873)			FJ432634 (HV873)
Udotea flabellum	AY942166				AF407270
Udotea glaucescens	FJ432650 (H0862)				
Udotea spinulosa	AY942160				
Ulothrix zonata	AF499683				Z47999
Ulva intestinalis	AB097617	AY454399			AJ000040

Table 2. Selection of partitioning strategy and model of sequence evolution using the Bayesian Information Criterion (BIC). The log-likelihood, number of parameters and BIC score are listed for various combination of partitioning strategies and models of sequence evolution. Each partition had its own copy of the model of sequence evolution with a separate set of model parameters. Lower BIC values indicate a better fit of the model to the data. The lowest BIC (in boldface) was observed for the partitioning strategy with four partitions (ribosomal DNA and separate codon positions for each gene), and GTR+Γ8 models applied to each partition.

model	lnL	# par	BIC
single partition			
F81	-56405.2	122	113863
F81+Γ_8	-51296.8	123	103655
HKY	-56009.6	123	113080
HKY+Γ_8	-50707.8	124	102486
GTR	-54913.6	127	110923
GTR+Γ_8	-50224.1	128	101553
2 partitions: ribosomal DNA, protein coding			
F81	-55852.3	126	112792
F81+Γ_8	-50702.4	128	102509
HKY	-55249.8	128	111604
HKY+Γ_8	-49969.8	130	101061
GTR	-54303.0	136	109779
GTR+Γ_8	-49774.3	138	100739
5 partitions: one for each locus (*atp*B, *rbc*L, *tuf*A, 16S, 18S)			
F81	-55772.5	138	112736
F81+Γ_8	-50769.0	143	102772
HKY	-55160.7	143	111555
HKY+Γ_8	-49913.6	148	101104
GTR	-54145.5	163	109697
GTR+Γ_8	-49483.9	168	100417
4 partitions: ribosomal DNA, separate codon positions for protein coding			
F81	-52199.0	134	105554
F81+Γ_8	-49807.4	138	100805
HKY	-50927.6	138	103046
HKY+Γ_8	-48121.3	142	97468
GTR	-50455.9	154	102241
GTR+Γ_8	-47763.5	158	96890
5 partitions: 16S, 18S, separate codon positions for protein coding			
F81	-52190.9	138	105572
F81+Γ_8	-49798.1	143	100830
HKY	-50918.0	143	103070
HKY+Γ_8	-48109.4	148	97496
GTR	-50160.4	163	101727
GTR+Γ_8	-47769.0	168	96988
11 partitions: 16S, 18S, separate codon positions for each gene			
F81	-52087.6	162	105573
F81+Γ_8	-49742.6	173	100978
HKY	-50825.1	173	103143
HKY+Γ_8	-48071.9	184	97731
GTR	-50188.9	217	102250
GTR+Γ_8	-47561.4	228	97090

Table 3. List of calibration points used to date the phylogenetic tree. The nodes to which the age constraints are applied (first column) are indicated on the phylogram in Fig. 3. If more than one age constraint was identified for a node, these are listed as different calibrations (e.g. calibrations a1, a2 and a3 all apply to node a). The 'age' column indicates the type of constraint (minimum vs. maximum) and the value of the constraint (in million years). Ages of epochs and stages follow the International Stratigraphic Chart (ICS 2008).

node	name	calibration	fossil	period	age (my)	reference
a	siphonous algae crown	a1	absence of siphonous fossils	Ediacaran	max 635	Zhang et al. (1998)
		a2	Yakutina aciculata	Middle Cambrian	min 499	Korde (1957)
		a3	Chaetocladus plumula	Middle Ordovician	min 460.9	Whitfield (1894)
b	Batophoreae stem	b1	Uncatoella verticillata	Lower Devonian	min 397.5	Kenrick and Li (1998)
		b2	Archaeobatophora typa	Upper Ordovician	min 443.7	Nitecki (1976)
		b3	Uncatoella verticillata	Lower Devonian	max 416.0	Kenrick and Li (1998)
c	Bornetelleae stem	c	Zittelina hispanica	Hauterivian	min 130.0	Masse et al. (1993)
d	Neomereae crown	d1	Neomeris cretacea	Albian	min 99.6	Granier and Deloffre (1993)
		d2	Neomeris cretacea	Barremian	min 125.0	Sotak & Misik (1993)
		d3	Pseudocymopolia jurassica	Portlandian	min 142.0	Dragastan (1968)
e	Acetabularieae stem	e1	Acicularia heberti	Danian	min 61.1	Morellet and Morellet (1922)
		e2	Acicularia boniae	Middle Triassic	min 228.7	Iannace et al. (1998)
f	core Halimedineae stem	f	Dimorphosiphon rectangulare	Middle Ordovician	min 460.9	Hoeg (1927)
g	Caulerpaceae stem	g	Caulerpa sp.	Wolfcampian	min 280.0	Gustavson and Delevoryas (1992)
h	Halimeda stem	h1	Halimeda marondei	Norian	min 203.6	Flügel (1988)
		h2	Halimeda soltanensis	Upper Permian	min 251.0	Poncet (1989)

Appendix

Alternative sets of calibration points used to constrain relaxed molecular clock analyses. Node ages were inferred for each of these conditions to evaluate the impact of some potentially erroneous fossil assignments.

condition	combination	description
01	a1 a2 b1 c d2 e1 f g h1	moderately conservative
02	a1 a3 b1 c d1 e1 f g h1	ultraconservative youngest
03	a1 a2 b2 c d3 e2 f g h2	ultraliberal oldest
04	a1 a2 b1 c d2 e1 f h1	moderately conservative, no *Caulerpa*
05	a1 a2 b1 c d2 e1 g h1	moderately conservative, no *Dimorphosiphon*
06	a1 a3 b1 c d1 e1 g h1	ultraconservative youngest, no *Dimorphosiphon*
07	a1 a2 b2 c d3 e2 g h2	ultraliberal oldest, no *Dimorphosiphon*
08	a1 a2 b1 c d2 e1 h1	moderately conservative, no *Caulerpa*, no *Dimorphosiphon*
09	a2 b1 b3 c d2 e1 h1	very conservative young: strongly constrained age for crown Dasycladaceae based on *Uncatoella*
10	a2 b1 b3 c d2 e1 h2	very conservative young: strongly constrained age for crown Dasycladaceae based on *Uncatoella*
11	a1 a2 c d2 e1 h1	condition 08 without calibrations on node b

As mentioned in the text, analyses were also repeated with different maximum age constraints on the root node to evaluate the sensitivity to this particular assumption. These analyses were variants of condition 08 in which the maximum age constraint a1 was replaced with 800 my and 500 my.

The chronogram resulting from analysis 08 is shown in Fig. 4.

The chronogram resulting from analysis 09 is shown in the online appendix (i.e. below).

Alternative chronogram inferred with node b constrained to lie within the Lower Devonian (416.0–397.5 my) based on occurrence of fossils with thalli comparable to those of Dasycladaceae in this epoch. The calibration points used for analysis are a2, b1, b3, c, d2, e1 and h1 (Table 3). This chronogram can be seen as a conservative (i.e. young) alternativ the analysis presented in Fig. 4.

7
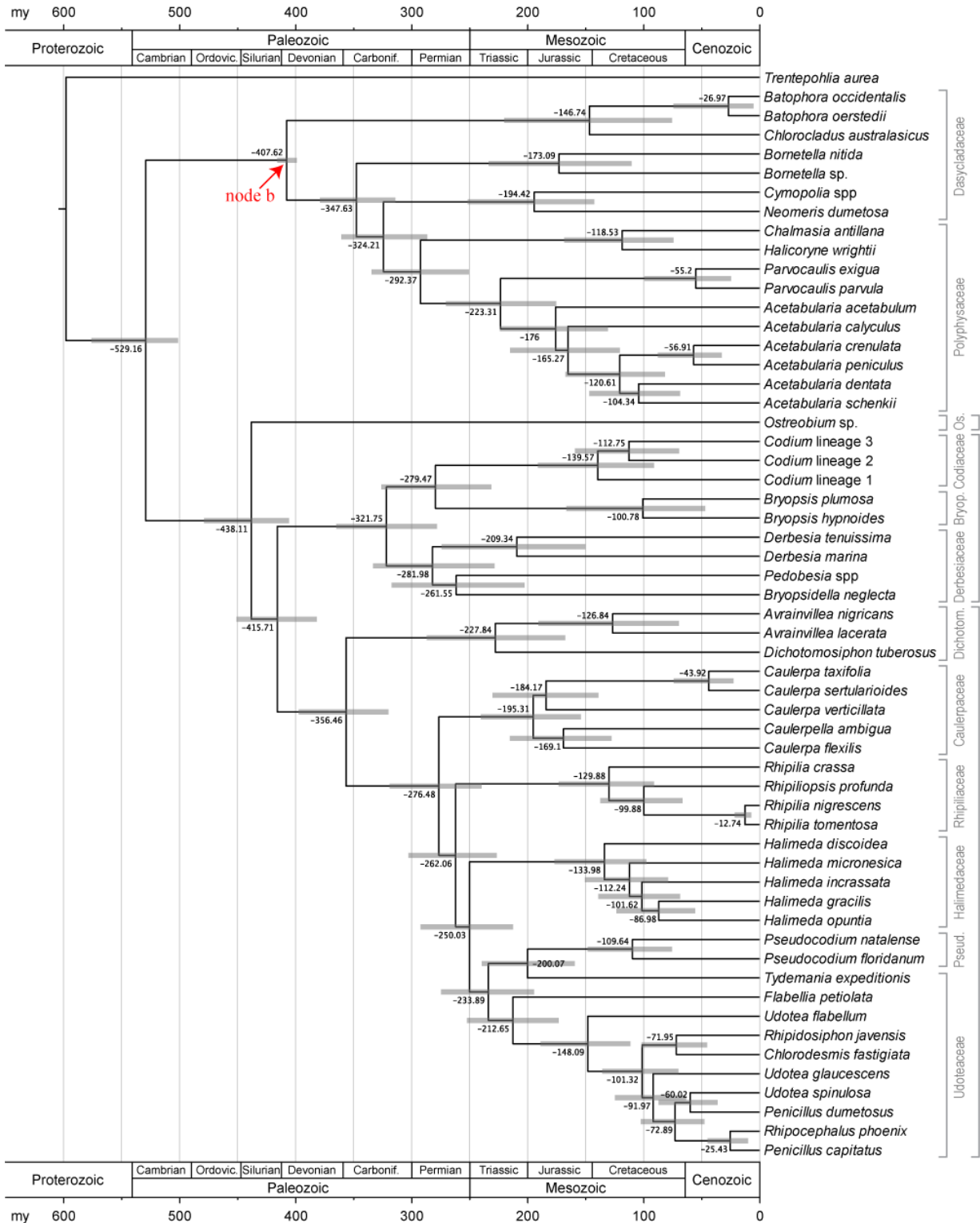
Systematics of the marine microfilamentous green algae *Uronema curvatum* and *Urospora microscopica* (Chlorophyta) [1]

Abstract

The microfilamentous green alga *Uronema curvatum* is widely distributed along the western and eastern coasts of the north Atlantic Ocean where it typically grows on crustose red algae and on haptera of kelps in subtidal habitats. The placement of this marine species in a genus of freshwater Chlorophyceae had been questioned. Molecular phylogenetic analysis of nuclear-encoded small and large subunit rDNA sequences reveal that *U. curvatum* is closely related to the ulvophycean order Cladophorales with which it shares a number of morphological features, including a siphonocladous level of organization and zoidangial development. The divergent phylogenetic position of *U. curvatum*, sister to the rest of the Cladophorales, along with a combination of distinctive morphological features, such as the absence of pyrenoids, the diminutive size of the unbranched filaments and the discoid holdfast, warrants the recognition of a separate genus, *Okellya*, within a new family of Cladophorales, Okellyaceae. The epiphytic *Urospora microscopica* from Norway, which has been allied with *U. curvatum*, is revealed as a member of the cladophoralean genus *Chaetomorpha* and is herein transferred to that genus as *C. norvegica* nom. nov.

Introduction

Green algae display a wide diversity of thallus organization, ranging from flagellate or coccoid unicells to colonial forms and various levels of multicellular organization. This morphological variation has been the basis for conventional green algal classification (Round 1984). For example flagellates were commonly grouped in the order Volvocales, coccoids in the Chlorococcales, and unbranched filaments in the Ulotrichales (Bold and Wynne 1985). Ultrastructural work, comparative biochemistry and life-history studies have demonstrated that a filamentous nature (and various other vegetative features) are independently derived in different lineages within the green algae (Mattox and Stewart 1984). Molecular systematics have largely corroborated these findings, showing that convergent evolution is responsible for the presence of unbranched filaments in distantly related green algal lineages, such as the chlorophytan classes Ulvophyceae (*Ulothrix*, *Urospora* and *Chaetomorpha*) and Chlorophyceae (*Microspora*, *Oedogonium*, *Uronema*), and the streptophytan classes Klebsormidiophyceae (*Klebsormidium*) and Zygnematophyceae (*Spirogyra* and other genera) (Lewis and McCourt 2004, Pröschold and Leliaert 2007). Most notably, the genus *Ulothrix*, which is often regarded as the morphological archetype of the unbranched uniseriate filamentous morphology, has been shown to be polyphyletic, with its various members belonging to different green algal classes (O'Kelly 2007).

Although the phylogenetic position of numerous unbranched filamentous species has now been resolved based on ultrastructural and molecular evidence (e.g. Booton et al. 1998, Leliaert et al. 2003, O'Kelly et al. 2004), several taxa have remained largely unstudied. Amongst them are the marine microfilamentous species *Uronema curvatum* Printz and *Urospora microscopica* Levring.

The genus *Uronema* Lagerheim (1887) is characterized by unbranched filaments of uninucleate cells with a single chloroplast and 1-4 pyrenoids (Chaudhary 1979). The filaments are attached by a basal discoid gelatinous holdfast, and apical cells are typically acuminate. Thalli reproduce asexually by a single zoospore (sometimes two) formed per cell. The genus is currently placed in the chlorophycean order Chaetophorales, based on ultrastructural evidence and 18S rDNA sequence data (Booton et al. 1998). *Uronema* includes about 17 species (Guiry and Guiry 2009), all but two restricted to freshwater or damp terrestrial habitats. The only marine species are the North Atlantic *U. curvatum* and the south-west Pacific *U. marinum* Womersley, both inconspicuous algae that have probably passed unnoticed in many investigations. *Uronema curvatum* was originally described from Trondheim Fjord, Norway (Printz 1926), and has since been reported from scattered localities along the eastern and western coasts of the north Atlantic Ocean (Feldmann 1954, Rueness 1977, South and Tittley 1986, Rueness 1992, Kornmann and Sahling 1994, Maggs and O' Kelly 2007). The species grows in subtidal habitats on non-calcified crustose red algae (such as *Peyssonnelia* and *Cruoria*), crustose Cyanobacteria on pebbles, and on haptera of kelps. *Uronema curvatum* differs from the freshwater representatives of the genus in zoosporganial and apical cell morphology, and its generic placement has been debated by Rueness (1992) and Kornmann and Sahling (1994) who suggested a relationship with Cladophorales or Ulotrichales (Acrosiphoniales).
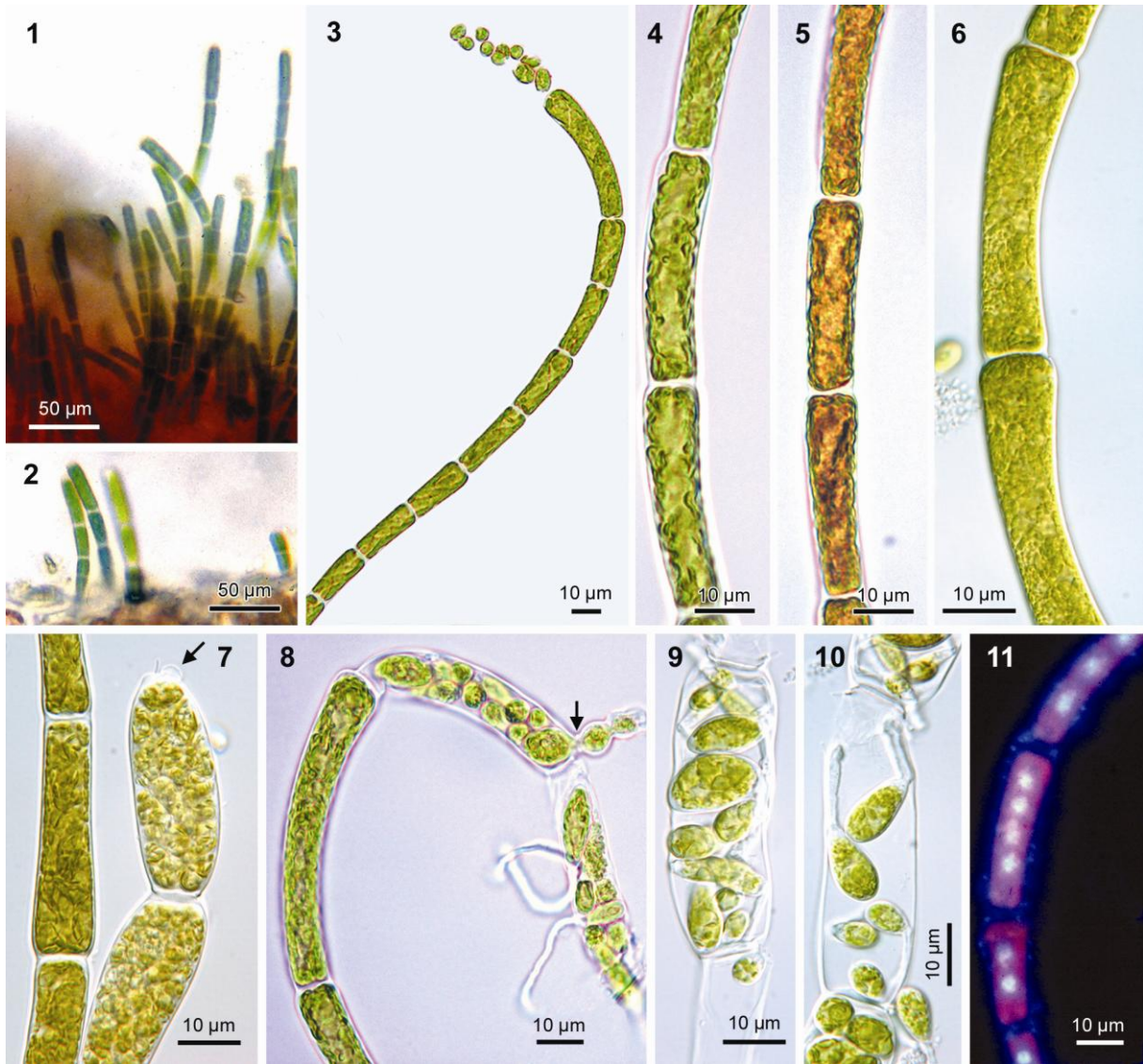
Figures 1-11. *Uronema curvatum* (= *Okellya curvata*, comb. nov.). Figs 1, 2. Field-collected sample, growing as an epiphyte on a red crust, on the haptera of *Laminaria hyperborea* (diameter of filaments 7-8 μm). Figs 3-11. Cultures. Fig. 3. Terminal sporangium with spores. Fig 4, 5. Vegetative filaments showing chloroplasts before and after staining with iodine solution (no pyrenoids are visible). Fig. 6. Irregular surface of chloroplast. Figs 7-10. Terminal sporangia with spores, some of which germinate *in situ*. Arrows indicate exit pore. Fig. 11. DAPI-stained vegetative cells with two to four nuclei.

Urospora Areschoug (1866) is a genus of cold water marine green algae characterized by an unbranched filamentous gametophyte composed of multinucleate cells with a parietal reticulate chloroplast, and a unicellular, club-shaped, uninucleate sporophyte (*Codiolum* phase). The uniseriate gametophytes are attached to the substrate by multicellular rhizoids arising from the basal cells. *Urospora* has a complex nomenclatural history and the taxonomic position of the genus has long been uncertain (Lokhorst and Trask 1981). The genus has been placed in the Cladophorales or Siphonocladales based on the multinucleate cells (Wille 1890, Rosenvinge 1893, Setchell and Gardner 1920, Printz 1932) but is now recognized as a close relative of *Acrosiphonia* and *Spongomorpha* in the

ulvophycean order Ulotrichales based on morphological, ultrastructural and life-history features, and molecular data (Jorde 1933, Kornmann 1963, Floyd and O' Kelly 1984, van Oppen et al. 1995, Jónsson 1999, Lindstrom and Hanic 2005). The genus includes about 12 species worldwide (Guiry and Guiry 2009), four of which are common along western European shores (Lokhorst and Trask 1981). The epiphytic *Urospora microscopica* from Norway has been distinguished from other species in the genus by its minute filaments (Levring 1937). Since its description, the species has remained largely unnoticed and its systematic position uncertain (Rueness 1992, Lein et al. 1999).

In the present study we aim to assess the phylogenetic position of the enigmatic microfilamentous species *Uronema curvatum* and *Urospora microscopica* by molecular phylogenetic analysis of nuclear-encoded small and large subunit rDNA sequences.

Material and methods

Specimens of *Uronema curvatum*, growing as epiphytes on *Peyssonnelia dubyi* P.L. Crouan & H.M. Crouan, were collected from Vega (county of Nordland, Norway) on 26 October 1990 (Rueness 1992). Specimens of *Urospora microscopica*, growing epiphytic on *Cystoclonium purpureum* at a depth of 3-5 m, were collected from Busepollen in Austevoll (county of Hordaland, Norway) in September 1994. Unialgal cultures of both species were obtained as described in Rueness (1992) and have been deposited in the Culture Collection of Algae and Protozoa (CCAP) as CCAP 455/1 (*Uronema curvatum*) and CCAP 504/1 (*Urospora microscopica*). Specimens were examined with a Nikon Eclipse TE 300 (Nikon Co., Tokyo, Japan) and Olympus BX51 (Olympus Co., Tokyo, Japan) bright field light microscopes. Photographs were taken with a Nikon DS-5M or Olympus E410 digital camera mounted on the microscope. Pyrenoids were stained with Lugol's iodine. DAPI nuclear staining was performed as described by Rueness (1992).

Molecular phylogenetic analyses were based on nuclear-encoded small subunit (SSU) and partial large subunit (LSU) rDNA sequences. DNA extraction, PCR amplification and sequencing were performed as described in Leliaert *et al.* (2007). Taxa for which new sequences were generated are listed in Table S1 of the online supplementary material. Sequences have been deposited to EMBL/GenBank under accession numbers FN257507- FN257512.

Two alignments were created for phylogenetic analyses. The first one was assembled to assess the phylogenetic position of *Uronema curvatum* and *Urospora microscopica* within the Chlorophyta. This alignment consisted of 30 SSU rDNA sequences, including other *Uronema* and *Urospora* representatives and exemplar taxa from a broad representation of chlorophytan classes for which SSU rDNA sequences have been deposited in GenBank. Two prasinophycean algae were used as outgroup taxa. Although no sequence data are available for the type species of *Uronema* (*U. confervicola* Lagerheim) and *Urospora* (*U. mirabilis* Areschoug), there is indirect evidence that *Uronema belkae* and *Urospora neglecta* (included in our phylogenetic analyses) are related to the types of the respective genera. Schlösser (1987) showed that the autolysin of *U. confervicola* reacts in bioassays on strains of *U. belkae,* suggesting that the two species are closely allied (Pröschold and Leliaert 2007). *Urospora mirabilis* is currently regarded as a synonym of *Urospora penicilliformis*

(Roth) J.E. Areschoug, which has been found to be related to *Urospora neglecta* based on 18S rRNA gene sequence data (Lindstrom and Hanic 2005).

Based on the results of the phylogenetic analysis inferred from the SSU rDNA alignment, a second dataset of partial LSU rDNA sequences was assembled and analysed to examine the phylogenetic position of the two species within the Cladophorales with more confidence. In the Cladophorales, partial LSU rDNA sequences (first ca. 500 bp) are known to be more phylogenetically informative than SSU rDNA sequences (Leliaert et al. 2003). The LSU rDNA alignment consisted of 20 cladophoralean sequences with *Caulerpa* and *Chlorodesmis* (Bryopsidales), *Acrosiphonia* (Ulotrichales) and *Ulva* (Ulvales) as outgroup taxa. Sequences were aligned using MUSCLE (Edgar 2004), and visually inspected.

Evolutionary models for the two alignments were determined by the Akaike Information Criterion in PAUP/Modeltest 3.6 (Posada and Crandall 1998, Swofford 2002). Both datasets were analysed with maximum likelihood (ML) and Bayesian inference (BI), using PhyML v2.4.4 (Guindon and Gascuel 2003) and MrBayes v3.1.2 (Ronquist and Huelsenbeck 2003) respectively. The SSU rDNA dataset was analysed under a general time-reversible model with a proportion of invariable sites and gamma distribution split into 4 categories (GTR+I+G4). The LSU rDNA alignment was analysed under a general time-reversible model with gamma distribution split into 4 categories and no separate rate class for invariable sites (GTR+G4). BI analyses consisted of two parallel runs of four incrementally heated chains each, and 4 million generations with sampling every 1000 generations. A burnin sample of 2000 trees was removed before constructing the majority rule consensus tree. For the ML trees, the reliability of internal branches was evaluated with non-parametric bootstrapping (1000 replicates).

Results

Morphology

Thalli of *Uronema curvatum* form minute epiphytic turfs of curved, uniseriate, unbranched filaments, composed of 3-10 cells, 100-180 μm long (in culture, filaments may grow up to 100 cells and 700 μm long), with an increasing diameter towards the apex (Figs 1, 2). Thallus is dull, yellowish green in colour. Filaments are attached to the substrate by a basal discoid holdfast. Vegetative cells are subcylindrical, 3.5-6 μm in diameter at the base, increasing to 7-10 μm at the apex, 1.5-6 times as long as broad, up to 21 μm long. The chloroplast is parietal and lobed and occupies most of the cell wall (Figs 4-6); transmission electron microscopy showed that more than one chloroplast might be present per cell (Rueness 1992); pyrenoids are absent (Fig. 5). Cells are multinucleate, containing (1-) 2-4 (-8) nuclei (Fig. 11). Thalli become reproductive before the filaments reach about 10 cells (in culture, unattached filaments may become longer). Prior to differentiation into sporangia, cells contain 8-16 (-32) nuclei (Rueness 1992). Zoids develop by transformation of apical and subapical cells into slightly swollen zoosporangia (Figs 3, 7); (8-) 16-32 zoids are formed per cell, which emerge through a domed pore in the upper part of the cell, on the outer face relative to the curvature of filaments (Figs 7, 8) (Rueness 1992). In culture, spores may germinate within the parent cell (Figs 9,

10). Filaments that were isolated into unialgal culture in 1990 have since been reproducing asexually by spores.
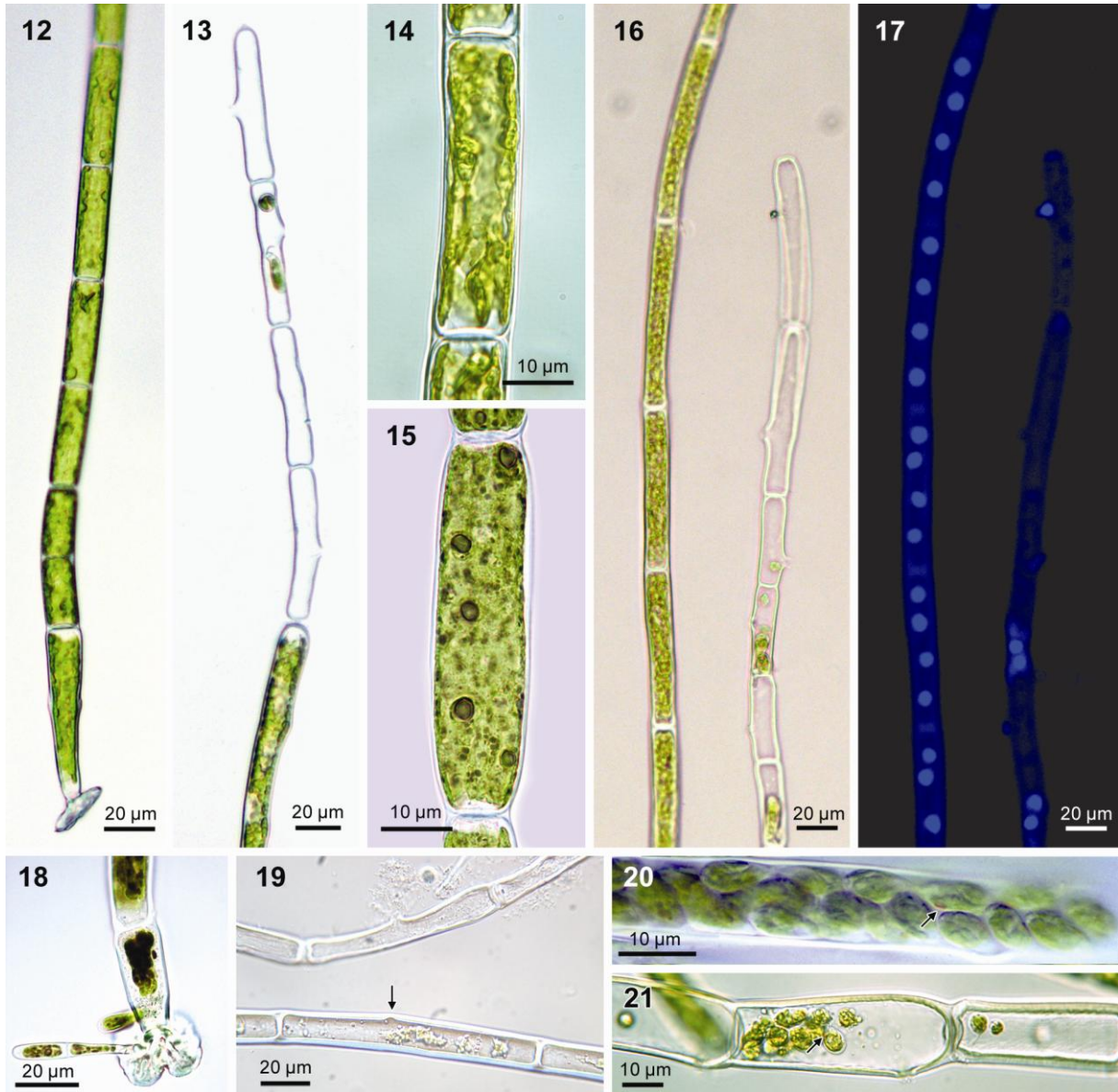
Figures 12-21. *Urospora microscopica* (= *Chaetomorpha norvegica*, nom. nov.), culture. Fig. 12. Filament with basal cell with attachment disc. Fig. 13. Apical portion of filament with sporangia and lateral exit pores. Fig. 14. Parietal, lobed chloroplast. Fig. 15. Cell showing pyrenoids following staining with iodide solution. Figs 16, 17. The same filaments after DAPI-staining as seen under light field and fluorescence microscopy, showing multinucleate cells with four nuclei in axial arrangement (note one spore attached outside the filament and a few spores left in sporangium). Fig. 18. Lobed attachment disc with germinating spores. Fig. 19. Sporangia with exit pore (top filament) and start of exit pore formation (bottom filament, arrow) (fixed material, without cell contents). Figs 20, 21. Sporangia with spores, red eye spot visible (arrows).

Thalli of *Urospora microscopica* (Figs 12-21) form straight or curved, uniseriate, unbranched filaments up to 1750 μm long, composed of cylindrical cells, 10-20 μm in diameter (Fig. 12). Thallus is bright, grass green in colour. Filaments are attached to the substrate by a basal hyaline lobed holdfast (Figs 12, 18). Cells contain a parietal, lobed chloroplast with several pyrenoids (ca. 5) (Figs 14, 15). Cells are multinucleate, containing 4 nuclei in axial arrangement (Figs 16, 17). Zoids develop by transformation of apical and subapical cells into zoosporangia; 10-35 zoids are formed per cell, which emerge through a domed pore in the middle part of the cell (sometimes or subapical or sub-basal) (Figs 19-21). Filaments that were isolated into unialgal culture in 1994 have since been reproducing asexually by spores.

Molecular phylogeny

Specifications of the SSU and LSU rDNA sequence alignments and evolutionary models applied are given in Table S2 (online supplementary material).

Phylogenetic analysis of the SSU rDNA dataset resulted in a chlorophytan tree with a poorly resolved backbone, in which the monophyly of the Ulvophyceae, Chlorophyceae and Trebouxiophyceae, and the relationships among these classes were weakly supported (Fig. 22). Even so, the phylogenetic positions of *U. curvatum* and *U. microscopica* could be determined with high support. *Uronema curvatum* is unrelated to the freshwater *Uronema belkae* (or any other member of the chlorophycean order Chaetophorales), but instead sister to the Cladophorales. *Urospora microscopica* is not allied with the *Urospora neglecta* or any other member of Ulvales but is placed within the Cladophorales clade.

Phylogenetic analysis of the Cladophorales LSU rDNA alignment resulted in three well-supported clades, termed *Cladophora*, *Siphonocladus* and *Aegagropila* clades (Fig. 23). Concordant with the SSU rDNA tree, *U. curvatum* is sister to the Cladophorales with high support. *U. microscopica* falls within the *Cladophora* clade. It is most closely related to *Chaetomorpha*, although its exact phylogenetic position could not be determined with satisfactory statistical support.

Discussion

The placement of the marine species *Uronema curvatum* in a genus of freshwater Chlorophyceae had been questioned. Rueness (1992) examined the species in culture, and suggested a relationship with the cladophoralean genera *Chaetomorpha* and *Rhizoclonium*, or with the ulotrichalean *Urospora* based on the multinucleate cells. Kornmann and Sahling (1994) formally transferred the species to *Urospora* based on zoospore morphology, but this transfer was not widely adopted (Bartsch and Kuhlenkamp 2000, Nielsen and Gunnarsson 2001, Maggs and O' Kelly 2007). The present study shows that *U. curvatum* is closely allied to the ulvophycean order Cladophorales.

U. curvatum shares a number of ecological and morphological features with the green macroalgal order Cladophorales. Like most members of this order, *U. curvatum* occurs in benthic marine coastal habitats. The assumption that the Cladophorales are an originally marine clade, which successfully

invaded freshwater habitats at least two times independently (Hanyuda et al. 2002), is reinforced by the phylogenetic position of *U. curvatum*, sister to the rest of the Cladophorales.

Morphologically, *U. curvatum* shares the typical siphonocladous level of organization of the Cladophorales, i.e. multicellular thalli composed of multinucleate cells. The number of nuclei in cladophoralean species is highly variable and generally proportional to cell size. Most cladophoralean taxa have relatively large cells (ranging from several μm to several mm across) with hundreds or even thousands of nuclei, arranged in cytoplasmic domains (Kapraun and Nguyen 1994). The cells of *U. curvatum* typically contain 2-4 nuclei (Rueness 1992, Maggs and O' Kelly 2007), comparable to numbers found in *Rhizoclonium riparium* (Roth) Harvey, which has cell dimensionsof the same order of magnitude as *U. curvatum* (mostly 5-20 μm in diameter) (Leliaert and Boedeker 2007).

Figure 22. ML tree of the core chlorophytes (Ulvophyceae, Trebouxiophyceae, Chlorophyceae), rooted with two prasinophytes, inferred from SSU rDNA sequences. The phylogenetic position of *Uronema curvatum* (= *Okellya curvata*, comb. nov.) and *Urospora microscopica* (= *Chaetomorpha norvegica*, nom. nov.), along with other members of the two genera are shown. ML bootstrap values (> 50) and BI posterior probabilities (> .90) are indicated at branches. The branches leading to *Acetabularia* and *Caulerpa* are scaled 50% (*) and 25% (**).

Thallus organization in the Cladophorales ranges from branched or unbranched uniseriate filaments to more complex architectural types (one notable exception being the coccoid Spongiochrysis in the Aegagropila clade, Rindi et al. 2006). Unbranched filamentous thalli, assigned to *Chaetomorpha* or *Rhizoclonium*, have evolved from branched forms several times independently within the Cladophorales (Hanyuda et al. 2002, Leliaert et al. 2003). *U. curvatum* thus represents another unbranched filamentous lineage of Cladophorales. *U. curvatum* attaches to the substrate by a basal discoid holdfast and hence differs from most Cladophorales, which are attached to the substrate by branched or unbranched rhizoids that develop from basal or intercalary cells. The diminutive *Cladophora pygmaea* Reinke also attaches by a similar basal discoid holdfast. Based on this feature it was placed in a separate section of the genus by van den Hoek (1963), but a molecular phylogenetic study refuted the separate placement of *C. pygmaea* and showed that mode of attachment is not an evolutionarily conserved character and has little taxonomic value above the species level (Leliaert et al. 2009).

U. curvatum shares the typical zoidangial development and exit aperture of the Cladophorales: after vegetative growth ceases, the apical cells slightly swell and the cytoplasm is divided into zoids that are dispensed through a domed pore at the upper end of the cell. Some cladophoralean taxa display variation in zoidangial morphology. For example, in *Wittrockiella* (*Aegagropila* clade) the spores are released through extremely elongated exit tubes, resembling colourless hairs (Leliaert and Boedeker 2007), and many taxa of the *Siphonocladus* clade have large cells that form numerous lateral exit pores (Hori 1994).

The chloroplast of *U. curvatum* is parietal and lobed and lines most of the cell wall. In contrast, the cells of most other Cladophorales contain numerous chloroplasts interconnecting by delicate strands to form a continuous layer or a parietal network. In some taxa, such as *Rhizoclonium riparium*, cells contain a single or few parietal, lobed chloroplasts, similar to *U. curvatum* (Leliaert and Boedeker 2007). *Uronema curvatum* differs from other Cladophorales by the lack of pyrenoids, although TEM observations by Rueness (1992) suggest the presence of starch grains inside the chloroplasts. In other Cladophorales most of the chloroplasts in a cell contain a single pyrenoid. The majority of species have bilenticular pyrenoids, i.e. each pyrenoid consists of two hemispheres, separated by a single thylakoid and each hemisphere is capped by a bowl-shaped starch grain. This pyrenoid structure was initially thought to be uniform within the order (Jónsson 1962, van den Hoek et al. 1995), but several exceptions to this pattern have been reported, mainly in species of the *Aegagropila* clade (Matsuyama et al. 1998, Miyaji 1999, Hanyuda et al. 2002).

Another marine species of *Uronema*, *U. marinum*, has been described as an epiphyte on green and (non-crustose) red seaweeds in shallow subtidal habitats from southern and western Australia, the Great Barrier Reef, Lord Howe Island, Micronesia and Hawaii (Womersley 1984, Abbott and Huisman 2004, Kraft 2007). This species resembles *U. curvatum* in size, cell dimensions and mode of attachment, but differs in having one or two pyrenoids per cell. The taxonomic position of *Uronema marinum* is left undecided at this stage, and the name is retained pending molecular investigations.

The micro-filamentous species, *Urospora microscopica*, was described from Osund, Norway, growing epiphytically on *Nitophyllum* and *Cystoclonium* (Levring 1937). Since its original description, the species has remained largely unnoticed (Lein et al. 1999). Rueness (1992), who re-examined the type material, found that *U. microscopica* differed from *U. curvatum* in cell dimensions, the straight

filaments, and the position of the sporangial pore, which is lateral or sub-basal in *U. microscopica* versus subapical in *U. curvatum*. In the present study we examined recent collections and cultures of *U. microscopica*, which made it possible to investigate this species in more detail and assess its phylogenetic position based on molecular data. The phylogenetic analysis clearly shows that *U. microscopica* is a member of the *Cladophora* clade. The species seems to be most closely related to *Chaetomorpha*, although the presence of four nuclei and sporangia with a single lateral pore would suggest a relationship with *Rhizoclonium riparium* (Leliaert and Boedeker 2007).
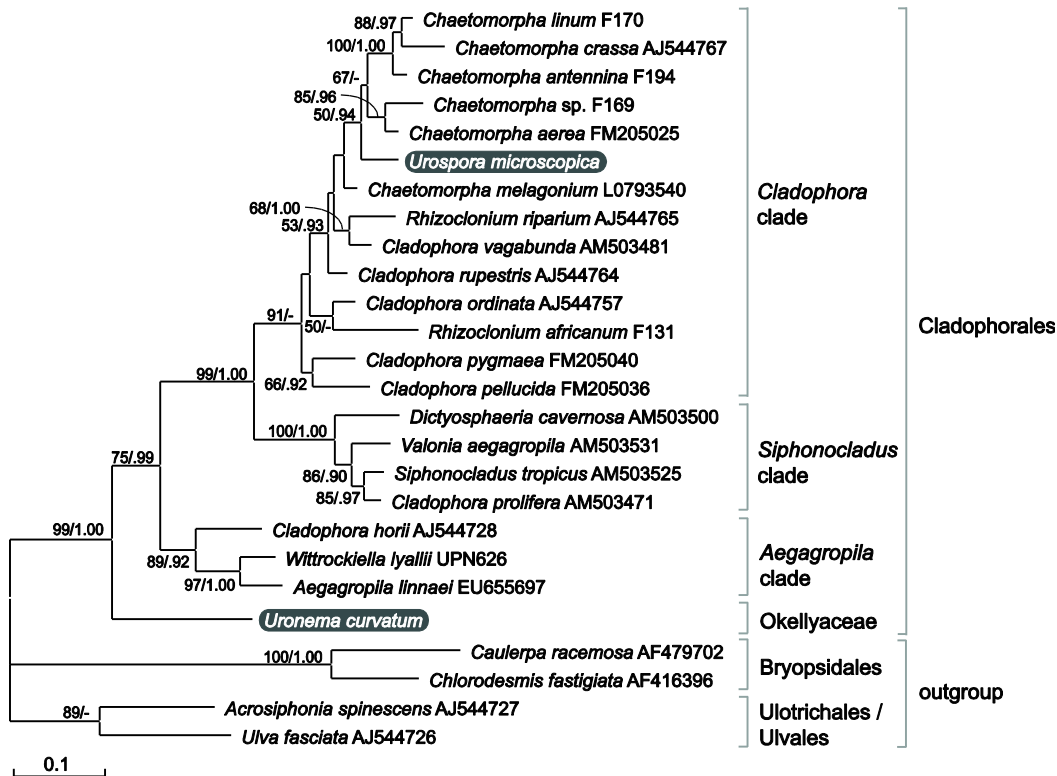
Figure 23. ML tree of the Cladophorales inferred from partial LSU rDNA sequences, showing the phylogenetic position of *Uronema curvatum* (= *Okellya curvata*, comb. nov.) and *Urospora microscopica* (= *Chaetomorpha norvegica*, nom. nov.). ML bootstrap values (> 50) and BI posterior probabilities (> .90) are indicated at branches.

Chaetomorpha is a marine genus of attached or unattached unbranched macro-filaments. More than 200 species and infraspecific taxa have been described worldwide (Index Nominum Algarum), of which only about 50 are currently accepted (Guiry and Guiry 2009). Morphological features used to delimit species within the genus are growth form, cell dimensions and shape of the basal attachment cell. However, the extensive variability of these morphological characters depending on environmental conditions accounts for a great deal of taxonomic confusion, and the genus is clearly in need of revision (Leliaert and Boedeker 2007).

Urospora microscopica differs from other members of the genus *Chaetomorpha* in the diminutive growth form, small cell size (10-20 μm in diameter) and low number of nuclei per cell. Most

Chaetomorpha species are much more robust, forming macroscopic thalli with cells ranging from ca. 60 μm in diameter in *C. ligustica* (Kützing) Kützing to one or several mm across (e.g. *C. melagonium* (Weber & Mohr) Kützing, *C. coliformis* (Montagne) Kützing). Only a few other minute *Chaetomorpha* species have been described. *Chaetomorpha sphacelariae* Foslie (1881) is a diminutive epiphyte on *Sphacelaria* from Norway, which has remained unnoticed since its original description. *Chaetomorpha minima* Collins & Hervey (1917) has been morphologically associated with *C. sphacelariae*. This species from the north-west Atlantic Ocean and Caribbean Sea also grows epiphytically on algae, seagrasses and salt-marsh plants, and forms inconspicuous filaments, composed of long cells which are 10-27 μm across (Schneider and Searles 1991, Dawes and Mathieson 2008, Littler et al. 2008). Another minute species, *C. recurva* Scagel (1966), is found along the Pacific coast of North America and forms minute thalli with filaments 6-10 μm across. Given that morphological features, especially cell dimensions, are now known to be poor indicators of phylogenetic relationships in the Cladophorales (and green algae in general) (Leliaert et al. 2007), these species will need to be further examined using molecular tools to assess their phylogenetic affinity.

Taxonomic conclusions

The divergent phylogenetic position of *U. curvatum*, sister to the rest of the Cladophorales, along with a combination of distinctive morphological features such as the absence of pyrenoids, the diminutive size of the unbranched filaments and a discoid holdfast, warrants the recognition of a separate genus within a new family of Cladophorales:

Okellyaceae Leliaert et Rueness, familia nov.

Algae benthicae marinae, filamentibus simplicibus erectis curvatis, disco basali ad substratum affixae. Cellularum divisio intercalaris. Filamenta 3-8 cellulibus, usque 200 μm longa (in cultura usque 100 cellulibus, 1 mm longa), diametro dilatato apicem versus. Cellulae apicales cylindricae obtusae, cellulae intercalares cylindricae, 3-20 μm latae, (1-) 2-4 (-8) nucleatae. Chloroplastus parietalis, pyrenoide absenti autem granis amyloideis. Zoosporangia apicalia et subapicalia, zoosporibus pluribus (8-32), e sporangioporo singulari emergentibus. Thalli epiphytici super algas crustosas in zona sublitorali.

Marine benthic algae forming minute tufts of slightly curved, erect unbranched uniseriate filaments, attached by a basal discoid holdfast. Growth by intercalary cell divisions. Filaments composed of 3-8 cells, up to 200 μm long (in culture sometimes up to 100 cells and 1 mm or more), diameter increasing towards the apex. Apical cell cylindrical with obtuse tip, intercalary cells cylindrical, 3-20 μm in diameter, containing (1-) 2-4 (-8) nuclei. Chloroplast parietal, lacking pyrenoids but including starch grains. Multiple zoospores (8-32) developing by transformation of apical and subapical cells into zoosporangia, emerging through a single pore in the upper part of the cell. Thalli epiphytic on various crustose algae in the subtidal zone.

Type genus: ***Okellya*** Leliaert et Rueness, gen. nov.

Cum characteribus familia. Characters as for family.

Type species: *Okellya curvata* (Printz) Leliaert et Rueness, comb. nov.

Basionym: *Uronema curvatum* Printz, *Algenveg. Trondhjemsfj.*: 233, pl. VII: figs 105-114 (1926).

Etymology: named in honour of Charles J. O'Kelly for his pioneering and influential work on green algal systematics.

Urospora microscopica is most closely related to *Chaetomorpha* (type: *Chaetomorpha linum*) in the Cladophorales and is therefore transferred to that genus. Since the combination *Chaetomorpha microscopica* has already been made by Meyer (1927) for a freshwater filamentous cladophoralean species (later transferred to *Cladochaete* and *Chaetocladiella*), a new species epithet must be chosen.

Chaetomorpha norvegica Leliaert et Rueness, nom. nov.

Basionym: *Urospora microscopica* Levring, *Lunds Univ. Årsskr.* 2, 33(8): 30, fig. 2e-k (1937).

Acknowledgements

Supplementary material

Table S1. Specimens for which new sequences were generated with collection data (location, collector, date of collection and voucher information) and EMBL accession numbers.

Species	Collection and voucher information	SSU rDNA	partial LSU rDNA
Uronema curvatum Printz	Norway: county of Nordland: Vega, subtidal, epiphytic on *Peyssonnelia dubyi* (Jan Rueness, 26 Oct 1990) [1]	FN257508	FN257507
Urospora microscopica Levring	Norway: county of Hordaland: Busepollen in Austevoll, subtidal 3-5 m deep, epiphytic on *Cystoclonium purpureum* (Jan Rueness, Sep 1994) [1]	FN257510	FN257509
Chaetomorpha melagonium (F. Weber & Mohr) Kützing	Norway: Spitsbergen: Kongsfjord, subtidal, epilithic (Kai Bischof, 14 July 2005, L: L0793540 / A88)		FN257511
Wittrockiella lyallii (Harvey) Hoek, Ducker & Womersley	New Zealand: Fiordland: Rum River estuary, high intertidal, on fallen tree (Svenja Heesch, 30 Sept 2005, WELT: UPN626 / H67) [2]		FN257512

[1] Cultures maintained by JR and FL, and deposited in Culture Collection of Algae and Protozoa (CCAP): http://www.ccap.ac.uk/

[2] UPN stands for "Ulva project number": http://www.maf.govt.nz/mafnet/publications/biosecurity-technical-papers/ulva/

Table S2. Specification of the SSU and LSU rDNA sequence alignments and summary of models and model parameters obtained.

	SSU rDNA (Chlorophyta)	LSU rDNA (Cladophorales + ulvophycean outgroups)
Ingroup	28: Ulvophyceae, Chlorophyceae, Trebouxiophyceae	20: Cladophorales
Outgroup	2: Prasinophyceae	4: Ulvophyceae
Alignment length / variable sites / parsimony informative sites	1787 / 867 / 557	647 / 388 / 311
Model estimated by the Akaike information criterion (AIC)	GTR+I+G4	GTR+G4
Estimated base frequencies (A/C/G/T)	0.24 / 0.21 / 0.28 / 0.27	0.21 / 0.24 / 0.32 / 0.23
Estimated substitution rates (AC / AG / AT / CG / CT / GT)	1.00 / 2.60 / 1.16 / 1.16 / 4.76 / 1.00	0.65 / 2.15 / 1.28 / 0.75 / 5.11 / 1.00
Among-site rate variation: proportion of invariable sites (I) / gamma distribution shape parameter (G)	0.27 / 0.46	0 / 0.39

8

General discussion

This thesis focuses on phylogenetic relationships among green algae and molecular evolutionary processes such as gain-loss dynamics of elongation factor genes, distribution of a non-canonical code and patterns in genome bias.

Phylogeny of the green algae

The green algae have long been recognized as a natural group, well differentiated from all other groups of algae by a number of shared characters, including intraplastidial starch storage, double-membraned plastids containing chlorophylls a and b, stacked thylakoids, isokont zoids, and a stellate transition zone between the flagellar axoneme and the basal body. All these characters are also shared by land plants (at least by the ones with flagellate stages in their life cycles), and hence the close relationship between green algae and land plants has been noted for many decades. The first molecular phylogenetic studies based on SSU nrDNA sequences have confirmed the monophyly of the green plants (Sogin et al. 1986, Gunderson et al. 1987). More recently, phylogenies based on nuclear and plastid genes have provided clear evidence that the green plants, red algae and glaucophytes diverged after primary endosymbiosis of a non-photosynthetic eukaryotic host cell and a cyanobacterium had taken place (Rodriguez-Ezpeleta et al. 2005).

Traditional evolutionary schemes of green algae have been largely based on thallus organization and it was assumed that ancestral coccoid or flagellated unicells have given rise to distinct lines of increasing size and complexity (Fott 1971, see Chapter 1). Since the 1970's a vast amount of ultrastructural information has gradually been accumulated. The hypotheses formulated from comparative ultrastructure studies were highly inconsistent with the traditional systems of classification; they showed that various vegetative features have evolved numerous times in the green algae (Irvine and John 1984). Molecular phylogenetic studies, mainly based on SSU nrDNA sequences, by and large corroborated the ultrastructural data, showing that convergent evolution is responsible for the presence of near-identical morphologies in distantly related green algal lineages (see Chapter 7). DNA sequence data have further revolutionized our understanding of green algal evolution. Current hypotheses of the evolution of the Viridiplantae posit the early divergence of two main lineages from an ancestral green flagellated unicell (Lewis and McCourt 2004, Becker and Marin 2009). One lineage, the Streptophyta, includes a paraphyletic assemblage of green algae and their descendents, the land plants. The second lineage, the Chlorophyta, includes the majority of green algae. This dichotomy is supported by fundamental differences in the ultrastructure of the flagellated cells (e.g. cruciate flagellar roots in Chlorophyta vs. unilateral flagellar root with multilayered structures in Streptophyta). Considerable progress has been made in clarifying the relationships among the streptophyte green algae and land plants based on multi-marker and genome-scale datasets (Parkinson et al. 1999, Turmel et al. 2003, Turmel et al. 2006, Lemieux et al. 2007, Moore et al. 2007, Rodriguez-Ezpeleta et al. 2007, Saarela et al. 2007). Conversely, the evolutionary history of

the Chlorophyta has received much less attention. Phylogenetic relationships among three chlorophytan clades (Trebouxiophyceae, Chlorophyceae and Ulvophyceae) have been the subject of long-standing debates, and the relationships within several major clades (e.g. the Ulvophyceae) have been difficult to elucidate.
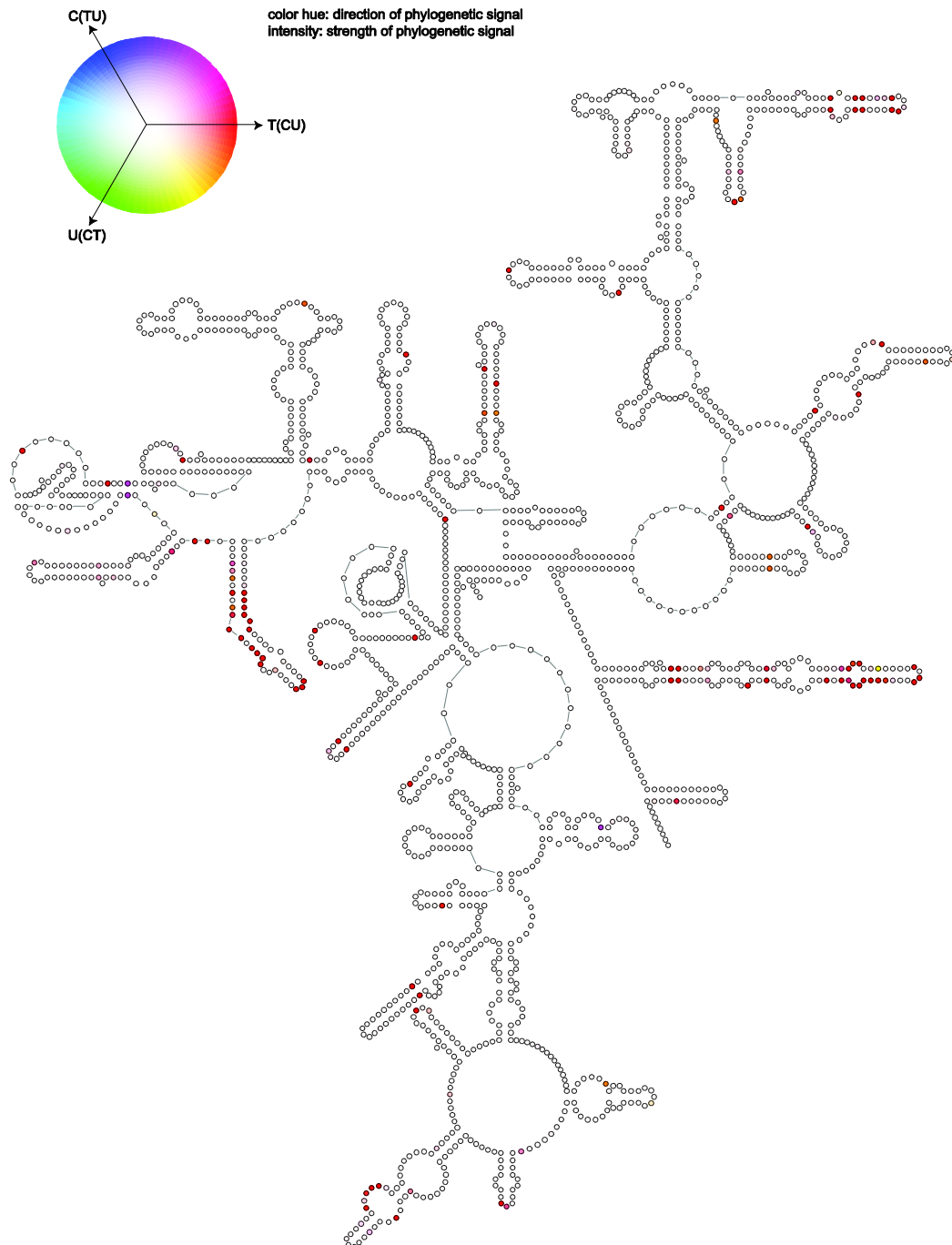
Figure 1. Phylogenetic signal of each site in the SSU alignment for the three alternative UTC topologies. Only a few sites contain phylogenetic signal concerning UTC topologies, i.e. colored sites, and all these sites favor a sister relationship between Chlorophyceae and Ulvophyceae, T(CU) topology. The phylogenetic signal of the 5% strongest signals is decreased to the 95 percentile of signal strength.

SSU nrDNA phylogenies

SSU nrDNA phylogenies showed conflicting results regarding the relationships between Ulvophyceae, Trebouxiophyceae and Chlorophyceae (UTC classes). A majority of SSU nrDNA molecular phylogenies showed a sister relationship between the Trebouxiophyceae and Chlorophyceae (Friedl 1995, Bhattacharya et al. 1996, Krienitz et al. 2001, Lopez-Bautista and Chapman 2003). Alternatively, the Chlorophyceae and Ulvophyceae are resolved as sisters clades in other studies (Friedl and O'Kelly 2002, Lewis and Lewis 2005, Watanabe and Nakayama 2007). Although the latter studies are more trustworthy due to a broader taxon sampling and the use of likelihood based methods with more realistic models of sequence evolution, these studies only included a small number of ulvophycean representatives, and almost never resulted in a comprehensive taxon sampling (i.e. including the Dasycladales, Bryopsidales, Cladophorales and Trentepohliales).

We checked the phylogenetic signal of each site in the SSU alignment for the three alternative UTC topologies (Fig. 1). This shows that only a few sites contain phylogenetic signal concerning UTC topologies, i.e. colored sites, and that the great majority of these sites favor a sister relationship between Chlorophyceae and Ulvophyceae. This strong phylogenetic signal is contradictory to SSU phylogenies showing conflicting results. The source of these conflicting topologies can possibly be attributed to alignment differences and the use of other phylogenetic methods. Comparison of the signal maps in Fig. 1 with rate maps (Ben Ali et al. 2001, Wuyts et al. 2001) indicates that signal concerning the UTC relationships is mainly found in the faster-evolving parts of the molecule.

Chloroplast phylogenomics

Organellar genomes have been shown to be particularly useful for phylogenomic reconstruction because of their relatively high gene content, condensed in comparison to nuclear genomes. Also, organellar genes are typically single-copy, in contrast to many nuclear genes that are multi-copy in nature, which can have a confounding effect on phylogenetic reconstruction if paralogous copies are being analyzed. Although chloroplast genes have been widely employed to reconstruct phylogenetic relationships among relatively closely related green algae (e.g. within orders or genera, Verbruggen et al. 2007, De Clerck et al. 2008, Verbruggen et al. 2009), only a few studies have used chloroplast genes to reconstruct phylogenies across the entire green plant lineage (e.g. Daugbjerg et al. 1995). During the last decade a large number of complete chloroplast genomes from a wide range of green algae have been sequenced, and chloroplast phylogenomic studies have been valuable to resolve problematic relationships among green algae (Qiu et al. 2006, Jansen et al. 2007, Lemieux et al. 2007, Turmel et al. 2008, Turmel et al. 2009). Despite the plethora of answers these datasets have provided, they have not been able to resolve the branching order of the UTC classes conclusively. A phylogenetic analysis of 58 concatenated chloroplast genes support a sister relationship between Ulvophyceae and Trebouxiophyceae, while Ulvophyceae and Chlorophyceae were sisters in a phylogenetic analysis of seven mitochondrial genes (Pombert et al. 2004, Pombert et al. 2005). The latter topology is also supported by chloroplast gene order data and genomic structural features (shared gene losses and rearrangements within conserved gene clusters). These phylogenomic analyses included two Ulvophyceae, *Pseudendoclonium akinetum* and *Oltmannsiellopsis viridis*, that belong to or are allied with the Ulvales—Ulotrichales group, respectively. Only recently, 23

chloroplast genes of the siphonous ulvophycean seaweed *Caulerpa filiformis* (Bryopsidales) have been sequenced (Zuccarello et al. 2009). Phylogenetic analyses of these new data suggested that neither Ulvophyceae nor Trebouxiophyceae are monophyletic, and showed that *Caulerpa* is more closely related to the trebouxiophyte *Chlorella* than to the ulvophytes *Oltmannsiellopsis* and *Pseudendoclonium* (Zuccarello et al. 2009). Explanations for this unsuspected relationship have not yet been provided. Another remarkable observation is that it has proven impossible to amplify chloroplast genes in the Cladophorales despite considerable efforts in various labs. Chloroplast isolation has also proven difficult.

Mitochondrial phylogenomics

Complete mitochondrial genomes are available for a number of green algae (see Chapter 1), providing an independent opportunity to study problematic relationships among green algae. Phylogenetic analysis of seven mitochondrial genes supported a sister relationship between Ulvophyceae and Chlorophyceae (Pombert et al. 2004). Similar to chloroplast phylogenomic studies, this study is based on a limited sample of taxa: one prasinophyte, one trebouxiophyte, a few Chlorophyceae and *Pseudendoclonium* as the sole ulvophyte. *Mesostigma* was revealed as the oldest lineage within the Streptophyta in a phylogenetic study of 33 mitochondrial genes (Rodriguez-Ezpeleta et al. 2007).

Amplifying nuclear markers: not a sinecure

Only a few studies have employed nuclear markers other than SSU rDNA to reconstruct green algal phylogenies. Single gene phylogenies have been made for actin (An et al. 1999, Bhattacharya et al. 2000), glucose-6-phosphate isomerase (Grauvogel et al. 2007), glyceraldehyde-3-phosphate dehydrogenase (Petersen et al. 2006, Robbens et al. 2007) and elongation factor genes (Keeling and Inagaki 2004, Noble et al. 2007). Recently, a multigene phylogeny (Rodriguez-Ezpeleta et al. 2007) has been made. In all case, these studies included only a few chlorophyte green algae and often no ulvophyte.

We chose to amplify nuclear genes in the first place because phylogenetic analyses inferred from SSU nrDNA, chloroplast or mitochondrial genes showed conflicting results with respect to the radiation of UTC classes and relationships between ulvophycean orders. The amplification failure of chloroplast genes in Cladophorales poses an additional problem hampering the use of chloroplast genes for green algal phylogenetics. Furthermore, nuclear genes offer the opportunity to study multiple independent loci.

We can think of four principal reasons why nuclear genes are barely used in green algal phylogenetic studies: (1) the limited availability of genomic data for green algae pose a restriction on the range of taxa which can be used directly as a data source or indirectly for primer design; (2) amplification and characterization of single copy nuclear genes is more difficult compared with chloroplast and SSU nrDNA sequences that can be relatively easily amplified with universal primers (Small et al. 2004). For SSU nrDNA there are numerous copies per cell present, whereas only a single copy of each singly

copy nuclear gene is present in each nucleus; (3) presence of multiple, large introns at variable positions hampering PCR amplification at the genomic DNA level, and (4) assessment of orthology is difficult due to the presence of evolutionary processes such as gene duplications, incomplete lineage sorting and lateral gene transfer.

We anticipated all these challenges and have succeeded to amplify ten nuclear markers among the different green algal classes: actin, glucose-6-phosphate isomerase (GPI), glyceraldehyde-3-phosphate dehydrogenase (GapA), oxygen-evolving enhancer protein (OOE1), 40S ribosomal protein S9, 60S ribosomal proteins L3 and L17, histone and elongation factor-1 alpha (EF-1α) or elongation factor-like (EFL). To overcome the limited availability of genomic data among green algae, an EST library for the siphonocladous ulvophyte *Cladophora coelothrix* was generated. The obtained EST sequences considerably contributed to our success since they allowed primer design and successful amplification of OOE1, three ribosomal proteins, histone and EF-1α genes among green algae.

We have not been able to amplify nuclear genes starting from DNA material, most likely due to the presence of large introns (e.g. observed in available actin genes). Therefore, we worked on mRNA of actively dividing algal cells. Amplification of the targeted nuclear genes may have been facilitated by their high expression and mRNA levels. Actin is the monomeric subunit of microfilaments, one of the three major component of the cytoskeleton, which provides cell shape and mechanical support and which is involved during intracellular transport and cell division. Histones are the chief protein components of chromatin, act as spools around which DNA winds and play an important role in gene regulation. Elongation factor-1 alpha (EF-1α) and elongation factor-like (EFL) are two key genes of the translational apparatus. GapA is an important enzyme of the Calvin cycle and glycolysis. Glucose-6-phosphate isomerase (GPI) is an essential enzyme for gluconeogenesis and glycolysis. Oxygen-evolving enhancer protein (OOE1) is an auxiliary component of the photosystem II manganese cluster. Eukaryotic ribosomes are made of four ribosomal RNAs and approximately 80 ribosomal proteins, of which we amplified three (40S ribosomal protein S9 and 60S ribosomal proteins L3 and L17).

A phylogenetic tree for each individual gene was made to check the orthology assumption. **Histone** genes were excluded from our final concatenated dataset because the orthology assumption was violated for the phylogenetic tree inferred from this gene. This is consistent with the observation that histone genes are highly duplicated across genomes (Nei and Rooney 2005; Wahlberg and Wheat 2008). **Elongation factor genes** EF-1α and EFL were also excluded from the concatenated alignment because of their mutually exclusive distribution in the green plant lineage (Cocquyt et al. 2009). All other individual gene trees did not violate the orthology assumption, although some of the genes were known to be duplicated in the green plant lineage (e.g. GapA/B, Petersen et al. 2006). Conventional **actin** genes form complex gene families in various groups of complex multicellular organism (e.g., animals and land plants), but are single copy in most green algae. The common ancestor of the Viridiplantae most likely contained a single actin gene, followed by independent duplications of actin genes in the Ulvophyceae: *Acetabularia cliftonii* contains 3 actin genes, and in the Trebouxiophyceae: *Nannochloris maculate*, *N. atomus* and *Chlorella vulgaris* each contain 2 actin genes (An et al. 1999, Yamamoto et al. 2001, Yamamoto et al. 2003). Despite the presence of multiple actin gene duplications events in green algae, we believe that actin is a useful marker for deep phylogenetic reconstruction because these duplications appear to have occurred relatively

recently, often within a single species (e.g. several *Nannochloris* species only contain a single actin gene). The Actin Related Proteins Annotation server (ARPAnno) was used to check if all actin sequences were conventional actins (Goodson and Hawse 2002, Muller et al. 2005). Land plants have cytosolic and plastid **glucose-6-phoshate isomerase** (GPI). The former is thought to be vertically inherited while the latter is of cyanobacterial origin. The chlorophytes *Chlamydomonas* and *Volvox* contain a single, nuclear encoded GPI that is related to the cytosolic GPI of land plants, which subsequently acquired a transit peptide for transport to the plastids (Grauvogel et al. 2007). All our GPI sequences are related to the cytosolic GPI genes of land plants and to the plastid-targeted GPI gene of *Chlamydomonas* and *Volvox*. **Glyceraldehyde-3-phosphate dehydrogenases** (GAPDH) are prominent examples of homologous isoenzymes. In green algae, there are several nuclear encoded GAPDHs with different evolutionary origins. Glycolytic GapC genes probably have a proteobacterial (mitochondrial) origin, whereas photosynthetic GapA was acquired from the cyanobacterial ancestor of the chloroplast. Both genes have been duplicated: GapCp and GapB, respectively. GapB originated from a GapA gene duplication, but clearly differs from GapA by the presence of a specific C-terminal extension and several GapB specific amino acid insertions and positions. GapB is found in Streptophytes and in the genome of the unicellular prasinophyte *Ostreoccoccus*, but is absent in the genomes of *Chlamydomonas* and *Volvox* (Petersen et al. 2006, Robbens et al. 2007). For all green algae sequenced in this study, we only retained homologous GapA sequences, which are clearly different from GapC sequences. GapB sequences were not amplified in any of the sampled green algae.

Resolving green algal phylogenies: contribution of nuclear genes

In our opinion, the progress in our understanding of the phylogeny of green algae achieved in this study is due to (1) a good balance between taxon and gene sampling; (2) the use of model-based techniques (ML and BI), paying careful attention to the selection of suitable partitioning strategies and models of sequence evolution; and (3) the removal of fast-evolving sites to improve phylogenetic signal. The dataset on which our analyses were based consisted of seven single-copy nuclear markers (actin, GPI, GapA, OOE1, 40S ribosomal protein S9 and 60S ribosomal proteins L3 and L17), SSU nrDNA and two plastid genes (*rbc*L and *atp*B) for 43 taxa representing the major lineages of the Viridiplantae.

We obtained high support across the topology of the Viridiplantae (see Chapter 2). We reveal the monophyly of the UTC classes and provided evidence for a sister relationship between the Ulvophyceae and Chlorophyceae with high support. We also inferred the relationships among the Ulvophyceae, which was found to consist of two main clades. The first clade contains the orders Ulvales and Ulotrichales. The second clade includes the early diverging genus *Ignatius* and a clade comprising the orders Trentepohliales, Bryopsidales, Dasycladales and Cladophorales, along with *Blastophysa*. In this clade, the Bryopsidales and Dasycladales are sisters, *Blastophysa* is most closely related to the Cladophorales, while the phylogenetic position of the Trentepohliales remains uncertain.
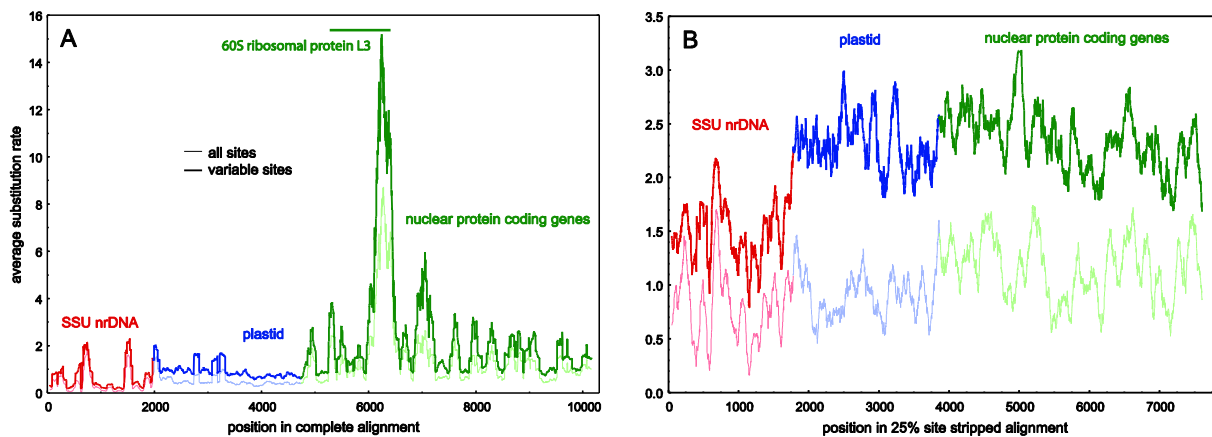
Figure 2. Substitution rates for SSU nrDNA, plastid and nuclear protein coding genes, measured as average substition rate in a sliding window across the alignment were calculated with HyPhy. Average substitution rates calculated only using variable sites (thick lines) are higher than average substitution rates based on all sites (thin lines) because the latter includes a number of constant sites that decrease the average substitution rate.

(A) Average substitution rates based on the complete alignment. Nuclear genes have the highest substitution rates (the end of 60S ribosomal protein L3 is extremely fast), followed by plastid genes. The substitution rates for SSU nrDNA are generally low, except for a few positions.

(B) Average substitution rates based on the 25% site stripped dataset, when looking at the rates of variable sites, average substitution rates for plastid and nuclear genes are higher than those for SSU nrDNA.
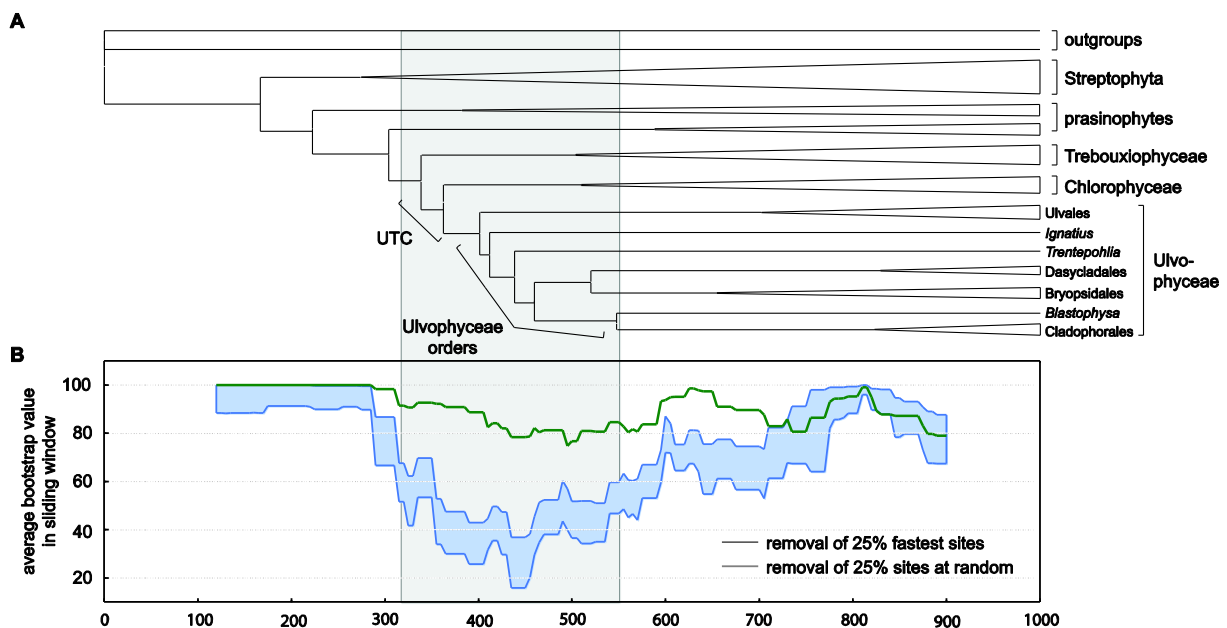
Removal of fast-evolving sites: site stripping

The rationale of removing fast-evolving sites (site stripping) is to remove noise from the data by removing those sites that are most likely to contain homoplasy and focusing on the more informative slow-evolving positions for phylogeny reconstruction (Philip et al. 2005). The usefulness of site stripping has been demonstrated in a number of studies that aimed to resolve deep nodes, for example a phylogenomic study of bilaterian animals (Delsuc et al. 2005). ML analysis of the complete bilaterian dataset (almost 150 genes) support the Coelomata hypothesis (arthropods + deuterostomes), most likely due to a long branch attraction artifact between the fast-evolving nematodes and the distant fungal outgroup. Progressive removal of fast-evolving sites resulted in increasing support for the Ecdysozoa hypothesis (arthropods + nematodes). This study showed that even advanced phylogenetic methods such as ML can be misled by differences in evolutionary rates among species.

An analysis of sitewise substitution rates in our complete alignment shows that nuclear genes evolve fastest, followed by plastid genes and SSU nrDNA, which generally has very low substitution rates (Fig. 2 A). This is consistent with Fig. 1, which shows that the phylogenetic signal at many sites of the SSU nrDNA molecule is too low for resolving deep phylogenetic relationships. On the other hand, phylogenetic signal in some parts of the nuclear genes may be lost due to high sitewise substitution rates (e.g. end of 60S ribosomal protein L3). To counteract this erosion of ancient phylogenetic signal in our dataset, we removed the 25% fastest-evolving sites to increase the signal to noise ratio. This did not change phylogenetic relationships but it improved the phylogenetic signal for the branching

order of the UTC classes and among the Ulvophyceae. The inferred relationships between the radiation of the UTC classes and the fast-evolving Bryopsidales—Dasycladales—Cladophorales orders are thus stable even when only the most reliable phylogenetic characters are used. Figure 2 B shows that for the 25% site-stripped dataset sitewise substitution rates for variable sites are higher in plastid and nuclear genes than in SSU nrDNA. Likewise plastid and nuclear genes may contain more phylogenetic signal than SSU nrDNA.

The comparison of the signal maps in Fig. 1 with rate maps (Ben Ali et al. 2001; Wuyts, Van de Peer, and Wachter 2001) indicates that signal concerning the UTC relationships is mainly found in the faster-evolving parts of the molecule. Because our site stripping approach removes these fast-evolving sites, one may expect that most sites containing phylogenetic signal concerning the UTC relationships will be removed from the SSU nrDNA alignment. However, about 60% of the sites containing phylogenetic signal concerning the UTC relationships are retained in the 25% site stripped dataset. This relatively high number of retained fast-evolving sites can be due to the lower sitewise substitution rates for SSU nrDNA compared to plastid and nuclear genes (Fig. 2). Nevertheless, SSU nrDNA alone cannot resolve the UTC relationship but it is generally accepted that the information contained in a single genes is not enough to resolve deep relationships (Philippe et al. 2005), which is one of the reason why we based this study on 10 markers and not on a single gene.

Figure 3. Difference between removal of the 25% fastest-evolving sites (site stripping) and random removal of 25% of the sites.

(A) Rate-smoothed tree with indication of the epoch in which the relationships of interest are situated (gray box across figure).

(B) The strength of the phylogenetic signal in the site-stripped alignment (black line) is much higher than that of the randomly stripped alignments (blue band) in the epoch of interest. The gray area represents the 95 percentile of signal strength (measured as average bootstrap values in a sliding window across the tree) obtained from the 100 alignments from which 25% of the sites were removed at random.

Our site stripping approach is also validated by a comparison between random removal of sites and selective removal of fast-evolving site. This exercise shows that only in the latter case phylogenetic signal is increased (Fig. 3). Random removal of one fourth of the positions in the alignment results in a global decrease of phylogenetic signal, as measured by bootstrap support.

Dating the green tree of life

To interpret the ecological, morphological and cytological diversification of the green algae in a geological timeframe, we dated our phylogeny of the Viridiplantae using relaxed molecular clock methods and a combination of the fossil record and previous molecular clock results (Fig. 4). We used the lognormal clock relaxation approach (Thorne et al. 1998) in a Bayesian framework as implemented in PhyloBayes (Lartillot et al. 2007). This method assumes autocorrelated rates of molecular evolution, more specifically that the logarithm of the rate of molecular evolution can be described by a Brownian motion process. This model indicates a rapid radiation of the UTC classes within a timeframe of ca. 20 my during the first period of the Neoproterozoic (between 900 my [852-957] and 881 my [835-935]). This rapid radiation of the UTC classes complicates the interpretation of ecological diversifications. A broader taxon sampling within the Trebouxiophyceae and Chlorophyceae may allow predictions about the number of times transitions from marine to freshwater and terrestrial habitats occurred. The inferred radiation time for the UTC classes precedes the Cryogenian period (850-635 my), in the middle of the Neoproterozoic, which is known as a period of severe glaciations during which ice sheets reached the equator, events better known as "Snowball Earth".

We show a Neoproterozoic diversification of all ulvophycean orders in a timeframe of ca. 130 my (between 826 my [792-870] and 596 my [500-686]). Originating from an ancestral marine prasinophyte, our phylogenetic tree indicates that transitions to freshwater and terrestrial habitats not only occurred in the Chlorophyceae and Trebouxiophyceae, but also several times independently within the predominantly marine Ulvophyceae (Lopez-Bautista et al. 2007, Becker and Marin 2009). A number of representatives within the Ulvales, Ulotrichales and Cladophorales are adapted to freshwater environments. *Ignatius*, all members of the Trentepohliales and two species of Cladophorales (*Cladophorella* and *Spongiochrysis*) are growing in subaerial habitats (Fritsch 1944, Lopez-Bautista and Chapman 2003, Rindi et al. 2006, Watanabe and Nakayama 2007).

Multicellularity evolved independently in the UTC classes. Within the Ulvophyceae, multicellularity arose independently in the Ulvales—Ulotrichales. In chapter 6, we calibrated a five-locus phylogeny of the siphonous orders Dasycladales and Bryopsidales using relaxed molecular clock methods calibrated with the fossil record. These models indicate a late Neoproterozoic or early Cambrian origin of the Dasycladales and Bryopsidales (571 million years [628–510]) and a diversification of the orders into their families during the Paleozoic. The siphonous thallus structure, which is essentially composed of a single giant cell containing numerous nuclei, is most likely derived from a multinucleate and multicellular common ancestor of the Bryopsidales—Dasycladales—Cladophorales (BCD clade). Additional support for the evolution of siphonous algae from a multicellular ancestor is provided by the occurrence of a cross wall at the base of each reproductive structure in one of the two major bryopsidalean lineages (the Bryopsidineae).
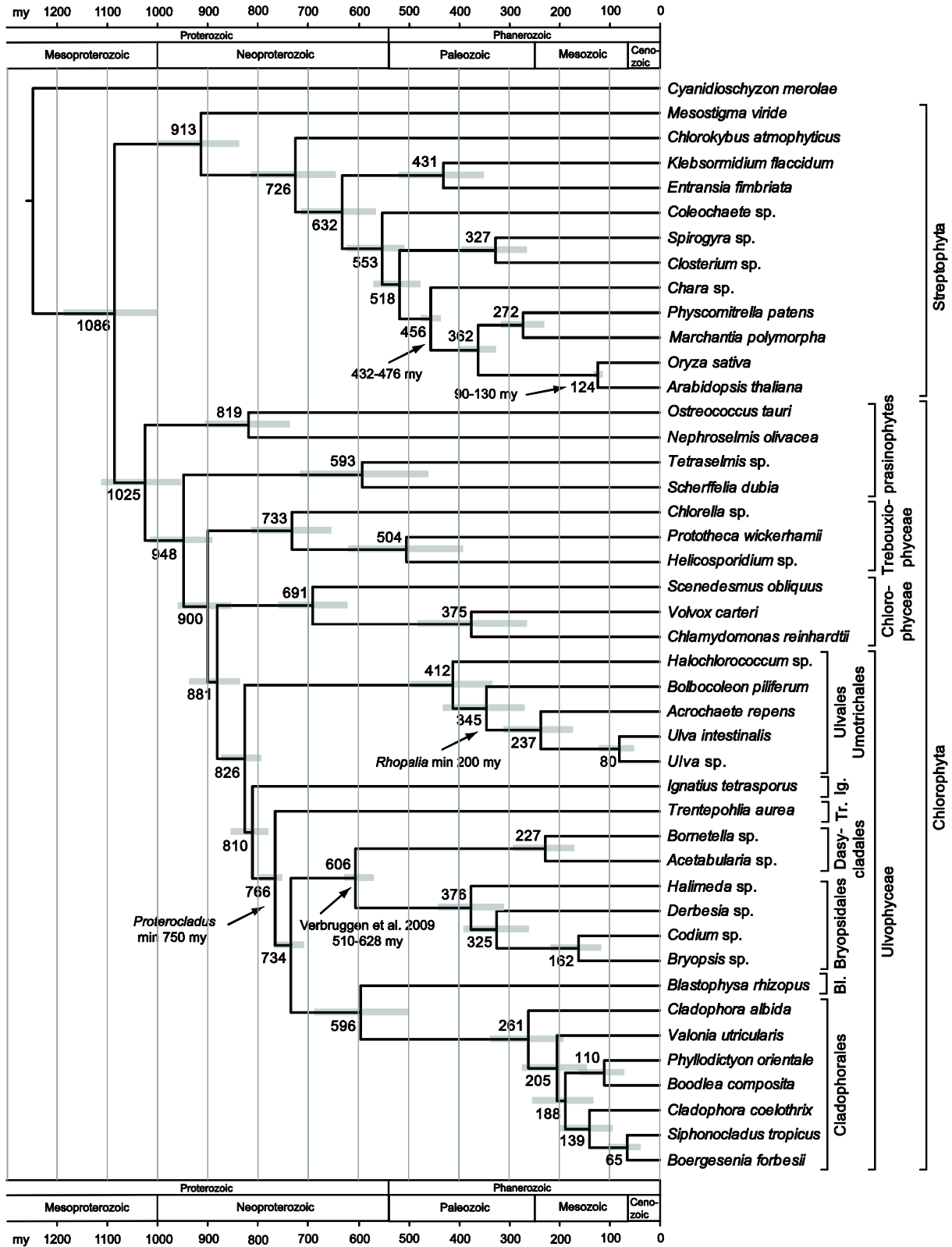
Figure 4. Chronogram of the Viridiplantae. Node ages were inferred using Bayesian inference assuming a relaxed molecular clock and a set of node age constraints derived from the fossil record and from Verbruggen et al. (2009). These calibration points are indicated on the tree. Values at nodes indicate average node ages and bars represent 95% confidence intervals.

In the light of this interpretation, a reinterpretation of the fossil *Proterocladus* seems in place. This fossil, which is characterized by large cells with cross-walls, has been interpreted as an ancestor of the Cladophorales (Butterfield et al. 1994). Our results now suggest that the ancestors of the BCD clade also featured this morphology and we propose that *Proterocladus* should be regarded as an ancestor of the BCD clade rather than of the Cladophorales. This is the interpretation we have used in our molecular clock study.

Phylogenetic position of *Blastophysa* warrants the recognition of a new order

The enigmatic, endophytic green alga *Blastophysa*, is usually placed in the cladophoralean family Chaetosiphonaceae based on morphological, ultrastructural, cytological, and biochemical features (O'Kelly and Floyd 1984). However, the phylogenetic position of *Blastophysa* has been doubted since its establishment. First, the monophyly of the family Chaetosiphonaceae is questionable due to differences in ultrastructural features of motile cells between *Blastophysa* and *Chaetosiphon*. Ultrastructural features of the motile cells of *Blastophysa* are almost identical to these of the Cladophorales and Dasycladales and have less features in common with the Bryopsidales (Chappell et al. 1991). *Blastophysa* has often been allied with the Bryopsidales in taxonomic treatments of seaweed flora's (e.g. Burrows 1991, Brodie et al. 2007, Kraft 2007). Although, Parker (1970) suggested that *Blastophysa* is not related to the Bryopsidales based on the presence of cellulose I in the cell walls (Burrows 1991). Finally, a phylogeny based on SSU and LSU nrDNA placed *Blastophysa* at the base of the Cladophorales, Dasycladales and Bryopsidales (Zechman et al. 1990).

Our multi-locus phylogeny of the Viridiplantae (Chapter 2), shows a sister relation between *Blastophysa* and the order Cladophorales. The unique ultrastructural features, along with the divergent phylogenetic position of *Blastophysa* would warrant the recognition of a separate order of Ulvophyceae.

Molecular evolution in green algae

There are several indications that profound changes to the translational system of the Ulvophyceae have occurred, more specifically in the siphonous and siphonocladous seaweed Bryopsidales, Cladophorales and Dasycladales (BCD clade) and their sister lineages *Ignatius* and Trentepohliales. First, we showed that their elongation factors, key genes of the translational apparatus that bind to the tRNAs before they attach to the ribosome, are fundamentally different than those of other green algae. In Chapter 3, we showed that **elongation factor-1 alpha** (EF-1α) and **elongation factor-like** (EFL) have an almost mutually exclusive distribution in the green plant lineage. The Streptophyta posses EF-1α except *Mesostigma*, which has EFL. All Chlorophyta encode EFL except the Ulvophyceae of the BCD clade and *Ignatius*. Due to amplification failure we could not find out which gene is present in *Trentepohlia* but its phylogenetic position strongly suggests the presence of EF-1α. The fact that the BCD clade and *Ignatius* use another elongation factor gene is in itself an indication for a change in the translational system. In Chapter 4, we discussed the presence and distribution of a **non-canonical code**, which translates stop codons TAG and TAA to glutamine. This non-canonical

code is found in Dasycladales, Cladophorales + *Blastophysa* and Trentepohliales. The remaining Ulvophyceae and green algae use the standard code. The presence of a different code requires changes in translation termination, tRNA-species and tRNA synthetase specificity, all factors influencing the translational system. In Chapter 5, we analyzed the evolution of **codon usage bias** and showed that codon usage is less biased in multicellular and macroscopic species compared to small unicells, indicating a more uniform use of synonymous codons in the former organisms. Within the Chlorophyta, the prasinophytes, Trebouxiophyceae and Chlorophyceae have strong codon usage bias, whereas codon usage bias is low in all Ulvophyceae, except for the small colonial *Ignatius*. The low codon usage bias observed for most Ulvophyceae indicates a uniform use of synonymous codons, implying that the tRNA pool is diverse, which impacts the efficiency of the translational system. In Chapter 5, we also showed that **GC content** shows a similar trend as codon usage bias, the Ulvophyceae having lower GC content than the prasinophytes, Trebouxiophyceae and Chlorophyceae. This trend is most pronounced in the Trentepohliales, Dasycladales and Bryopsidales. The link with the translational system is not directly clear but a low GC content may facilitate denaturation of the DNA during replication and transcription. Finally, **rates of molecular evolution** are high in the Trentepohliales and BCD clade as indicated by the long root-to-tip path lengths in our phylogeny. These high rates of molecular evolution suggest that more mutations are fixed in these groups. It is striking that root-to-tip paths are particularly longer for the Trentepohliales and BCD clade in SSU nrDNA phylogenies compared to plastid gene phylogenies (data not shown). Because SSU nrDNA is the structural RNA for the small component of eukaryotic cytoplasmic ribosomes, this provides another indication for changes in the translational apparatus. Taken together, this evidence points to profound changes in the translational apparatus.

Perspectives for future research

The approach applied in this study started with mining the available green algal genomic data (e.g. EST, complete genomes) and a newly generated cDNA library of a siphonocladous ulvophyte. This allowed designing primers that could amplify nuclear genes in a wide range of green algae. Amplification starting from genomic DNA was difficult, if not impossible, mainly due to the omnipresence of introns. The introns hampered the design of primers applicable to a wide range of taxa at the genomic DNA level because length and position of introns are highly variable between different green algal species (e.g. actin). De novo sequencing of nuclear genomes of a selection of green algae (e.g. with Roche 454 technology) could alleviate this need by permitting the identification of intronless genes and regions suitable for exon-primed intron crossing (EPIC).

Next-generation sequencing technologies (NGST) parallelize the sequencing procedure, allowing rapid sequencing of billions of nucleotides simultaneously (e.g. Roche 454 technology, Rokas and Abbot 2009). NGST data for a balanced set of green algae is likely to provide a major boost to the assembly of the green tree of life. This brute-force approach is probably more efficient than primer design and amplification of nuclear genes as was attempted in this study. In addition to serving the resolution of the tree of life, NGST data can also be used to explore several other aspects of genome evolution. NGST data are appropriate to evaluate the importance and relative contributions of processes such as lineage sorting, hybridization and lateral gene transfer in shaping the respective

genomes (Rokas and Abbot 2009). Also the origin of new genes or gene clusters and the evolution of introns and repetitive elements can be studied.

NGST data for Ulvophyceae, especially siphonous and siphonocladous species, as well as Trebouxiophyceae would be particularly useful since genomic data for these groups are lacking or sparse. Most available green algal genomic data come from prasinophytes and Chlorophyceae, e.g. the recently released or nearly finished genomes of *Chlamydomonas* and *Volvox* (Chlorophyceae) and *Bathycoccus* and *Micromonas* (prasinophytes).

With respect to my study, NGST data would allow testing the reliability of the observed differences in codon usage bias and GC content among green algae (Chapter 5). A comparison of genes that regulate translation (e.g. eukaryotic release factor, glutamine tRNA genes, genes responsible for nonsense-mediated mRNA decay) among the various ulvophyte lineages would be particularly interesting because this will further increase our knowledge about the profound changes in translation machinery that occurred in these lineages as is indicated by the presence of a non-canonical genetic code, a different elongation factor gene, fast rates of molecular evolution especially of the ribosomal RNA (SSU nrDNA) and a more balanced codon usage.

Despite that NGST data for a balanced set of green algae would provide a lot of useful information, continuing the approach followed in this study can in the short term increase our knowledge about green algal phylogenetic. In a next stage it is advisable to incorporate more prasinophytess, Trebouxiophyceae and Chlorophyceae. A broader taxon sampling within the Trebouxiophyceae and Chlorophyceae would allow definitive conclusion about the monophyletic nature of these classes and could demonstrate whether the relationship between the UTC classes is robust to differences in taxon sampling. Although the class Ulvophyceae was covered by 21 species, also here improvements in taxon sampling can be beneficial. For example, *Oltmannsiellopsis* is considered to be an ulvophyte based on SSU nrDNA and complete chloroplast genome sequences (Pombert et al. 2006) despite having ultrastructural features reminiscent of prasinophytes, especially of the genus *Tetraselmis* (O'Kelly personal communication). Incorporation of *Oltmannsiellopsis* would therefore provide further insight into the evolution of the ulvophyte lineages.

With respect to the non-canonical genetic code in some ulvophyceaen lineages, studying selenoproteins could also prove interesting. The amino acid selenocysteine is encoded by the stop codon TGA in conjunction with a selenocysteine insertion sequence (SECIS) element in the 3'UTR (Lobanov et al. 2007). The non-canonical code of some Ulvophyceae involves the reassignment of stop codons TAG and TAA to glutamine, leaving TGA as the only stop codon. Because the Ulvophyceae only use TGA as stop codon, the last stop codon in a selenoprotein should still terminate translation while the in-frame TGA codons should encode selenocysteine in the presence of a SECIS element. In this context, it would be useful to know whether selenoproteins are converted back to conventional cysteine-containing proteins in Ulvophyceae with a non-canonical code.

References

Abascal F, Zardoya R, and Posada D. 2005. ProtTest: selection of best-fit models of protein evolution. Bioinformatics **21**:2104-2105.

Abbott IA and Huisman JM. 2004. Marine green and brown algae of the Hawaiian Islands. Bishop Museum Press, Honolulu.

Adl SM, Simpson AGB, Farmer MA, Andersen RA, Anderson OR, Barta JR, Bowser SS, Brugerolle G, Fensome RA, Fredericq S, James TY, Karpov S, Kugrens P, Krug J, Lane CE, Lewis LA, Lodge J, Lynn DH, Mann DG, McCourt RM, Mendoza L, Moestrup Ø, Mozley-Standridge SE, Nerad TA, Shearer CA, Smirnov AV, Spiegel FW, and Taylor M. 2005. The new higher level classification of eukaryotes with emphasis on the taxonomy of protists. Journal of Eukaryotic Microbiology **52**:399-451.

Akashi H. 1995. Inferring weak selection from patterns of polymorphism and divergence at silent sites in *Drosophila* DNA. Genetics **139**:1067-1076.

An SS, Mopps B, Weber K, and Bhattacharya D. 1999. The origin and evolution of green algal and plant actins. Molecular Biology and Evolution **16**:275-285.

Andersen R. 2005. Algal culturing techniques. Elsevier Academic Press.

Andersson JO, Sjogren AM, Davis LAM, Embley TM, and Roger AJ. 2003. Phylogenetic analyses of diplomonad genes reveal frequent lateral gene transfers affecting eukaryotes. Current Biology **13**:94-104.

Andersson JO. 2005. Lateral gene transfer in eukaryotes. Cellular and Molecular Life Sciences **62**:1182-1197.

Angellotti MC, Bhuiyan SB, Chen G, and Wan XF. 2007. CodonO: codon usage bias analysis within and across genomes. Nucleic Acids Research **35**:W132-W136.

Archibald JM. 2005. Jumping genes and shrinking genomes - Probing the evolution of eukaryotic photosynthesis with genomics. IUBMB Life **57**:539-547.

Archibald JM. 2008. Plastid evolution: Remnant algal genes in ciliates. Current Biology **18**:R663-R665.

Areschoug JE. 1866. Observationes phycologicae. Particula prima. De Confervaceis nonnullis. Nova Acta Regiae Societatis Scientiarum Upsaliensis **Ser 3, 6(2)**:1-26.

Badour SS. 1981. The inorganic carbon requirements of *Chlamydomonas segnis* (Chlorophyceae) for cell development in synchronous cultures. Journal of Phycology **17**:293-299.

Baker KE and Parker R. 2004. Nonsense-mediated mRNA decay: terminating erroneous gene expression. Current Opinion in Cell Biology **16**:293-299.

Baldauf SL, Palmer JD, and Doolittle WF. 1996. The root of the universal tree and the origin of eukaryotes based on elongation factor phylogeny. Proceedings of the National Academy of Sciences of the United States of America **93**:7749-7754.

Baldauf SL and Doolittle WF. 1997. Origin and evolution of the slime molds (Mycetozoa). Proceedings of the National Academy of Sciences of the United States of America **94**:12007-12012.

Baldauf SL. 2008. An overview of the phylogeny and diversity of eukaryotes. Journal of Systematics and Evolution **46**:263-273.

Barker D, Meade A, and Pagel M. 2007. Constrained models of evolution lead to improved prediction of functional linkage from correlated gain and loss of genes. Bioinformatics **23**:14-20.

Bartsch I and Kuhlenkamp R. 2000. The marine macroalgae of Helgoland (North Sea): an annotated list of records between 1845 and 1999. Helgoland Marine Research **54**:160-189.

Becker B and Marin B. 2009. Streptophyte algae and the origin of embryophytes. Annals of Botany **103**:999–1004.

Beier H and Grimm M. 2001. Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. Nucleic Acids Research **29**:4767-4782.

Beltran M, Jiggins CD, Brower AVZ, Bermingham E, and Mallet J. 2007. Do pollen feeding, pupal-mating and larval gregariousness have a single origin in *Heliconius* butterflies? Inferences from multilocus DNA sequence data. Biological Journal of the Linnean Society **92**:221-239.

Ben Ali A, De Baere R, Van der Auwera G, De Wachter R, and Van de Peer Y. 2001. Phylogenetic relationships among algae based on complete large-subunit rRNA sequences. International Journal of Systematic and Evolutionary Microbiology **51**:737-749.

Berger S and Liddle LB. 2003. The life cycle of *Acetabularia* (Dasycladales, Chlorophyta): textbook accounts are wrong. Phycologia **42**:204-207.

Berney C and Pawlowski J. 2006. A molecular time-scale for eukaryote evolution recalibrated with the continuous microfossil record. Proceedings of the Royal Society B-Biological Sciences **273**:1867-1872.

Bhattacharya D, Stickel SK, and Sogin ML. 1993. Isolation and molecular phylogenetic analysis of actin coding regions from *Emiliania huxleyi*, a prymnesiophyte alga, by reverse transcriptase and PCR methods. Molecular Biology and Evolution **10**:689-703.

Bhattacharya D, Friedl T, and Damberger S. 1996. Nuclear-encoded rDNA group I introns: Origin and phylogenetic relationships of insertion site lineages in the green algae. Molecular Biology and Evolution **13**:978-989.

Bhattacharya D, Weber K, An SS, and Berning-Koch W. 1998. Actin phylogeny identifies *Mesostigma viride* as a flagellate ancestor of the land plants. Journal of Molecular Evolution **47**:544-550.

Bhattacharya D, Aubry J, Twait EC, and Jurk S. 2000. Actin gene duplication and the evolution of morphological complexity in land plants. Journal of Phycology **36**:813-820.

Bold HC and Wynne MJ. 1985. Introduction to the Algae, 2nd edition. Prentice-Hall, New Jersey.

Booton GC, Floyd GL, and Fuerst PA. 1998. Origins and affinities of the filamentous green algal orders Chaetophorales and Oedogoniales based on 18S rRNA gene sequences. Journal of Phycology **34**:312-318.

Brinkmann H, Van der Giezen M, Zhou Y, De Raucourt GP, and Philippe H. 2005. An empirical assessment of long-branch attraction artefacts in deep eukaryotic phylogenomics. Systematic Biology **54**:743-757.

Brodie J, Maggs CA, and John DM. 2007. Green seaweeds of Britain and Ireland. British Phycological Society, London.

Brown JM and Lemmon AR. 2007. The importance of data partitioning and the utility of bayes factors in Bayesian phylogenetics. Systematic Biology **56**:643-655.

Bulmer M. 1991. The selection-mutation-drift theory of synonymous codon usage. Genetics **129**:897-907.

Burrows EM. 1991. Seaweeds of the British Isles. Natural History Museum Publications, London.

Butterfield NJ, Knoll AH, and Swett K. 1994. Paleobiology of the Neoproterozoic Svanbergfjellet formation, Spitsbergen. Fossils and Strata **34**:1-84.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution **17**:540-552.

Cavalier-Smith T. 1981. Eukaryote kingdoms: Seven or nine? Biosystems **14**:461-481.

Chappell DF, O'Kelly CJ, and Floyd GL. 1991. Flagellar apparatus of the biflagellate zoospores of the enigmatic marine green alga *Blastophysa rhizopus*. Journal of Phycology **27**:423-428.

Chaudhary BT. 1979. Some observations on the morphology, reproduction and cytology of the genus *Uronema* Lagh. (Ulotrichales, Chlorophyceae) (Note). Phycologia **18**:299-302.

Chen SL, Lee W, Hottes AK, Shapiro L, and McAdams HH. 2004. Codon usage between genomes is constrained by genome-wide mutational processes. Proceedings of the National Academy of Sciences of the United States of America **101**:3480-3485.

Chisholm JRM, Dauga C, Ageron E, Grimont PAD, and Jaubert JM. 1996. 'Roots' in mixotrophic algae. Nature **381**:382-382.

Cocquyt E, Verbruggen H, Leliaert F, Zechman FW, Sabbe K, and De Clerck O. 2009. Gain and loss of elongation factor genes in green algae. BMC Evolutionary Biology **9**:39.

Cocquyt E, Gile GH, Leliaert F, Verbruggen H, Keeling P, and De Clerck O. in prep. Complex phylogenetic distribution of a non-canonical genetic code in green algae.

Cocquyt E, Verbruggen H, Leliaert F, and De Clerck O. submitted. Ancient relationships among green algae inferred from nuclear and chloroplast genes.

Cohen J and Adoutte A. 1995. Why does the genetic code deviate so easily in ciliates? Biology of the Cell **85**:105-108.

Collins FS and Hervey AB. 1917. The algae of Bermuda. Proceedings of the National Academy of Sciences of the United States of America **53**.

Coppejans E. versie 1998-1999. Syllabus Morfologie en Systematiek van de "Lagere Planten". Partim Algologie. Universiteit Gent.

Cutter AD, Wasmuth JD, and Blaxter ML. 2006. The evolution of biased codon and amino acid usage in nematode genomes. Molecular Biology and Evolution **23**:2303-2315.

Daugbjerg N, Moestrup Ø, and Arctander P. 1995. Phylogeny of genera of Prasinophyceae and Pedinophyceae (Chlorophyta) deduced from molecular analysis of the *rbc*L gene. Phycological Research **43**:203–213.

Dawes CJ and Mathieson AC. 2008. The seaweeds of Florida. University Press of Florida, Gainesville, Florida.

De Clerck O, Verbruggen H, Huisman JM, Faye EJ, Leliaert F, Schils T, and Coppejans E. 2008. Systematics and biogeography of the genus *Pseudocodium* (Bryopsidales, Chlorophyta), including the description of *P. natalense* sp nov from South Africa. Phycologia **47**:225-235.

de Koning AP, Noble GP, Heiss AA, Wong J, and Keeling PJ. 2008. Environmental PCR survey to determine the distribution of a non-canonical genetic code in uncultivable oxymonads. Environmental Microbiology **10**:65-74.

De Rijk P and De Wachter R. 1993. DCSE, an interactive tool for sequence alignment and secondary structure research. Comput. Appl. Biosci. **9**:735-740.

De Rijk P, Robbrecht E, de Hoog S, Caers A, Van de Peer Y, and De Wachter R. 1999. Database on the structure of large subunit ribosomal RNA. Nucleic Acids Research **27**:174-178.

Delsuc F, Brinkmann H, and Philippe H. 2005. Phylogenomics and the reconstruction of the tree of life. Nature Reviews Genetics **6**:361-375.

Derelle E, Ferraz C, Rombauts S, Rouze P, Worden AZ, Robbens S, Partensky F, Degroeve S, Echeynie S, Cooke R, Saeys Y, Wuyts J, Jabbari K, Bowler C, Panaud O, Piegu B, Ball SG, Ral JP, Bouget FY, Piganeau G, De Baets B, Picard A, Delseny M, Demaille J, Van de Peer Y, and Moreau H. 2006. Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. Proceedings of the National Academy of Sciences of the United States of America **103**:11647-11652.

DOE Joint Genome Institute. www.jgi.doe.gov

dos Reis M and Wernisch L. 2009. Estimating translational selection in eukaryotic genomes. Molecular Biology and Evolution **26**:451-461.

Douzery EJP, Snell EA, Bapteste E, Delsuc F, and Philippe H. 2004. The timing of eukaryotic evolution: Does a relaxed molecular clock reconcile proteins and fossils? Proceedings of the National Academy of Sciences of the United States of America **101**:15386-15391.

Drew EA and Abel KM. 1992. Studies on *Halimeda* IV. An endogenous rhythm of chloroplast migration in the siphonous green-alga, *H. distorta*. Journal of Interdisciplinary Cycle Research **23**:128-135.

Driskell AC, Ane C, Burleigh JG, McMahon MM, O'Meara BC, and Sanderson MJ. 2004. Prospects for building the tree of life from large sequence databases. Science **306**:1172-1174.

Duret L and Mouchiroud D. 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis, Drosophila, Arabidopsis*. Proceedings of the National Academy of Sciences of the United States of America **96**:4482-4487.

Dybvig K, Cao Z, French CT, and Yu HL. 2007. Evidence for type III restriction and modification systems in *Mycoplasma pulmonis*. Journal of Bacteriology **189**:2197-2202.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research **32**:1792-1797.

Fawley MW, Yun Y, and Qin M. 2000. Phylogenetic analyses of 18S rDNA sequences reveal a new coccoid lineage of the Prasinophyceae (Chlorophyta). Journal of Phycology **36**:387-393.

Feldmann J. 1954. Inventaire de la flore marine de Roscoff. Algues, champignons, lichens et spermatophytes. Trav. St. Biol. Roscoff, Nouv. Sér. Suppl. **6**:152.

Felsenstein J. 1985. Confidence-limits on phylogenies - an approach using the bootstrap. Evolution **39**:783-791.

Felsenstein J. 2005. PHYLIP (Phylogeny Inference Package) version 3.6. http://evolution.genetics.washington.edu/phylip/

Fenchel T and Finlay BJ. 2004. The ubiquity of small species: Patterns of local and global diversity. Bioscience **54**:777-784.

Finlay BJ. 2002. Global dispersal of free-living microbial eukaryote species. Science **296**:1061-1063.

Floyd GL and O' Kelly CJ. 1984. Motile cell ultrastructure and the circumscription of the orders Ulotrichales and Ulvales (Ulvophyceae, Chlorophyta). American Journal of Botany **71**:111–120.

Foslie M. 1881. Om nogle nye arctiske havalger. Christiania Vidensk.-Selsk. Forh. **14**:1-14.

Fott B. 1971. Algenkunde. 2nd Ed. VEB Fischer, Jena.

Friedl T. 1995. Inferring taxonomic positions and testing genus level assignments in coccoid green lichen algae - a phylogenetic analysis of 18S ribosomal-RNA sequences from *Dictyochloropsis reticulata* and from members of the genus *Myrmecia* (Chlorophyta, Trebouxiophyceae cl. nov.). Journal of Phycology **31**:632-639.

Friedl T and O'Kelly CJ. 2002. Phylogenetic relationships of green algae assigned to the genus *Planophila* (Chlorophyta): evidence from 18S rDNA sequence data and ultrastructure. European Journal of Phycology **37**:373-384.

Fritsch FE. 1944. *Cladophorella calcicola* nov. gen. et sp., a terrestrial member of the Cladophorales. Annals of Botany **N.S. 8**:157-171.

Galtier N. 2007. A model of horizontal gene transfer and the bacterial phylogeny problem. Systematic Biology **56**:633-642.

Geuten K, Massingham T, Darius P, Smets E, and Goldman N. 2007. Experimental design criteria in phylogenetics: where to add taxa. Systematic Biology **56**:609-622.

Gile GH, Patron NJ, and Keeling PJ. 2006. EFL GTPase in cryptomonads and the distribution of EFL and EF-1α in chromalveolates. Protist **157**:435-444.

Gile GH, Novis P, Cragg D, Zuccarello GC, and Keeling P. 2009. The distribution of elongation factor-1alpha (EF-1α), elongation factor-like (EFL), and a non-canonical genetic code in the Ulvophyceae: discrete genetic characters support a consistent phylogenetic framework. Journal of Eukaryotic Microbiology **in press**.

Gogarten JP. 2003. Gene transfer: Gene swapping craze reaches eukaryotes. Current Biology **13**:R53-R54.

Goodson HV and Hawse WF. 2002. Molecular evolution of the actin family. Journal of Cell Science **115**:2619-2622.

Graham LE, Graham JM, and Wilcox LW. 2009. Algae. Benjamin Cummings, San Francisco, Boston, New York, Cape Town, Hong Kong, London, Madrid, Mexico City, Montreal, Munich, Paris, Singapore, Sydney, Tokyo, Toronto.

Grauvogel C, Brinkmann H, and Petersen J. 2007. Evolution of the glucose-6-phosphate isomerase: The plasticity of primary metabolism in photosynthetic eukaryotes. Molecular Biology and Evolution **24**:1611-1621.

Guillou L, Eikrem W, Chretiennot-Dinet MJ, Le Gall F, Massana R, Romari K, Pedros-Alio C, and Vaulot D. 2004. Diversity of picoplanktonic prasinophytes assessed by direct nuclear SSU rDNA sequencing of environmental samples and novel isolates retrieved from oceanic and coastal marine ecosystems. Protist **155**:193-214.

Guindon S and Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology **52**:696-704.

Guiry MD and Guiry GM. 2009. AlgaeBase. World-wide electronic publication. National University of Ireland, Galway. http://www.algaebase.org

Gunderson JH, Elwood HJ, Ingold A, Kindle KL, and Sogin ML. 1987. Phylogenetic relationships between chlorophytes, chrysophytes, and Oomycetes. Proceedings of the National Academy of Sciences of the United States of America **84**:5823-5827.

Hall JD and Delwiche CF. 2007. In the shadow of giants: systematics of the charophyte green algae. Pp. 155-169 *in* J. Brodie, and J. Lewis, eds. Unravelling the algae: the past, present, and future of algal systematics. The Systematics Association Special Volume Series, 75. CRC Press, Boca Raton, FL, USA.

Hanyu N, Kuchino Y, Nishimura S, and Beier H. 1986. Dramatic events in ciliate evolution - alteration of UAA and UAG termination codons to glutamine codons due to anticodon mutations in two *Tetrahymena* tRNAs-Gln. EMBO Journal **5**:1307-1311.

Hanyuda T, Arai S, and Ueda K. 2000. Variability in the *rbc*L introns of Caulerpalean algae (Chlorophyta, Ulvophyceae). Journal of Plant Research **113**:403-413.

Hanyuda T, Wakana I, Arai S, Miyaji K, Watano Y, and Ueda K. 2002. Phylogenetic relationships within Cladophorales (Ulvophyceae, Chlorophyta) inferred from 18S rRNA gene sequences, with special reference to *Aegagropila linnaei*. Journal of Phycology **38**:564-571.

Hashimoto T, Nakamura Y, Nakamura F, Shirakura T, Adachi J, Goto N, Okamoto K, and Hasegawa M. 1994. Protein phylogeny gives a robust estimation for early divergences of eukaryotes: phylogenetic place of a mitochondria-lacking protozoan, *Giardia lamblia*. Molecular Biology and Evolution **11**:65-71.

Hedges SB, Blair JE, Venturi ML, and Shoe JL. 2004. A molecular timescale of eukaryote evolution and the rise of complex multicellular life. BMC Evolutionary Biology **4**.

Herron MD, Hackett JD, Aylward FO, and Michod RE. 2009. Triassic origin and early radiation of multicellular volvocine algae. Proceedings of the National Academy of Sciences of the United States of America **106**:3254-3258.

Hershberg R and Petrov DA. 2008. Selection on codon bias. Annual Review of Genetics **42**:287-299.

Higgs PG and Ran WQ. 2008. Coevolution of codon usage and tRNA genes leads to alternative stable states of biased codon usage. Molecular Biology and Evolution **25**:2279-2291.

Holder M and Lewis PO. 2003. Phylogeny estimation: Traditional and Bayesian approaches. Nature Reviews Genetics **4**:275-284.

Hori T. 1994. An illustrated atlas of the life history of algae. Volume 1. Green Algae. Uchida Rokakuho Publishing Co., Ltd., Tokyo.

Horowitz S and Gorovsky MA. 1985. An An unusual genetic code in nuclear genes of *Tetrahymena*. Proceedings of the National Academy of Sciences of the United States of America **82**:2452-2455.

Inagaki Y, Blouin C, Doolittle WF, and Roger AJ. 2002. Convergence and constraint in eukaryotic release factor 1 (eRF1) domain 1: the evolution of stop codon specificity. Nucleic Acids Research **30**:532-544.

Index Nominum Algarum UH, University of California, Berkeley. Compiled by Paul Silva. Available online at http://ucjeps.berkeley.edu/INA.html.

Ingvarsson PK. 2008. Molecular evolution of synonymous codon usage in *Populus*. BMC Evolutionary Biology **8**:307.

Irvine D and John D. 1984. The Systematics of Green Algae. Academic Press, London.

James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, Celio G, Gueidan C, Fraker E, Miadlikowska J, Lumbsch HT, Rauhut A, Reeb V, Arnold AE, Amtoft A, Stajich JE, Hosaka K, Sung GH, Johnson D, O'Rourke B, Crockett M, Binder M, Curtis JM, Slot JC, Wang Z, Wilson AW, Schussler A, Longcore JE, O'Donnell K, Mozley-Standridge S, Porter D, Letcher PM, Powell MJ, Taylor JW, White MM, Griffith GW, Davies DR, Humber RA, Morton JB, Sugiyama J, Rossman AY, Rogers JD, Pfister DH, Hewitt D, Hansen K, Hambleton S, Shoemaker RA, Kohlmeyer J, Volkmann-Kohlmeyer B, Spotts RA, Serdani M, Crous PW, Hughes KW, Matsuura K, Langer E, Langer G, Untereiner WA, Lucking R, Budel B, Geiser DM, Aptroot A, Diederich P, Schmitt I, Schultz M, Yahr R, Hibbett DS, Lutzoni F, McLaughlin DJ, Spatafora JW, and Vilgalys R. 2006. Reconstructing the early evolution of Fungi using a six-gene phylogeny. Nature **443**:818-822.

Jansen RK, Cai Z, Raubeson LA, Daniell H, Depamphilis CW, Leebens-Mack J, Muller KF, Guisinger-Bellian M, Haberle RC, Hansen AK, Chumley TW, Lee SB, Peery R, McNeal JR, Kuehl JV, and Boore JL. 2007. Analysis of 81 genes from 64 plastid genomes resolves relationships in

angiosperms and identifies genome-scale evolutionary patterns. Proceedings of the National Academy of Sciences of the United States of America **104**:19369-19374.

Jobb G, von Haeseler A, and Strimmer K. 2004. TREEFINDER: a powerful graphical analysis environment for molecular phylogenetics. BMC Evolutionary Biology **4**.

Jobb G. 2008. TREEFINDER version of March 2008. www.treefinder.de

Jónsson S. 1962. Recherches sur des Cladophoracées marines: structure, reproduction, cycles comparés, conséquences systématiques. Annales des Sciences Naturelles, Botanique **12**:25-263.

Jónsson S. 1999. The status of the Acrosiphoniales (Chlorophyta). Rit Fiskideildar **16**:187–196.

Jorde I. 1933. Untersuchungen über den Lebenszyklus von *Urospora* Aresch. und *Codiolum* A. Braun. Nyt Magazin for Naturvidensk **73**:1-19.

Jukes TH and Osawa S. 1993. Evolutionary changes in the genetic code. Comparative Biochemistry and Physiology - Part B: Biochemistry & Molecular Biology **106**:489-494.

Kamikawa R, Inagaki Y, and Sako Y. 2008. Direct phylogenetic evidence for lateral transfer of elongation factor-like gene. Proceedings of the National Academy of Sciences of the United States of America **105**:6965-6969.

Kapraun DF and Nguyen MN. 1994. Karyology, nuclear DNA quantification and nucleus-cytoplasmic domain variations in some multinucleate green algae (Siphonocladales, Chlorophyta). Phycologia **33**:42-52.

Karol KG, McCourt RM, Cimino MT, and Delwiche CF. 2001. The closest living relatives of land plants. Science **294**:2351-2353.

Karsten U, Friedl T, Schumann R, Hoyer K, and Lembcke S. 2005. Mycosporine-like amino acids and phylogenies in green algae: *Prasiola* and its relatives from the Trebouxiophyceae (Chlorophyta). Journal of Phycology **41**:557-566.

Kass RE and Raftery AE. 1995. Bayes Factors. Journal of the American Statistical Association **90**:773-795.

Kawahara Y and Imanishi T. 2007. A genome-wide survey of changes in protein evolutionary rates across four closely related species of *Saccharomyces* sensu stricto group. BMC Evolutionary Biology **7**:9.

Keeling PJ and Doolittle WF. 1996. A non-canonical genetic code in an early diverging eukaryotic lineage. EMBO Journal **15**:2285-2290.

Keeling PJ and Leander BS. 2003. Characterisation of a non-canonical genetic code in the oxymonad *Streblomastix strix*. Journal of Molecular Biology **326**:1337-1349.

Keeling PJ and Inagaki Y. 2004. A class of eukaryotic GTPase with a punctate distribution suggesting multiple functional replacements of translation elongation factor 1α. Proceedings of the National Academy of Sciences of the United States of America **101**:15380-15385.

Keeling PJ and Palmer JD. 2008. Horizontal gene transfer in eukaryotic evolution. Nature Reviews Genetics **9**:605-618.

Kim GH, Klotchkova TA, and Kang YM. 2001. Life without a cell membrane: regeneration of protoplasts from disintegrated cells of the marine green alga *Bryopsis plumosa*. Journal of Cell Science **114**:2009-2014.

Knight R, Freeland S, and Landweber L. 2001a. A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. Genome Biology **2**:1-13.

Knight RD, Freeland SJ, and Landweber LF. 2001b. Rewiring the keyboard evolvability of the genetic code. Nature Reviews Genetics **2**:49-58.

Kolisko M, Cepicka I, Hampl V, Leigh J, Roger AJ, Kulda J, Simpson AGB, and Flegr J. 2008. Molecular phylogeny of diplomonads and enteromonads based on SSU rRNA, alpha-tubulin and HSP90 genes: Implications for the evolutionary history of the double karyomastigont of diplomonads. BMC Evolutionary Biology **8**:205.

Kornmann P. 1963. Die Ulotrichales, neu geordnet auf der Grundlage entwicklungsgeschichtlicher Befunde. Phycologia **3**:60-68.

Kornmann P and Sahling P-H. 1994. Meeresalgen von Helgoland: Zweite Ergänzung. Helgoländer Meeresuntersuchungen **48**:365-406.

Kraft GT. 2007. Algae of Australia: marine benthic algae of Lord Howe Island and the Southern Great Barrier Reef, 1: Green Algae. Australian Biological Resources Study & CSIRO Publishing, Canberra & Melbourne.

Krienitz L, Ustinova I, Friedl T, and Huss VAR. 2001. Traditional generic concepts versus 18S rRNA gene phylogeny in the green algal family Selenastraceae (Chlorophyceae, Chlorophyta). Journal of Phycology **37**:852-865.

Krienitz L, Hegewald E, Hepperle D, and Wolf M. 2003. The systematics of coccoid green algae: 18S rRNA gene sequence data versus morphology. Biologia **58**:437-446.

Lagerheim G. 1887. Note sur l'*Uronema*, nouveau genre des algues d'eau douce. Malpigia **1**:517-523.

Lake JA. 2008. Reconstructing evolutionary graphs: 3D parsimony. Molecular Biology and Evolution **25**:1677-1682.

Lane CE and Archibald JM. 2008. The eukaryotic tree of life: endosymbiosis takes its TOL. Trends in Ecology & Evolution **23**:268-275.

Lartillot N, Blanquart S, and Lepage T. 2007. PhyloBayes v2.3. http://www.lirmm.fr/mab/article.php3?id_article=329

Lein TE, Bruntse G, Gunnarsson K, and Nielsen R. 1999. New records of benthic marine algae for Norway, with notes on some rare species from the Florø district, western Norway. Sarsia **84**:39-53.

Leliaert F, Rousseau F, De Reviers B, and Coppejans E. 2003. Phylogeny of the Cladophorophyceae (Chlorophyta) inferred from partial LSU rRNA gene sequences: is the recognition of a separate order Siphonocladales justified? European Journal of Phycology **38**:233-246.

Leliaert F and Boedeker C. 2007. Cladophorales. Pp. 131–183 *in* J. Brodie, C. A. Maggs, and D. M. John, eds. Green seaweeds of Britain and Ireland. British Phycological Society, London, UK.

Leliaert F, De Clerck O, Verbruggen H, Boedeker C, and Coppejans E. 2007. Molecular phylogeny of the Siphonocladales (Chlorophyta : Cladophorophyceae). Molecular Phylogenetics and Evolution **44**:1237-1256.

Leliaert F, Boedeker C, Pena V, Bunker F, Verbruggen H, and De Clerck O. 2009. *Cladophora rhodolithicola* sp. nov. (Cladophorales, Chlorophyta), a diminutive species from European maerl beds. European Journal of Phycology **43: in press**.

Lemieux C, Otis C, and Turmel M. 2000. Ancestral chloroplast genome in *Mesostigma viride* reveals an early branch of green plant evolution. Nature **403**:649-652.

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 Biology **5**:2.

Levring T. 1937. Zur Kenntnis der Algenflora der Norwegischen Westküste. Lunds Univ. Årsskr. **33**:1-147.

Lewis LA and McCourt RM. 2004. Green algae and the origin of land plants. American Journal of Botany **91**:1535-1556.

Lewis LA and Lewis PO. 2005. Unearthing the molecular phylodiversity of desert soil green algae (Chlorophyta). Systematic Biology **54**:936–947.

Li CH, Lu GQ, and Orti G. 2008. Optimal data partitioning and a test case for ray-finned fishes (Actinopterygii) based on ten nuclear loci. Systematic Biology **57**:519-539.

Lindstrom SC and Hanic LA. 2005. The phylogeny of North American *Urospora* (Ulotrichales, Chlorophyta) based on sequence analysis of nuclear ribosomal genes, introns and spacers. Phycologia **44**:194-201.

Littler DS, Littler MM, and Hanisak MD. 2008. Submersed plants of the Indian River Lagoon: A floristic inventory and field guide. OffShore Graphics, Inc., Washington, DC.

Littler MM, Littler DS, and Lapointe BE. 1988. A comparison of nutrient-limited and light-limited photosynthesis in psammophytic versus epilithic forms of *Halimeda* (Caulerpales, Halimedaceae) from the Bahamas. Coral Reefs **6**:219-225.

Lobanov AV, Fomenko DE, Zhang Y, Sengupta A, Hatfield DL, and Gladyshev VN. 2007. Evolutionary dynamics of eukaryotic selenoproteomes: large selenoproteomes may associate with aquatic life and small with terrestrial life. Genome Biology **8**:R198.

Lokhorst GM and Trask BJ. 1981. Taxonomic studies on *Urospora* (Acrosiphoniales, Chlorophyceae) in western Europe. Acta Botanica Neerlandica **30**:353-431.

Lopez-Bautista JM and Chapman RL. 2003. Phylogenetic affinities of the Trentepohliales inferred from small-subunit rDNA. International Journal of Systematic and Evolutionary Microbiology **53**:2099-2106.

Lopez-Bautista JM, Rindi F, and Casamatta D. 2007. The subaerial algae. Pp. 601-617 *in* J. Seckbach, ed. Algae and Cyanobacteria in Extreme Environments. Springer.

Lozupone CA, Knight RD, and Landweber LF. 2001. The molecular basis of nuclear genetic code change in ciliates. Current Biology **11**:65-74.

Maggs CA and O' Kelly CJ. 2007. *Uronema*. Pp. 209-210 *in* J. Brodie, C. A. Maggs, and D. M. John, eds. Green seaweeds of Britain and Ireland. British Phycological Society, London, UK.

Mann DG. 1996. Crossing the Rubicon: the effectiveness of the marine/freshwater interface as a barrier to the migration of diatom germplasm. Pp. 1-21 *in* S. Mayama, M. Idei, and I. Koizumi, eds. Fourteenth International Diatom Symposium. Koeltz Scientific Books, Tokyo, Japan.

Maquat LE. 2004. Nonsense-mediated mRNA decay: Splicing, translation and mRNP dynamics. Nature Reviews Molecular Cell Biology **5**:89-99.

Marin B and Melkonian M. 1999. Mesostigmatophyceae, a new class of streptophyte green algae revealed by SSU rRNA sequence comparisons. Protist **150**:399-417.

Martins EP. 1999. Estimation of ancestral states of continuous characters: A computer simulation study. Systematic Biology **48**:642-650.

Matsuyama K, Matsuoka T, Miyaji K, Tanaka J, and Aruga Y. 1998. Ultrastructure of the pyrenoid in the family Cladophoraceae (Cladophorales, Chlorophyta). Journal of Japanese Botany **73**:279-286.

Mattox K and Stewart K. 1984. Classification of the green algae: a concept based on comparative cytology. Pp. 29–72 *in* D. Irvine, and D. John, eds. Systematics of the green algae. The Systematics Association Special Volume 27, Academic Press, London and Orlando.

McCourt RM, Delwiche CF, and Karol KG. 2004. Charophyte algae and land plant origins. Trends in Ecology & Evolution **19**:661-666.

Melkonian M. 1989. Flagellar Apparatus Ultrastructure in *Mesostigma viride* (Prasinophyceae). Plant Systematics and Evolution **164**:93-122.

Menzel D. 1988. How do giant plant cells cope with injury? - The wound response in siphonous green algae. Protoplasma **144**:73-91.

Meyer KI. 1927. Algae of the northern end of lake Baikal. Archives Russes de Protisologie **6**:93-118 [In Russian, with French summary].

Mishler BD, Lewis LA, Buchheim MA, Renzaglia KS, Garbary DJ, Delwiche CF, Zechman FW, Kantz TS, and Chapman RL. 1994. Phylogenetic relationships of the "green algae" and "bryophytes". Annals of the Missouri Botanical Garden **81**:451-483.

Miyaji K. 1999. A new type of pyrenoid in the genus *Rhizoclonium* (Cladophorales, Chlorophyta). Phycologia **38**:267-276.

Moestrup Ø. 1978. Phylogenetic validity of flagellar apparatus in green algae and other chlorophyll a and b containing plants. Biosystems **10**:117-144.

Moore MJ, Bell CD, Soltis PS, and Soltis DE. 2007. Using plastid genome-scale data to resolve enigmatic relationships among basal angiosperms. Proceedings of the National Academy of Sciences of the United States of America **104**:19363-19368.

Muller J, Oma Y, Vallar L, Friederich E, Poch O, and Winsor B. 2005. Sequence and comparative genomic analysis of actin-related proteins. Molecular Biology of the Cell **16**:5736-5748.

Negrutskii BS and El'skaya AV. 1998. Eukaryotic translation elongation factor 1α : Structure, expression, functions, and possible role in aminoacyl-tRNA channeling. Pp. 47-78. Progress in Nucleic Acid Research and Molecular Biology, Volume 60.

Nei M and Rooney AP. 2005. Concerted and birth-and-death evolution of multigene families. Annual Review of Genetics **39**:121-152.

Nielsen R and Gunnarsson K. 2001. Seaweeds of the Faroe Islands. An annotated checklist. Fróðskaparrit **49**:45-108.

Noble GP, Rogers MB, and Keeling PJ. 2007. Complex distribution of EFL and EF-1α proteins in the green algal lineage. BMC Evolutionary Biology **7**:82.

Nosenko T, Lidie KL, Van Dolah FM, Lindquist E, Cheng JF, and Bhattacharya D. 2006. Chimeric plastid proteome in the florida "red tide" dinoflagellate *Karenia brevis*. Molecular Biology and Evolution **23**:2026-2038.

Nozaki H, Matsuzaki M, Takahara M, Misumi O, Kuroiwa H, Hasegawa M, Shin-i T, Kohara Y, Ogasawara N, and Kuroiwa T. 2003a. The phylogenetic position of red algae revealed by multiple nuclear genes from mitochondria-containing eukaryotes and an alternative hypothesis on the origin of plastids. Journal of Molecular Evolution **56**:485-497.

Nozaki H, Ohta N, Matsuzaki M, Misumi O, and Kuroiwa T. 2003b. Phylogeny of plastids based on cladistic analysis of gene loss inferred from complete plastid genome sequences. Journal of Molecular Evolution **57**:377-382.

Nylander JAA, Ronquist F, Huelsenbeck JP, and Nieves-Aldrey JL. 2004. Bayesian phylogenetic analysis of combined data. Systematic Biology **53**:47-67.

O'Kelly CJ and Floyd GL. 1984a. Flagellar apparatus absolute orientations and the phylogeny of the green algae. Biosystems **16**:227-251.

O'Kelly CJ and Floyd GL. 1984b. Correlations among patterns of sporangial structure and development, life histories, and ultrastructural features in the Ulvophyceae. Pp. 121-156 *in* D. Irvine, and D. John, eds. Systematics of the green algae. The Systematics Association Special Volume 27, Academic Press, London and Orlando.

O'Kelly CJ, Wysor B, and Bellows WK. 2004. Gene sequence diversity and the phylogenetic position of algae assigned to the genera *Phaeophila* and *Ochlochaete* (Ulvophyceae, Chlorophyta). Journal of Phycology **40**:789-799.

O'Kelly CJ. 2007. The origin and early evolution of green plants. Pp. 287-309 *in* P. G. Falkowski, and A. H. Knoll, eds. Evolution of primary producers in the sea. Elsevier Academic Press, Amsterdam, The Netherlands.

O'Neil RM and La Claire JW. 1984. Mechanical wounding induces the formation of extensive coated membranes in giant algal cells. Science **225**:331-333.

Pagel M. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. Proceedings of the Royal Society of London Series B-Biological Sciences **255**:37-45.

Pagel M. 1999. The maximum likelihood approach to reconstructing ancestral character states of discrete characters on phylogenies. Systematic Biology **48**:612-622.

Pagel M and Meade A. 2004. A phylogenetic mixture model for detecting pattern-heterogeneity in gene sequence or character-state data. Systematic Biology **53**:571-581.

Pagel M and Meade A. 2006. Bayesian analysis of correlated evolution of discrete characters by reversible-jump Markov chain Monte Carlo. American Naturalist **167**:808-825.

Palmer JD. 2003. The symbiotic birth and spread of plastids: How many times and whodunit? Journal of Phycology **39**:4-11.

Paradis E, Claude J, and Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics **20**:289-290.

Parker B. 1970. Significance of cell wall chemistry to phylogeny in the green algae. Annals of the New York Academy of Sciences **175**:417-428.

Parkinson CL, Adams KL, and Palmer JD. 1999. Multigene analyses identify the three earliest lineages of extant flowering plants. Current Biology **9**:1485-1488.

Petersen J, Teich R, Becker B, Cerff R, and Brinkmann H. 2006. The GapA/B gene duplication marks the origin of streptophyta (Charophytes and land plants). Molecular Biology and Evolution **23**:1109-1118.

Philip GK, Creevey CJ, and McInerney JO. 2005. The Opisthokonta and the Ecdysozoa may not be clades: stronger support for the grouping of plant and animal than for animal and fungi and stronger support for the Coelomata than Ecdysozoa. Molecular Biology and Evolution **22**:1175-1184.

Philippe H, Snell EA, Bapteste E, Lopez P, Holland PWH, and Casane D. 2004. Phylogenomics of eukaryotes: Impact of missing data on large alignments. Molecular Biology and Evolution **21**:1740-1752.

Philippe H, Delsuc F, Brinkmann H, and Lartillot N. 2005a. Phylogenomics. Annual Review of Ecology Evolution and Systematics **36**:541-562.

Philippe H, Lartillot N, and Brinkmann H. 2005b. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. Molecular Biology and Evolution **22**:1246-1253.

Pickett-Heaps JD and Marchant HJ. 1972. The phylogeny of the green algae: a new proposal. Cytobios **6**:255-264.

Pombert JF, Otis C, Lemieux C, and Turmel M. 2004. The complete mitochondrial DNA sequence of the green alga *Pseudendoclonium akinetum* (Ulvophyceae) highlights distinctive evolutionary trends in the Chlorophyta and suggests a sister-group relationship between the Ulvophyceae and Chlorophyceae. Molecular Biology and Evolution **21**:922-935.

Pombert JF, Otis C, Lemieux C, and Turmel M. 2005. Chloroplast genome sequence of the green alga *Pseudendoclonium akinetum* (Ulvophyceae) reveals unusual structural features and new insights into the branching order of chlorophyte lineages. Molecular Biology and Evolution **22**:1903-1918.

Pombert JF, 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 Biology **4**.

Pond SLK, Frost SDW, and Muse SV. 2005. HyPhy: Hypothesis testing using Phylogenies. Bioinformatics **21**:676-679.

Posada D and Crandall KA. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics **14**: 817-818.

Powell JR and Moriyama EN. 1997. Evolution of codon usage bias in *Drosophila*. Proceedings of the National Academy of Sciences of the United States of America **94**:7784-7790.

Printz H. 1926. Die Algenvegetation des Trondjemsfjordes. Skr. Norske Vidensk.-Akad. Oslo, Mat.-naturv. Kl. 1926 **5**:1- 273.

Printz H. 1932. Observations on the structure and reproduction in *Urospora*. Aresch. Nyt. Mag. Naturvidensk. **70**:274-287.

Pröschold T and Leliaert F. 2007. Systematics of the green algae: conflict of classic and modern approaches. Pp. 123-153 *in* B. J., and L. J.M., eds. Unravelling the algae: the past, present, and future of algal systematics. The Systematics Association Special Volume Series, 75. CRC Press, Boca Raton, FL, USA.

Qiu YL, Li LB, Wang B, Chen ZD, Knoop V, Groth-Malonek M, Dombrovska O, Lee J, Kent L, Rest J, Estabrook GF, Hendry TA, Taylor DW, Testa CM, Ambros M, Crandall-Stotler B, Duff RJ, Stech M, Frey W, Quandt D, and Davis CC. 2006. The deepest divergences in land plants inferred from phylogenomic evidence. Proceedings of the National Academy of Sciences of the United States of America **103**:15511-15516.

Rambaut A and Drummond A. 2007. Tracer: MCMC Trace Analysis Tool. http://tree.bio.ed.ac.uk/software/tracer/

Reyes-Prieto A and Bhattacharya D. 2007. Phylogeny of nuclear-encoded plastid-targeted proteins supports an early divergence of glaucophytes within plantae. Molecular Biology and Evolution **24**:2358-2361.

Reyes-Prieto A, Weber APM, and Bhattacharya D. 2007. The origin and establishment of the plastid in algae and plants. Annual Review of Genetics **41**:147-168.

Rindi F, Guiry MD, and Lopez-Bautista JM. 2006a. New records of Trentepohliales (Ulvophyceae, Chlorophyta) from Africa. Nova Hedwigia **83**:431-449.

Rindi F, López-Bautista JM, Sherwood AR, and Guiry MD. 2006b. Morphology and phylogenetic position of *Spongiochrysis hawaiiensis* gen. et sp. nov., the first known terrestrial member of the order Cladophorales (Ulvophyceae, Chlorophyta). International Journal of Systematic and Evolutionary Microbiology **56**:913-922.

Robbens S, Petersen J, Brinkmann H, Rouze P, and Van de Peer Y. 2007. Unique regulation of the Calvin cycle in the ultrasmall green alga *Ostreococcus*. Journal of Molecular Evolution **64**:601-604.

Rodriguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G, Löffelhardt W, Bohnert HJ, Philippe H, and Lang BF. 2005. Monophyly of primary photosynthetic eukaryotes: green plants, red algae, and glaucophytes. Current Biology **15**:1325-1330.

Rodriguez-Ezpeleta N, Philippe H, Brinkmann H, Becker B, and Melkonian M. 2007. Phylogenetic analyses of nuclear, mitochondrial, and plastid multigene data sets support the placement of *Mesostigma* in the Streptophyta. Molecular Biology and Evolution **24**:723-731.

Roger AJ, Sandblom O, Doolittle WF, and Philippe H. 1999. An evaluation of elongation factor 1 alpha as a phylogenetic marker for eukaryotes. Molecular Biology and Evolution **16**:218-233.

Roger AJ and Hug LA. 2006. The origin and diversification of eukaryotes: problems with molecular phylogenetics and molecular clock estimation. Philosophical Transactions of the Royal Society B-Biological Sciences **361**:1039-1054.

Rogers MB, Watkins RF, Harper JT, Durnford DG, Gray MW, and Keeling PJ. 2007. A complex and punctate distribution of three eukaryotic genes derived by lateral gene transfer. BMC Evolutionary Biology **7**:89.

Rokas A and Abbot P. 2009. Harnessing genomics for evolutionary insights. Trends in Ecology & Evolution **24**:192-200.

Ronquist F and Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19**:1572-1574.

Rosenvinge LK. 1893. Grønlands Havalger. Medd. Grønland **3**:763-981.

Round FE. 1984. The systematics of the Chlorophyta: an historical review leading to some modern concepts. *in* I. D.E.G., and D. M. John, eds. The Systematics of Green Algae. The Systematics Association Special Volume 27, Academic Press, London and Orlando.

Rueness J. 1977. Norsk Algeflora. Universitetsforlaget, Olso.

Rueness J. 1992. Field and culture observations on *Uronema curvatum* Printz (Chlorophyta). Acta Phytogeogr. Suec. **78**:125-130.

Saarela JM, Rai HS, Doyle JA, Endress PK, Mathews S, Marchant AD, Briggs BG, and Graham SW. 2007. Hydatellaceae identified as a new branch near the base of the angiosperm phylogenetic tree. Nature **446**:312-315.

Sakaguchi M, Suzaki T, Khan S, and Hausmann K. 2002. Food capture by kinetocysts in the heliozoon *Raphidiophrys contractilis*. European Journal of Protistology **37**:453-458.

Sakaguchi M, Takishita K, Matsumoto T, Hashimoto T, and Inagaki Y. 2008. Tracing back EFL gene evolution in the cryptomonads–haptophytes assemblage: Separate origins of EFL genes in haptophytes, photosynthetic cryptomonads, and goniomonads. Gene **in press**.

Salim HMW and Cavalcanti ARO. 2008. Factors influencing codon usage bias in genomes. Journal of the Brazilian Chemical Society **19**:257-262.

Sambrook J, Fritch EF, and Maniatis T. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, New York.

Sanderson MJ. 2002. Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. Molecular Biology and Evolution **19**:101-109.

Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics **19**:301-302.

Santos MAS, Moura G, Massey SE, and Tuite MF. 2004. Driving change: the evolution of alternative genetic codes. Trends in Genetics **20**:95-102.

Scagel RF. 1966. Marine algae of British Columbia and northern Washington, Part I: Chlorophyceae (green algae). Bulletin National Museum of Canada **207**:1-257.

Schlösser UG. 1987. Action of cell wall autolysins in asexual reproduction of filamentous green algae: evidence and species specificity. Pp. 75–80 *in* W. Wiessner, D. G. Robinson, and S. R.C., eds. Algal Development: Molecular and Cellular Aspects. Springer, Berlin, Heidelberg, Germany.

Schluter D, Price T, Mooers AO, and Ludwig D. 1997. Likelihood of ancestor states in adaptive radiation. Evolution **51**:1699-1711.

Schluter PM, Stuessy TF, and Paulus HF. 2005. Making the first step: practical considerations for the isolation of low-copy nuclear sequence markers. Taxon **54**:766-770.

Schneider CW and Searles RB. 1991. Seaweeds of the southeastern United States. Cape Hatteras to Cape Canaveral. Duke University Press, Durham & London.

Schneider SU, Leible MB, and Yang XP. 1989. Strong homology between the small subunit of ribulose-1,5-bisphosphate carboxylase/oxygenase of two species of *Acetabularia* and the occurrence of unusual codon usage. Molecular & General Genetics **218**:445-452.

Schneider SU and de Groot EJ. 1991. Sequences of two *rbcS* cDNA clones of *Batophora oerstedii*: structural and evolutionary considerations. Current Genetics **20**:173-175.

Schöniger M and Von Haeseler A. 1994. A stochastic model for the evolution of autocorrelated DNA sequences. Molecular Phylogenetics and Evolution **3**:240-247.

Schultz DW and Yarus M. 1994. Transfer RNA mutation and the malleability of the genetic code. Journal of Molecular Biology **235**:1377-1380.

Sengupta S and Higgs PG. 2005. A unified model of codon reassignment in alternative genetic codes. Genetics **170**:831-840.

Seo TK and Kishino H. 2008. Synonymous substitutions substantially improve evolutionary inference from highly diverged proteins. Systematic Biology **57**:367-377.

Setchell WA and Gardner NL. 1920. The marine algae of the Pacific coast of North America. Part II. Chlorophyceae. University of California Publications in Botany **8**:139-374.

Shapiro B, Rambaut A, and Drummond AJ. 2006. Choosing appropriate substitution models for the phylogenetic analysis of protein-coding sequences. Molecular Biology and Evolution **23**:7-9.

Shimodaira H and Hasegawa M. 2001. CONSEL: for assessing the confidence of phylogenetic tree selection. Bioinformatics **17**:1246-1247.

Shimodaira H. 2002. An approximately unbiased test of phylogenetic tree selection. Systematic Biology **51**:492-508.

Sluiman HJ. 1989. The green algal class Ulvophyceae. An ultrastructural survey and classification. Cryptogamic botany **1**:83-94.

Small RL, Cronn RC, and Wendel JF. 2004. Use of nuclear genes for phylogeny reconstruction in plants. Australian Systematic Botany **17**:145-170.

Sogin ML, Elwood HJ, and Gunderson JH. 1986. Evolutionary diversity of eukaryotic small subunit rRNA genes. Proceedings of the National Academy of Sciences of the United States of America **83**:1383-1387.

Soltis PS and Soltis DE. 2003. Applying the bootstrap in phylogeny reconstruction. Statistical Science **18**:256-267.

South GR and Tittley I. 1986. A checklist and distributional index of the benthic marine algae of the North Atlantic Ocean. British Museum (Natural History) and Huntsman Marine Laboratory, London and St. Andrews, New Brunswick, Canada.

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. Journal of Phycology **30**:340-345.

Stiller JW, Riley J, and Hall BD. 2001. Are red algae plants? A critical evaluation of three key molecular data sets. Journal of Molecular Evolution **52**:527-539.

Stiller JW and Harrell L. 2005. The largest subunit of RNA polymerase II from the Glaucocystophyta: functional constraint and short-branch exclusion in deep eukaryotic phylogeny. BMC Evolutionary Biology **5**.

Subramanian S. 2008. Nearly neutrality and the evolution of codon usage bias in eukaryotic genomes. Genetics **178**:2429-2432.

Sung GH, Sung JM, Hywel-Jones NL, and Spatafora JW. 2007. A multi-gene phylogeny of Clavicipitaceae (Ascomycota, Fungi): Identification of localized incongruence using a combinational bootstrap approach. Molecular Phylogenetics and Evolution **44**:1204-1223.

Swofford DL. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Tartar A, Boucias DG, Adams BJ, and Becnel JJ. 2002. Phylogenetic analysis identifies the invertebrate pathogen *Helicosporidium sp*. as a green alga (Chlorophyta). International Journal of Systematic and Evolutionary Microbiology **52**:273-279.

Telford MJ, Wise MJ, and Gowri-Shankar V. 2005. Consideration of RNA secondary structure significantly improves likelihood-based estimates of phylogeny: Examples from the Bilateria. Molecular Biology and Evolution **22**:1129-1136.

Than C, Ruths D, Innan H, and Nakhleh L. 2006. Identifiability issues in phylogeny-based detection of horizontal gene transfer. Pp. 215-229. Comparative Genomics, Proceedings.

Thorne JL, Kishino H, and Painter IS. 1998. Estimating the rate of evolution of the rate of molecular evolution. Molecular Biology and Evolution **15**:1647-1657.

Tourancheau AB, Tsao N, Klobutcher LA, Pearlman RE, and Adoutte A. 1995. Genetic code deviations in the ciliates: evidence for multiple and independent events. EMBO Journal **14**:3262-3267.

Turmel M, Ehara M, Otis C, and Lemieux C. 2002a. Phylogenetic relationships among streptophytes as inferred from chloroplast small and large subunit rRNA gene sequences. Journal of Phycology **38**:364-375.

Turmel M, Otis C, and Lemieux C. 2002b. The complete mitochondrial DNA sequence of *Mesostigma viride* identifies this green alga as the earliest green plant divergence and predicts a highly compact mitochondrial genome in the ancestor of all green plants. Molecular Biology and Evolution **19**:24-38.

Turmel M, Otis C, and Lemieux C. 2003. The mitochondrial genome of *Chara vulgaris*: Insights into the mitochondrial DNA architecture of the last common ancestor of green algae and land plants. Plant Cell **15**:1888-1903.

Turmel M, Otis C, and Lemieux C. 2006. The chloroplast genome sequence of *Chara vulgaris* sheds new light into the closest green algal relatives of land plants. Molecular Biology and Evolution **23**:1324-1338.

Turmel M, Brouard JS, Gagnon C, Otis C, and Lemieux C. 2008. Deep division in the Chlorophyceae (Chlorophyta) revealed by chloroplast phylogenomic analyses. Journal of Phycology **44**:739-750.

Turmel M, Gagnon M-C, O'Kelly CJ, Otis C, and Lemieux C. 2009. 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. Molecular Biology and Evolution **26**:631-648.

Van de Peer Y, Taylor JS, Braasch I, and Meyer A. 2001. The ghost of selection past: Rates of evolution and functional divergence of anciently duplicated genes. Journal of Molecular Evolution **53**:436-446.

van den Hoek C. 1963. Revision of the European species of *Cladophora*, Brill, Leiden, The Netherlands.

van den Hoek C, Mann DG, and Jahns HM. 1995. Algae. An Introduction to Phycology. Cambridge University Press, Cambridge.

van Oppen MJH, Olsen JL, and Stam WT. 1995. Genetic variation within and among North Atlantic and Baltic populations of the benthic alga *Phycodrys rubens* (Rhodophyta). European Journal of Phycology **30**:251-260.

Venditti C, Meade A, and Pagel M. 2008. Phylogenetic mixture models can reduce node-density artifacts. Systematic Biology **57**:286-293.

Verbruggen H, Leliaert F, Maggs CA, Shimada S, Schils T, Provan J, Booth D, Murphy S, De Clerck O, Littler DS, Littler MM, and Coppejans E. 2007. Species boundaries and phylogenetic relationships within the green algal genus Codium (Bryopsidales) based on plastid DNA sequences. Molecular Phylogenetics and Evolution **44**:240-254.

Verbruggen H. 2008. TreeGradients v1.02. http://www.phycoweb.net

Verbruggen H and Theriot EC. 2008. Building trees of algae: some advances in phylogenetic and evolutionary analysis. European Journal of Phycology **43**.

Verbruggen H. 2009. TreeGradients v1.02. http://www.phycoweb.net

Verbruggen H, Ashworth M, LoDuca ST, Vlaeminck C, Cocquyt E, Sauvage T, Zechman FW, Littler DS, Littler MM, Leliaert F, and De Clerck O. 2009. A multi-locus time-calibrated phylogeny of the siphonous green algae. Molecular Phylogenetics and Evolution **50**:642-653.

Vetsigian K and Goldenfeld N. 2009. Genome rhetoric and the emergence of compositional bias. Pp. 215-220.

Vroom PS and Smith CM. 2003. Reproductive features of Hawaiian *Halimeda velasquezii* (Bryopsidales, Chlorophyta), and an evolutionary assessment of reproductive characters in *Halimeda*. Cryptogamie Algologie **24**:355-370.

Waddell PJ, 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. Systematic Biology **48**:31-53.

Wahlberg N and Wheat CW. 2008. Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of Lepidoptera. Systematic Biology **57**:231-242.

Wan XF, Zhou J, and Xu D. 2006. CodonO: a new informatics method for measuring synonymous codon usage bias within and across genomes. International Journal of General Systems **35**:109-125.

Watanabe S, Kuroda N, and Maiwa F. 2001. Phylogenetic status of *Helicodictyon planctonicum* and *Desmochloris halophila* gen. et comb. nov and the definition of the class Ulvophyceae (Chlorophyta). Phycologia **40**:421-434.

Watanabe S and Nakayama T. 2007. Ultrastructure and phylogenetic relationships of the unicellular green algae *Ignatius tetrasporus* and *Pseudocharacium americanum* (Chlorophyta). Phycological Research **55**:1-16.

Wiens JJ. 2005. Can incomplete taxa rescue phylogenetic analyses from long-branch attraction? Systematic Biology **54**:731-742.

Wiens JJ, Kuczynski CA, Smith SA, Mulcahy DG, Sites JW, Townsend TM, and Reeder TW. 2008. Branch lengths, support, and congruence: testing the phylogenomic approach with 20 nuclear loci in snakes. Systematic Biology **57**:420-431.

Wiens JJ and Moen DS. 2008. Missing data and the accuracy of Bayesian phylogenetics. Journal of Systematics and Evolution **46**:307-314.

Wille N. 1890. Cladophoraceae. Pp. 114-119 *in* A. Engler, and K. Prantl, eds. Die natürlichen Pflanzenfamilien. I.Teil, Abt. 2, Leipzig.

Wolf PG. 1997. Evaluation of *atp*B nucleotide sequences for phylogenetic studies of ferns and other pteridophytes. American Journal of Botany **84**:1429-1440.

Womersley HBS. 1984. The marine benthic flora of southern Australia. Part I. Government Printer, South Australia, Adelaide.

Worden AZ, Lee J-H, Mock T, Rouze P, Simmons MP, Aerts AL, Allen AE, Cuvelier ML, Derelle E, Everett MV, Foulon E, Grimwood J, Gundlach H, Henrissat B, Napoli C, McDonald SM, Parker MS, Rombauts S, Salamov A, Von Dassow P, Badger JH, Coutinho PM, Demir E, Dubchak I, Gentemann C, Eikrem W, Gready JE, John U, Lanier W, Lindquist EA, Lucas S, Mayer KFX, Moreau H, Not F, Otillar R, Panaud O, Pangilinan J, Paulsen I, Piegu B, Poliakov A, Robbens S, Schmutz J, Toulza E, Wyss T, Zelensky A, Zhou K, Armbrust EV, Bhattacharya D, Goodenough UW, Van de Peer Y, and Grigoriev IV. 2009. Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. Science **324**:268-272.

Wuyts J, Van de Peer Y, and Wachter RD. 2001. Distribution of substitution rates and location of insertion sites in the tertiary structure of ribosomal RNA. Nucleic Acids Research **29**:5017-5028.

Wuyts J, Perriere G, and Van de Peer Y. 2004. The European ribosomal RNA database. Nucleic Acids Research **32**:D101-D103.

Yamamoto M, Nozaki H, and Kawano S. 2001. Evolutionary relationships among multiple modes of cell division in the genus *Nannochloris* (Chlorophyta) revealed by genome size, actin gene multiplicity, and phylogeny. Journal of Phycology **37**:106-120.

Yamamoto M, Nozaki H, Miyazawa Y, Koide T, and Kawano S. 2003. Relationship between presence of a mother cell wall and speciation in the unicellular microalga *Nannochloris* (Chlorophyta). Journal of Phycology **39**:172-184.

Yoon HS, Hackett JD, Ciniglia C, Pinto G, and Bhattacharya D. 2004. A molecular timeline for the origin of photosynthetic eukaryotes. Molecular Biology and Evolution **21**:809-818.

Zechman FW, Theriot EC, Zimmer EA, and Chapman RL. 1990. Phylogeny of the Ulvophyceae (Chlorophyta): cladistic analysis of nuclear-encoded rRNA sequence data. Journal of Phycology **26**:700-710.

Zhang HC and Qiao GX. 2007. Systematic status of the genus *Formosaphis* Takahashi and the evolution of galls based on the molecular phylogeny of Pemphigini (Hemiptera: Aphididae: Eriosomatinae). Systematic Entomology **32**:690-699.

Zhang RF, Cui ZL, Zhang XZ, Jiang JD, Gu JD, and Li SP. 2006. Cloning of the organophosphorus pesticide hydrolase gene clusters of seven degradative bacteria isolated from a methyl parathion contaminated site and evidence of their horizontal gene transfer. Biodegradation **17**:465-472.

Zimmer A, Lang D, Richardt S, Frank W, Reski R, and Rensing SA. 2007. Dating the early evolution of plants: detection and molecular clock analyses of orthologs. Molecular Genetics and Genomics **278**:393-402.

Zuccarello GC, Price N, Verbruggen H, and Leliaert F. 2009. Analysis of a plastid multigene dataset and the phylogenetic position of the marine macroalga *Caulerpa filiformis* (Chlorophyta). Journal of Phycology **in press**.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Research **31**:3406-3415.

Zwickl DJ and Hillis DM. 2002. Increased taxon sampling greatly reduces phylogenetic error. Systematic Biology **51**:588-598.

Summary

Green algae are distributed worldwide and can be found in almost every habitat, ranging from polar to tropical marine, freshwater and terrestrial environments, and as symbionts. They exhibit a remarkable morphological and cytological diversity ranging from unicells and colonial forms, over multicellular filaments and foliose blades, to siphonous life forms that are essentially composed of a single giant cell containing countless nuclei. Together with land plants, green algae form the green lineage or Viridiplantae. Morphological and molecular studies have identified a major split within the Viridiplantae giving rise to two monophyletic lineages, the Chlorophyta and the Streptophyta. The Streptophyta consists of several lineages of freshwater green algae from which land plants evolved approximately 470 million years ago. Whereas considerable progress has been made during the past decade in clarifying the relationships among the streptophyte green algae and land plants, the phylogeny and evolutionary history of the Chlorophyta has been more difficult to elucidate.

The Chlorophyta consists of a paraphyletic assemblage of early diverging unicellular green algae, termed the prasinophytes, which gave rise to three main clades, the classes Ulvophyceae, Trebouxiophyceae and Chlorophyceae (UTC). The relationships among these three classes have been at the center of a long-standing debate. Based on ultrastuctural characters, and SSU nrDNA, chloroplast and mitochondrial phylogenies all possible relationships between UTC classes have been hypothesized. The unstable relationships exhibited among these three classes are likely due to a combination of their ancient age and the short time span over which they diverged from one another. Determining relationships among the different orders of the Ulvophyceae poses a similar problem. Up to now, green algal phylogenies were either based on limited gene or taxon sampling. The slow progress in green algal phylogenetics is in part due to difficulties in amplifying single-copy nuclear markers, as a consequence of the limited availability of genomic data for green algae and the presence of large introns in their genes.

In **Chapter 2**, we performed phylogenetic analyses (ML and BI) of seven nuclear genes, SSU nrDNA and two plastid markers with carefully chosen partitioning strategies and models of sequence evolution. We obtained high support across the topology of the Chlorophyta, show the monophyly of the UTC classes and reveal a sister relationship between Chlorophyceae and Ulvophyceae. Even though topology tests (AU) do not exclude an alternative branching order of UTC classes, we showed that moderate removal of fast-evolving sites improves the phylogenetic signal in the desired epoch.

We also inferred the relationships among the Ulvophyceae, which was found to consist of two main clades. The first clade contains the orders Ulvales and Ulotrichales. The second clade includes the early diverging genus *Ignatius* and a clade comprising the orders Trentepohliales, Bryopsidales, Dasycladales and Cladophorales, along with *Blastophysa*. In this clade, the Bryopsidales and Dasycladales are sisters, *Blastophysa* is most closely related to the Cladophorales, while the phylogenetic position of the Trentepohliales remains uncertain. The inferred relationships provide novel insights into the evolution of multicellularity and multinucleate cells in the green tree of life. Multicellularity evolved multiple times independently in the Streptophyta and Chlorophyta, and at least twice in the two main clades of the Ulvophyceae. Siphonous thallus structures are most likely

derived from a multinucleate and multicellular common ancestor of the Bryopsidales—Dasycladales—Cladophorales clade.

In order to better understand the timeframe in which key evolutionary events took place in the green algae, we dated our multilocus phylogeny using relaxed molecular clock methods and the fossil record (**Chapters 6** and **8**). This dated phylogeny shows a rapid radiation of the UTC classes in ca. 20 my during the first period of the Neoproterozoic (between 900—881 my) and a Neoproterozoic diversification of all ulvophycean orders in a timeframe of ca. 130 my (between 826—596 my).

Guided by this improved green algal phylogenetic tree, we addressed various topics relating to molecular evolution of the Chlorophyta.

In **Chapter 3**, we studied the distribution and gain-loss patterns of elongation factor genes. Elongation factor-1 alpha (EF-1α) and elongation factor-like (EFL), two key genes of the translational apparatus, have an almost mutually exclusive distribution in eukaryotes. All streptophytes except *Mesostigma* encode EF-1α. In the Chlorophyta, the prasinophytes, Trebouxiophyceae, Chlorophyceae and ulvophycean orders Ulvales and Ulotrichales have EFL. We showed that not only the ulvophyte *Acetabularia* (Dasycladales) but also closely related lineages, i.e. *Ignatius*, Cladophorales + *Blastophysa*, Bryopsidales and other Dasycladales, possess EF-1α.

In order to gain more insight in the evolution of EF-1α and EFL in the Viridiplantae we analyzed their gain-loss dynamics in a maximum likelihood framework using continuous-time Markov models. These models revealed that the presence of EF-1α, EFL or both genes along the backbone of the green plant phylogeny is highly uncertain due to sensitivity to branch lengths and lack of prior knowledge about ancestral states or rates of gene gain and loss. Model refinements based on insights gained from the EF-1α phylogeny reduce uncertainty but still imply several equally likely possibilities: a primitive EF-1α state with multiple independent EFL gains or coexistence of both genes in the ancestor of the Viridiplantae or Chlorophyta followed by differential loss of one or the other gene in the various lineages.

The genetic code, which translates nucleotide triplets (codons) into amino acids, is virtually identical in all living organisms. However, a small number of eubacterial, and eukaryotic nuclear and mitochondrial genomes have evolved slight variations on this universal code. A non-canonical code, in which TAG and TAA have been reassigned from stop codons to glutamine, has previously been reported for the green algal order Dasycladales. Based on the amplified housekeeping nuclear genes, we demonstrate in **chapter 4** that this non-canonical genetic code is shared with the related clades Trentepohliales, Cladophorales and *Blastophysa*, but not with the sister clade of the Dasycladales, the Bryopsidales. We favor a stepwise acquisition model for the evolution of a non-canonical code, whereby the alternative codes observed in these green algal orders share a single origin. Stop codon reassignment is a gradual process requiring changes to tRNA and eukaryotic release factor (eRF1) genes. We suggest that mutations in the anticodons of canonical glutamine tRNAs occurred once along the branch leading to the orders Trentepohliales, Dasycladales, Bryopsidales, Cladophorales and the genus *Blastophysa*. The presence of these mutated tRNAs allow TAG and TAA codons to be

translated to glutamine instead of terminating translation. At this step, the mutated tRNAs compete with eRF1 for the TAA and TAG codons. To complete the transition to the non-canonical code, a subsequent mutation of eRF1 that prevents binding of eRF1 with TAG and TAA is required. The complex distribution of the non-canonical code in the Ulvophyceae could then be explained by three independent mutations of eRF1 in the Trentepohliales, Dasycladales and Cladophorales + *Blastophysa*, along with a decrease in importance of the mutated tRNAs or their extinction through selection or drift in the branch leading to the Bryopsidales. A detailed comparison of eukaryotic release factors (eRF1) and glutamine tRNAs in the respective clades of the Ulvophyceae is, however, needed to test this evolutionary scenario.

In **Chapter 5**, we studied differences in synonymous codon usage bias and GC content among green algae based on our housekeeping nuclear genes and analyzed their evolution in a phylogenetic framework. We observe stronger codon usage bias in the ancestral streptophytes *Mesostigma* and *Chlorokybus* than in the remainder of the Streptophyta. Within the Chlorophyta, the prasinophytes, Trebouxiophyceae and Chlorophyceae have markedly stronger codon usage bias than the Ulvophyceae. One exception is the ulvophyte *Ignatius*, which has a markedly stronger codon usage bias than other members of the Ulvophyceae. GC content patterns show congruent trends, species with strong codon usage bias having high GC content.

We interpret these results along with the biology of the organisms in the framework of two models: the mutation-selection-drift model and the co-evolutionary model of genome composition and resource allocation. It is remarkable that unicellular organisms and colony-forming species have much more pronounced GC and codon usage biases as compared to multicellular and macroscopic species. This may follow from unicells having large population sizes, which leads to more codon usage bias due to stronger selection as compared to species with smaller population sizes where drift can more rapidly fix mutation. Their large population sizes are revealed by characteristics for r-selected species: small body size, fast growth rate and short generation times. We observe a negative correlation between rate of molecular evolution and codon usage bias and a positive correlation between GC content and codon usage bias.

Chapters 6 and 7 focus on the phylogenetic relationships and evolution within two specific ulvophycean clades. In **Chapter 6**, we dated a five-locus phylogeny of the siphonous orders Dasycladales and Bryopsidales using relaxed molecular clock methods calibrated with the fossil record. These models indicate a late Neoproterozoic or early Cambrian origin of the Dasycladales and Bryopsidales (571 million years [628–510]) and a Paleozoic diversifications of the different families within each order. In **Chapter 7** we studied phylogenetic relationships within the Cladophorales based on small and large subunit nrDNA sequences. We found that *Uronema curvatum*, a marine microfilamentous species placed in the Chlorophyceae, is sister to the rest of the Cladophorales. Based on the divergent phylogenetic position of *U. curvatum* we described a new genus and family of Cladophorales (*Okellya*, Okellyaceae).

Samenvatting

Groenwieren worden wereldwijd aangetroffen in ongeveer elk mogelijke habitat, zowel in polaire gebieden alsook in tropische zeeën, zoetwatermeren, terrestrische omgevingen en zelfs als symbiont. Morfologische en cytologische diversiteit tussen groenwieren is groot, gaande van eencelligen en kolonies naar meercellige filamenten en bladachtige thalli, tot sifonale levensvormen die bestaan uit één enkele reusachtige cel met een ontelbaar aantal kernen.

Samen met de landplanten vormen de groenwieren de groene planten of Viridiplantae. Morfologische en moleculaire studies toonden aan dat de Viridiplantae opgesplitst worden in 2 monofyletische groepen, de Chlorophyta en Streptophyta. De Streptophyta bestaan uit enkele groepen van zoetwater groenwieren die ongeveer 470 miljoen jaar geleden aan de oorsprong lagen van de landplanten. De laatste 10 jaar is er veel vooruitgang geboekt in de opheldering van de verwantschap tussen de streptophyte groenwieren en de landplanten. De fylogenie en evolutionaire geschiedenis van de Chlorophyta blijken echter moeilijker te onderzoeken.

De Chlorophyta bestaan uit een parafyletische groepering van reeds vroeg gedivergeerde eencellige groenwieren, de zogenaamde prasinophyten, die aan de basis liggen van drie belangrijke groepen: de klassen Ulvophyceae, Trebouxiophyceae and Chlorophyceae (UTC). De verwantschap tussen deze drie klassen is niet eenduidig en vormt de basis voor een langdurig debat.

Gebaseerd op ultrastructurele kenmerken en fylogenieën op basis van SSU nrDNA, chloroplast en mitochondriale genen, zijn alle mogelijke verwantschappen tussen de UTC klassen vastgesteld. Deze onstabiele verwantschappen worden waarschijnlijk veroorzaakt door een combinatie van de ouderdom van deze klassen en de korte tijdsspanne waarin ze van elkaar divergeerden. Het bepalen van de verwantschap tussen de verschillende ordes binnen de Ulvophyceae vormt een gelijkaardig probleem. Totnogtoe waren groenwier fylogenieën steeds gebaseerd op een beperkt aantal genen of taxa. De trage vooruitgang in het ophelderen van de groenwier fylogenie is waarschijnlijk te wijten aan moeilijkheden bij de amplificatie van nucleaire merkers, waarvan slechts een kopie per genoom aanwezig is als gevolg van de beperkte beschikbaarheid van genomische data voor groenwieren en de aanwezigheid van lange introns.

In **hoofdstuk 2**, voeren we een fylogenetische analyse (ML and BI) uit op basis van zeven nucleaire genen, SSU nrDNA en twee chloroplast merkers waarbij veel aandacht gaat naar de partitioneringsstrategie en de keuze van evolutionaire sequentiemodellen. We verkrijgen zo een goede ondersteuning van de volledige fylogenie, tonen de monofylie van de UTC klassen aan en bewijzen een zusterverwantschap tussen de Chlorophyceae en Ulvophyceae. Alhoewel dat topologie testen (AU) een alternatieve vertakkingvolgorde tussen de UTC klassen niet uitsluit, tonen we aan dat het verwijderen van een beperkt aantal snel evoluerende posities het fylogenetisch signaal versterkt in de gewenste periode.

De studie toont eveneens de verwantschap aan tussen de verschillende ordes binnen de Ulvophyceae. Deze bestaan uit twee groepen. De eerste groep omvat de Ulvales en Ulotrichales. De tweede groep bestaat uit het reeds vroeg ontstane genus *Ignatius* en uit een groep die de ordes Trentepohliales, Bryopsidales, Dasycladales en Cladophorales, samen met het genus *Blastophysa*

omvat. Binnen deze laatste groep zijn de Bryopsidales en Dasycladales zusters, *Blastophysa* is het nauwst verwant aan de Cladophorales, terwijl de fylogenetische positie van de Trentepohliales onzeker blijft.

De afgeleide verwantschappen verschaffen ons nieuwe inzichten over de evolutie van meercellige organismen en cellen met meerdere kernen. Meercellige organismen zijn verschillende keren onafhankelijk van elkaar ontstaan: minstens eenmaal in de Streptophyta en eenmaal in Chlorophyta en tenminste eenmaal in beide hoofdgroepen binnen de Ulvophyceae. Sifonale thallus structuren zijn hoogstwaarschijnlijk afgeleid van een multinucleate, meercellige voorouder van de Bryopsidales—Dasycladales—Cladophorales groep.

Om de evolutionaire tijdschaal waarbinnen deze gebeurtenissen plaatsvonden beter te begrijpen, hebben we onze multilocus groenwier fylogenie gedateerd aan de hand van 'relaxed molecular clock' methodes en fossiele data (**hoofdstukken 6** en **8**). Deze gedateerde fylogenie geeft een snelle radiatie van de UTC klassen aan in ca. 20 miljoen jaar (mj) gedurende de eerste periode van het Neoproterozoic (tussen de 900—881 mj) en een Neoproterozoische diversificatie van alle ulvophyceae ordes in een tijdsspannen van ca. 130 mj (826—596 mj).

In het kader van deze verbeterde groenwier fylogenie worden verschillende topics gerelateerd aan moleculaire evolutie binnen de Chlorophyta besproken.

In **hoofdstuk 3** bestuderen we de distributie en evolutie van 'elongation factor' genen. Elongation factor-1 alpha (EF-1α) en elongation factor-like (EFL), twee sleutelgenen van het translatie apparaat, hebben een onderling exclusief distributiepatroon in eukaryoten. Alle Streptophyta, behalve *Mesostigma*, hebben EF-1α. In de Chlorophyta hebben de prasinofyten, Trebouxiophyceae, Chlorophyceae en ulvophyte ordes Ulvales en Ulotrichales het EFL gen. Wij tonen aan dat niet alleen de ulvofyt *Acetabularia* (Dasycladales), maar ook nauw verwante groepen, in het bijzonder *Ignatius*, Cladophorales + *Blastophysa*, Bryopsidales en andere Dasycladales, EF-1α bezitten.

Om meer inzicht te krijgen in de evolutie van EF-1α en EFL binnen de Viridiplantae, hebben we de dynamiek van winst en verlies van deze genen in een maximum likelihood kader geanalyseerd met behulp van 'continuous-time Markov' modellen. Deze modellen tonen aan dat de aanwezigheid van EF-1α, EFL of beide genen langsheen de fylogenie van de groene planten onzeker is tengevolge van gevoeligheid voor taklengtes en een gebrek aan voorafgaande kennis omtrent voorouderlijke kenmerktoestanden en de snelheid van aanwinst en verlies van deze genen. Inzichten afgeleid van de EF-1α fylogenie maakten het mogelijk om ons model te verfijnen en verminderden de onzekerheid omtrent de aanwezigheid van EF-1α, EFL of beide genen. Desondanks blijven er nog steeds 3 mogelijke scenario's over: EF-1α als voorouderlijke toestand met meerdere onafhankelijke aanwinsten van EFL of een co-existentie van beide genen in de voorouder van de Viridiplantae of Chlorophyta gevolgd door differentieel verlies van een van beide genen in de verschillende evolutionaire lijnen.

De genetische code, die nucleotide triplets (codons) vertaalt naar aminozuren, is bijna identiek in alle levende organismen. Een klein aantal eubacteriële en eukaryote genomen ontwikkelden echter

kleine variaties op deze universele code. Een alternatieve code, waarbij TAG en TAA veranderen van stop codons naar glutamine coderende codons, was reeds aangetoond bij de groenwier orde Dasycladales. Op basis van de geamplificeerde nucleaire huishoudgenen, tonen we in **hoofdstuk 4** aan dat deze alternatieve code ook voorkomt bij de nauw verwanten groepen Trentepohliales, Cladophorales en *Blastophysa*, maar niet bij de zustergroep van de Dasycladales, de Bryopsisales.

We verkiezen een 'stepwise acquisition' model voor de evolutie van de alternatieve code, waarbij de alternatieve code, geobserveerd in deze groenwier ordes, een gemeenschappelijk oorsprong heeft. De verandering van stop codon naar aminozuur coderend codon is een geleidelijk proces dat tevens verandering in tRNAs en 'eukaryotic release factor' (eRF1) genen vergt. We veronderstellen dat mutaties in het anticodon van gewone glutamine tRNAs eenmaal plaatsvond in de tak leidend naar de groepen Trentepohliales, Dasycladales, Bryopsidales, Cladophorales and the genus *Blastophysa*. De aanwezigheid van gemuteerde tRNAs (met anticodons complementair aan TAG en TAA) maakt het dan mogelijk dat TAG en TAA codons worden vertaald in glutamine in plaats van de translatie te beëindigen. Op dat moment zijn de gemuteerde tRNAs in competitie met eRF1 om te binden op TAG en TAA codons. Om de overgang naar een alternatieve code te vervolledigen is een mutatie in het eRF1 gen, die de binding met TAG en TAA verhindert, noodzakelijk. De complexe distributie van de alternatieve code binnen de Ulvophyceae kan verklaard worden door drie onafhankelijk mutaties van eRF1 in de Trentepohliales, Dasycladales, Cladophorales + *Blastophysa* tezamen met een afname in belang of verlies van de gemuteerde tRNAs door selectie of genetische drift in de tak leidend naar de Bryopsidales. Een nauwgezette vergelijking van eRF1 genen en glutamine tRNAs in de verschillende groepen binnen de Ulvophyceae is echter noodzakelijk om dit scenario te testen.

In **hoofdstuk 5** bestuderen we verschillen in codon gebruik en GC % tussen groenwieren gebaseerd op onze huishoudgenen en analyseren we de evolutie van codon gebruik en GC % in een fylogenetisch kader. We stellen een sterkere bias in codon gebruik vast bij de streptofyten *Mesostigma* en *Chlorokybus* dan bij de rest van de Streptophyta. Binnen de Chlorophyta hebben de prasinofyten, Trebouxiophyceae en Chlorophyceae veel meer bias in codon gebruik dan de Ulvophyceae. Een uitzondering is de ulvofyt *Ignatius* die veel meer bias in codon gebruik heeft dan de rest van de Ulvophyceae. GC waarden vertonen een gelijkaardig patroon, soorten met veel bias in codon gebruik hebben een hoger GC gehalte.

Deze resultaten worden gerelateerd aan de biologie van de organismen door middel van twee modellen: het mutatie-selectie-drift model en het co-evolutie model tussen genoomsamenstelling en bouwsteen allocatie. Het valt op dat unicellulaire organismen en kolonievormende soorten veel meer bias in GC % en codon gebruik vertonen dan multicellulaire en macroscopische organismen. Dit kan het gevolg zijn van de grotere populatiegrootte van unicellulaire organismen waarbij een hogere selectiedruk leidt tot een sterkere bias in codon gebruik. Bij soorten met een kleine populatiegrootte zal de genetische drift er namelijk voor zorgen dat mutaties snel fixeren. De grote populatiegrootte is kenmerkend voor r-geselecteerde soorten. R-geselecteerde sooten zijn tevens klein, groeien snel en hebben een korte generatietijd. We stellen een negatieve correlatie vast tussen de snelheid van moleculaire evolutie en bias in codon gebruik enerzijds en een positieve correlatie tussen GC % en bias in codon gebruik anderzijds.

In hoofdstukken 6 en 7 ligt de nadruk op de fylogenetisch verwantschap en evolutie binnen enkele ulvophyt groepen. Gebaseerd op 5 loci, dateren we in **hoofdstuk 6** een fylogenie van de sifonale ordes Dasycladales en Bryopsidales gebruik makend van 'relaxed molecular clock' methodes gekalibreerd met fossiele gegevens. Volgens deze modellen ontstonden de Dasycladales en Bryopsidales in het late Neoproterozoic of het vroege Cambrium (571 miljoen jaar [628–510]) en diversifieerden de verschillende families binnen deze ordes gedurende het Paleozoic. In **hoofdstuk 7** besturen we de fylogenetische verwantschappen binnen de Cladophorales gebaseerd op small en large subunit nrDNA sequenties. We stellen vast dat *Uronema curvatum*, een mariene soort bestaande uit kleine filamenten die normaalgezien binnen de Chlorophyceae wordt geplaatst, nauw verwant is aan de rest van de Cladophorales. Gebaseerd op deze afwijkende fylogenetische positie beschrijven we een nieuw genus en familie binnen de Cladophorales voor *U. curvatum* (*Okellya*, Okellyaceae).

UNIVERSITEIT
GENT