



# Diversity and ecology of green microalgae in marine systems: an overview based on 18S rRNA gene sequences

Margot Tragin<sup>1</sup>, Adriana Lopes dos Santos<sup>1</sup>, Richard Christen<sup>2,3</sup> and Daniel Vaultot<sup>1\*</sup>

<sup>1</sup> Sorbonne Universités, UPMC Univ Paris 06, CNRS, UMR 7144, Station Biologique, Place Georges Teissier, 29680 Roscoff, France

<sup>2</sup> CNRS, UMR 7138, Systématique Adaptation Evolution, Parc Valrose, BP71. F06108 Nice cedex 02, France

<sup>3</sup> Université de Nice-Sophia Antipolis, UMR 7138, Systématique Adaptation Evolution, Parc Valrose, BP71. F06108 Nice cedex 02, France

\* Corresponding author: vaultot@sb-roscoff.fr

With 5 figures in the text and an electronic supplement

**Abstract:** Green algae (Chlorophyta) are an important group of microalgae whose diversity and ecological importance in marine systems has been little studied. In this review, we first present an overview of Chlorophyta taxonomy and detail the most important groups from the marine environment. Then, using public 18S rRNA Chlorophyta sequences from culture and natural samples retrieved from the annotated Protist Ribosomal Reference (PR<sup>2</sup>) database, we illustrate the distribution of different green algal lineages in the oceans. The largest group of sequences belongs to the class Mamiellophyceae and in particular to the three genera *Micromonas*, *Bathycoccus* and *Ostreococcus*. These sequences originate mostly from coastal regions. Other groups with a large number of sequences include the Trebouxiophyceae, Chlorophyceae, Chlorodendrophyceae and Pyramimonadales. Some groups, such as the undescribed prasinophytes clades VII and IX, are mostly composed of environmental sequences. The 18S rRNA sequence database we assembled and validated should be useful for the analysis of metabarcoding datasets acquired using next generation sequencing.

**Keywords:** Chlorophyta; Prasinophytes; diversity; distribution; 18S rRNA gene; phylogeny; ecology; marine systems

## Introduction

Throughout history, the Earth has witnessed the appearance and disappearance of organisms adapted to their contemporary environments and sometimes these organisms have deeply modified the environment (Kopp et al. 2005, Scott et al. 2008). The best example is provided by the oxygenation of the ocean and the atmosphere by photosynthetic bacteria that first began about 3,500 million years ago (Yoon et al. 2004). Eukaryotic phytoplankton subsequently acquired a chloroplast, a membrane-bound organelle resulting from the phagocytosis without degradation of a cyanobacterium by a heterotrophic host cell (Margulis 1975), 1,500–1,600 million years ago (Hedges et al. 2004, Yoon et al. 2004). This event marked the origin of oxygenic photosynthesis in eukaryotes. During the course of evolution, endosymbiosis has been repeated several times, new hosts engulfing a eukaryote with an existing plastid, leading to secondary and tertiary endosymbioses (McFadden 2001). Early in their evolutionary history photosynthetic eukaryotes separated into two major lineages: the green lineage (which includes green algae

and land plants) and the red lineage (including diatoms and dinoflagellates) (Falkowski et al. 2004). These two lineages diverged approximately 1,100 million years ago according to molecular clock estimates (Yoon et al. 2004), marking the beginning of algal diversification in the ocean. A number of fundamental differences exist between the members of these two lineages (Falkowski et al. 2004), in particular with respect to pigment content, cellular trace-element composition and plastid gene composition. Green algae possess chlorophyll *b* as the main accessory chlorophyll, while algae from the red lineage mainly harbour chlorophyll *c* (i.e. their chloroplast evolved from a Rhodophyta algae after secondary endosymbiosis), influencing their respective light absorption properties and ultimately their distribution in aquatic environments. Algae from the red lineage are often derived from secondary or tertiary endosymbioses and have a chloroplast surrounded by three or four membranes, while algae from the green lineage originate mostly from primary endosymbiosis and have a chloroplast surrounded by only two membranes. The evolutionary history of these lineages is probably much more complex than originally

thought since it has been suggested that the nuclear genome of diatoms contain green genes (Moustafa et al. 2009), although this has been challenged (Deschamps & Moreira 2012). Fossil evidence suggests that during the Palaeozoic Era the eukaryotic phytoplankton was dominated by green algae allowing the colonization of terrestrial ecosystems by charophytes, a branch of the green lineage, ultimately leading to the appearance of land plants (Harholt et al. 2015). However, since the Triassic, the major groups of eukaryotic phytoplankton belong to the red lineage (Tappan & Loeblich 1973, Falkowski et al. 2004).

Green microalgae constitute the base of the green lineage (Nakayama et al. 1998), leading to the hypothesis that the common ancestor of green algae and land plants could be an ancestral green flagellate (AGF) closely related to Chlorophyta (Leliaert et al. 2012). A detailed knowledge of the diversity of green microalgae is necessary to reconstruct phylogenetic relationships within the green lineage. In the marine environment, the diversity, ecology and distribution of green phytoplankton is poorly known since most studies have focused on groups such as diatoms or dinoflagellates. Finally, green algae could become economically important because in recent years potential applications have developed in industrial sectors such as aquaculture, pharmacy and biofuels (Gómez & González 2004, Mishra et al. 2008).

This review summarizes current information on the phylogenetic, morphological and ecological diversity of unicellular marine and halotolerant Chlorophyta (we also include some freshwater groups such as the Monomastigales that are very closely related to marine groups). We used around 9,000 Chlorophyta 18S rRNA gene sequences from culture and environmental samples available in public databases to assess the extent of their diversity and, based on a subset of 2,400 sequences for which geographical information is available, their oceanic distribution. We first present the current state of green algae taxonomy. Then, we detail what is known about each class, and finally we analyze their distribution in oceanic systems from available 18S rRNA sequences. These public sequences were extracted from the annotated and expert validated PR<sup>2</sup> database (Guillou et al. 2013), as detailed in the methodology section at the end of the review.

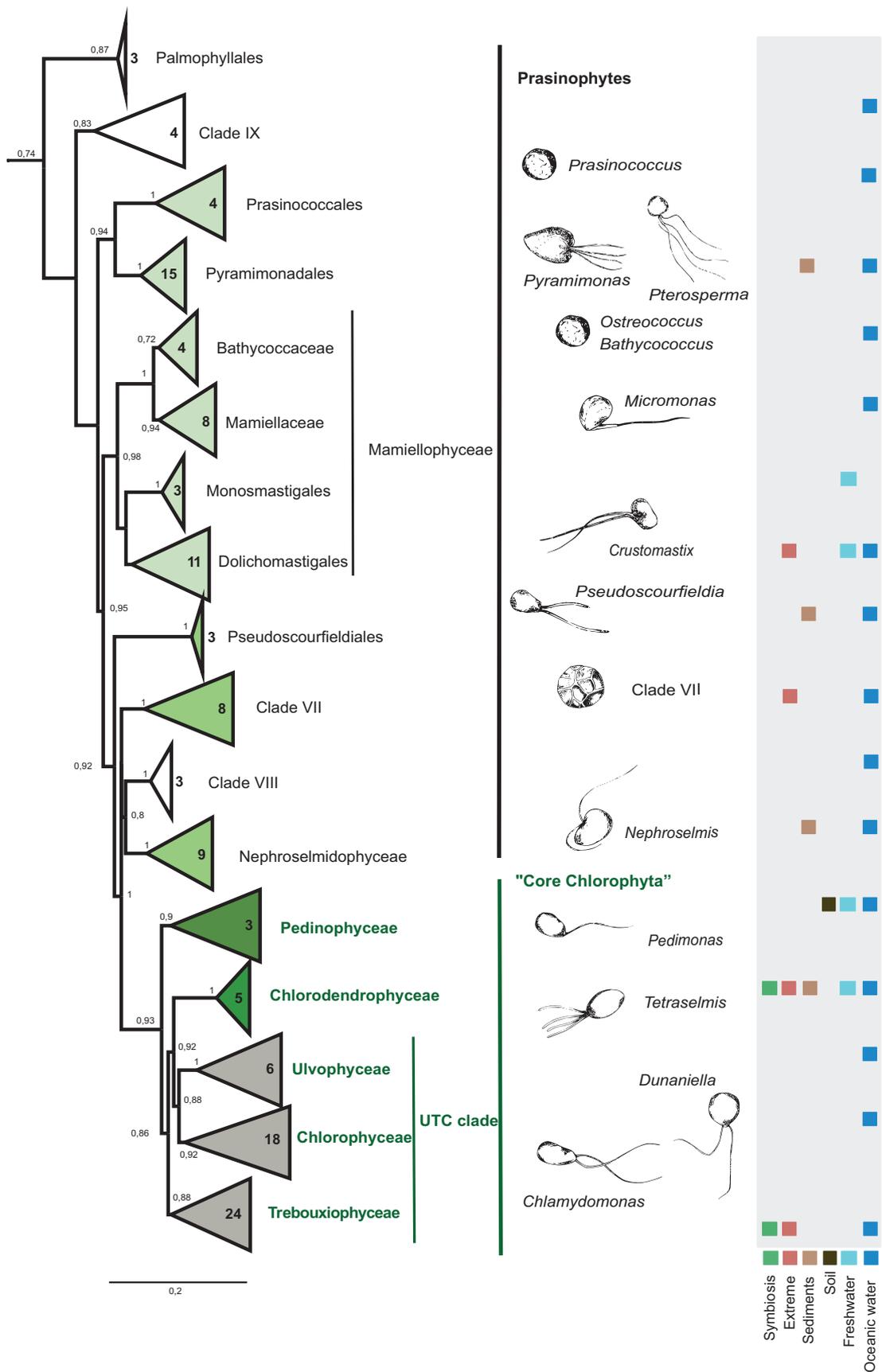
## The present state of Chlorophyta classification

The first description of tiny green cells growing in aquatic environments and the first ideas about the classification of microalgae occurred in the middle of the 19<sup>th</sup> century (Nägeli 1849). This was followed by a large number of descriptions of green microalgae, leading scientists to reflect on the ecological significance of these organisms. Gaarder (1933) discovered the importance of green microalgae in the food web by looking at the source of oyster food in Norway. Twenty years later, the first marine picoeukaryotic phyto-

plankter to be described (*Chromulina pusilla*, later renamed *Micromonas pusilla*) was a tiny green alga (Butcher 1952).

In the 1960s and early 1970s, Round (1963, 1971), reviewing available morphological information, divided the green algae into four divisions: Euglenophyta, Charophyta, Chlorophyta and Prasinophyta. While Round classified the Prasinophyta in a separate phylum, other authors (Bourrelly 1966, Klein & Cronquist 1967) included them in the order Volvocales within the Chlorophyta. The division Chlorophyta was reorganized by Mattox and Stewart (1975) mainly based on ultrastructural characteristics such as the type of mitosis (Sluiman et al. 1989), presence/absence of an interzonal spindle, the structure of the flagellar apparatus (O'Kelly & Floyd 1983), and the presence of extracellular features such as scales and thecae. They proposed the division of Chlorophyta into four major groups: the Prasinophyceae, Charophyceae, Ulvophyceae and Chlorophyceae (Stewart & Mattox 1978). This has been partly confirmed by molecular phylogenetic analyses over the years (Chapman et al. 1998), although it was recognized from the beginning (Christensen 1962) that prasinophytes constitute a polyphyletic assemblage (i.e. phylogenetic branches without a common ancestor). Therefore the class name Prasinophyceae is no longer used and the generic term prasinophytes, that has no phylogenetic meaning, has replaced it (Leliaert et al. 2012). At present, the Chlorophyta is viewed as composed of two major groups: the prasinophytes and the “core” chlorophytes (Leliaert et al. 2012, Fučíková et al. 2014).

The prasinophytes currently consist of nine major lineages of microalgae corresponding to different taxonomic levels (order, class, undescribed clades) that will probably all be raised to the class level in the future (Leliaert et al. 2012). These lineages share ancestral features such as flagella and organic scales. The number of prasinophyte lineages has been increasing following the availability of novel environmental sequences. Ten years ago, prasinophyte clade VII was introduced using sequences from cultured strains and environmental clone libraries (Guillou et al. 2004). Four years later, two additional clades, VIII and IX, were reported (Viprey et al. 2008) that are only known so far from environmental sequences. Prasinophytes may be divided into three informal groups (Marin & Melkonian 2010): a group of “basal” lineages (Prasinococcales, Pyramimonadales, Mamiellophyceae), a group of “intermediate” lineages (Pseudosourfieldiales, clade VII, Nephroselmidophyceae) and a group of “late” diverging lineages (Pedinophyceae and Chlorodendrophyceae). Recently, these two “late” diverging lineages have been merged with the Ulvophyceae-Trebouxiophyceae-Chlorophyceae (UTC) clade into the “core” chlorophytes (Fučíková et al. 2014): the Chlorodendrophyceae based on common features, in particular a mode of cell division mediated by a phycoplast (Mattox & Stewart 1984, Leliaert et al. 2012) and the Pedinophyceae based on strong phylogenetic support (Marin 2012, Fučíková et al. 2014).



**Fig. 1.** Phylogenetic tree of a set of 132 rRNA 18S reference sequences (Supplementary Table S1) constructed by FastTree (options used: General Time Reverse model, optimized gamma likelihood, rate categories of sites 20), rooted with *Oryza sativa* (AACV01033636) based on an edited MAFFT 1752 bp alignment stripped at 50% (columns of the alignment counting more than 50% gaps were deleted, Supplementary data). The phylogenetic tree was validated by MrBayes phylogeny which provided a similar result. Fast Tree bootstrap values larger than 70% are reported. The number of reference sequences for each group is also reported. Triangle colors correspond to the different groups defined by Marin & Melkonian (2010). Groups labelled in green correspond to "core" chlorophytes. Symbols on the right side of the tree indicate the habitat of each group.

## Major lineages within marine Chlorophyta

We extracted a set of 132 reference sequences from the PR<sup>2</sup> database that were used to build a phylogenetic tree of marine Chlorophyta (Fig. 1). In this tree, each triangle (except for the Mamiellophyceae for which we have represented the different families) corresponds to a “lineage” which currently corresponds to either a class, an order or a “clade” *sensu* Guillou et al. (2004). In this section, we review what is known about the major Chlorophyta lineages in marine waters following the order used in Guillou et al. (2004).

**Pyramimonadales** (prasinophyte clade I, Guillou et al. 2004) are pyramidal, oval or heart-shaped cells (10 to 400  $\mu\text{m}$  long on average) with generally 4, rarely 8 or even 16 flagella (Chadefaud 1950, Hori et al. 1985). In *Pyramimonas*, cells possess three layers of different organic scales on the cell body, two layers on the flagella (Pennick 1982, 1984) and flagellar hairs (Moestrup 1982). Twenty-two genera have been described, with *Pyramimonas*, *Pterosperma* and *Halosphaera* containing most species. For the genus *Pyramimonas*, almost 50 species (Suda et al. 2013, Harðardóttir et al. 2014, Bhuiyan et al. 2015) and six sub-genera (Hori et al. 1995) have been described, but the low number of ribosomal rRNA sequences from described species in public sequence databases is an obstacle to the resolution of the phylogeny of this genus (Table S1, Suda et al. 2013). Novel species have recently been described from isolates from the North Pacific Ocean (Fig. 3A, Suda et al. 2013, Bhuiyan et al. 2015) and polar regions (Moro et al. 2002, Harðardóttir et al. 2014). In Disko Bay (Greenland), *Pyramimonas* has been found to be important in the sea ice and in the water column and plays an important role in the spring phytoplankton bloom (Harðardóttir et al. 2014). Pyramimonadales have been recorded in coastal waters as well as in confined environments such as tide pools (Chisholm & Brand 1981, Lee 2008). *Halosphaera* occurs in two forms, one flagellated and one coccoid; the latter can be up to 800  $\mu\text{m}$  in size and may sediment quickly. In the Mediterranean Sea, high abundances of *Halosphaera* have been recorded at depths between 1,000 and 2,000 meters (Wiebe et al. 1974).

**Mamiellophyceae** (clade II, Guillou et al. 2004) are characterized by a wide morphological diversity. They are split into three orders: Mamiellales, which is composed of two families (Mamiellaceae and Bathycoccaceae), Dolichomastigales and Monomastigales (Fig. 1, Marin & Melkonian 2010).

The Mamiellaceae contain three genera that are ecologically important. *Micromonas* are ellipsoid to pyriform naked cells (1 to 3  $\mu\text{m}$ ) with a single emergent flagellum (Butcher 1952). Phylogenetic and ecological studies on the micro-diversity of *Micromonas* suggest that this genus may consist of at least three cryptic species (Šlapeta et al. 2006, Foulon et al. 2008). *Micromonas* is a ubiquitous genus with cultures originating from a wide range of environments extending from the poles to the tropics, but more prevalent in coastal

waters. *Mamiella* and *Mantoniella* are reniform cells (up to 10  $\mu\text{m}$ ) covered by two types of body scales: large, more or less square, and small, less regular (Barlow & Cattolico 1980, Moestrup 1984). *Mamiella* have two long flagella and spined flagellar scales, while *Mantoniella* has one long and one very short flagella with flagellar scales lacking spines (Marin & Melkonian 1994). Environmental sequences from the latter two genera have been found in the Arctic Ocean and the Mediterranean Sea using Chlorophyta specific primers or sorted samples (Viprey et al. 2008, Balzano et al. 2012).

Bathycoccaceae are spherical or elliptical coccoid cells and contain two genera, *Bathycoccus*, which is covered by spider-web-like scales (1.5 to 2.5  $\mu\text{m}$ ) (Eikrem & Throndsen 1990), and *Ostreococcus*, which is naked and the smallest known photosynthetic eukaryote to date, with a typical size of 0.8  $\mu\text{m}$  (Chrétiennot-Dinet et al. 1995). *Ostreococcus* was first isolated from a Mediterranean Sea lagoon (Courties et al. 1994) and then from many mesotrophic oceanic regions (Rodríguez et al. 2005, Viprey et al. 2008). Four clades of *Ostreococcus* have been described (Guillou et al. 2004) leading to the formal description of 2 species (Subirana et al. 2013). *Bathycoccus* does not seem to show micro-diversity based on sequences of the 18S rRNA gene (Guillou et al. 2004), although ITS (internal transcribed spacers) sequence and genomic evidence points to the existence of two different ecotypes (Vaultot et al. 2012, Monier et al. 2013). *Bathycoccus* was first isolated from Mediterranean Sea and Norwegian waters (Eikrem & Throndsen 1990), but sequences have now been recovered from many regions (Viprey et al. 2008).

Monomastigales cells are oblong (3.5 to 15  $\mu\text{m}$  long) and covered by proteinaceous scales. Cells have a single flagellum and a second immature basal body (Heimann et al. 1989). The only member of this order is the freshwater genus *Monomastix*. Sequences have been recorded only in freshwater on four continents (Europe, North America, Asia, Australia) (Scherffel 1912, Marin & Melkonian 2010).

Dolichomastigales cells are round or bean-shaped (2 to 5  $\mu\text{m}$  long), biflagellate, naked or covered by spider-web-like scales or a crust. This order regroups the genera *Dolichomastix*, isolated in the Arctic, South Africa and Mediterranean Sea (Manton 1977, Throndsen & Zingone 1997) and *Crustomastix*, first isolated in the Mediterranean Sea (Nakayama et al. 1998, Zingone et al. 2002, Marin & Melkonian 2010).

**Nephroselmidophyceae** (clade III, Guillou et al. 2004) is a class of flattened or bean-shaped cells, with two unequal flagella. The cell body (4.5 to 7  $\mu\text{m}$  long) is covered by 5 different types of scales (squared and stellate) and the flagella by 3 types (Nakayama et al. 2007). Eleven genera have been described in this class and the genus *Nephroselmis* counts the largest number of described species (14 according to AlgaeBase, Table S2) of which 6 new species have recently been described from coastal South African and Pacific waters (Faria et al. 2011, 2012, Yamaguchi et al. 2011, 2013).

**Pseudoscourfieldiales** (clade V, Guillou et al. 2004) are coccoid cells (1.5 to 5  $\mu\text{m}$  in diameter) without scales but with a cell wall (*Pycnococcus provasolii*, Guillard et al. 1991) or with scales and biflagellate (*Pseudoscourfieldia marina*, Moestrup & Thronsen 1988). *Pycnococcus* was initially isolated from the North Atlantic Ocean (Guillard et al. 1991), but cultures have also been recovered from other environments such as the South-East Pacific Ocean (Le Gall et al. 2008). The 18S rRNA sequences of the two species are 100% identical, leading to the hypothesis that they could represent different life cycle stages, or growth forms, of the same species (Fawley et al. 1999).

**Prasinococcales** (clade VI, Guillou et al. 2004) is an order composed of coccoid cells (2.5 to 5.5  $\mu\text{m}$  in diameter) without scales, surrounded in general by a multilayer gelatinous matrix made of polysaccharides (Hasegawa et al. 1996, Sieburth et al. 1999). The two main genera are *Prasinococcus* (Miyashita et al. 1993) and *Prasinoderma* (Hasegawa et al. 1996). One species, *Prasinoderma singularis*, lacks the gelatinous matrix (Jouenne et al. 2011). Prasinococcales have been isolated from coastal and open oceanic waters in the North Atlantic (Sieburth et al. 1999) and Pacific Oceans (Miyashita et al. 1993) as well as in the Mediterranean Sea (Viprey et al. 2008). One novel environmental *Prasinoderma* clade has been found using Chlorophyta specific primers (Viprey et al. 2008).

**Prasinophyte clade VII** has been identified from environmental and culture sequences (Guillou et al. 2004). The first isolate of prasinophyte clade VII, CCMP1205 (= RCC15), was reported by Potter et al. (1997). Since then, the lack of distinct morphological characters has kept these small (3 to 5  $\mu\text{m}$ ) coccoid cells without a formal description despite their importance in oceanic waters in particular in the South Pacific Ocean, Mediterranean Sea and South China Sea (Moon-van der Staay et al. 2001, Viprey et al. 2008, Shi et al. 2009, Wu et al. 2014). Prasinophyte clade VII is divided into three well-supported lineages, A, B and C, the latter being formed by *Picocystis salinarum*, a small species found in hypersaline lakes (Lewin et al. 2000, Krienitz et al. 2012). The large number of clade VII strains and environmental sequences now present in public databases has allowed further delineation of at least 10 sub-clades (Lopes dos Santos et al. submitted) within the two major marine lineages A and B described by Guillou et al. (2004).

**Prasinophyte clade VIII** is a clade known purely from environmental sequences, specifically 3 sequences from the picoplankton size fraction (i.e. cells passing through a 3  $\mu\text{m}$  pore-size filter) found at a single sampling location station in the Mediterranean Sea (Viprey et al. 2008).

**Prasinophyte clade IX** is also an environmental clade. This clade was initially found using either Chlorophyta-specific primers or from flow cytometry sorted samples (Viprey et al. 2008, Shi et al. 2009). Sequences originate mostly from picoplankton samples collected in oligotrophic

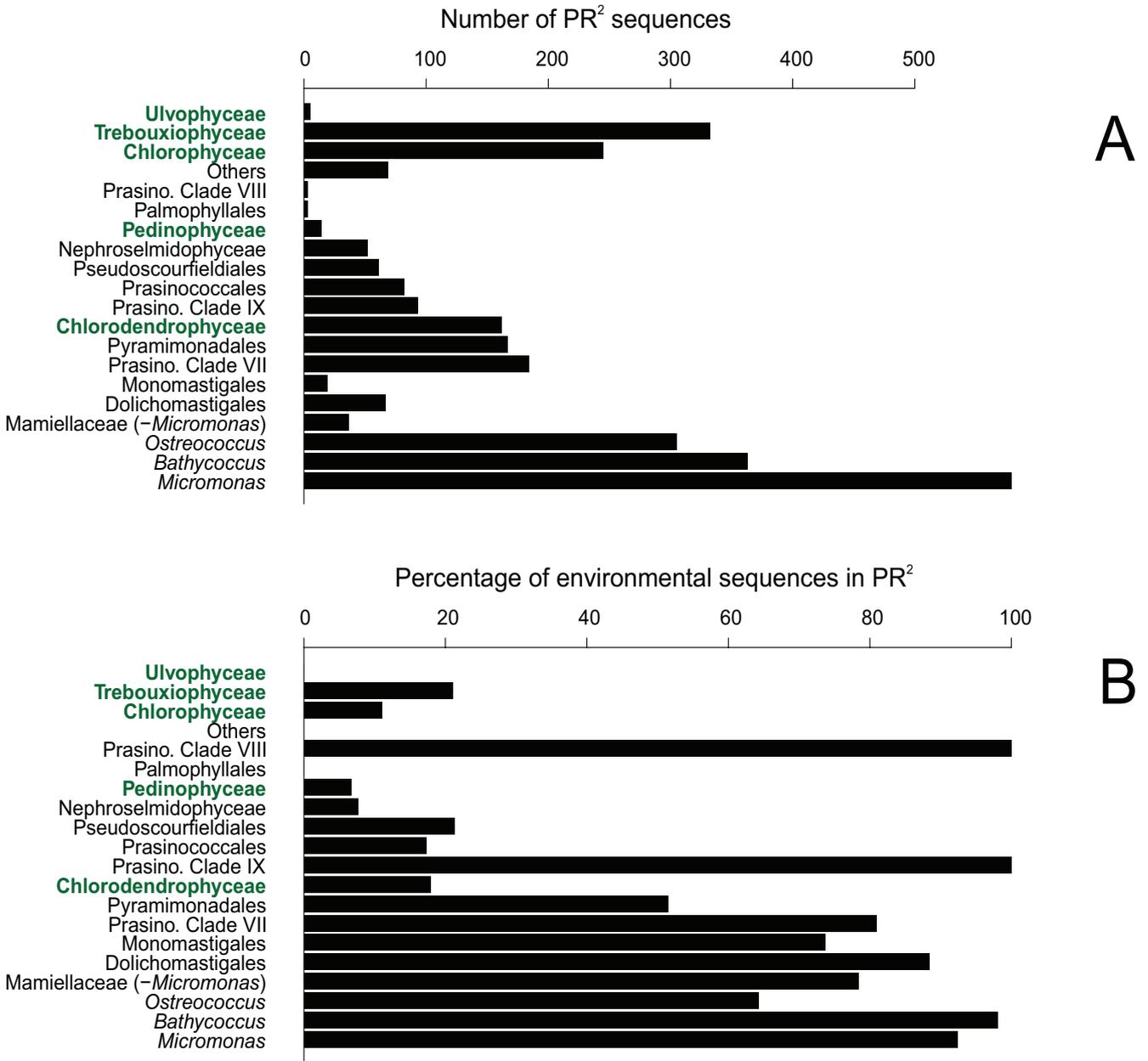
areas from the Pacific Ocean (Shi et al. 2009, Wu et al. 2014) and the Mediterranean Sea (Viprey et al. 2008).

**Palmophyllales** is an order of poorly known colonial algae with a thalli formed by isolated spherical cells in a gelatinous matrix (Zechman et al. 2010, Leliaert et al. 2012). These green algae have been isolated from moderately deep waters. The genus *Palmophyllum* was described from cells (6–7  $\mu\text{m}$ ) growing at 70 m depth near New Zealand (Nelson & Ryan 1986), while *Verdigellas* (Ballantine & Norris 1994) may live below 100 m and was isolated from the tropical Atlantic Ocean (Zechman et al. 2010). Phylogenetic studies based on the 18S rRNA (3 sequences from isolates are available) and two plastid genes suggested that this lineage is deep branching within the Chlorophyta (Zechman et al. 2010).

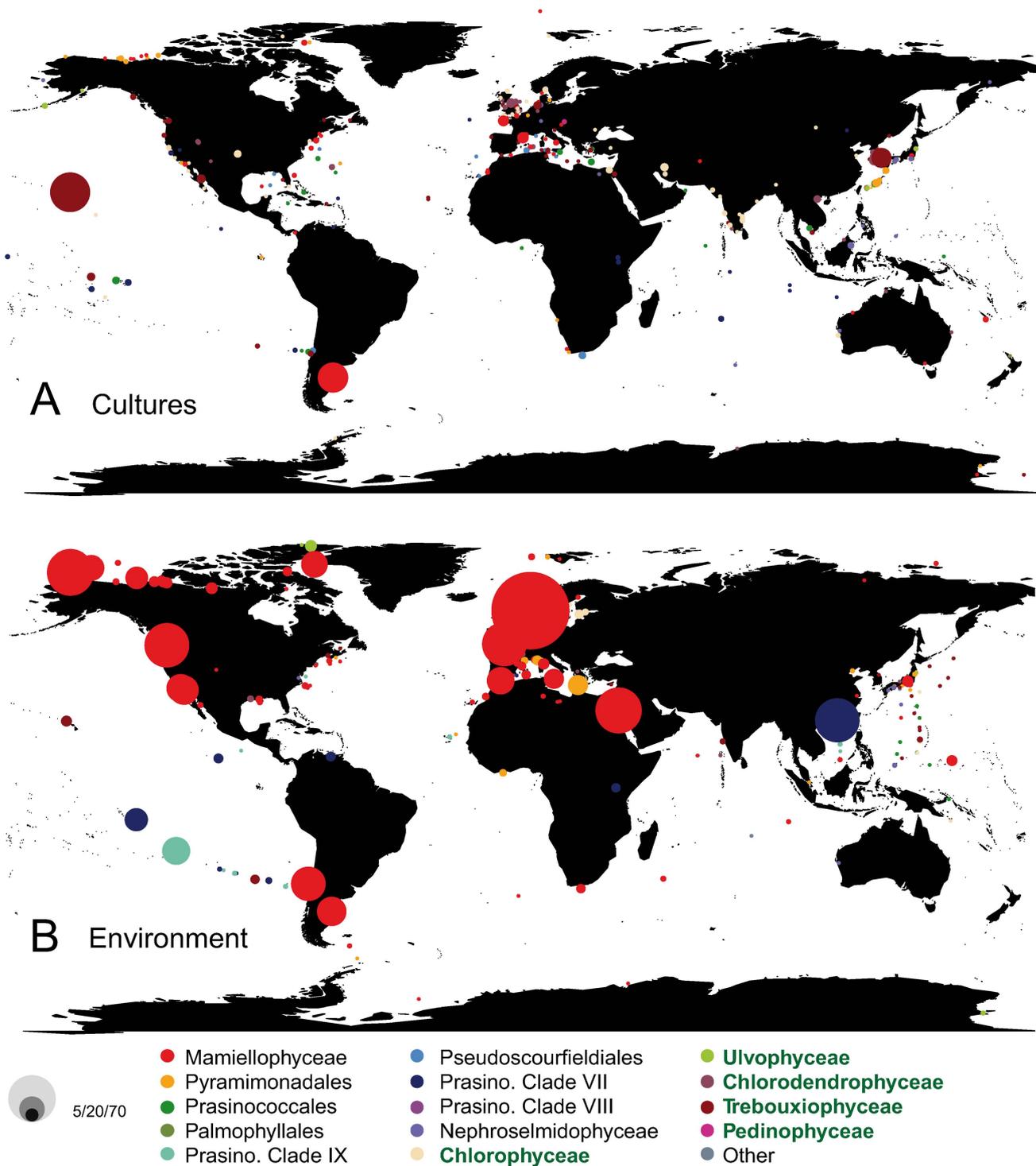
**Pedinophyceae** cells are asymmetrical, ovoid or ellipsoid (about 3  $\mu\text{m}$  long), uniflagellate and naked (Moestrup 1991). This class consists of two orders, the Pedinomonadales and the Marsupiomonadales (Marin 2012) and six genera (Table S2). One genus of Marsupiomonadales (*Resultomonas*) does not have any 18S sequence available. Marsupiomonadales are marine, while Pedinomonadales live in soil and freshwater (Fig. 1, Marin 2012).

**Chlorodendrophyceae** (clade IV, Guillou et al. 2004) are quadriflagellate elliptical cells (on average 15 to 20  $\mu\text{m}$  long). The cell body is covered by a theca (resulting from the fusion of stellar scales) and the flagella, thick and shorter than the cell, are covered by 2 layers of scales and hairs (Hori et al. 1982). This class contains four genera (Table S2) (Lee & Hur 2009). The genus *Tetraselmis*, for which several species have been isolated from brackish lagoons, has been divided into four sub-genera (Hori et al. 1982, 1983, 1986), but molecular studies using the 18S rRNA gene fail to resolve the phylogeny of this genus (Arora et al. 2013).

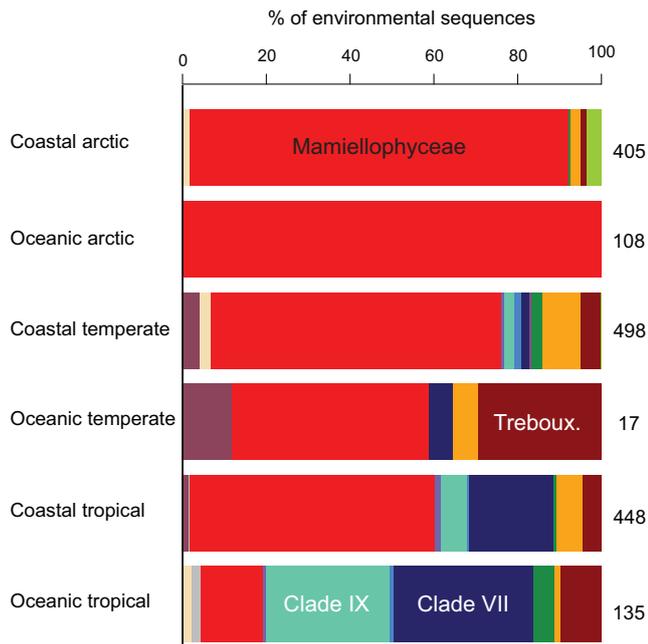
**The UTC (Ulvophyceae, Trebouxiophyceae and Chlorophyceae) clade** shows a wide morphologic diversity. Most UTC representatives are macroalgae or originate from freshwater or terrestrial environments. We only focus here on unicellular marine representatives. Unicellular marine Ulvophyceae are represented by one genus, *Halochlorococcus*, with very few sequences from cultures, all originating from Japan (Table S3). Trebouxiophyceae are mostly represented by coccoid cells in coastal marine environments belonging to the genera *Picochlorum* (2  $\mu\text{m}$  in diameter), *Chlorella* (1.5 to 10  $\mu\text{m}$  in diameter), *Elliptochloris* (5 to 10  $\mu\text{m}$  in diameter) and *Chloroidium* (~15  $\mu\text{m}$  in diameter) (Andreoli et al. 1978, Henley et al. 2004, Letsch et al. 2009, Darienko et al. 2010). Chlorophyceae are morphologically diverse (de Reviers 2003) from non-motile coccoid cells to flagellates. Most sequences from marine Chlorophyceae strains, isolated from coastal waters or salt pools, belong to the genera *Asteromonas* (12 to 22  $\mu\text{m}$  long), *Chlamydomonas* (7 to 11  $\mu\text{m}$  long), and *Dunaliella* (8 to 18  $\mu\text{m}$  long) (Hoshaw & Ettl 1966, Peterfi & Manton 1968, Preetha 2012).



**Fig. 2.** Number of Chlorophyta 18S rRNA sequences in the PR<sup>2</sup> database (A) and percentage of environmental sequences in PR<sup>2</sup> (B) for each clade. The number of sequences for Mamiellaceae does not include *Micromonas* which is reported separately. Groups labelled in green correspond to “core” chlorophytes (see Fig. 1).



**Fig. 3.** Oceanic distribution of PR<sup>2</sup> 18S rRNA sequences for major Chlorophyta lineages for cultures (A) and environmental samples (B). The color of the circle corresponds to the most abundant lineage and the surface of the circle is proportional to the number of sequences for this lineage obtained at the location. Groups labelled in green correspond to “core” chlorophytes (see Fig. 1).



**Fig. 4.** Distribution of PR<sup>2</sup> Chlorophyta 18S rRNA environmental sequences according to three latitudinal zones (90° to 60°, 60° to 35° and 35° to 0°) and to the distance to the nearest shore (locations closer than 200 km were considered as coastal and the rest as oceanic). Distances to the coast were computed for each sequence using the R packages *rgdal* and *rgeos*. Antarctica is not represented because of the very low number of sequences from this area. Colours correspond to Chlorophyta classes and are the same as in Figure 3. The number of sequences in each group is indicated on the right.

## Environmental distribution of Chlorophyta in marine ecosystems from 18S rRNA sequences

The number of publicly available 18S rRNA gene sequences (Fig. 2A, based on the PR<sup>2</sup> database, Guillou et al. 2013) varies widely between the Chlorophyta groups from 3 for the Palmophyllales up to 560 for the sole genus *Micromonas*. Mamiellophyceae and in particular *Micromonas*, *Bathycoccus* and *Ostreococcus* are the most represented green algal taxa in public sequence databases, followed by Chlorophyceae and Trebouxiophyceae, two groups which were previously mostly seen as continental (Fig. 2). The proportion between sequences from cultures and environmental samples is also highly variable (Fig. 2B). Some groups are mostly represented by sequences from cultures (e.g. Nephroselmidophyceae and “core” chlorophytes) while others are predominantly or wholly uncultured (e.g. prasinophyte clade IX). The geographic distribution obtained from cultures and from environmental sequences is quite different (compare A and B in Fig. 3 and S1). While Mamiellophyceae dominate environmental sequences, this is not true for culture sequences, which offer a better balance between the

different Chlorophyta groups (Fig. S1). The contribution of different classes to environmental sequences differs between latitudinal bands and coastal vs. oceanic stations (Fig. 4).

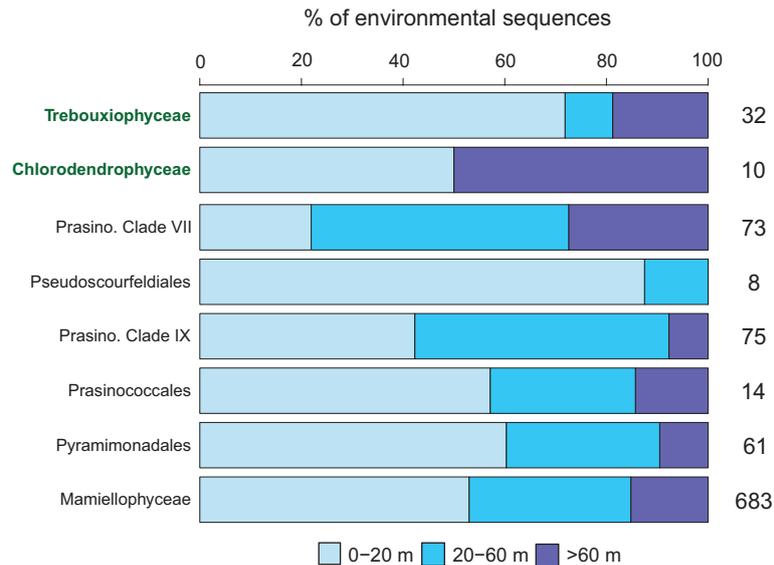
Polar waters, whether oceanic or coastal, are totally dominated by Mamiellophyceae (Fig. 4), in particular the arctic *Micromonas* clade (Lovejoy et al. 2007, Balzano et al. 2012). The diversity of classes recovered is minimal (Supplementary Fig. S1), with representatives of the Pyramimonadales, Ulvophyceae and Prasinococcales in addition to the Mamiellophyceae. It is noteworthy that few sequences have been recovered from the Southern Ocean in comparison to the Arctic (Supplementary Fig. S1).

The dominance of Mamiellophyceae is less marked for temperate waters where other classes such as the Trebouxiophyceae can be important, especially away from the coast (Fig. 4). Indeed, it is in temperate waters that Chlorophyta environmental sequence diversity is maximal, in particular in the North-West Atlantic and North-East Pacific Oceans with more than 10 Chlorophyta classes recovered (Supplementary Fig. S1). Chlorophyceae, Prasinococcales Pseudoscourfieldiales, and clade IX sequences have also been recovered from coastal temperate areas including the Mediterranean Sea (Fig. 3B and 4). Nephroselmidophyceae have been repeatedly isolated from Japanese coastal waters (Fig. 3). Chlorodendrophyceae, Trebouxiophyceae, Pyramimonadales and clade VII have been found in both coastal and oceanic temperate waters (North Pacific Ocean and Mediterranean Sea, Fig. 3 and 4).

The decrease in the dominance of Mamiellophyceae is even more marked in tropical waters. While it shares dominance with prasinophyte clade VII in coastal waters, it becomes a minor component offshore where it is replaced by clade VII and the uncultured clade IX. Trebouxiophyceae, Prasinococcales, Pyramimonadales and Chlorophyceae have also been found at some locations in the subtropical Pacific and Atlantic Oceans (Fig. 3B and 4).

With respect to depth distribution, both Mamiellophyceae, Pyramimonadales, as well as prasinophyte clade VII and IX sequences have been found throughout the photic zone, even below 60 meters (Fig. 5). Pseudoscourfieldiales, Trebouxiophyceae and Prasinococcales sequences seem to be restricted to surface waters, while Chlorodendrophyceae sequences appear to be preferentially found at the bottom of the photic zone, below 60 m (Fig. 5). The deepest Mamiellophyceae sequences have been recovered from 500 m depth for *Micromonas* and down to 2500 m depth for *Ostreococcus* (Lie et al. 2014).

Mamiellophyceae have been found to dominate environmental sequences in some anoxic waters, as, for example, near Saanich Inlet off Vancouver (Orsi et al. 2012). Sediments also constitute environments where green microalgal sequences have been recovered (Fig. 1). For example Dolichomastigales (Mamiellophyceae), Chlorodendrophyceae and prasinophyte clade VII have been found in anoxic sediments (Edgcomb et al. 2011) or in cold methane sediments (Takishita et al. 2007).



**Fig. 5.** Number of environmental Chlorophyta 18S rRNA sequences in PR<sup>2</sup> according to depth range for each lineage (only sequences for which depth is reported in the GenBank record are included and lineages for which less than 5 sequences were available were omitted).

Cultures of Nephroselmidophyceae, Chlorodendrophyceae, Pseudoscourfieldiales and Pyramimonadales have also been isolated from sediments (Fig. 1). However, Chlorophyta found in sediments may not correspond to truly benthic species, but could result from cell sedimentation down the water column.

### Advantages and limitations of 18S rRNA as a marker gene for Chlorophyta

Our analysis was based on Chlorophyta sequences for the 18S rRNA gene that are publicly available. Using this gene, we were able to recover (Fig. 1) the three diverging groups described by Marin and Melkonian (Marin & Melkonian 2010) using both nuclear (18S) and plastid (16S) encoded rRNA: the early diverging group (Pyramimonadales, Mamiellophyceae and Prasinococcales), the intermediate group (Nephroselmidophyceae, Pseudoscourfieldiales and clade VII) and the late-diverging group (Chlorodendrophyceae and Pedinophyceae). Further investigation of the phylogenetic relationships between the different Chlorophyta lineages would require multiple markers. For example, Fučíková et al. (2014) used 8 genes, including *rcbL* (the large subunit of the ribulose-1,5-biphosphate carboxylase-oxygenase gene), *tufA* (translation unstable factor) and the 18S rRNA to address the relationship within “core” chlorophytes. In order to explore microdiversity at the species level or below, the LSU (large ribosomal subunit) or the ITS seems to be more suitable (Coleman 2003). For example, the four

*Ostreococcus* clades (Mamiellophyceae) are better discriminated with ITS than 18S rRNA (Rodríguez et al. 2005).

In the course of this work, Chlorophyta 18S rRNA sequences were verified and re-annotated. The resulting updated database contains 8554 sequences (Supplementary data S1) and will be useful to annotate Chlorophyta metabarcoding sequences from the V4 or V9 regions of the 18S rRNA gene obtained with High Throughput Sequencing (de Vargas et al. 2015, Massana et al. 2015). The level of similarity within each phylogenetic lineage varies depending on the Chlorophyta lineage and on the region of the gene considered (Supplementary Fig. S2). For the full 18S rRNA gene, it varies from 83.6% for Ulvophyceae to 99.9% for Pseudoscourfieldiales.

Within most of the lineages, the V9 region (2,416 sequences) seems more divergent than the V4 region (6,530 sequences), but identity levels are more variable for the former (Supplementary Fig. S2). The V9 region therefore appears to be a good marker for Chlorophyta, although the larger size of the V4 region could be advantageous to reconstruct the phylogeny of novel groups without representatives in the reference database.

A number of caveats have, however, to be considered. Some sequences do not cover the full length of the 18S rRNA gene. For example, only 2,416 sequences (28% of sequences analyzed) cover the full V9 region. Some environmental clades (e.g. prasinophyte clade VIII) are represented only by short sequences, and this clade would be missed in metabarcoding studies using the V9 region. Moreover, not all described species have published 18S sequences.

For example, two Pedinophyceae genera are known to live in marine waters, *Resultomonas* and *Marsupiomonas*, but sequences are only available for the latter genus. *Resultomonas* will therefore be “invisible” in surveys based on environmental DNA. Some groups have only cultured sequences, e.g. Nephroselmidophyceae, with an overrepresentation off Japan, because scientists from this country have a keen interest in this group (Nakayama et al. 2007, Faria et al. 2011, 2012, Yamaguchi et al. 2011, 2013). Other groups, such as prasinophyte clades VIII and IX, have completely escaped cultivation and obtaining environmental sequences from these groups is difficult because of competition among different templates when using universal PCR primers. Two methods have been used to increase recovery of Chlorophyta sequences: the use of Chlorophyta specific primers and flow cytometry sorting of photosynthetic organisms (Viprey et al. 2008, Shi et al. 2009). Another limitation is that metadata available in GenBank are far from complete and even sometimes not accurate. For example, only ~ 2,500 sequences are associated with geographical coordinates (Fig. 3 and Fig. S1) and less than 1,000 environmental sequences have depth information (Fig. 5).

## Conclusion and perspectives

Despite being neglected in comparison to other groups such as diatoms and dinoflagellates, marine green algae are very diverse and are distributed worldwide. Some groups, such as the Mamiellophyceae, are ubiquitous (Fig. 4) and are starting to be well characterized from the physiological and genomic points of view, while other groups, such as prasinophyte clade IX, still remain uncultured. In the future, metabarcoding will make it possible to improve our knowledge of the worldwide distribution of each clade and identify their ecological niches.

## Methodology

In order to determine the extent of molecular diversity of marine Chlorophyta, we used the Protist Ribosomal Reference (PR<sup>2</sup>) database (Guillou et al. 2013). This database contains public eukaryotic 18S rRNA sequences from cultured isolates as well as from environmental samples that have been quality controlled and annotated. All Chlorophyta sequences were extracted, yielding a final dataset of around 9,000 sequences. For each sequence, we extracted metadata from GenBank (such as sampling coordinates and date, publication details), when available. Other metadata were obtained from the literature or from culture collection websites. This information was entered into a Microsoft Access database. In particular, the sampling coordinates were used to map the sequence distribution using the packages maps v2.3–9 and mapdata v2.2–3 of the R3.0.2 software ([\[www.R-project.org/\]\(http://www.R-project.org/\)\). The database and the metadata have been deposited to Figshare \(see Supplementary data\).](http://</a></p>
</div>
<div data-bbox=)

The assignment of sequences was checked down to the species level. For this purpose, we aligned sequences for each phylogenetic group (in general at the class level) using MAFFT v1.3.3 (Kato 2002). Phylogenetic trees were constructed using FastTree v1.0 (Price et al. 2009) run within the Geneious software v7.1.7 (Kearse et al. 2012). Phylogenetic trees were compared with those found in the literature. We defined phylogenetic clades as monophyletic groups of sequences that were supported by bootstrap values higher than 70%, with 2 or 3 different phylogenetic methods (Grosillier et al. 2006, Guillou et al. 2008). If more than 2 strain sequences from the same species belonged to a given clade, then the other sequences in this clade were assigned to that species in the database. When the tree was not clear enough, for example for groups represented by a large number of sequences, signatures in the alignment were used to validate the assignment. Chimeric sequences were filtered out by assigning the first 300 and last 300 base pairs of the sequences with the software mothur v1.35.1 (Schloss et al. 2009). If a conflict of assignment between the beginning and the end of the sequences was detected then, they were BLASTed against GenBank to confirm whether they were chimeras and in the latter case, removed from any further analysis.

Reference sequences for each Chlorophyta class were selected and a reference Chlorophyta tree containing 132 sequences was built using Maximum Likelihood and Bayesian methods (Fig. 1, Table S1, Supplementary data). When possible, the reference sequences were full-length 18S rRNA sequences from culture strains and already used as references in the literature. Moreover, they were chosen to be distributed in the major clades of each lineage and as a result the number of reference sequences was a function on the micro-diversity within each class.

**Acknowledgements:** Financial support for this work was provided by the European Union projects MicroB3 (UE-contract-287589) and MaCuMBA (FP7-KBBE-2012-6-311975) and the ANR Project PhytoPol. MT was supported by a PhD fellowship from the Université Pierre et Marie Curie and the Région Bretagne. We would like to thank Adriana Zingone, Bente Edvardsen, Fabrice Not, Ian Probert and two anonymous reviewers for their constructive comments during the preparation of this review.

## References

- Andreoli, C., Rascio, N. & Casadoro, G. (1978): *Chlorella nana* sp. nov. (Chlorophyceae): a new marine *Chlorella*. – Bot. Mar. 21: 253–256.
- Arora, M., Anil, A.C., Leliaert, F., Delany, J. & Mesbahi, E. (2013): *Tetraselmis indica* (Chlorodendrophyceae, Chlorophyta), a new species isolated from salt pans in Goa, India. – Eur. J. Phycol. 48: 61–78.

- Ballantine, D. & Norris, J. (1994): *Verdigellas*, a new deep water genus (Tetrasporales, Chlorophyta) from the Tropical Western Atlantic. – *Cryptogam. Bot.* 4: 368.
- Balzano, S., Marie, D., Gourvil, P. & Vault, D. (2012): Composition of the summer photosynthetic pico and nanoplankton communities in the Beaufort Sea assessed by T-RFLP and sequences of the 18S rRNA gene from flow cytometry sorted samples. – *ISME J.* 6: 1480–1498.
- Barlow, S.B. & Cattolico, R.A. (1980): Fine structure of the scale-covered green flagellate *Mantoniella squamata* (Manton et Parke) Desikachary. – *Br. Phycol. J.* 15: 321–333.
- Bhuiyan, M.A.H., Faria, D.G., Horiguchi, T., Sym, S.D. & Suda, S. (2015): Taxonomy and phylogeny of *Pyramimonas vacuolata* sp. nov. (Pyramimonadales, Chlorophyta). – *Phycologia* 54: 323–332.
- Bourrelly, P. (1966): Les Algues d'eau douce: Initiation a la systematique. I les algues vertes. Boubée. Paris.
- Butcher, R.W. (1952): Contributions to our knowledge of the smaller marine algae. – *J. Mar. Biol. Assoc. United Kingdom* 31: 175.
- Chadefaud, M. (1950): Les cellules nageuses des algues dans l'embranchement des Chromophycées. – *C. R. Acad. Sci. Paris* 231: 788–790.
- Chapman, R.L., Buchheim, M.A., Delwiche, C.F., Friedl, T., Huss, V.A.R., Karol, K.G., Lewis, L.A. et al. (1998): Molecular systematics of the Green Algae. – In: Soltis, D.E., Soltis, P.S. & Doyle, J.J. (eds.), *Molecular Systematics of Plants II*. Springer US; Boston, MA, pp. 508–540.
- Chisholm, S.W. & Brand, L.E. (1981): Persistence of cell division phasing in marine phytoplankton in continuous light after entrainment to light: dark cycles. – *J. Exp. Mar. Bio. Ecol.* 51: 107–118.
- Chrétiennot-Dinet, M.-J., Courties, C., Vaquer, A., Neveux, J., Claustre, H., Lautier, J. & Machado, M.C. (1995): A new marine picoeucaryote: *Ostreococcus tauri* gen. et sp. nov. (Chlorophyta, Prasinophyceae). – *Phycologia* 34: 285–292.
- Christensen, T. (1962): Systematisk Botanik, Alger. – In: Bocher, T.W., Lange, M. & Sorensen, T. (eds.), *Botanik*. Munksgaard, Copenhagen, pp. 1–178.
- Coleman, A.W. (2003): ITS2 is a double-edged tool for eukaryote evolutionary comparisons. – *Trends Genet.* 19: 370–375.
- Courties, C., Vaquer, A., Troussellier, M., Lautier, J., Chrétiennot-Dinet, M.J., Neveux, J., Machado, C. et al. (1994): Smallest eukaryotic organism. – *Nature* 370: 255–255.
- Darienko, T., Gustavs, L., Mudimu, O., Menendez, C.R., Schumann, R., Karsten, U., Friedl, T. et al. (2010): *Chloroidium*, a common terrestrial coccoid green alga previously assigned to *Chlorella* (Trebouxiophyceae, Chlorophyta). – *Eur. J. Phycol.* 45: 79–95.
- de Reviers, B. (2003): *Biologie et phylogénie des algues Tome 2*. Belin, 255 pp.
- de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahe, F., Logares, R., Lara, E. et al. (2015): Eukaryotic plankton diversity in the sunlit ocean. – *Science* 348: 1261605.
- Deschamps, P. & Moreira, D. (2012): Reevaluating the green contribution to diatom genomes. – *Genome Biol. Evol.* 4: 683–688.
- Edgcomb, V., Orsi, W., Bunge, J., Jeon, S., Christen, R., Leslin, C., Holder, M. et al. (2011): Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. – *ISME J.* 5: 1344–1356.
- Eikrem, W. & Throndsen, J. (1990): The ultrastructure of *Bathycoccus* gen. nov. and *B. prasinus* sp. nov., a non-motile picoplanktonic alga (Chlorophyta, Prasinophyceae) from the Mediterranean and Atlantic. – *Phycologia* 29: 344–350.
- Falkowski, P.G., Schofield, O., Katz, M.E., Van de Schootbrugge, B. & Knoll, A.H. (2004): Why is the land green and the ocean red? – In: Thierstein, H.R. & Young, J.R. (eds.), *Coccolithophores: From Molecular Processes to Global Impact*. Springer. Berlin, pp. 427–453.
- Faria, D.G., Kato, A., de la Peña, M.R. & Suda, S. (2011): Taxonomy and phylogeny of *Nephroselmis clavistella* sp. nov. (Nephroselmidophyceae, Chlorophyta). – *J. Phycol.* 47: 1388–1396.
- Faria, D.G., Kato, A. & Suda, S. (2012): *Nephroselmis excentrica* sp. nov. (Nephroselmidophyceae, Chlorophyta) from Okinawajima, Japan. – *Phycologia* 51: 271–282.
- Fawley, M.W., Qin, M. & Yun, Y. (1999): The relationship between *Pseudoscourfieldia marina* and *Pycnococcus provasolii* (Prasinophyceae, Chlorophyta): evidence from 18S rDNA sequence data. – *J. Phycol.* 843: 838–843.
- Foulon, E., Not, F., Jalabert, F., Cariou, T., Massana, R. & Simon, N. (2008): Ecological niche partitioning in the picoplanktonic green alga *Micromonas pusilla*: Evidence from environmental surveys using phylogenetic probes. – *Environ. Microbiol.* 10: 2433–2443.
- Fučíková, K., Leliaert, F., Cooper, E.D., Škaloud, P., D'Hondt, S., Clerck De, O., Gurgel, C.F.D. et al. (2014): New phylogenetic hypotheses for the core Chlorophyta based on chloroplast sequence data. – *Front. Ecol. Evol.* 2: 63.
- Gaarder, T. (1933): Untersuchungen über Produktions und Lebensbedingungen in norwegischen Austern-Pollen. – In: *Naturvidenskapelig Rekke 3*. Bergens Museum, pp. 1–64.
- Gómez, P.I. & González, M.A. (2004): Genetic variation among seven strains of *Dunaliella salina* (Chlorophyta) with industrial potential, based on RAPD banding patterns and on nuclear ITS rDNA sequences. – *Aquaculture* 233: 149–162.
- Groisillier, A., Massana, R., Valentin, K., Vault, D. & Guillou, L. (2006): Genetic diversity and habitats of two enigmatic marine alveolate lineages. – *Aquat. Microb. Ecol.* 42: 277–291.
- Guillard, R.R.L., Keller, M.D., O'Kelly, C.J. & Floyd, G.L. (1991): *Pycnococcus provasolii* gen. et sp. nov., a coccoid prasinonanthin-containing phytoplankton from the western north Atlantic and Gulf of Mexico. – *J. Phycol.* 27: 39–47.
- Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C. et al. (2013): The Protist Ribosomal Reference database (PR2): A catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. – *Nucleic Acids Res.* 41: 597–604.
- Guillou, L., Eikrem, W., Chrétiennot-Dinet, M.-J., Le Gall, F., Massana, R., Romari, K., Pedrós-Alió, C. et al. (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.
- Guillou, L., Viprey, M., Chambouvet, A., Welsh, R.M., Kirkham, A.R., Massana, R., Scanlan, D.J. et al. (2008): Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). – *Environ. Microbiol.* 10: 3349–3365.
- Harðardóttir, S., Lundholm, N., Moestrup, Ø. & Nielsen, T.G. (2014): Description of *Pyramimonas diskoicola* sp. nov. and the importance of the flagellate *Pyramimonas* (Prasinophyceae) in Greenland sea ice during the winter – spring transition. – *Polar Biol.* 37: 1479–1494.

- Harholt, J., Moestrup, Ø. & Ulvskov, P. (2015): Why plants were terrestrial from the beginning. – *Trends Plant Sci.* 21: 1–6.
- Hasegawa, T., Miyashita, H., Kawachi, M., Ikemoto, H., Kurano, N., Miyachi, S. & Chihara, M. (1996): *Prasinoderma coloniale* gen. et sp. nov., a new pelagic coccoid prasinophyte from the western Pacific Ocean. – *Phycologia* 35: 170–176.
- Hedges, S.B., Blair, J.E., Venturi, M.L. & Shoe, J.L. (2004): A molecular timescale of eukaryote evolution and the rise of complex multicellular life. – *BMC Evol. Biol.* 4: 2.
- Heimann, K., Benting, J., Timmermann, S. & Melkonian, M. (1989): The flagellar developmental cycle in algae. Two types of flagellar development in uniflagellated algae. – *Protoplasma* 153: 14–23.
- Henley, W.J., Hironaka, J.L., Guillou, L., Buchheim, M.A., Buchheim, J.A., Fawley, M.W. & Fawley, K.P. (2004): Phylogenetic analysis of the “*Nannochloris*-like” algae and diagnoses of *Picochlorum oklahomensis* gen. et sp. nov. (Trebouxiophyceae, Chlorophyta). – *Phycologia* 43: 641–652.
- Hori, T., Inouye, I., Horiguchi, T. & Boalch, G.T. (1985): Observations on the motile stage of *Halosphaera minor* Ostenfeld (Prasinophyceae) with special reference to the cell structure. – *Bot. Mar.* 28: 529–538.
- Hori, T., Moestrup, Ø. & Hoffman, L.R. (1995): Fine structural studies on an ultraplanktonic species of *Pyramimonas*, *P. virginica* (Prasinophyceae), with a discussion of subgenera within the genus *Pyramimonas*. – *Eur. J. Phycol.* 30: 219–234.
- Hori, T., Norris, R.E. & Mitsuo, A.N. (1982): Studies on the ultrastructure and taxonomy of the genus *Tetraselmis* (Prasinophyceae) – I. Subgenus *Tetraselmis*. – *Bot. Mag. Tokyo* 95: 49–61.
- Hori, T., Norris, R.E., Mitsuo, A.N. & Chihara, M. (1983): Studies on the ultrastructure and taxonomy of the genus *Tetraselmis* (Prasinophyceae) – II. Subgenus *Prasinocladia*. – *Bot. Mag. Tokyo* 96: 385–392.
- Hori, T., Norris, R.E. & Chihara, M. (1986): Studies on the ultrastructure and taxonomy of the genus *Tetraselmis* – III. Subgenus *Parviselmis* 99: 123–135.
- Hoshaw, R.W. & Ettl, H. (1966): *Chlamydomonas smithii* sp. nov. a Chlamydomonad interfertile with *Chlamydomonas reinhardtii*. – *J. Phycol.* 2: 93–96.
- Jouenne, F., Eikrem, W., Le Gall, F., Marie, D., Johnsen, G. & Vaultot, D. (2011): *Prasinoderma singularis* sp. nov. (Prasinophyceae, Chlorophyta), a solitary coccoid prasinophyte from the South-East Pacific Ocean. – *Protist* 162: 70–84.
- Katoh, K. (2002): MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. – *Nucleic Acids Res.* 30: 3059–3066.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S. et al. (2012): Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. – *Bioinformatics* 28: 1647–1649.
- Klein, R.M. & Cronquist, A. (1967): A consideration of the evolutionary and taxonomic significance of some biochemical, micro-morphological, and physiological characters in the Thallophytes. – *Q. Rev. Biol.* 42: 108–296.
- Kopp, R.E., Kirschvink, J.L., Hilburn, I.A. & Nash, C.Z. (2005): The Paleoproterozoic snowball Earth: a climate disaster triggered by the evolution of oxygenic photosynthesis. – *Proc. Natl. Acad. Sci. U. S. A.* 102: 11131–11136.
- Krienitz, L., Bock, C., Kotut, K. & Luo, W. (2012): *Picocystis salinarum* (Chlorophyta) in saline lakes and hot springs of East Africa. – *Phycologia* 51: 22–32.
- Le Gall, F., Rigaut-Jalbert, F., Marie, D., Garczarek, L., Viprey, M., Gobet, A. & Vaultot, D. (2008): Picoplankton diversity in the South-East Pacific Ocean from cultures. – *Biogeosciences* 5: 203–14.
- Lee, H.-J. & Hur, S.-B. (2009): Genetic relationships among multiple strains of the genus *Tetraselmis* based on partial 18S rDNA sequences. – *Algae* 24: 205–212.
- Lee, R.E. (2008): *Phycology*. Cambridge University Press 547 pp.
- Leliaert, F., Smith, D.R., Moreau, H., Herron, M.D., Verbruggen, H., Delwiche, C.F. & De Clerck, O. (2012): Phylogeny and molecular evolution of the green algae. – *CRC. Crit. Rev. Plant Sci.* 31: 1–46.
- Letsch, M.R., Muller-Parker, G., Friedl, T. & Lewis, L.A. (2009): *Elliptochloris marina* sp. nov. (Trebouxiophyceae, Chlorophyta), symbiotic green alga of the temperate Pacific sea anemones *Anthopleura xanthogrammatica* and *A. elegantissima* (Anthozoa, Cnidaria). – *J. Phycol.* 45: 1127–1135.
- Lewin, R.A., Krienitz, L., Goericke, R., Takeda, H. & Hepperle, D. (2000): *Picocystis salinarum* gen. et sp. nov. (Chlorophyta) – a new picoplanktonic green alga. – *Phycologia* 39: 560–565.
- Lie, A.A.Y., Liu Z., Hu, S.K., Jones, A.C., Kim, D.Y., Countway, P.D. et al. (2014): Investigating microbial eukaryotic diversity from a global census: insights from a comparison of pyrotag and full-length sequences of 18S rRNA gene sequences. – *Appl. Environ. Microbiol.* 80: 4363–4373.
- Lopes dos Santos, A., Tragin, M., Gourvil, P., Noël, M.-H., Decelle, J., Romac, S. & Vaultot, D. (2016): Prasinophytes clade VII, an important group of green algae in oceanic waters: diversity and distribution. – *ISME J.* submitted.
- Lovejoy, C., Vincent, W.F., Bonilla, S., Roy, S., Martineau, M.J., Terrado, R., Potvin, M. et al. (2007): Distribution, phylogeny, and growth of cold-adapted picoprasinophytes in arctic seas. – *J. Phycol.* 43: 78–89.
- Manton, I. (1977): *Dolichomastix* (Prasinophyceae) from arctic Canada, Alaska and South Africa: a new genus of flagellates with scaly flagella. – *Phycologia* 16: 427–438.
- Margulis, L. (1975): Symbiotic theory of the origin of eukaryotic organelles; criteria for proof. – *Symp. Soc. Exp. Biol.* 29: 21–38.
- Marin, B. (2012): Nested in the Chlorellales or independent class? Phylogeny and classification of the Pedinophyceae (Viridiplantae) revealed by molecular phylogenetic analyses of complete nuclear and plastid-encoded rRNA operons. – *Protist* 163: 778–805.
- Marin, B. & Melkonian, M. (1994): Flagellar hairs in Prasinophytes (Chlorophyta): ultrastructure and distribution on the flagellar surface. – *J. Phycol.* 30: 659–678.
- Marin, B. & Melkonian, M. (2010): Molecular phylogeny and classification of the Mamiellophyceae class. nov. (Chlorophyta) based on sequence comparisons of the nuclear- and plastid-encoded rRNA operons. – *Protist* 161: 304–336.
- Massana, R., Gobet, A., Audic, S., Bass, D., Bittner, L., Boutte, C., Chambouvet, A. et al. (2015): Marine protist diversity in European coastal waters and sediments as revealed by high-throughput sequencing. – *Environ. Microbiol.* 17: 4035–4049.
- Mattox, K.R. & Stewart, K.D. (1984): Classification of the green algae: a concept based on comparative cytology. – In: Irvine, D.E.G. & John, D.M. (eds.), *Systematics of the Green Algae*. Academic Press. London and Orlando, pp. 29–72.
- McFadden, G.I. (2001): Primary and secondary endosymbiosis and the origin of plastids. – *J. Phycol.* 37: 951–959.
- Mishra, A., Mandoli, A. & Jha, B. (2008): Physiological characterization and stress-induced metabolic responses of *Dunaliella*

- salina* isolated from salt pan. – J. Ind. Microbiol. Biotechnol. 35: 1093–1101.
- Miyashita, H., Ikemoto, H., Kurano, N., Miyachi, S. & Chihara, M. (1993): *Prasinococcus capsulatus* gen. et sp. nov., a new marine coccoid prasinophyte. – J. Gen. App. Microbio. 39: 571–582.
- Moestrup, Ø. (1982): Flagellar structure in algae: a review, with new observations particularly on the Chrysophyceae, Phaeophyceae (Fucophyceae), Euglenophyceae, and *Reckertia*. – Phycologia 21: 427–528.
- Moestrup, Ø. (1984): Further studies on *Nephroselmis* and its allies (Prasinophyceae). II. *Mamiella* gen. nov., Mamiellaceae fam. nov., Mamiellales ord. nov. – Nord. J. Bot. 4: 109–121.
- Moestrup, Ø. (1991): Further Studies of presumed primitive green algal, including the description of Pedinophyceae Class. Nov. and *Resultor* Gen. Nov. – J. Phycol. 27: 119–133.
- Moestrup, Ø. & Thronsdén, J. (1988): Light and electron microscopical studies on *Pseudoscourfieldia marina*, a primitive scaly green flagellate (Prasinophyceae) with posterior flagella. – Can. J. Bot. 66: 1415–1434.
- Monier, A., Sudek, S., Fast, N.M. & Worden, A.Z. (2013): Gene invasion in distant eukaryotic lineages: discovery of mutually exclusive genetic elements reveals marine biodiversity. – ISME J. 7: 1764–1774.
- Moon-van der Staay, S.Y., De Wachter, R. & Vaalot, D. (2001): Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. – Nature 409: 607–610.
- Moro, I., La Rocca, N., Dalla Valle, L., Moschin, E., Negrisolo, E. & Andreoli, C. (2002): *Pyramimonas australis* sp. nov. (Prasinophyceae, Chlorophyta) from Antarctica: fine structure and molecular phylogeny. – Eur. J. Phycol. 37: 103–114.
- Moustafa, A., Beszteri, B., Maier, U.G., Bowler, C., Valentin, K. & Bhattacharya, D. (2009): Genomic footprints of a cryptic plastid endosymbiosis in diatoms. – Science 324: 1724–1726.
- Nägeli, C. (1849): Gattungen einzelliger Algen physiologisch und systematisch bearbeitet. Zürich. 139 pp.
- Nakayama, T., Marin, B., Kranz, H.D., Surek, B., Huss, V.A., Inouye, I. & Melkonian, M. (1998): The basal position of scaly green flagellates among the green algae (Chlorophyta) is revealed by analyses of nuclear-encoded SSU rRNA sequences. – Protist 149: 367–380.
- Nakayama, T., Suda, S., Kawachi, M. & Inouye, I. (2007): Phylogeny and ultrastructure of *Nephroselmis* and *Pseudoscourfieldia* (Chlorophyta), including the description of *Nephroselmis anterostigmatica* sp. nov. and a proposal for the Nephroselmiales ord. nov. – Phycologia 46: 680–697.
- Nelson, W.A. & Ryan, K.G. (1986): *Palmophyllum umbracola* sp. nov. (Chlorophyta) from offshore islands of northern New Zealand. – Phycologia 25: 168–177.
- O’Kelly, C.J. & Floyd, G.L. (1983): Flagellar apparatus absolute orientations and the phylogeny of the green algae. – Biosystems 16: 227–251.
- Orsi, W., Song, Y.C., Hallam, S. & Edgcomb, V. (2012): Effect of oxygen minimum zone formation on communities of marine protists. – ISME J. 6: 1586–1601.
- Pennick, N.C. (1982): Studies of the External Morphology of *Pyramimonas* 6. *Pyramimonas cirolanae* sp. nov. – Arch. für Protistenkd. 125: 87–94.
- Pennick, N.C. (1984): Comparative ultrastructure and occurrence of scales in *Pyramimonas* (Chlorophyta, Prasinophyceae). – Arch. für Protistenkd. 128: 3–11.
- Peterfi, L.S. & Manton, I. (1968): Observations with the electron microscope on *Asteromonas gracilis* Artari emend. (*Stephanoptera gracilis* (Artari) wisl.), with some comparative observations on *Dunaliella* sp. – Br. Phycol. Bull. 3: 423–440.
- Potter, D., Lajeunesse, T.C., Saunders, G.W. & Andersen, R.A. (1997): Convergent evolution masks extensive biodiversity among marine coccoid picoplankton. – Biodivers. Conserv. 6: 99–107.
- Preetha, K., John, L., Subin, C.S. & Vijayan, K.K. (2012): Phenotypic and genetic characterization of *Dunaliella* (Chlorophyta) from Indian salinas and their diversity. – Aquat. Biosyst. 8: 27.
- Price, M.N., Dehal, P.S. & Arkin, A.P. (2009): Fasttree: Computing large minimum evolution trees with profiles instead of a distance matrix. – Mol. Biol. Evol. 26: 1641–1650.
- Rodríguez, F., Derelle, E., Guillou, L., Le Gall, F., Vaalot, D. & Moreau, H. (2005): Ecotype diversity in the marine picoeukaryote *Ostreococcus* (Chlorophyta, Prasinophyceae). – Environ. Microbiol. 7: 853–859.
- Round, F.E. (1963): The taxonomy of the Chlorophyta. – Br. Phycol. Bull. 2: 224–235.
- Round, F.E. (1971): The taxonomy of the Chlorophyta. II. – Br. Phycol. J. 6: 235–264.
- Scherffel, A. (1912): Zwei neue trichocystenartige Bildungen führende Flagellaten. – Arch. für Protistenkd. 27: 94–128.
- Schloss, P.D., Westcott, S.L., Ryabin, T., Hall, J.R., Hartmann, M., Hollister, E.B., Lesniewski, R.A. et al. (2009): Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. – Appl. Environ. Microbiol. 75: 7537–7541.
- Scott, C., Lyons, T.W., Bekker, A., Shen, Y., Poulton, S.W., Chu, X. & Anbar, A.D. (2008): Tracing the stepwise oxygenation of the Proterozoic ocean. – Nature 452: 456–459.
- Shi, X.L., Marie, D., Jardillier, L., Scanlan, D.J. & Vaalot, D. (2009): Groups without cultured representatives dominate eukaryotic picophytoplankton in the oligotrophic South East Pacific Ocean. – PLoS One 4: e7657.
- Sieburth, J.M., Keller, M.D., Johnson, P.W. & Myklestad, S.M. (1999): Widespread occurrence of the oceanic ultraplankter, *Prasinococcus capsulatus* (Prasinophyceae), the diagnostic “Golgi-decapore complex” and the newly described polysaccharide “Capsulan.” – J. Phycol. 35: 1032–1043.
- Šlapeta, J., López-García, P. & Moreira, D. (2006): Global dispersal and ancient cryptic species in the smallest marine eukaryotes. – Mol. Biol. Evol. 23: 23–29.
- Sluiman, H.J., Kouwets, F.A.C. & Blommers, P.C.J. (1989): Classification and definition of cytokinetic patterns in green algae: Sporulation versus (vegetative) cell division. – Arch. für Protistenkd. 137: 277–290.
- Stewart, K.D. & Mattox, K.R. (1975): Comparative cytology, evolution and classification of the green algae with some consideration of the origin of other organisms with chlorophylls A and B. – Bot. Rev. 41: 104–135.
- Stewart, K.D. & Mattox, K.R. (1978): Structural evolution in the flagellated cells of green algae and land plants. – Biosystems 10: 145–152.
- Subirana, L., Péquin, B., Michely, S., Escande, M.L., Meilland, J., Derelle, E., Marin, B. et al. (2013): Morphology, genome plasticity, and phylogeny in the genus *Ostreococcus* reveal a cryptic species, *O. mediterraneus* sp. nov. (Mamiellales, Mamiellophyceae). – Protist 164: 643–659.
- Suda, S., Bhuiyan, M.A.H. & Faria, D.G. (2013): Genetic diversity of *Pyramimonas* from Ryukyu Archipelago, Japan

- (Chlorophyceae, Pyramimonadales). – *J. Mar. Sci. Technol.* 21: 285–296.
- Takishita, K., Yubuki, N., Kakizoe, N., Inagaki, Y. & Maruyama, T. (2007): Diversity of microbial eukaryotes in sediment at a deep-sea methane cold seep: Surveys of ribosomal DNA libraries from raw sediment samples and two enrichment cultures. – *Extremophiles* 11: 563–576.
- Tappan, H. & Loeblich, A.R. (1973): Evolution of the oceanic plankton. – *Earth-Science Rev.* 9: 207–240.
- Thronsdon, J. & Zingone, A. (1997): *Dolichomastix tenuilepis* sp. nov., a first insight into the microanatomy of the genus *Dolichomastix* (Mamiellales, Prasinophyceae, Chlorophyta). – *Phycologia* 36: 244–254.
- Vaultot, D., Lepère, C., Toulza, E., de la Iglesia, R., Poulain, J., Gaboyer, F., Moreau, H. et al. (2012): Metagenomes of the picoalga *Bathycoccus* from the Chile coastal upwelling. – *PLoS One* 7: e39648.
- Viprey, M., Guillou, L., Ferréol, M. & Vaultot, D. (2008): Wide genetic diversity of picoplanktonic green algae (Chloroplastida) in the Mediterranean Sea uncovered by a phylum-biased PCR approach. – *Environ. Microbiol.* 10: 1804–1822.
- Wiebe, P.H., Remsen, C.C. & Vaccaro, R.F. (1974): *Halosphaera viridis* in the Mediterranean sea: size range, vertical distribution, and potential energy source for deep-sea benthos. – *Deep Sea Res. Oceanogr. Abstr.* 21: 657–667.
- Wu, W., Huang, B., Liao, Y. & Sun, P. (2014): Picoeukaryotic diversity and distribution in the subtropical-tropical South China Sea. – *FEMS Microbiol. Ecol.* 89: 563–579.
- Yamaguchi, H., Nakayama, T. & Inouye, I. (2013): Proposal of *Microsquama* subgen. nov. for *Nephroselmis pyriformis* (Carter) Ettl. – *Phycol. Res.* 61: 268–269.
- Yamaguchi, H., Suda, S., Nakayama, T., Pienaar, R.N., Chihara, M. & Inouye, I. (2011): Taxonomy of *Nephroselmis viridis* sp. nov. (Nephroselmidophyceae, Chlorophyta), a sister marine species to freshwater *N. olivacea*. – *J. Plant Res.* 124: 49–62.
- Yoon, H.S., Hackett, J.D., Ciniglia, C., Pinto, G. & Bhattacharya, D. (2004): A molecular timeline for the origin of photosynthetic eukaryotes. – *Mol. Biol. Evol.* 21: 809–818.
- Zechman, F.W., Verbruggen, H., Leliaert, F., Ashworth, M., Buchheim, M.A., Fawley, M.W., Spalding, H. et al. (2010): An unrecognized ancient lineage of green plants persists in deep marine waters. – *J. Phycol.* 46: 1288–1295.
- Zingone, A., Borra, M., Brunet, C., Forlani, G., Kooistra, W.H.C.F. & Procaccini, G. (2002): Phylogenetic position of *Crustomastix stigmatica* sp. nov. and *Dolichomastix tenuilepis* in relation to Mamiellales (Prasinophyceae, Chlorophyta). – *J. Phycol.* 38: 1024–1039.

Manuscript received: 18 December 2015

Accepted: 25 April 2016

Handling editor: Ramon Massana

### The pdf version of this paper includes an electronic supplement

Please save the electronic supplement contained in this pdf-file by clicking the blue frame above. After saving rename the file extension to .zip (for security reasons Adobe does not allow to embed .exe, .zip, .rar etc. files).

**Fig. S1.** Pie chart distribution by oceanic areas of sequences of major Chlorophyta lineages for cultures (A) and environmental samples (B). Sequences were regrouped in rectangular regions using the R software (Table S5). Areas may overlap in regions where no sequences were recorded.

**Fig. S2.** Range of variation of the percentage of sequence identity for the full 18S rRNA, the V4 and V9 region sequences for each Chlorophyta lineage (maximum, third quartile, median, first quartile, minimum). Identity matrixes were calculated using the function *seqidentity* of the R package bio3d. Number of sequences is indicated between parentheses.

**Table S1.** List of reference sequences used for building the tree of Fig. 1.

**Table S2.** Number of species per described genus for each prasinophyte clade according to AlgaeBase (<http://www.algaebase.org/>). Number of strain sequences considered for each genus.

**Table S3.** Members of the UTC clades that are found in the marine environment with the number of sequences considered.

**Table S4.** Sequences from the Roscoff Culture Collection recently deposited to GenBank and used in this work.

**Table S5.** Oceanic regions used to regroup sequences presented in Figure S2.

### Supplementary data (available from Figshare):

**Data S1.** Annotated Chlorophyta GenBank sequences. Two files containing sequences in fasta format and their annotated taxonomy. These files can be used to assign metabarcoding data using software such as mothur or Qiime: <https://figshare.com/s/c5edff05cd551e466320>

**Data S2.** Metadata for Chlorophyta sequences: <https://figshare.com/s/a72fabf6add26f4c98ce>

**Data S3.** Reference alignment for Fig. 1: <https://figshare.com/s/3ebc93f306f3935cab80>