environmental microbiology

Environmental Microbiology (2019) **21**(10), 3885–3895 doi:10.1111/1462-2920.14738

Global distribution of Trebouxiophyceae diversity explored by high-throughput sequencing and phylogenetic approaches

Sebastian Metz,[1] **David Singer**,[2,3] **Isabelle Domaizon**,[4] **Fernando Unrein**[1] **and Enrique Lara**[5*]

[1]*Instituto Tecnológico de Chascomús (INTECH), UNSAM-CONICET, Chascomús, Buenos Aires, Argentina.*

[2]*Laboratory of Soil Biodiversity, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland.*

[3]*Department of Zoology, Institute of Biosciences, University of São Paulo, Butantã, São Paulo, Brazil.*

[4]*CARRTEL, INRA, Université Savoie Mont Blanc, Thonon, France.*

[5]*Real Jardín Botánico de Madrid, CSIC, Madrid, Spain.*

Summary

Trebouxiophyceae are a ubiquitous class of Chlorophyta encountered in aquatic and terrestrial environments. Most taxa are photosynthetic, and many acts as photobionts in symbiotic relationships, while others are free-living. Trebouxiophyceae have also been widely investigated for their use for biotechnological applications. In this work, we aimed at obtaining a comprehensive image of their diversity by compiling the information of 435 freshwater, soil and marine environmental DNA samples surveyed with Illumina sequencing technology in order to search for the most relevant environments for bioprospecting. Freshwater and soil were most diverse and shared more than half of all operational taxonomic units (OTUs), however, their communities were significantly distinct. Oceans hosted the highest genetic novelty, and did not share any OTUs with the other environments; also, marine samples host more diversity in warm waters. Symbiotic genera usually found in lichens such as *Trebouxia*, *Myrmecia* and *Symbiochloris* were also abundantly detected in the ocean, suggesting either free-living lifestyles or unknown symbiotic relationships with marine planktonic organisms. Altogether, our study opens the way to new prospection for trebouxiophycean strains, especially in understudied environments like the ocean.

Received 12 February, 2019; revised 22 June, 2019; accepted 10 July, 2019. *For correspondence. E-mail enrique.lara@rjb.csic.es; Tel. 34 914 20 30 17; Fax 34 914 20 01 57.

Introduction

Trebouxiophyceae are a ubiquitous class of photosynthetic unicellular Chlorophyta (='green algae') encountered in many environments, including extreme ones such as hot (Büdel *et al*., 2009; Flechtner *et al*., 2013; Fučíková *et al*., 2014) and cold (Cavacini, 2001; Fermani *et al*., 2007; Hodač *et al*., 2016) deserts, or hyper acidic waters (Juárez *et al*., 2011). They are also common in mesophilic environments like soil and continental waters, some of them can also be found in the sea (Heesch *et al*., 2012; Tragin and Vaulot, 2018). But Trebouxiophyceae are mostly known for their symbiotic relationships with other organisms (Pröschold *et al*., 2011), where they act as photobionts, providing the by-products of their photosynthesis to the host. Most lichenizing fungi are indeed associated with Trebouxiophyceae (Thüs *et al*., 2011). Other mutualistic symbioses are also known, including diverse amoeboid protists (Gomaa *et al*., 2014; Lara and Gomaa, 2017), Metazoa (Aboal and Werner, 2011) and even vascular plants (Trémouillaux-Guiller and Huss, 2007). Some species lost secondarily their photosynthetic ability to evolve as heterotrophic free-living organisms or parasites of diverse organisms (de Koning and Keeling, 2006; Pombert and Keeling, 2010), including humans (Lass-Flörl and Mayr, 2007). The variety of symbiotic associations including Trebouxiophyceae has motivated research for over 100 years (Pröschold *et al*., 2011).

Trebouxiophyceae have also attracted the interest of applied research. Indeed, green algae strains are generally simple to cultivate and allow the retrieval of a high biomass, which designated them for sustainable biogas production (Huss *et al*., 1999; Safi *et al*., 2014). Other strains also produce secondary metabolites of medical interest, for instance anticancer drugs (Safi *et al*., 2014). These individual examples reflect the economic and biotechnological importance of Trebouxiophyceae and

justifies the research effort in the prospection for new strains with potentially promising properties. In 2012, the online reference site for algae (http://www.algaebase.org) documented 546 described Trebouxiophyceae species (Guiry and Guiry, 2018). This number has been raised 3 years later (October 2018) to 857 (Guiry and Guiry, 2018), certainly fostered by the generalization of genetic barcoding to discriminate between morphologically similar strains.

This prospection is, however, not straightforward. Indeed, identifying strains based on their morphology is often impossible, and while certain Trebouxiophyceae present conspicuous morphologies (Lewis and McCourt, 2004), the bulk of their diversity cannot be readily distinguished from each other (i.e. 'green balls', most of all formerly classified into a highly paraphyletic genus '*Chlorella*') despite being genetically and physiologically very heterogeneous. The lack of flagellar apparatus reduces even more the number of possible morphological discrimination criteria to apply, which resulted in an underestimation of their diversity (Krienitz *et al*., 2015). The advent of molecular methods permitted establishing a taxonomic framework, resulting in splitting the vast genus *Chlorella* into many monophyletic genera, providing a framework for the classification of strains of evolutionary, ecological or biotechnological interest (e.g. Bock *et al*., 2011; Krienitz *et al*., 2015; Neofotis *et al*., 2016).

Environmental DNA approaches based on high throughput sequencing of amplicons of a chosen marker gene are nowadays the silver bullet technique for prospecting microbial diversity. This type of approach can be applied to detect peaks in diversity and evaluate the genetic divergence of the environmental sequences as compared to reference databases ('genetic novelty'; Mahé *et al*., 2017). It provides therefore the needed information on where to look for undescribed organisms. In the case of the Trebouxiophyceae, given the size of the reference database and the fact that many deposited strains have been barcoded, it can be safely assumed that most novel sequences correspond to potentially undescribed organisms.

Here, we applied an environmental DNA approach based on the V9 variable region of the gene coding for the SSU rRNA to a wide sampling design including samples taken from freshwater (FW) and soil (SO) samples, plus all sequences obtained in the course of the TARA Oceans expedition (de Vargas *et al*., 2015), for a total of 435 samples spanning a vast range of climates, and environmental parameters. Trebouxiophyceae-related operational taxonomic units (OTUs) were placed in a reference phylogenetic tree based on both cultured organisms and environmental sequences and assigned to the existing genera using evolutionary placement algorithm (EPA) (Zhang *et al*., 2013). Our aim was to obtain an overview of the distribution of trebouxiophycean diversity, infer which environments would host the largest unknown diversity, and also the most genetically diverse clades. Also, we aimed at detecting the presence of symbiotic clades in environments where they have not been reported, suggesting unreported biotic associations. Because of the importance of the salinity barrier (Logares *et al*., 2009), and also due to the stress associated with desiccation, we hypothesize that freshwater, soil and marine communities will differ in their composition. We also expect finding more genetic novelty in marine environments, as knowledge on oceanic Trebouxiophyceae is nowadays still very limited. In line with better-known soil and freshwater systems, we expect that marine environments should also host symbiotic associations; this would be suggested by the frequent presence of known symbiotic genera such as *Symbiochloris* for instance.

Results

Phylogenetic tree construction and annotation

In total, we extracted 2018 sequences from the public databases SILVA and NCBI that belonged unambiguously to Trebouxiophyceae (Table S1). Amongst them, we used 1814 high quality (no errors or ambiguous nucleotides), long (i.e. >1000 nt) and non-redundant sequences to build the reference phylogenetic tree. Most of the main Trebouxiophyceae clades were supported with high bootstrap values (>70%) and annotated at the genus level. Those that have bootstrap values lower than 70% were annotated at the order level according to the bibliography (Leliaert *et al*., 2012; Lemieux *et al*., 2014). Three formerly uncharacterized deep clades were retrieved and annotated as TREB01-TREB03 (Fig. 1). These clades were composed exclusively by environmental sequences that had been already published previously.

The reference phylogenetic tree was used to classify 961 OTUs from our environmental DNA survey. We identified 11 new, deep branching, clades. Amongst them, nine could be potentially assigned to new genera as compared with the average genetic distances that separate trebouxiophycean lineages in the phylogenetic tree (Fig. 1), and the two remaining others branched that are at the deepest nodes could be assigned to a suprageneric status. Three lineages were marine and the remaining eight were shared between SO and FW.

General diversity patterns

The final data set included 63 samples from FW, 33 from SO and 107 samples from MA (Fig. 2, Table S2 for FW and SO, Supplementary Material from de Vargas *et al*., 2015 for MA). The distribution of Trebouxiophyceae across samples and environments were not homogeneous. Their contribution to the totality of sequences per sample was, in average, 1% in FW samples, but ranged between 0.01% and 24%;

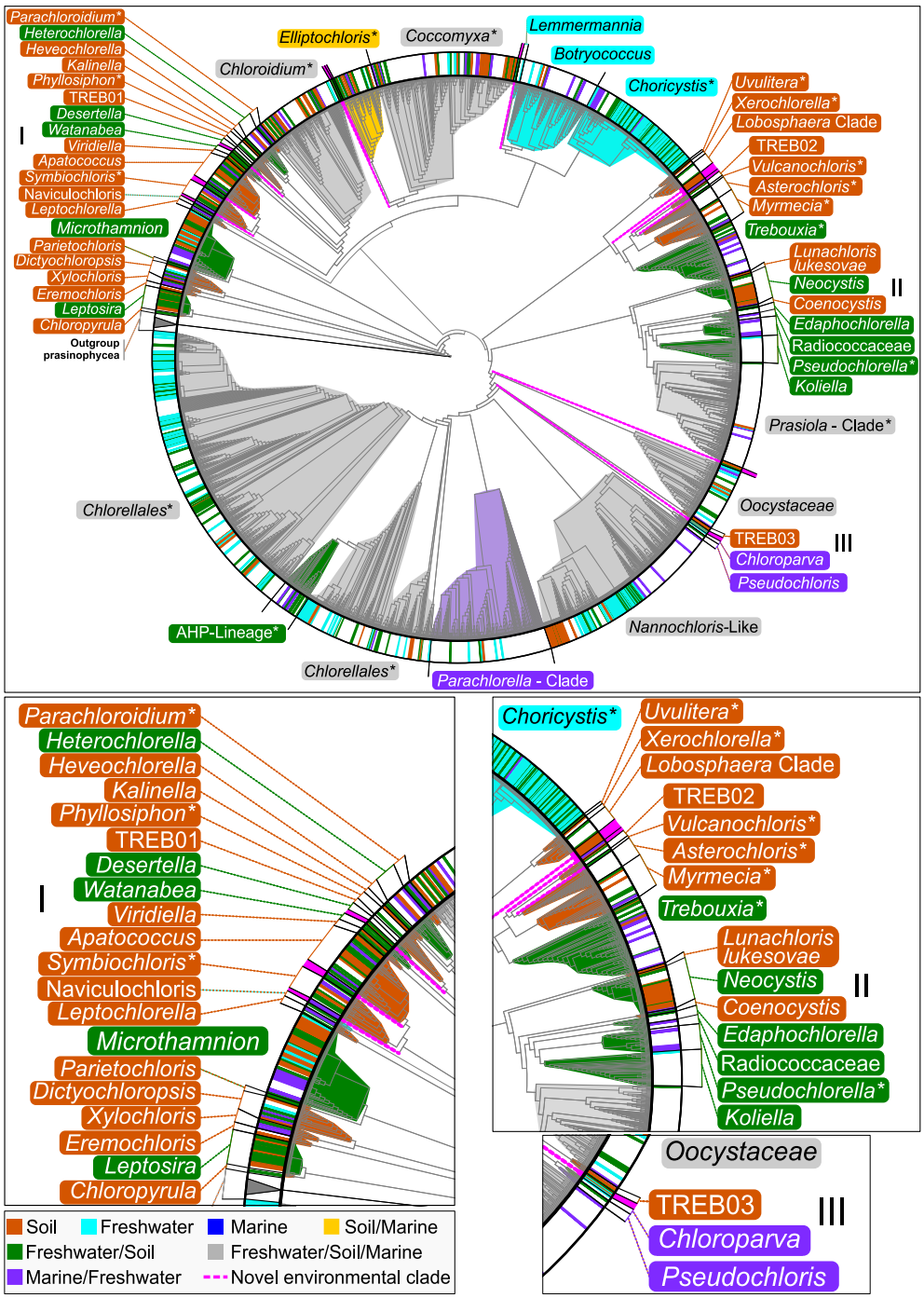

Fig. 1. Reference tree with sequences annotations. Colours of the clade names illustrate the environments types (SO, FW and MA) where the corresponding organisms have been found before this study. Inner circle: reference tree; colours indicate environment types from where the reference sequences (isolates and environmental clones) were found. The asterisks correspond to clades that contain symbiotic organisms. Branches in purple correspond to novel deep branching lineages discovered in this study. Outer circle: lines correspond to OTUs found in this study. Colours indicate the environment type from which they originate. Numbers I, II and III correspond to regions of the tree that have been illustrated in detail, notably illustrating the OTUs from symbiotic groups that have been found in MA (genera *Trebouxia*, *Myrmecia*, *Pseudochlorella* and *Symbiochloris*). [Color figure can be viewed at wileyonlinelibrary.com]

0.44% in SO samples, with a minimum and maximum contribution between 0.006% and 2.95%; and 0.16% in MA samples ranging between 0.01% and 3%. Within these samples, a total of 200,747 sequences were affiliated to Trebouxiophyceae. FW contributed with 80.73% of the total sequences, SO and MA with 15.79% and 3.48%

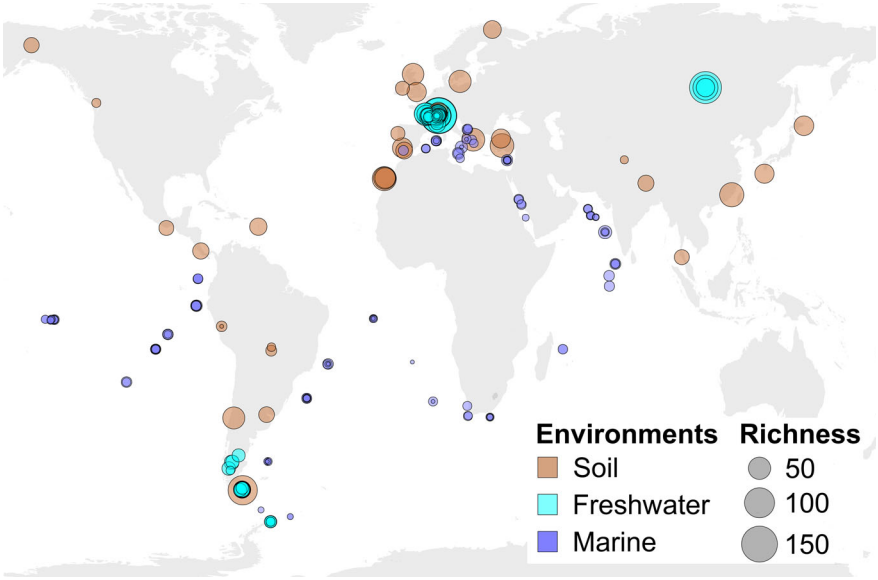

Fig. 2. Map including all sites in the study. Numbers of Trebouxiophyceae OTUs sequences found on each sample are indicated by the diameter of the circles. Colours indicate the environment types from where the sample has been taken. [Color figure can be viewed at wileyonlinelibrary.com]

respectively. All sequences clustered into 961 OTUs (Tables S3 and S4). From these, 298 were found in FW, 227 in SO and 90 in MA, while 346 OTUs were found in both FW and SO. MA did not share any single OTU with any other environments. The non-metric multidimensional scaling (NMDS) analysis showed a clear separation between SO and FW (Fig. 3). Further, irrigated and mostly wet soils were placed in an intermediate position between FW and SO experimenting drought (Fig. 3). Two wet/irrigated soils respectively from a bog in Pto. Williams in extreme Southern Chile (SO_10) and an agricultural soil from the Danish Island Bornholm (SO_90) fell between soils experimenting drought. A careful inspection of their composition showed that an overrepresentation of common OTUs from dry soil, *Leptosira* (OTU_89645) and *Neocystis* (OTU_89623) over-represented in SO_10; and *Chloroidium (OTU_89851)* in SO_90 (Table S4).

FW was dominated by Chlorellales (48.7% of sequences), *Choricystis* (31% of sequences), TREB03HTS (14.3% of sequences) and *Nanochloris*-like sequences (1.8% of sequences) (Table S5). Most of these groups were highly diverse. Chlorellales contributed with 208 OTUs, *Choricystis* with 138 and *Nannochloris*-like with 77 OTUs. The most widespread clade was TREB03HTS, a previously undetected clade present in 58% of all FW samples, and represented by four OTUs in total. The most widespread OTU belonged to symbiotic genus *Coccomyxa* (OTU_89751) and was present in 77.7% of all samples, though its average contribution was low (0.4% of total FW Trebouxiophyceae sequences). The most abundant OTU was the Chlorellales affiliated OTU_14045 that contributed with 42% of the total FW Trebouxiophyceae sequences and was present in 66.7% of all FW samples (Table S5).

SO was dominated by Chlorellales (18% of all SO sequences; 90 OTUs), genus *Apatococcus* (12% of sequences; 27 OTUs) and genus *Chloroidium* (11% of sequences; 33 OTUs). Interestingly, the most widespread OTU in SO was OTU_89751 (genus *Coccomyxa*), the same as in FW, and was present in 87.9% of total SO samples, also in low abundance (6.4% of all SO sequences).

MA was dominated by genera *Chloroidium* (37.4%, 4 OTUs) and *Apatoccocus* (8.3%, 2 OTUs) respectively. The most frequent OTUs belonged to *Heterochlorella*

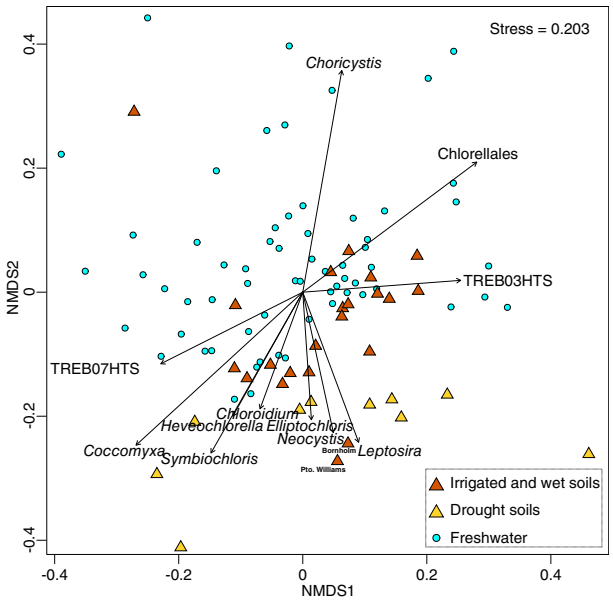

Fig. 3. NMDS indicating the relationships between soil and freshwater communities. [Color figure can be viewed at wileyonlinelibrary.com]

(43.9% of the samples), newly detected environmental clade TREB10HTS (40.2% of the samples), *Coccomyxa* and *Chloroidium* (30.8% and 24.3% of the samples respectively). Marine Trebouxiophyceae peaked in diversity in the Mediterranean and were distributed almost evenly between surface waters (53%) and deep chlorophyll maximum (47%). Most of the OTUs from FW and MA (638 OTUs of 734) derive from the pico sized fraction, in the samples where plankton was sequentially filtered (0.8–5 μm).

Some OTUs were observed exclusively in cold environments in several distant locations, suggesting adaptations to cold temperatures (Fig. 2, Table S4). For instance, OTU_109033 (genus *Coccomyxa*) was well represented in Antarctic samples and was also present in other cold environments such as Argentinean Patagonia and Alaska, both in SO and FW. Other clades containing OTUs possibly exclusive to cold environments were TREB06HTS and TREB07HTS. On the other hand, OTUs affiliated with TREB04HTS (OTU_89644 and OTU_89739) and TREB08HTS (OTU_117567) were found exclusively but consistently in all samples from Lanzarote (Canary Islands), i.e. under desertic climate. Some clades containing many (or mostly) symbiotic forms usually found in lichens were also found in MA samples. These clades correspond to genera *Apatococcus*, *Heterochlorella*, *Myrmecia*, *Symbiochloris*, *Trebouxia*, family Radioccocaceae and the order Prasiolales.

Genetic novelty

About 80% of all OTUs shared more than 95% identity with published sequences in the PR^2 eukaryotic ribosomal database (Guillou *et al*., 2013). However, present knowledge on Trebouxiophyceae genetic diversity was not evenly distributed. Only 20% of MA OTUs had more than 95% of identity with PR^2 reference sequences. In contrast, this percentage rose above 89% in FW and SO (Fig. 4, Table S3).

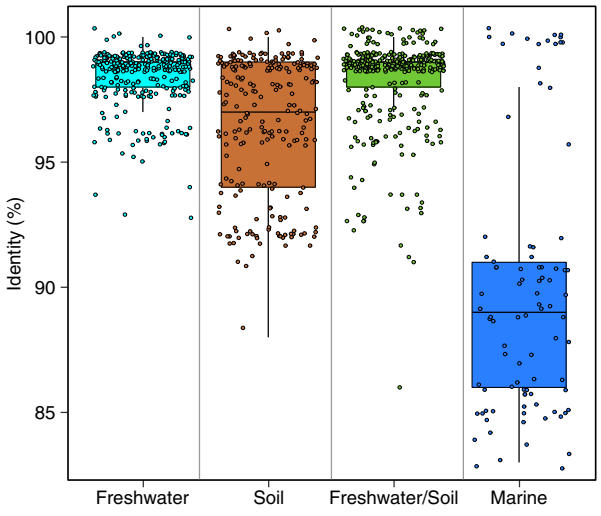

Fig. 4. An illustration of genetics novelty of OTUs in all three environment types, showing the percentage of identity for each OTU as aligned with all PR^2 entries. [Color figure can be viewed at wileyonlinelibrary.com]

Discussion

A reference phylogenetic tree to classify Trebouxiophyceae environmental molecular diversity

Our reference phylogenetic tree is a compilation of all knowledge on Trebouxiophyceae diversity available on public databases; indeed, all genera described to date have at least one representative species barcoded for the 18S SSU rRNA gene (Leliaert *et al*., 2012; Lemieux *et al*., 2014). Leliaert *et al*. (2012) predicted the existence of seven major groups amongst Trebouxiophyceae. Even if the limitation of the SSU rRNA gene for phylogenetic Trebouxiophyceae studies has been previously reported (Neustupa *et al*., 2013; Fučíková *et al*., 2014), our analysis retrieved the seven predicted groups as well-supported clades most often with high bootstrap values (>70%). Moreover, our study reveal three uncharacterized clades from soil, evidencing the existence of an unsuspected deep diversity (i.e. arguably at the genus level or above). Our reference phylogenetic tree can be therefore considered as a useful tool to approach new ecological and evolutionary questions (Pan *et al*., 2017) and classify the unknown sequences (Mahé *et al*., 2017) like it has already been done for other groups of eukaryotic microorganisms (del Campo *et al*., 2018).

Environmental Trebouxiophyceae diversity and promising environments for new strain prospection

Despite our large sampling effort our rarefaction curves reached saturation only in MA (Fig. S1). This indicates that we did not cover all diversity in FW and SO, and that environmental prospection may still reveal more undetected lineages. Still, the number of OTUs estimated for the whole study (961) was higher than the number of species estimated for class Trebouxiophyceae (857 species; Guiry and Guiry, 2018), knowing that the V9 region of the 18S rRNA gene from many published sequences are partial and not are resolutive enough to discriminate some organisms at the species level (Tragin *et al*., 2018).

Altogether, our findings show that Trebouxiophyceae diversity is highest in FW habitats followed by the SO habitat. This contrasts with MA, where only a limited number of OTUs have been detected despite a higher number of samples surveyed. The ocean is, however, of high interest for prospecting new forms, as genetic novelty in marine systems was highest, suggesting a high potential for new strains with novel properties. Indeed, only 20% of all sequences shared 95% or more similarity with sequences deposited in PR^2. In MA samples, subtropical and tropical oceans (Mediterranean, Red

Sea, central Indian Ocean) seemed to host a larger diversity than colder regions (Fig. 2) suggesting a latitudinal gradient of diversity as observed in marine bacteria (Fuhrman *et al*., 2008). Warm oceans are therefore most relevant environments for prospection for new Trebouxiophyceae strains. Filtering seems to be an adequate strategy to enrich environments in Trebouxiophyceae. Indeed, most (40.5%) sequences from MA and (98.7%) from FW, data sets have been retrieved from 'pico' fraction (0.8–5 μm) suggesting that a large part of the Trebouxiophyceae diversity is composed of tiny-sized organisms. This assumption is confirmed by several works based on flow cytometry that found Trebouxiophyceae to dominate picoeukaryotic diversity in certain freshwater assemblages (Pálffy *et al*., 2014; Metz *et al*., 2019).

The marked difference between MA and FW/SO floras confirms our hypothesis, and follows well-known patterns as the salinity barrier represent a major physiological divide for prokaryotic and eukaryotic microorganisms alike (Logares *et al*., 2009). The transition across the salinity barrier has been apparently achieved in the free-living genus *Microthamnion* where our analysis revealed a large, deep branching clade sister to all other species exclusively composed of marine sequences (Fig. 1). Further investigations directed towards the isolation of these organisms may lead to the splitting of *Microthamnion*, given the respective position of MA and FW sequences and reveal organisms with emerging properties. Other supposedly exclusive soil clades have been found in the sea, such as the exclusively free-living *Xylochloris*, *Dichtyochloropsis*, and family Radiococcaceae (Fig. 1). Genus *Picochlorum* (included within *Nannochloris* – Like; Table S1) well-known for its tolerance towards large variations in salinity has been isolated accordingly from marine and brackish environments alike (Foflonker *et al*., 2018). However, in our study, these organisms appeared only in MA samples, possibly; even if they tolerate freshwater, it is still unclear if they can form viable populations under low salinity conditions.

SO and FW are more similar to each other in their taxonomic composition, yet our NMDS analysis (Fig. 3) illustrates a trend for a separation between SO and FW. SO can be differentiated further in two main categories, 'wet' soils (agricultural soils that are irrigated, moist forest soils and peat), that are generally situated between FW and the second main category, 'dry' soils that experiment drought (most natural soils) as can be interpreted in Fig. 3. The differences between these communities shows that drought is also an important driver for Trebouxiophyceae diversity. Moreover, this suggests that intermediate environments like wet soils may have been used as an evolutionary stepping-stone towards transition between FW and dry soil. This evolutionary scenario has been proposed as a possible way for colonizing emerged lands by the ancestors of modern plants (de Vries and Archibald, 2018).

Typical 'dry' SO genera were the free-living *Leptosira* and *Neocystis*, and also the lichen symbionts *Symbiochloris* and *Coccomyxa*, while FW genera were *Choricystis* and Chlorellales (Fig. 3). There are, however, single OTUs that seem to transcend these barriers like OTU_13937 (uncharacterized clade TREB03HTS) and OTU_14045 (Chlorellales), both abundant in FW and SO, including 'dry' soils (Table S4). In contrast to MA samples, 89% of all FW and SO sequences have 95% or more similarity with sequences published in PR^2 (Fig. 4), which illustrates a better knowledge of the trebouxiophycean flora of these environments. Still, the most common and abundant OTUs in FW and SO environments correspond to undescribed organisms (OTU_14045, OTU_13937), which illustrates the fact that both SO and FW host also great potential for discovery.

In total, we found eight deep-branching, genus level clades. Two of these clades, namely TREB05HTS and TREB06HTS, were composed with OTUs that were exclusively found in cold environments such as Antarctica, Tierra del Fuego, Patagonia and, in the Northern Hemisphere, Northern Finland/Lappland (Table S4). In contrast to what has been found in MA, cold environments did not host significantly lower diversity than warmer environments in SO and FW, and the richest FW samples were alpine Lake Brévent (287 OTUs), a high altitude lake (2100 m) covered with ice 8 months per year, and Lake Baikal (198 OTUs); the richest soil sample was originated from subantarctic Puerto Williams (147 OTUs). A systematic prospection of cold environments may reveal new psychrophilic strains with potentially genomic adaptations to cold environments, like the polar Trebouxiophyceae *Coccomyxa subellipsoidea* (Blanc *et al*., 2012).

Desert environments, characterized by low humidity and extreme UV radiation levels may host also organisms of potential interest for biotechnology. Two OTUs were observed only in samples retrieved from these environments, OTU_89739 (affiliated to uncharacterized clade TREB04HTS) and OTU_89758 (*Botryoccocus* sp.); interestingly, this genus includes several strains that are well-known for their biotechnological application in the industry of biofuel production (Metzger *et al*., 1991; Metzger *et al*., 1990; Metzger and Largeau, 1999; Banerjee *et al*., 2002; Raja *et al*., 2008). A wider prospection of these regions would be most wanted to characterize their diversity and estimate their potential for novel diversity discovery.

Novel possible symbiotic associations

Several trebouxiophyceaen genera are well-known to have the tendency of building symbiotic relationships with

lichenizing fungi. These associations are generally found in subaerial environments like rocks, trunks and soil, but not in permanently wet environments (Hawksworth, 2000). Therefore, finding these organisms in aquatic environments, and especially the ocean, suggests symbiotic associations that have not been reported previously. Recently, the ability to live as a symbiont was proposed as a mechanism to adapt to saline environments (Hinojosa-Vidal *et al*., 2018). Although many marine protists are known to host symbiotic photosynthetic eukaryotes, there are little mentions of Trebouxiophyceae involved in such relationships. A probable case is the association between a '*Chlorella*' (to be considered as a Trebouxiophyceae *sensu lato*) and a foraminiferan (Lee *et al*., 1982), or between *Elliptochloris marina* and the sea anemone *Anthopleura* (Letsch *et al*., 2009). However, it is not unlikely that other similar associations remained overlooked given of the lack of characteristic morphological features of most Trebouxiophyceae.

Genera *Myrmecia*, *Symbiochloris* and *Trebouxia* are species-rich genera that include a majority of lichen photosymbionts (Honegger and Brunner, 1981; Rambold *et al*., 1998; Škaloud *et al*., 2016). Other genera, such as *Apatococcus*, *Coccomyxa*, *Elliptochloris* and *Pseudochlorella* comprise a variety of lichen symbionts and free-living forms. In this study, we detected 33 OTUs belonging to these clades present in marine systems. Some can be considered as abundant, representing over 1% of the whole community in certain MA samples (Table S4). We consider unlikely that these sequences correspond to organisms of edaphic origin brought in the sea as we systematically removed rare sequences. Moreover, the marine sampling was performed far away from the coast (de Vargas *et al*., 2015). Some of these organisms may possibly be free-living forms from these groups or, perhaps, free-living stages of symbiotic organisms. In support for this hypothesis, two OTUs (OTU_59ee09bd8eccf066d00d755548d60655 and OTU_a0cbafb71d700506213b5e4b3c3870dc) were classified within *Elliptochloris*; a future expansion of the trebouxiophyceaen genetic database will help determine how far these organisms are related to *Elliptochloris marina* or if they constitute a new symbiotic lineage. Althogether, these findings open the way for an accurate assessment of Trebouxiophycean symbiotic associations in marine systems, their range of possible hosts and the relevance of these associations at ecosystem level.

Concluding remarks

Environmental DNA surveys based on large sampling designs are excellent strategies to investigate diversity hotspots, especially for groups with homogeneous morphologies such as Trebouxiophyceae. Here, we identified the environments where the bulk of global diversity is to be found (freshwater, soil), but also where the potential for novel discovery is the highest (marine plankton), and where potentially unknown symbiotic associations take place. This sets a starting point for organism centered research, based on cultures, with potential implications ranging from evolutionary studies focused on symbiosis to the retrieval of new strains with relevant biotechnological properties.

Experimental procedures

Curated database and the reference phylogenetic tree reconstruction

The reference phylogenetic tree was built with a similar pipeline as proposed by del Campo *et al*. (2018). Briefly, sequences related to Trebouxiophyceae were extracted from SILVA v132 database (Quast *et al*., 2012). They were aligned with MAFFT v7.123 (Katoh and Standley, 2013). Introns and intergenic regions were excluded using Geneious v9.0.5 (Biomatters, Auckland, New Zealand). Sequences were clustered at 99% of identity with USEARCH v10.0.24 (Edgar, 2010). The centroids for each cluster were aligned with MAFFT and a phylogenetic tree inference with RaxML v8.2.8 (Stamatakis, 2014) using GTR + GAMMA model and 100 bootstraps, after the non-informative regions of the alignment excluded by trimAL v1.4.rev22 (Capella-Gutierrez *et al*., 2009) was performed. The phylogenetic tree was manually curated and these steps were iteratively repeated until obtaining a consistent database with sequences that represent most of the clades (Leliaert *et al*., 2012).

Sequences belonging to Trebouxiophyceae but not assigned as such in GenBank (like unassigned environmental clones) were searched into the NCBI database (September 2017). The search was performed with BLAST v 2.6.0 (Altschul *et al*., 1997). For each sequence, 100 hits with more than 97% sequence identity, e-value less than 1E-6 and bit-score higher than 1000 were retrieved. This step was repeated until no new sequences were obtained. Sequences with 18S mark region length shorter than 500 nucleotides were excluded using the filtration step proposed by Logares (2017). New sequences were manually revised using web BLAST with standard parameters in order to exclude non-Trebouxiophyceae sequences.

To build a reference phylogenetic tree, 1814 sequences with more than 1000 nucleotides and not redundant were used to made an initial alignment with MAFFT using the E-INS-I algorithm (Katoh and Standley, 2013), which we automatically trimmed poorly aligned regions with trimAL (Capella-Gutierrez *et al*., 2009). Sequences selected from each clade were aligned with MAFFT (Katoh and Standley, 2013) using E-INS-I algorithm and 1000 replicates, the alignment was edited and a phylogenetic tree was performed using RaxML (Stamatakis, 2014) with GTR + GAMMA model and 1000 bootstraps. This tree was used as a reference to

placement environmental sequences from high throughput sequencing into the different Trebouxiophycean clades or genera.

Phylogenetic tree annotation

The names of clades from the reference tree were assigned according to the bibliography (Leliaert *et al*., 2012; Fučíková *et al*., 2014; Hodač *et al*., 2016). We defined new environmental clades (designed as TREBXX) when environmental sequenced clustered together in a robust clade supported by a bootstrap value of at least 70%. We annotated original environment type (FW, MA or SO) and lifestyle (photosymbionts/parasites) based on information retrieved from Genbank and on the bibliography. Organisms found in brackish environments were classified as marine as the salinity barrier is considered as the greatest divide in microbial diversity (Logares *et al*., 2009). This information was deposited in Table S1. OTUs were added to the Trebouxiophyceae reference alignment using PaPaRa v2.5 (Berger and Stamatakis, 2011) and placed in the reference phylogenetic tree with EPA implemented in RaxML v 8.2 and – epa-accumulated-threshold = 0.999 command (Zhang *et al*., 2013; Mahé *et al*., 2017). The best placements for each OTU were selected using Gappa (https://github.com/lczech/gappa) and corroborated using BLASTN web service tool (Altschul *et al*., 1997). New clades obtained in this study were annotated as TREBXXHTS. The phylogenetic tree figure was performed using Itool (Letunic and Bork, 2007) and edited with inkscape 0.48.4 (https://inkscape.org).

Sampling, DNA extraction, amplification and sequencing

SO samples (0.5 g) were stored in a buffer appropriate for environmental nucleic acids preservation (Life Guard, Qiagen) prior to DNA extraction (see Seppey *et al*., 2017). FW samples consisted in plankton that were pre-filtered at different size fractions, with an initial pre-filtration through 50 μm mesh (Table S2). The filters were placed in cryovials with 1.8 ml of lysis buffer (40 mM EDTA, 50 mM Tris–HCl, 0.75 M sucrose) and stored at −80°C until DNA extraction (see Schiaffino *et al*., 2016). Nucleic acids extraction was conducted using the kit PowerSoil® DNA Isolation Kit (Qiagen). PCR amplification of the v9 region of the gene coding for the SSU rRNA gene was performed using the primers 1389F 5′- TTGTACACACCGCCC -3′ and 1510R 5′- CCTTCYGCAGGTTCACCTAC (Amaral-Zettler *et al*., 2009), and the same protocols as in de Vargas *et al*. (2015). For MA, the different size and depth fraction as described in de Vargas *et al*. (2015) were analysed individually.

Bioinformatic analysis

Our initial data set was constituted of 435 samples distributed in different three different environments: 36 Soil (SO), 65 Freshwater (FW) and 334 Marine (MA). For details about the geographical distribution and the reference of each sample see Table S2 for SO and FW, and supplementary material from de Vargas *et al*. (2015) for MA; all these correspond to open ocean samples. The data set and the sequences of de Vargas *et al*. (2015) are available in http://taraoceans.sb-roscoff.fr/EukDiv/. In order to compare the three environments, we processed the FW and the SO samples bioinformatically (i.e. assembling, curation, parsing) with the same pipeline developed by de Vargas *et al*. (2015) and already applied in (Singer *et al*., 2016). Likewise, Tara Ocean and our own samples were clustered with Swarm v2.2.2 (Mahé *et al*., 2014) with the recommended parameters. The swarm seeds (OTUs) were assigned using PR^2 v 4.10.0 (Guillou *et al*., 2013) and sequences affiliated with Trebouxiophyceae were extracted. New sequences were deposited in NCBI under the numbers MK988625-MK989495.

Generation of the community matrix (OTU table)

In order to estimate abundance for each OTU, we first drew a high-throughput environmental sequences (HTES) table, defined as a contribution of all dereplicated sequences to each sample. Finally, for each OTU we searched all HTES that compose them. Because we wanted to consider only samples that contained a significant population of Trebouxiophyceae, samples with less than 10 sequences in total were excluded. Indeed, the ecological relevance of rare sequences in Illumina-based environmental diversity survey can be subject to controversy (Schiaffino *et al*., 2016). All obtained OTUs with their respective affiliations are detailed in Table S3.

The rarefaction curve was performed by each sample, each environment (Freshwater, Soil and Marine) and for all OTUs together using the VEGAN version 2.5–3 package (Oksanen *et al*., 2008) and R (http://cran.r-project.org).

Ordination and global distribution of Trebouxiophyceae and genetic novelty

To highlight the heterogeneity between the environments, we performed a non-metric multidimensional scaling (NMDS). We used only the samples from SO and FW and exclude those from MA as they did not share any OTU with the others environments. Moreover, SO samples were divided into irrigated and mostly wet soils (like forest soil and peat) and others soils experimenting drought, to be displayed differentially in the figure. We normalized our data set with a Hellinger transformation (Legendre and Gallagher, 2001) and used a Bray-Curtis distance to compute the NMDS using VEGAN 2.5-3 package (Oksanen *et al*., 2008) package from R 3.5 (http://cran.r-project.org). Different biotic variables, calculated as a contribution for

each clade to each sample were fit in the NMDS with the function envfit from VEGAN 2.5-3, only variables with significant *p*-value (*p*-value<0.01) were shown in Fig. 3.

To assess the global distribution of Trebouxiophyceae, we calculated the specific richness of each samples, here defined as the total number of OTUs. We plotted the three environment independently on a world map, to see in which environment the samples were more diverse. We used GGMAP version 2.6.1 and GGPLOT version 3.1.0 package from R to compute the analysis. To determine the representation of the OTUs in references databases and in each environment, OTUs were separated according to their distribution, forming four classifications (Freshwater, Soil, Freshwater and Soil, and Marine). For each OTU we extracted the percentage of identity assigned in the classification step explained before using PR^2 (Guillou *et al*., 2013) reference database and plotted in boxplot graphs using a GGPLOT version 3.1.0 from R (http://cran.r-project.org).

Conflict of interest

The authors declare that there is no conflict of interest regarding the publication of this article.

Acknowledgements

We are indebted to Didier Debroas for providing several freshwater samples from French lakes. EL wishes to acknowledge the program 'Atracción de talentos de la Comunidad de Madrid' (grant 2017-T1/AMB-5210 4) and Swiss National Fund project SNF 31003A_163254 for funding. The project was funding by the ANPCyT 'Agencia Nacional de Promoción Científica y Tecnológica' (PICT-2014-1290; PICT-2016-1079). The Swiss NSF (P2NEP3_178543) to D.S. We would like to acknowledge also Sergio Pérez Ortega and Raquel Pino (Real Jardín Botánico de Madrid, CSIC, Spain) for the fruitful discussions on trebouxiophyceae and symbioses in lichens, and two anonymous reviewers for valuable comments on the manuscript.

References

Aboal, M., and Werner, O. (2011) Morphology, fine structure, life cycle and phylogenetic analysis of *Phyllosiphon arisari*, a siphonous parasitic green alga. *Eur J Phycol* **46**: 181–192.

Altschul, S., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Amaral-Zettler, L.A., McCliment, E.A., Ducklow, H.W., and Huse, S.M. (2009) A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS ONE*: **4**: e6372.

Banerjee, A., Sharma, R., Chisti, Y., and Banerjee, U.C. (2002) *Botryococcus braunii*: a renewable source of hydrocarbons and other chemicals. *Crit Rev Biotechnol* **22**: 245–279.

Berger, S.A., and Stamatakis, A. (2011) Aligning short reads to reference alignments and trees. *Bioinformatics* **27**: 2068–2075.

Blanc, G., Agarkova, I., Grimwood, J., Kuo, A., Brueggeman, A., Dunigan, D.D., *et al*. (2012) The genome of the polar eukaryotic microalga *Coccomyxa subellipsoidea* reveals traits of cold adaptation. *Genome Biol* **13**: R39.

Bock, C., Krienitz, L., and Pröschold, T. (2011) Taxonomic reassessment of the genus *Chlorella* (Trebouxiophyceae) using molecular signatures (barcodes), including description of seven new species. *Fottea* **11**: 293–312.

Büdel, B., Darienko, T., Deutschewitz, K., Dojani, S., Friedl, T., Mohr, K.I., *et al*. (2009) Southern African biological soil crusts are ubiquitous and highly diverse in drylands, being restricted by rainfall frequency. *Microb Ecol* **57**: 229–247.

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

Cavacini, P. (2001) Soil algae from northern Victoria Land (Antarctica). *Polar Biosci* **14**: 45–60.

de Koning, A.P., and Keeling, P.J. (2006) The complete plastid genome sequence of the parasitic green alga *Helicosporidium* sp. is highly reduced and structured. *BMC Biol* **4**: 12.

de Vargas, C., Audic, S., Henry, N., Decelle, J., Mahé, F., Logares, R., *et al*. (2015) Ocean plankton. Eukaryotic plankton diversity in the sunlit ocean. *Science* **348**: 1261605.

de Vries, J., and Archibald, J.M. (2018) Plant evolution: landmarks on the path to terrestrial life. *New Phytol* **217**: 1428–1434.

del Campo, J., Kolisko, M., Boscaro, V., Santoferrara, L.F., Nenarokov, S., Massana, R., *et al*. (2018) EukRef: phylogenetic curation of ribosomal RNA to enhance understanding of eukaryotic diversity and distribution. *PLoS Biol* **16**: e2005849.

Edgar, R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460–2461.

Fermani, P., Mataloni, G., and Van de Vijver, B. (2007) Soil microalgal communities on an antarctic active volcano (Deception Island, south Shetlands). *Polar Biol* **30**: 1381–1393.

Flechtner, V.R., Pietrasiak, N., and Lewis, L.A. (2013) Newly revealed diversity of green microalgae from wilderness areas of Joshua tree national park (JTNP). *Monogr West North Am Nat* **6**: 43–63.

Foflonker, F., Mollegard, D., Ong, M., Yoon, H.S., and Bhattacharya, D. (2018) Genomic analysis of *Picochlorum* species reveals how microalgae may adapt to variable environments. *Mol Biol Evol* **35**: 2702–2711.

Fučíková, K., Lewis, P.O., and Lewis, L.A. (2014) Widespread desert affiliation of trebouxiophycean algae (Trebouxiophyceae, Chlorophyta) including discovery of three new desert genera. *Phycol Res* **62**: 294–305.

Fuhrman, J.A., Steele, J.A., Hewson, I., Schwalbach, M.S., Brown, M.V., Green, J.L., and Brown, J.H. (2008) A latitudinal diversity gradient in planktonic marine bacteria. *Proc Natl Acad Sci U S A* **105**: 7774–7778.

Gomaa, F., Kosakyan, A., Heger, T.J., Corsaro, D., Mitchell, E.A.D., and Lara, E. (2014) One alga to rule them all: unrelated mixotrophic testate amoebae (Amoebozoa, Rhizaria and Stramenopiles) share the same symbiont (Trebouxiophyceae). *Protist* **165**: 161–176.

Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., *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**: D597–D604.

Guiry, M.D., and Guiry, G.M. (2018) *AlgaeBase*. National University of Ireland, Galway: World-wide electronic publication. http://www.algaebase.org.

Hawksworth, D. (2000) Freshwater and marine lichen-forming fungi. *Fungal Divers* **5**: 1–7.

Heesch, S., Sutherland, J.E., and Nelson, W.A. (2012) Marine Prasiolales (Trebouxiophyceae, Chlorophyta) from New Zealand and the Balleny Islands, with descriptions of *Prasiola novaezelandiae* sp. nov. and *Rosenvingiella australis* sp. nov. *Phycologia* **51**: 217–227.

Hinojosa-Vidal, E., Marco, F., Martínez-Alberola, F., Escaray, F.J., García-Breijo, F.J., Reig-Armiñana, J., *et al*. (2018) Characterization of the responses to saline stress in the symbiotic green microalga *Trebouxia* sp. *TR9. Planta* **248**: 1473–1486.

Hodač, L., Hallmann, C., Spitzer, K., Elster, J., Faßhauer, F., Brinkmann, N., *et al*. (2016) Widespread green algae *Chlorella* and *Stichococcus* exhibit polar-temperate and tropical-temperate biogeography. *FEMS Microbiol Ecol* **92**: 1–16.

Honegger, R., and Brunner, U. (1981) Sporopollenin in the cell walls of *Coccomyxa* and *Myrmecia* phycobionts of various lichens: an ultrastructural and chemical investigation. *Can J Bot* **59**: 2713–2734.

Huss, V.A.R., Frank, C., Hartmann, E.C., Hirmer, M., Kloboucek, A., Seidel, B.M., *et al*. (1999) Biochemical taxonomy and molecular phylogeny of the genus *Chlorella* sensu lato (Chlorophyta). *J Phycol* **35**: 587–598.

Juárez, Á.B., Vélez, C.G., Iñiguez, A.R., Martínez, D.E., Rodríguez, M.C., Vigna, M.S., and de Molina, M.d.C.R. (2011) A *Parachlorella kessleri* (Trebouxiophyceae, Chlorophyta) strain from an extremely acidic geothermal pond in Argentina. *Phycologia* **50**: 413–421.

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

Krienitz, L., Huss, V.A.R., and Bock, C. (2015) Chlorella: 125 years of the green survivalist. *Trends Plant Sci* **20**: 67–69.

Lara, E., and Gomaa, F. (2017) Symbiosis between testate amoebae and photosynthetic organisms. In *Algal and Cyanobacteria Symbioses*. New Jersey, USA: World Scientific (Europe), pp. 399–419.

Lass-Flörl, C., and Mayr, A. (2007) Human protothecosis. *Clin Microbiol Rev* **20**: 230–242.

Lee, J.J., Reidy, J., and Kessler, E. (1982) Symbiotic chlorella species from larger foraminifera. *Bot Mar* **25**: 171–176.

Legendre, P., and Gallagher, E.D. (2001) Ecologically meaningful transformations for ordination of species data. *Oecologia* **129**: 271–280.

Leliaert, F., Smith, D.R., Moreau, H., Herron, M.D., Verbruggen, H., Delwiche, C.F., and De Clerck, O. (2012) Phylogeny and molecular evolution of the Green algae. *CRC Crit Rev Plant Sci* **31**: 1–46.

Letunic, I., and Bork, P. (2007) Interactive tree of life (iTOL): an online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**: 127–128.

Lewis, L.A., and McCourt, R.M. (2004) Green algae and the origin of land plants. *Am J Bot* **91**: 1535–1556.

Lemieux, C., Otis, C., and Turmel, M. (2014) Chloroplast phylogenomic analysis resolves deep-level relationships within the green algal class Trebouxiophyceae. *BMC Evol Biol* **14**: 211.

Logares, R., Bråte, J., Bertilsson, S., Clasen, J.L., Shalchian-Tabrizi, K., and Rengefors, K. (2009) Infrequent marine–freshwater transitions in the microbial world. *Trends Microbiol* **17**: 414–422.

Logares R. ramalok/amplicon_processing: Workflow for Analysing MiSeq Amplicons based on Uparse. 2017, doi: https://doi.org/10.5281/ZENODO.259579.

Letsch, M.R., Muller-Parker, G., Friedl, T., and Lewis, L.A. (2009) *Elliptochloris marina* sp. nov.(Trebouxiophyceae, Chlorophyta), symbiotic green alga of the temperate pacific sea anemones *Anthopleura xanthogrammica* and *A. elegantissima* (Anthozoa, Cnidaria) 1. *J Phycol* **45**: 1127–1135.

Mahé, F., Rognes, T., Quince, C., de Vargas, C., and Dunthorn, M. (2014) Swarm: robust and fast clustering method for amplicon-based studies. *PeerJ* **2**: e593.

Mahé, F., de Vargas, C., Bass, D., Czech, L., Stamatakis, A., Lara, E., *et al*. (2017) Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. *Nat Ecol Evol* **1**: 1–8.

Metz, S., Lopes dos Santos, A., Berman, M.C., Bigeard, E., Licursi, M., Not, F., *et al*. (2019) Diversity of photosynthetic picoeukaryotes in eutrophic shallow lakes as assessed by combining flow cytometry cell-sorting and high throughput sequencing. *FEMS Microbiol Ecol* **95**: fiz038.

Metzger, P., Allard, B., Casadevall, E., Berkaloff, C., and Coute, A. (1990) Structure and chemistry of a new chemical race of *Botryococcus braunii* (chlorophyceae) that produces lycopadiene, a tetraterpenoid hydrocarbon. *J Phycol* **26**: 258–266.

Metzger, P., and Largeau, C. (1999) Chemicals of Botryococcus braunii. In: Chemicals from microalgae. Cohen, Z. (ed.). London: Taylor & Francis, pp. 205–260.

Metzger, P., Largeau, C., and Casadevall, E. (1991) Lipids and macromolecular lipids of the hydrocarbon-rich microalga *Botryococcus braunii*. Chemical structure and biosynthesis. In *Geochemical and Biotechnological Importance*. Vienna: Springer, pp. 1–70.

Neofotis, P., Huang, A., Sury, K., Chang, W., Joseph, F., Gabr, A., *et al*. (2016) Characterization and classification of highly productive microalgae strains discovered for biofuel and bioproduct generation. *Algal Res* **15**: 164–178.

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

Oksanen J, Kindt R, Legendre P et al. The vegan package: community ecology package, version 1.13–1. 2008. http://vegan.r-forge.r-project.org.

Pan, J., del Campo, J., and Keeling, P.J. (2017) Reference tree and environmental sequence diversity of Labyrinthulomycetes. *J Eukaryot Microbiol* **64**: 88–96.

Pálffy, K., Felföldi, T., Mentes, A., Horváth, H., Márialigeti, K., Boros, E., *et al*. (2014) Unique picoeukaryotic algal community under multiple environmental stress conditions in a shallow, alkaline pan. *Extremophiles* **18**: 111–119.

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

Pröschold, T., Darienko, T., Silva, P.C., Reisser, W., and Krienitz, L. (2011) The systematics of Zoochlorella revisited employing an integrative approach. *Environ Microbiol* **13**: 350–364.

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., *et al*. (2012) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res* **41**: D590–D596.

Raja, R., Hemaiswarya, S., Kumar, N.A., Sridhar, S., and Rengasamy, R. (2008) A perspective on the biotechnological potential of microalgae. *Crit Rev Microbiol* **34**: 77–88.

Rambold, G., Friedl, T., and Beck, A. (1998) Photobionts in lichens: possible indicators of phylogenetic relationships? *Bryologist* **101**: 392–397.

Schiaffino, M.R., Lara, E., Fernández, L.D., Balagué, V., Singer, D., Seppey, C.V.W., *et al*. (2016) Microbial eukaryote communities exhibit robust biogeographical patterns along a gradient of Patagonian and Antarctic lakes. *Environ Microbiol* **18**: 5249–5264.

Safi, C., Zebib, B., Merah, O., Pontalier, P.-Y., and Vaca-Garcia, C. (2014) Morphology, composition, production, processing and applications of *Chlorella vulgaris*: a review. *Renew Sust Energ Rev* **35**: 265–278.

Singer, D., Lara, E., Steciow, M.M., Seppey, C.V.W., Paredes, N., Pillonel, A., *et al*. (2016) High-throughput sequencing reveals diverse oomycete communities in oligotrophic peat bog micro-habitat. *Fungal Ecol* **23**: 42–47.

Škaloud, P., Friedl, T., Hallmann, C., Beck, A., and Dal Grande, F. (2016) Taxonomic revision and species delimitation of coccoid green algae currently assigned to the genus *Dictyochloropsis* (Trebouxiophyceae, Chlorophyta). *J Phycol* **52**: 599–617.

Seppey, C.V.W., Singer, D., Dumack, K., Fournier, B., Belbahri, L., Mitchell, E.A.D., and Lara, E. (2017) Distribution patterns of soil microbial eukaryotes suggests widespread algivory by phagotrophic protists as an alternative pathway for nutrient cycling. *Soil Biol Biochem* **112**: 68–76. https://doi.org/10.1016/j.soilbio.2017.05.002.

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

Tragin, M., Zingone, A., and Vaulot, D. (2018) Comparison of coastal phytoplankton composition estimated from the V4 and V9 regions of the 18S rRNA gene with a focus on photosynthetic groups and especially Chlorophyta. *Environ Microbiol* **20**: 506–520.

Tragin, M., and Vaulot, D. (2018) Green microalgae in marine coastal waters: the ocean sampling day (OSD) dataset. *Sci Rep* **8**: 14020.

Thüs, H., Muggia, L., Pérez-Ortega, S., Favero-Longo, S.E., Joneson, S., O'Brien, H., *et al*. (2011) Revisiting photobiont diversity in the lichen family Verrucariaceae (Ascomycota). *Eur J Phycol* **46**: 399–415.

Trémouillaux-Guiller, J., and Huss, V.A.R. (2007) A cryptic intracellular green alga in *Ginkgo biloba*: ribosomal DNA markers reveal worldwide distribution. *Planta* **226**: 553–557.

Zhang, J., Kapli, P., Pavlidis, P., and Stamatakis, A. (2013) A general species delimitation method with applications to phylogenetic placements. *Bioinformatics* **29**: 2869–2876.

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's web-site:

Fig. S1. Rarefaction curve for each environment and for all sequences together respectively.

Table S1. Sequences retrived from public databases with all their information obteined by genbank and culture collections databases.

Table S2. Final samples analysed in this study. For marine samples reference to de Vargas *et al*., 2015.

Table S3. Taxonomy assignation for each OTU using PR^2 as reference database. The reference corresponded to NCBI accession Number for new OTUs from this work (MK988625 to MK989495) and the reference paper where the OTUs were extracted.

Table S4. Trebouxiophyceae OTUs contribution to each sample.

Table S5 Representatives OTUs for each environment. Clade and OTUs that contribute with more than 1% at least one environment or in the total reads for each clade were showed. The number is the percentage of contribution for each OTUs to each environment.