Environmental Microbiology (2008) **10**(7), 1804–1822 doi:10.1111/j.1462-2920.2008.01602.x

Wide genetic diversity of picoplanktonic green algae (Chloroplastida) in the Mediterranean Sea uncovered by a phylum-biased PCR approach

Manon Viprey,[1] Laure Guillou,[1] Martial Ferréol,[2] and Daniel Vaulot[1]*

[1]*Station Biologique de Roscoff, UMR7144, CNRS et Université Pierre et Marie Curie, BP74, 29682 Roscoff Cedex, France.*

[2]*Public Research Center Gabriel Lippmann, Department Environment and Agro-biotechnologies, 41 rue du Brill, L-4422 Belvaux, Grand-Duchy of Luxembourg.*

Summary

The genetic diversity of picoplanktonic (i.e. cells that can pass through a 3 μm pore-size filter) green algae was investigated in the Mediterranean Sea in late summer by a culture-independent approach. Genetic libraries of the 18S rRNA gene were constructed using two different primer sets. The first set is commonly used to amplify the majority of eukaryotic lineages, while the second was composed of a general eukaryotic forward primer and a reverse primer biased towards the phylum Chloroplastida. A total of 3980 partial environmental sequences were obtained: 1668 using the general eukaryotic primer set and 2312 using the Chloroplastida-biased primer set. Of these sequences, 65 (4%) and 594 (26%) belonged to the Chloroplastida respectively. A 99.5% sequence similarity cut-off value allowed classification of these 659 Chloroplastida sequences into 74 different operational taxonomic units. A majority of the Chloroplastida sequences (99%) belonged to the prasinophytes. In addition to the seven independent prasinophyte lineages previously described, we discovered two new clades (clades VIII and IX), as well as a significant genetic diversity at the species and subspecies levels, notably among the genera *Crustomastix*, *Dolichomastix* and *Mamiella* (Mamiellales), but also within *Pyramimonas* and *Halosphaera* (Pyramimonadales). Such diversity within prasinophytes has not previously been observed by cloning approaches, illustrating the power of using targeted primers for clone library construction. Prasinophyte assemblages differed especially in relation to nutrient levels. *Micromonas* and *Ostreococcus* were mainly recovered from mesotrophic areas, whereas *Mamiella*, *Crustomastix* and *Dolichomastix* were mostly detected in oligotrophic surface waters. Within genera such as *Ostreococcus* or *Crustomastix* for which several clades were observed, depth seemed to be the main factor controlling differential distribution of genotypes.

Introduction

Picoplanktonic marine protists, i.e. unicellular eukaryote less than 3 μm in size, have a worldwide distribution and contribute significantly to planktonic biomass in many ecosystems (Zubkov *et al*., 1998; Worden *et al*., 2004). They are core members of marine microbial food webs and play a fundamental role in global elemental cycles. Phototrophs contribute to global carbon fixation, while heterotrophs play a role in prokaryote consumption and nutrient remineralisation. Despite their ecological importance, the genetic diversity of picoeukaryotes has only been recently explored. Over the last decade, molecular techniques have played a key role in revealing the complexity of these assemblages of small protists in marine pelagic waters (e.g. Moon-van der Staay *et al*., 2001). In particular, direct polymerase chain reaction (PCR) amplification of the 18S rDNA gene using general eukaryotic primers has led to the discovery of previously unknown lineages of marine stramenopiles and alveolates (Massana *et al*., 2004a) and radiolarians (Not *et al*., 2007).

This approach has proved less suitable, however, for highlighting the extent of genetic diversity within photosynthetic organisms and particularly of green algae (Chloroplastida), as this group contributes little to most 18S rDNA genetic libraries that are dominated by heterotrophic groups (e.g. Romari and Vaulot, 2004; Lovejoy *et al*., 2006). This contradicts the recognized contribution of green algae to marine picoplanktonic communities, at least in coastal systems. The importance of Chloroplastida in marine environments was first established by the pioneering work of Johnson and Sieburth (1982) using

Received 1 October, 2007; accepted 11 February, 2008. *For correspondence. E-mail vaulot@sb-roscoff.fr, Tel. (+33) 2 98 29 23 34; Fax (+33) 2 98 29 23 24.

electron microscopy on natural samples. Later, the detection of chlorophyll *b*, a characteristic pigment of green algae, in different oceanic and coastal environments (Rodríguez *et al*., 2002) confirmed the ubiquity of Chloroplastida in marine waters, although in very oligotrophic regions part of this chlorophyll *b* can in fact be divinyl chlorophyll *b* originating from the cyanobacterium *Prochlorococcus* (Partensky *et al*., 1999). More recently, quantitative analyses using specific molecular probes detected by fluorescent *in situ* hybridization (FISH) have demonstrated that green algae, and particularly prasinophytes, can be one of the principal components of picoeukaryotic communities (Not *et al*., 2004). An illustration of the mismatch between qualitative (cloning-based) and quantitative (probe-based) molecular approaches is provided by data obtained from English Channel coastal waters, where Chloroplastida constituted the dominant picoeukaryotic component (on average 85% of the total number of picoeukaryotes) throughout the year according to FISH counts (Not *et al*., 2004), while they represented only 20% of the environmental sequences in genetic libraries (Romari and Vaulot, 2004). In addition, despite the high sequence diversity of Chloroplastida from culture collections, sequences recovered from genetic libraries are mostly related to three prasinophyte genera (*Bathycoccus*, *Micromonas* and *Ostreococcus*) within the order Mamiellales (Lovejoy *et al*., 2006; Worden, 2006). However, a recent study using quantitative PCR (Q-PCR) demonstrated that these three genera represented only a small fraction (13%) of picoplanktonic Chloroplastida in Mediterranean Sea samples (Marie *et al*., 2006). This suggests that a large fraction of Chloroplastida diversity is overlooked using 18S rDNA genetic libraries constructed with universal eukaryotic primers. This bias may have multiple causes, some linked to the organisms themselves (e.g. tough cell walls inducing poor DNA yield, low 18S rRNA gene copy numbers) and some due to the PCR approach (primer specificity, cloning efficiency). In particular, it has recently been demonstrated that a single PCR primer set fails to recover the full diversity of protist taxa (Stoeck *et al*., 2006). To date, few studies have taken this into account by applying different primer sets to environmental samples (Dawson and Pace, 2002; Stoeck *et al*., 2006) or by using specific primers to target a particular eukaryotic taxon (Bass and Cavalier-Smith, 2004). To our knowledge, such a targeted approach has never been applied to the Chloroplastida.

The Mediterranean Sea constitutes an environment that is particularly interesting for ecological studies as it offers a range of trophic conditions including extreme oligotrophy, particularly in summer when the water column is strongly stratified (Berman *et al*., 1985). While N:P is close to the Redfield ratio (16:1) in most oceanic waters, Mediterranean waters have a higher ratio especially in the eastern Basin, leading to strong phosphorus limitation (Moutin and Raimbault, 2002). A complex thermohaline circulation coupled with regional hydrodynamic features also contributes to the establishment of many different oceanic regions throughout the Mediterranean Sea (Manca *et al*., 2004). As an example, the exchange of the Atlantic and Mediterranean water masses at the Strait of Gibraltar induces marked salinity and temperature gradients (Gascard and Richez, 1985). In this paper we analyse the genetic diversity of marine Chloroplastida using both general eukaryotic and Chloroplastida-biased primers on picoplanktonic communities collected in late summer along a Mediterranean transect sampled during the PROSOPE (PROductivité des Systèmes Océaniques PElagiques) cruise in 1999.

Results

Oceanographic context

During the PROSOPE cruise, seven different stations (stars in Fig. 1) were sampled at one to three depths. The main physico-chemical characteristics of these samples were analysed using a principal component analysis (PCA) including eight variables (temperature, salinity, dissolved oxygen, nitrate, nitrite, phosphate, silicate and chlorophyll *a* concentrations). This analysis allowed discrimination of the Morocco upwelling and three main Mediterranean regions (Fig. S1) corresponding to the Alboran/Algerian Basins (stations 1 and 3), the Strait of Sicily (station 5) and the Ionian/Tyrrhenian/Ligurian Seas (stations MIO, 9 and DYF). For each of these regions, surface (i.e. above 50 m) and deeper euphotic (i.e. between 50 and 100 m) waters could be separated. Nutrient levels were always lower near the surface except for the Strait of Sicily where samples from 25 and 55 m had similar nutrient levels. Surface waters from the Morocco upwelling and the Strait of Sicily displayed high and intermediate nutrient levels, respectively, while the other regions were nutrient depleted in surface waters. Salinity and temperature displayed eastward gradients in surface waters (Fig. S1), while oxygen decreased from the Alboran/Algerian Basins towards the Ionian Sea. In contrast, deep euphotic waters from the Alboran/Algerian Basins and the Ionian/Tyrrhenian/Ligurian Seas were relatively similar as they were both cold, enriched in nutrients and depleted in oxygen. Finally, two samples had unique characteristics: the first (Algerian Basin, station 3, 5 m) had nutrient concentrations and salinity similar to other surface waters from the Alboran/Algerian Basins, but was characterized by higher temperature and lower oxygen; the second (Ionian Sea, station MIO, 50 m) was similar to the surface waters from the Ionian/Tyrrhenian/Ligurian Seas but with lower salinity and higher oxygen.

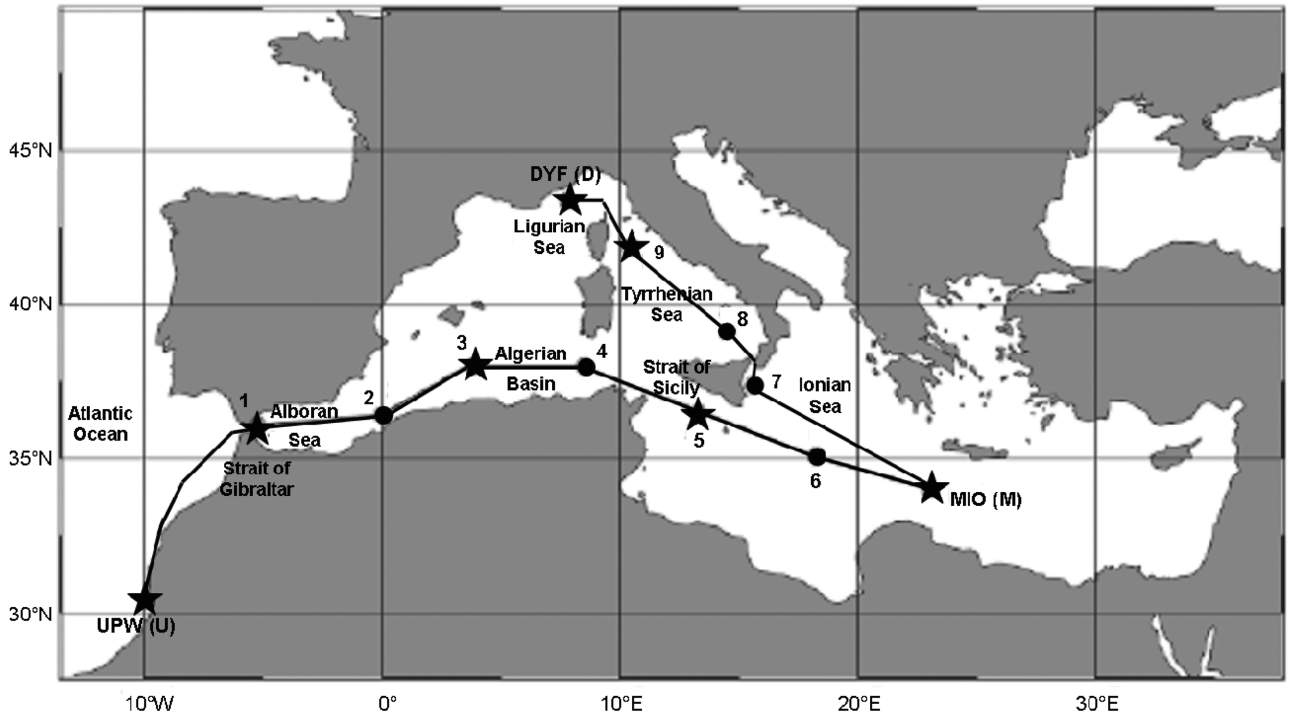


Fig. 1. Cruise track and stations investigated during the PROSOPE (1999) cruise. Stars indicate stations sampled for the present study.

Picoplanktonic diversity in the Mediterranean Sea

The genetic diversity of picoplanktonic Chloroplastida in the Mediterranean Sea was assessed by amplifying, cloning and sequencing the 18S rRNA gene from 15 samples collected along a transect during the PROSOPE cruise in 1999. For each sample, two primer sets were used, one targeting eukaryotes and the second designed to preferentially amplify Chloroplastida, yielding 1668 and 2312 environmental small subunit rDNA clones respectively (Table 1). As observed in previous studies, general eukaryotic libraries were dominated mainly by Alveolata (64%) with only 4% of the clones belonging to the Chloroplastida (65 clones). The Chloroplastida-biased genetic libraries were much more enriched in Chloroplastida sequences that represented 26% (594 clones) of total clones. Surprisingly, these genetic libraries also included sequences belonging to other photosynthetic groups such as the Chlorarachniophyta (11%), as well as to heterotrophic lineages such as the Radiolaria (37%) and Alveolata (9%). The 659 Chloroplastida sequences were grouped into either 49 or 74 operational taxonomic units (OTUs) using, respectively, 98.5% and 99.5% sequence identity cut-off levels (Fig. S2). Depending on the cut-off level, 80–82% of the OTUs were recovered only from Chloroplastida-biased libraries, while 4–6% were present only in the general eukaryotic libraries, indicating that the Chloroplastida-biased primer allows better sampling of the diversity of green algae. Rarefaction curves computed from the whole data set (659 Chloroplastida clones) never completely reached saturation (Fig. S2). The non-parametric estimators Ace and Chao1 yielded almost identical values of 62 and 68 unique OTUs for the 98.5% sequence identity cut-off and of 125 and 125 for the 99.5% sequence identity cut-off. Therefore, the recovered OTUs represented 72% and 59% of the estimated OTU richness (Ace indicator) at the 98.5% and 99.5% cut-off level, respectively, suggesting that an important part of the genetic diversity remains unsampled.

Phylogenetic diversity of prasinophytes

Of the 659 green algae sequences recovered from the genetic libraries, 99% belonged to the prasinophytes, with only one sequence of chlorophytes and six sequences of embryophytes (Streptophyta). The latter, also observed in previous studies (Countway *et al.*, 2005), were probably due to contamination by pollen or from the ship. For each OTU (99.5% cut-off), at least one clone from each genetic library was re-sequenced with different primers, yielding 174 18S rDNA consensus sequences of nearly 900 bp that were used for phylogenetic analyses.

The Chloroplastida clade was well supported (Fig. 2), but the lineage relationships within this group were not resolved. As previously reported by numerous authors (e.g. Nakayama *et al.*, 1998; Guillou *et al.*, 2004), prasinophytes were paraphyletic. Mediterranean environmen-

Table 1. Sample locations, chlorophyll *a* concentrations and clones libraries analysed in this study.

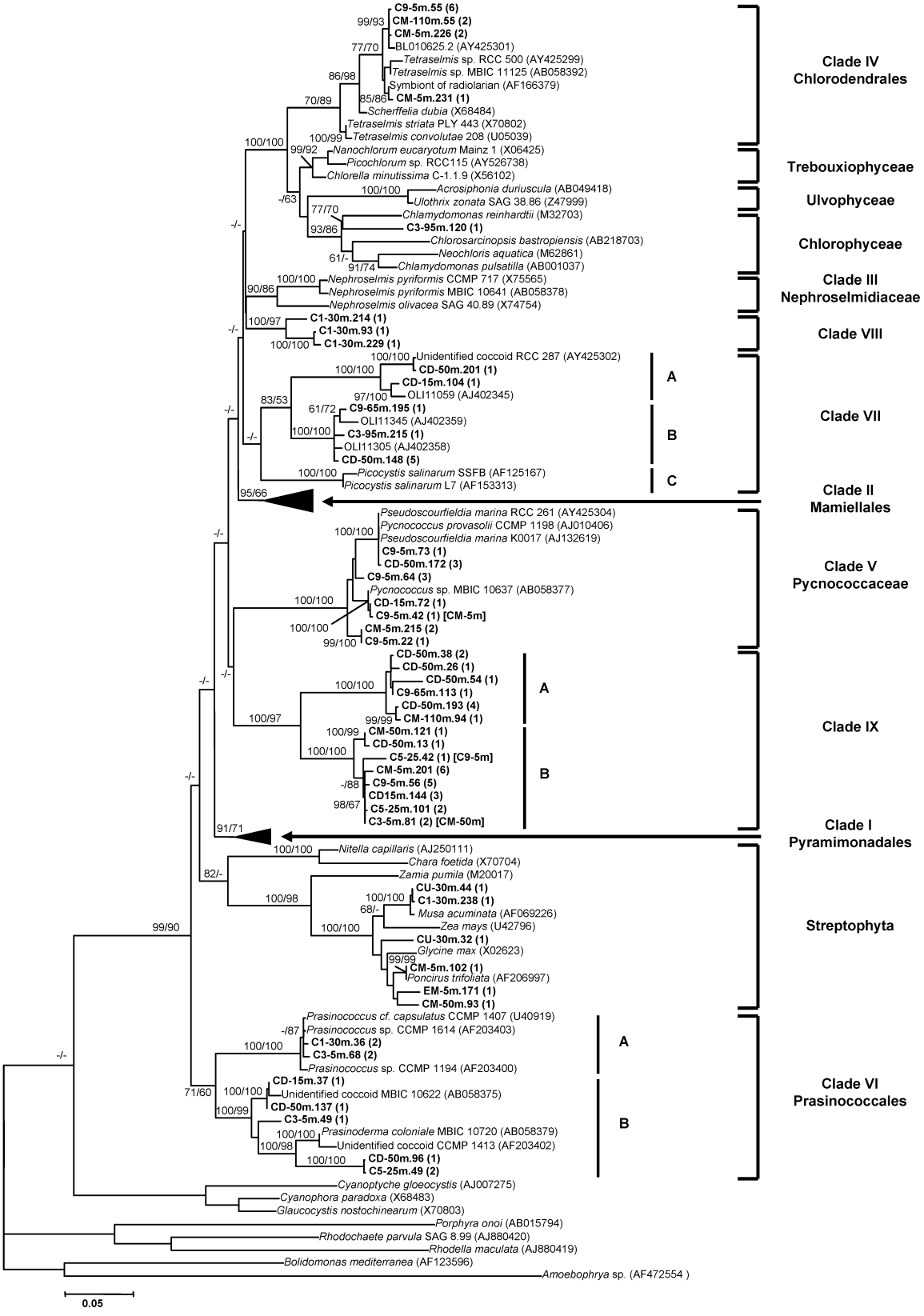
Station	Coordinates	Sampling depth (m)	Chl *a* (mg m^{-3})	No. of sequenced clones			No. of Chloroplastida clones		
				Euk	Chlo	Total	Euk	Chlo	Total
UPW	31°02′N 10°03′W	30	1.494	126	53	179	4	22	26
1	36°08′N 05°18′W	30	0.543	135	246	381	12	166	178
		50*	0.646	nd	nd	nd	nd	nd	nd
		80	0.340	69	195	264	4	3	7
3	37°98′N 03°83′E	5	0.071	46	165	211	0	28	28
		25	0.321	56	23	79	1	0	1
		40*	0.585	nd	nd	nd	nd	nd	nd
		95	0.176	27	173	200	0	5	5
5	36°47′N 13°32′E	25	0.072	168	151	319	3	81	84
		55*	0.766	185	131	316	28	69	97
MIO	33°98′N 22°02′E	5	0.024	62	173	235	1	24	25
		50	0.110	102	171	273	2	16	18
		90*	0.191	nd	nd	nd	nd	nd	nd
		110	0.128	115	134	249	0	14	14
9	41°88′N 10°43′E	5	0.082	144	101	245	3	30	33
		50*	0.298	nd	nd	nd	nd	nd	nd
		65	0.289	119	223	342	3	35	38
DYF	43°38′N 07°82′E	15	0.101	118	172	290	1	17	18
		50*	0.433	196	201	397	3	82	85
All samples				1668	2312	3980	65	594	659

Number of clones sequenced and number of *Chloroplastida* clones obtained for each sample in general eukaryotic (Euk) or in Chloroplastida-biased (Chlo) genetic libraries and in total. Depth of the deep chlorophyll maximum is indicated by an asterisk (*). nd indicates that no clone library was constructed for the corresponding depth.

tal sequences were distributed within six of the seven clades described earlier by Guillou and colleagues (2004), the Nephroselmidiaceae (clade III) not being represented. In addition, two new clades emerged from this study, named, respectively, VIII and IX. Mamiellales (clade II) was the best represented clade (78%), followed by Pyramimonadales (clade I, 9%) and the novel clade IX (5%). All prasinophyte clades were well supported (Fig. 2), with the exception of clade VII. This clade was composed of three strongly supported lineages (VII-A, VII-B and VII-C) that grouped together by maximum likelihood (ML), a topology not supported by neighbour joining (NJ) and maximum parsimony (MP) bootstrap analyses, as previously established by analysis of full-length sequences (Guillou *et al*., 2004). A limited number of Mediterranean sequences were recovered within subclades VII-A (two sequences) and VII-B (seven sequences).

The Mamiellales (clade II) were divided into seven lineages (Fig. 3) with strong bootstrap supports, corresponding to the level of the genus except for one lineage which included two genera: *Micromonas* and *Mantoniella*. All Mamiellales genera had representative clones in Mediterranean genetic libraries: *Dolichomastix* (4), *Crustomastix* (26), *Bathycoccus* (152), *Ostreococcus* (167), *Mamiella* (6), RCC strain 391 (10), *Micromonas* (125), and *Mantoniella* (14). Previous analyses have revealed phylogenetic heterogeneities within the genera *Ostreococcus* and *Micromonas*, leading to definition of four *Ostreococcus* clades (named A, B, C and D, Guillou *et al*., 2004) and five *Micromonas* clades (Guillou *et al*., 2004; Slapeta *et al*., 2006) recently renamed by Worden (2006) clades 1–5. *Micromonas* clade 5 and *Ostreococcus* clades C and D were missing in Mediterranean genetic libraries (Fig. 3). Clear genetic diversity within other genera such as *Dolichomastix* (two main clades), *Crustomastix* (three main clades) and *Mamiella* (two main clades) was also unveiled in the present study. In contrast, the genus *Bathycoccus* was confirmed to be very homogeneous (Fig. 3). Interestingly, we also retrieved one sequence closely related to *Mantoniella squamata* and 13 sequences related to *M. antarctica*.

The Pyramimonadales (clade I) were divided into two lineages. One contained the genera *Cymbomonas* and *Halosphaera*. *Halosphaera* was well represented in Mediterranean libraries (19 sequences) and was divided into two clades (Fig. 4): clade A included *Halosphaera* sp. and 11 Mediterranean sequences, while clade B was exclusively composed of environmental sequences. The second lineage corresponds to the *Pyramimonas–Pterosperma* complex (Fig. 4), including 41 Mediterranean sequences. While the monophyly of the genus *Pterosperma* was well supported, the genus *Pyramimonas* was paraphyletic (Fig. 4), as previously reported (Moro *et al*., 2002). The latter genus was therefore subsequently subdivided into several subgenera, *Trichocystis* (*P. parkae* and *P. australis*), *Vestigifera* (*P. disomota*), *Pyramimonas* (*P. propulsa*) and *Punctatae* (*P. aureus* and *P. olivaceae*) (Suda, 2004). Both the *Trichocystis* (33 sequences) and *Vestigifera* (three sequences) subgenera were represented in



Mediterranean genetic libraries. A fifth well-supported *Pyramimonas* lineage (two sequences) was observed (Fig. 4), weakly associated with the subgenus *Trichocystis* and which also included the environmental sequence BL010625.18 from the Spanish Mediterranean coast (Massana *et al*., 2004b).

Clades VIII and IX correspond probably to two novel prasinophyte orders (Fig. 2). Clade VIII was composed of three sequences obtained from the same sample (station 1, 30 m), while clade IX was composed of two subclades (IX-A and B) containing 10 and 23 sequences from three and seven different genetic libraries respectively. The exact position of these clades among prasinophytes was not resolved by the present analysis and will probably require full-length sequences.

Other environmental sequences from this study belonged to the Chlorodendrales, Prasinococcales and Pycnococcaceae (Fig. 2). Ten Mediterranean Chlorodendrales sequences were associated with the environmental sequence BL010625.2 originating from the Spanish coast (Massana *et al*., 2004b), while the eleventh sequence was related to a cultured free-living symbiont of the radiolarian *Spongodrymus* sp. (Fig. 2). The Prasinococcales were divided into two well-supported clades grouped together in ML analysis but with low NJ and MP bootstrap values (Fig. 2), probably as a consequence of the use of partial sequences as the Prasinococcales were strongly supported in previous full-length analyses (Fawley *et al*., 2000; Guillou *et al*., 2004). Prasinococcales clade A was composed of members of the genus *Prasinococcus* (together with four Mediterranean sequences), while clade B corresponds to the group formed by the genus *Prasinoderma* and the unidentified strain MBIC 10622 isolated from the Pacific Ocean (together with 6 Mediterranean sequences). Clade V (Pycnococcaceae) was composed of the species *Pycnococcus provasolii* and *Pseudoscourfieldia marina* and by 13 Mediterranean sequences (Fig. 2).

Biogeography of prasinophytes in the Mediterranean Sea

Among the eight prasinophyte clades recovered from this study, only the Mamiellales were present in all samples. The Pyramimonadales and clade IX were recovered in 10 and nine of the 15 samples, respectively, while the other orders were more sporadic (recovered in 1–5 samples, Fig. 5). The widespread distribution of Mamiellales was mainly due to the three genera *Bathycoccus*, *Ostreococcus* and *Micromonas*. However, none of these genera was present in all samples, suggesting that they may have distinct distributions and ecological niches. The same trend was evident at the species/infra-species level (Fig. 5). In order to determine the biogeographic trends observed at different taxonomic levels, separate correspondence analyses (CA) were performed for orders (Fig. 6A), genera (Fig. 6B) and species/infra-species (Fig. 6C). These analyses allowed determination of which sample(s) the different taxonomic assemblages preferentially occurred in (Fig. 7).

At the order level, clade VIII, recovered only from surface waters of the Strait of Gibraltar (station 1, 30 m) and the Mamiellales, that were much more abundant than all other orders, were removed from the analysis. A preliminary analysis including the Mamiellales (data not shown) confirmed their ubiquity and showed that this taxon was particularly well represented in surface coastal waters (station UPW, 30 m and station 1, 30 m) as well as in the Strait of Sicily (station 5, 25 m and 55 m). Of the other orders, three groups were discriminated (Fig. 6A). Chlorodendrales, clade V and clade IX were mainly recovered from surface waters of the Ionian and Tyrrhenian Basins (station MIO, 5 m and station 9, 5 m), characterized by very low nutrients as well as high temperature and salinity. Clade VII was more often observed in colder nutrient-rich deep euphotic zone samples of the Algerian and Tyrrhenian Basins (station 3, 95 m and station 9,

Fig. 2. SSU rDNA phylogenetic tree of prasinophytes inferred from 887 homologous positions of an alignment of 101 partial Chloroplastida sequences and eight partial outgroup sequences (three Glaucocystophyta, three Rhodophyta and two Chromalveolata). The phylogenetic tree was inferred by the ML method based on a TrNef (TrN equal base frequencies) model of DNA substitutions with a gamma distribution shape parameter of 0.5546 and substitution rates of R(b) [A–G] = 2.5501, R(e) [C–T] = 5.2552, and 1.0 for all other substitution rates. Total number of rearrangements tried = 174444. Clades I and II are represented, respectively, by four and eight sequences available in GenBank (AB017123, *Pyramimonas propulsa*; AB017125, *Halosphaera* sp.; AB017126, *Cymbomonas tetramitiformis* and AB017127, *Pterosperma cristatum* for clade I and AF509625, *Dolichomastix tenuilepis*; AB183628, *Crustomastix* sp.; AY425308, *Ostreococcus* sp. RCC 356; AY425315, *Bathycoccus prasinos*; AB017129, *Mamiella* sp.; AY425320, *Micromonas pusilla*; AY425321, prasinophyte strain RCC 391 and X73999, *Mantoniella squamata* for clade II). New sequences obtained in this study (99.5% sequence identity threshold) are in bold. One sequence per library was selected for each OTU. Phylotypes are labelled as follows: C or E (indicating the primer set used to construct the library, Chloroplastida-biased or Eukaryote-specific respectively) followed by station, depth in metres and finally name of the clone. The number of clones obtained for each phylotype is given in brackets. When an OTU has been observed from other genetic libraries than those of the corresponding phylotype, these genetic libraries are indicated in square brackets. The maximum likelihood (ML) tree is shown with corresponding significant (> 60) bootstrap values (percentage of 1000 replicates) on the internal branches obtained from neighbour joining (NJ) and maximum parsimony (MP) methods or a '-' if not significant (< 60). Scale bars indicate 0.05% divergence.

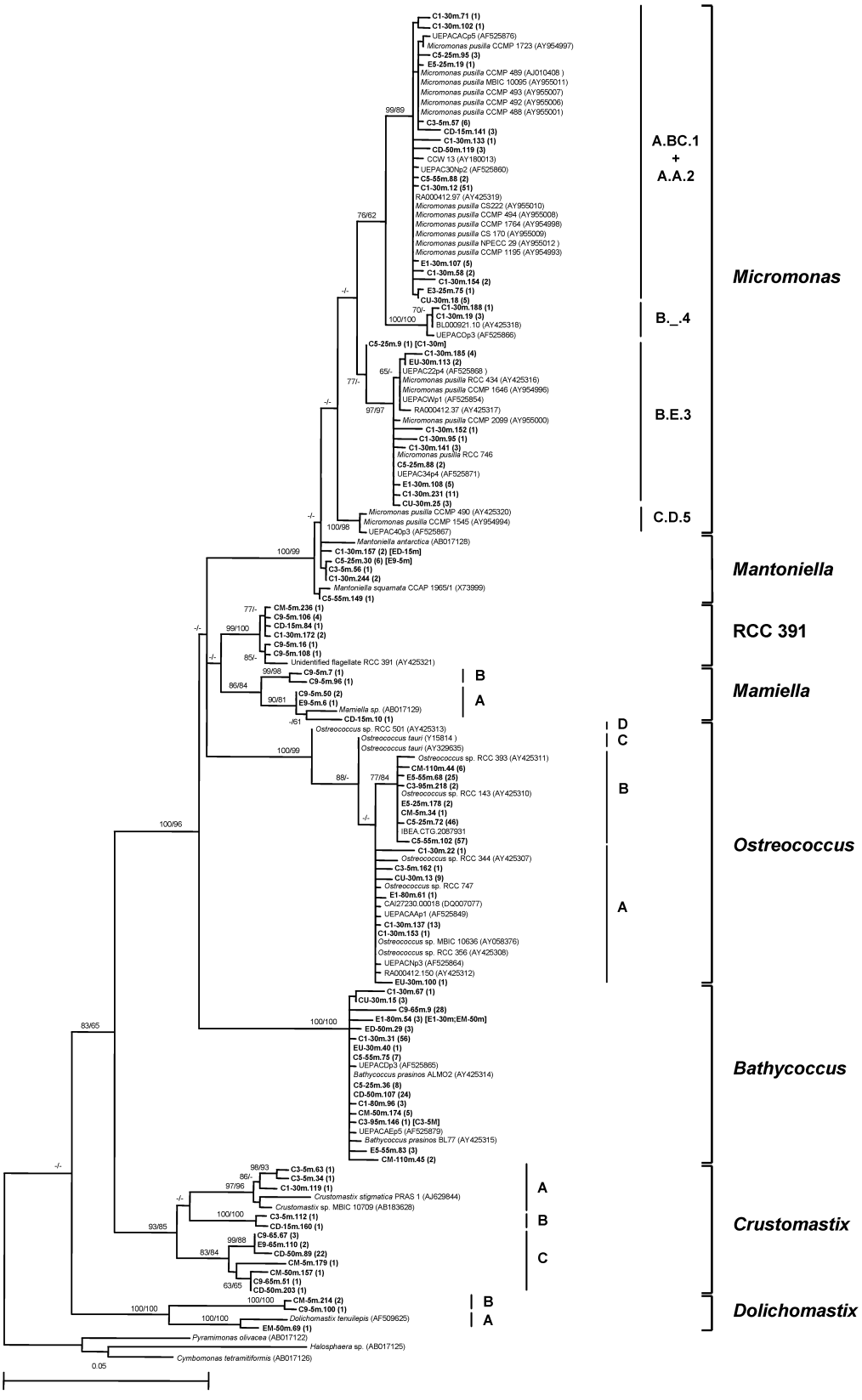


Fig. 3. SSU rDNA phylogenetic tree of Mamiellales inferred from 869 homologous positions of an alignment of 143 partial Mamiellales sequences and three partial outgroup sequences (Pyramimonadales). Designation of *Micromonas* clades follows the nomenclature of Worden (2006). The phylogenetic tree was inferred by the ML method based on a TrNef (TrN equal base frequencies) model of DNA substitutions with a gamma distribution shape parameter of 0.6195 and substitution rates of R(a) [A–C] = 0.7367, R(b) [A–G] = 3.0250, R(d) [C–G] = 1.5819, R(e) [C–T] = 5.9276 and 1.0 for all other substitution rates. Total number of rearrangements tried = 74440. IBEA.CTG.2087931 clone is from the Sargasso Sea project (Venter *et al*., 2004). *Micromonas pusilla* RCC 746 and *Ostreococcus* sp. RCC 747 sequences were provided by the Roscoff Culture Collection (RCC; Vaulot *et al*., 2004). Rest of legend as in Fig. 2.

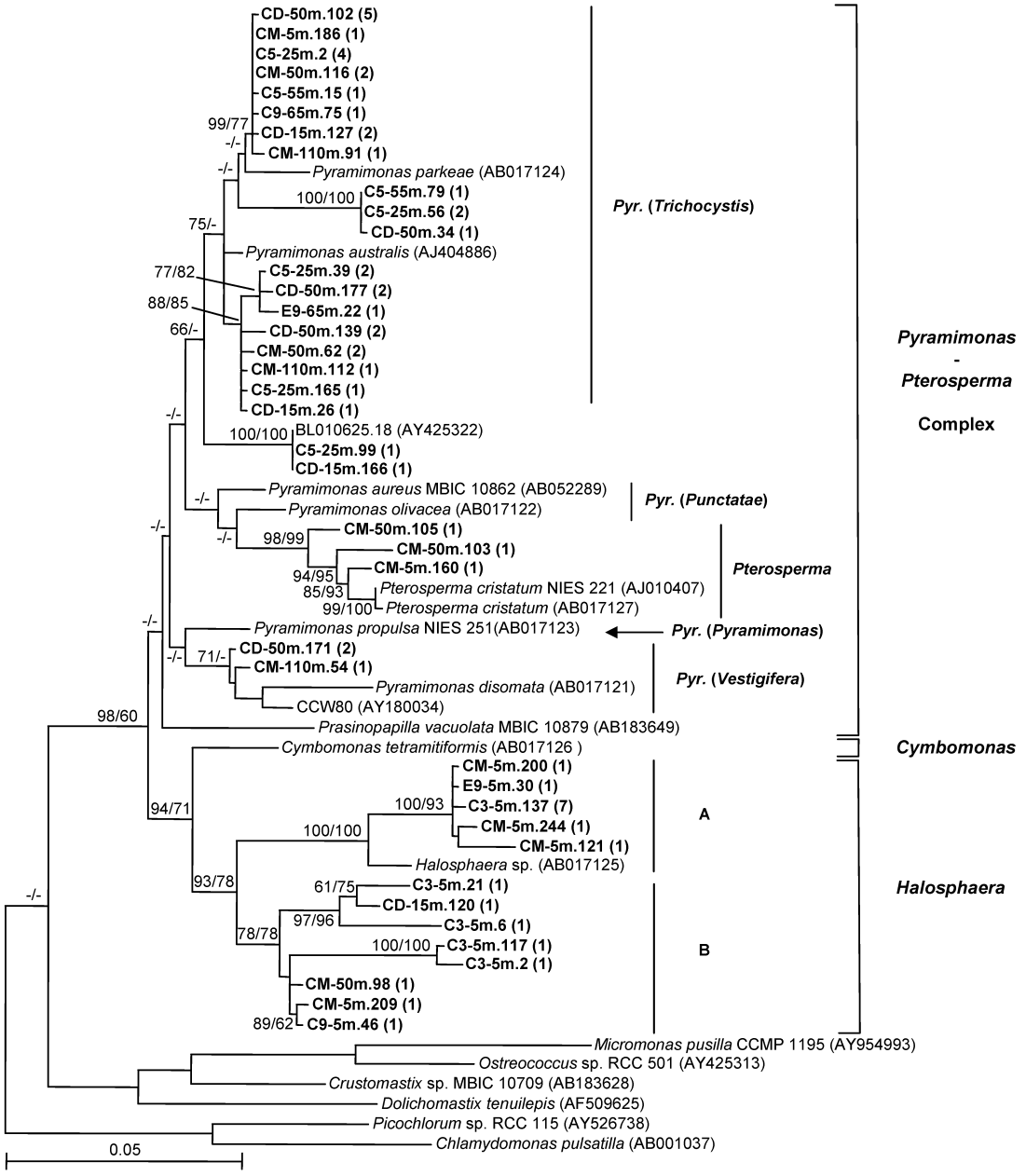


Fig. 4. SSU rDNA phylogenetic tree of Pyramimonadales inferred from 903 homologous positions of an alignment of 52 partial Pyramimonadales sequences and six partial outgroup sequences (four Mamiellales and two Chlorophyceae). The phylogenetic tree was inferred by the ML method based on a TrNef (TrN equal base frequencies) model of DNA substitutions with a gamma distribution shape parameter of 0.6155 and substitution rates of R(b) [A–G] = 2.2488, R(e) [C–T] = 4.5227, and 1.0 for all other substitution rates. Total number of rearrangements tried = 70557. Rest of legend as in Fig. 2.

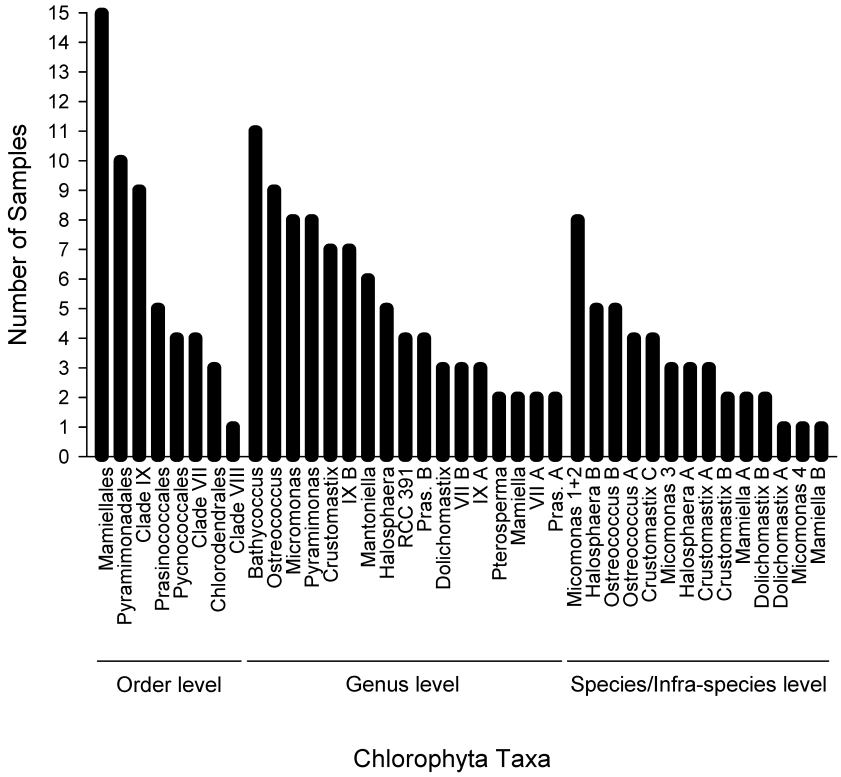


Fig. 5. For a given taxon, number of samples where this taxon was recovered based on the results from all genetic libraries. Pras. A and Pras B. correspond to the Prasinococcales clades.

65 m), but was also present in both surface and deep euphotic zone samples of the Ligurian Sea (station DYF, 15 m and 50 m). Finally, the Pyramimonadales and Prasinococcales seemed to be able to adapt to a range of conditions as they were present in both nutrient-depleted (station 1, 30 m; station 3, 5 m and station MIO, 50 m) and mesotrophic (station 5, 25 m and 55 m) samples and did not seem to be influenced by temperature. The only important parameter for these clades seemed to be depth as they were mainly recovered from the top 50 m of the water column.

At the genus level, *Ostreococcus* (Mamiellales) appeared mostly in the mesotrophic waters of the Strait of Sicily both in surface waters and at depth (station 5, 25 m and 55 m) (Fig. 6B). *Mamiella*, RCC 391, *Dolichomastix* (Mamiellales), and clade IX-B showed similar preferences for the oligotrophic surface waters of the Ionian and Tyrrhenian Seas (station MIO, 5 m and station 9, 5 m), characterized by low nutrients and high temperatures and salinities. *Crustomastix* (Mamiellales), clade IX-A and clade VII-B seemed to prefer colder waters with more nutrients as they were principally recovered from deep euphotic waters of the Tyrrhenian and Ligurian Seas (station 9, 65 m and station DYF, 50 m). Finally, *Micromonas* (Mamiellales) displayed an important contribution in the Morocco upwelling (station UPW, 30 m) as well as in surface waters of the Alboran/Algerian Basins (station 1, 30 m and station 3, 25 m), characterized by high oxygen and low nutrient content. The distributions of the other genera were not resolved (Fig. 6B).

The different species/infra-species within *Mamiella*, *Micromonas* (Mamiellales) and *Halosphaera* (Pyramimonadales) had similar biogeographic trends. However, this was not the case for all genera (Fig. 6C). *Ostreococcus* clade B was recovered from mesotrophic surface waters of the Strait of Sicily (station 5, 25 m) as well as in deeper euphotic waters with higher nutrient concentrations on a transect from the Algerian Basin to the Ionian Sea (station 3, 95 m; station 5, 55 m; station MIO, 110 m). *Ostreococcus* clade A was recovered together with the three *Micromonas* clades in the western part of the cruise transect, i.e. in the nutrient rich waters from the Morocco upwelling (station UPW, 30 m), at depth in the Strait of Gibraltar (station 1, 80 m) as well as in the more nutrient-depleted and well-oxygenated surface waters of the Strait of Gibraltar and the Algerian Basin (station 1, 30 m and station 3, 25 m). *Crustomastix* clade C was mainly present in the cold nutrient-rich deep euphotic waters of the Tyrrhenian and Ligurian Seas (station 9, 65 m and station DYF, 50 m), whereas *Crustomastix* clades A and B occurred together in the surface waters of the Algerian Basin and Ligurian Sea (station 3, 5 m and station DYF, 15 m), characterized by high temperature and low nutrient concentrations. *Dolichomastix* clades A and B were both

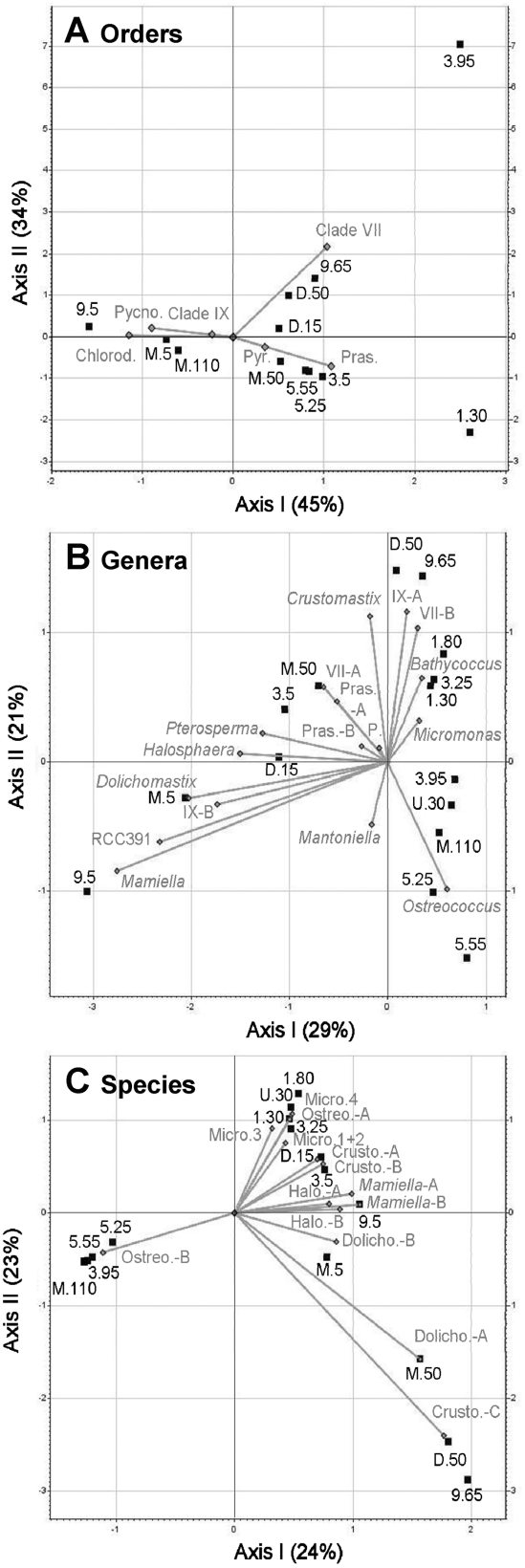


Fig. 6. Correspondence analysis calculated using a relative-abundance matrix obtained for:
A. The prasinophyte orders Chlorodendrales (Chlorod.), clade V or Pycnococcaceae (Pycno.), Pyramimonadales (Pyr.), Prasinococcales (Pras.), clade VII and clade IX.
B. The different genera of Mamiellales, Pyramimonadales (*P* = *Pyramimonas*), clade VII, clade IX and Prasinococcales (Pras.).
C. The different species/infra-species of *Micromonas* (Micro.), *Ostreococcus* (Osteo.), *Mamiella*, *Crustomastix* (Crusto.), *Dolichomastix* (Dolicho.) and *Halosphaera* (Halo.). Samples (square) and variables (arrows) were plotted using axis I and II with the percentage of variability explained by each axis reported. Samples are labelled as follows: Station.Depth. Stations M and D correspond to MIO and Dyfamed respectively.

recovered preferentially in nutrient-depleted waters from the Ionian Sea. *Dolichomastix* clade B was present in warm surface waters (station MIO, 5 m), while clade A was recovered mainly from 50 m.

Discussion

Methodological aspects and overview of genetic diversity of the Chloroplastida

In order to preferentially amplify Chloroplastida, we used both a general eukaryotic primer and a specific primer, CHLO02r (5′-ctt cga gcc ccc aac ttt c-3′). The CHLO02 probe was initially designed for whole-cell hybridization (Simon *et al*., 2000), and later slightly modified at the 3′ end (addition of one base, CHLO02r) for Q-PCR applications (Zhu *et al*., 2005). Both *in silico* analysis (Table S1) and gene amplification of cultured strains (Zhu *et al*., 2005) confirmed the specificity of the CHLO02r primer. In the present work, the percentage of clones falling within the Chloroplastida was six times higher in the Chloroplastida-biased genetic libraries than in general eukaryotic libraries. However, more than 70% of the clones obtained were not from Chloroplastida despite optimization of PCR conditions. In fact, the CHLO02r primer has only a few mismatches to certain other non-target taxonomic groups (Table S1). Chlorarachniophyta (members of the Cercozoa) and Radiolaria, which are particularly abundant in the Chloroplastida-biased genetic libraries, displayed on average about one mismatch to CHLO02r (Table S1). Nevertheless, the 659 Chloroplastida clones reported here constitute the most extensive data set obtained for this group to date and include taxa never previously found in environmental samples.

Curiously, despite the use of a primer biased towards the Chloroplastida, we failed to recover sequences from the classes Trebouxiophyceae and Pedinophyceae. For the latter class, the absence of sequences from cultures in public databases makes it impossible to verify whether the CHLO02r primer actually targets these organisms, and conversely it cannot be excluded that part

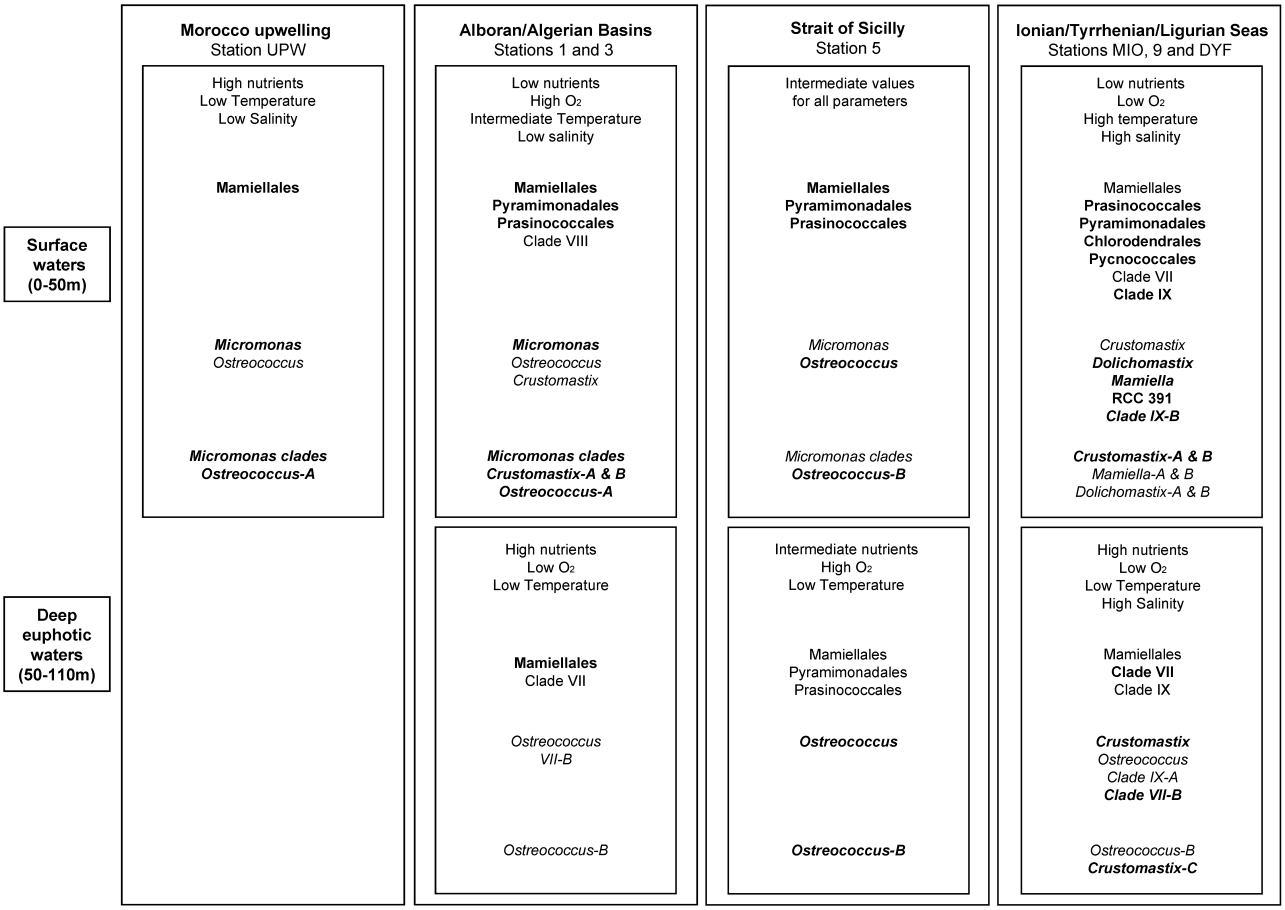


Fig. 7. Summary of the environmental factors influencing the distribution of Prasinophyceae during the PROSOPE cruise. Surface (0–50 m) and deep euphotic zone (50–100 m) samples were differentiated. For each water mass, the main prasinophyte taxa (at different taxonomic levels) recovered are given. For a given taxon, bold characters indicate its biogeographical preferences (in terms of relative abundance) according to the statistical analyses. Only taxa for which the distribution was resolved by the statistical analysis are included.

of the uncultured diversity observed here belongs to this class. The Pedinophyceae includes two free-living marine picoplanktonic species: *Marsupiomonas pelliculata*, isolated from a salt marsh in England (Jones *et al*., 1994), and *Resultor mikron* which is more ubiquitous (Thomsen and Buck, 1998). However, to our knowledge, these two species have never been observed in the Mediterranean Sea. For the Trebouxiophyceae, the situation is clearly different as the CHLO02r primer has been demonstrated to successfully amplify cultured strains (Zhu *et al*., 2005). A single sequence was recovered from more than 300 sequences from seasonal samples collected at a Mediterranean coastal site (Massana *et al*., 2004b) and the class was absent in wider Mediterranean sample sets analysed by denaturing gradient gel electrophoresis (DGGE), temporal temperature gradient gel electrophoresis (TTGE) or cloning (Díez *et al*., 2001; Marie *et al*., 2006). Surprisingly, numerous trebouxiophytes have been isolated into culture from the Mediterranean Sea during the MINOS and the PROSOPE oceanographic cruises (http://www.sb-roscoff.fr/Phyto/RCC/, Vaulot *et al*., 2004). For instance, strain RCC 475 (*Picochlorum*, Trebouxiophyceae) was isolated from a sample from 80 m in the Strait of Gibraltar, but no trebouxiophyte sequences were obtained from this sample. The same discrepancy between diversity from culture and genetic analysis was observed several times during the PROSOPE cruise. Strain RCC 391, which belongs to a new Mamiellales genus (W. Eikrem, pers. comm.), was isolated from the Morocco upwelling but not observed in the corresponding genetic library. Similarly, two Prasinococcales strains (RCC 896 and RCC 520) were isolated from deep euphotic waters in the Ionian and Tyrrhenian Seas, but no Prasinococcales sequences were recovered from these areas.

All of the prasinophyte clades recently reported by Guillou and colleagues (2004) were recovered, except for clade III corresponding to the Nephroselmidiaceae (Pseudoscourfieldiales). The Nephrosemidaceae and the Chlorodendrales are the only prasinophyte clades with no picoplankton sized representatives in culture (D. Vaulot, W. Eikrem, M. Viprey, and H. Moreau, submitted). In contrast to the Nephroselmidiaceae,

however, Chlorodendrales sequences have previously been obtained from picoplanktonic genetic libraries (Massana *et al*., 2004b). At lower taxonomic levels (i.e. genus and species/infra-species), all previously described prasinophyte lineages were detected in Mediterranean genetic libraries. In addition, a substantial number of sequences from undescribed lineages were obtained. Some of these, e.g. clades VII-A and B, had previously been recovered from the English Channel (Romari and Vaulot, 2004) or the Equatorial Pacific Ocean (Moon-van der Staay *et al*., 2001). Others were detected for the first time, such as the two novel environmental prasinophyte clades (VIII and IX), as well as many species/infra-species lineages within *Halosphaera* (Pyramimonadales), *Dolichomastix*, *Crustomastix* and *Mamiella* (Mamiellales). Therefore, despite prasinophytes being relatively well represented in culture, further isolation efforts and formal description of existing isolates are clearly required given the very wide diversity existing in the field.

At the 98.5% cut-off level (corresponding more or less to the genus level) the observed prasinophyte diversity was quite large and fairly well sampled. The libraries were obviously under-sampled at the lowest taxonomic levels (i.e. species/infra-species, corresponding to the 99.5% threshold), as indicated by the absence of saturation of the rarefaction curve and the low observed OTU richness compared with those estimated by the non-parametric Ace and Chao1 diversity estimators (Fig. S2). Seventy-four unique phylotypes out of more than 600 sequences (11%) were observed using a 99.5% sequence identity threshold, a lower proportion than that observed for picoplanktonic alveolates or stramenopiles (Massana *et al*., 2004a; Groisillier *et al*., 2006). This may reflect higher genetic heterogeneity in heterotrophic lineages compared with photosynthetic lineages resulting from the complexity of their roles in marine microbial food webs (predation, remineralisation, parasitism . . . Vaulot *et al*., 2002).

One of the main objectives of the present study was to investigate physico-chemical factors influencing biogeography of the Chloroplastida. Given the biases inherent to PCR-based methods such as differential extraction of DNA due to the nature of the cell wall or differential amplification (for a review see von Wintzingerode *et al*., 1997), it is now well established that the contribution of different taxonomic groups inferred from a random selection of clones may deviate significantly from the real composition of communities in the environment. At best, the diversity retrieved from genetic libraries reflects that of 18S rDNA molecules which can be influenced by the variability in rRNA operon copy number from one cell to another (Zhu *et al*., 2005). However, whole-genome sequencing of the picoplanktonic prasinophyte *Ostreococcus tauri* (Derelle *et al*., 2006) revealed that its different 18S rRNA gene copies were 100% identical (E. Derelle, pers. comm.), suggesting that intragenome variability does not bias diversity estimates, at least for picoplanktonic prasinophytes. The sampling approach used here does not allow distinction of whether the 18S rDNA sequences originate from free-living organisms, symbionts, parasites or even free DNA attached to particles (Corinaldesi *et al*., 2005). Furthermore, the biogeography of prasinophytes is probably not only explained by variations in the physico-chemical environment, but also by top-down regulation from predators and viruses. All of these points have to be kept in mind when considering biogeographic trends obtained from genetic libraries, and the hypotheses we put forward will need to be tested by quantitative approaches such as FISH (Not *et al*., 2002) or Q-PCR (Zhu *et al*., 2005).

Biogeography of prasinophytes

Mediterranean picoplanktonic Chloroplastida communities appear to be dominated by prasinophytes which represented 99% of Chloroplastida sequences and were present throughout the investigated transect. This confirms reports of the ubiquity of this photosynthetic group in various environments including coastal sites (e.g. Massana *et al*., 2004b), oligotrophic systems (Moon-van der Staay *et al*., 2001) and polar waters (Lovejoy *et al*., 2006). In this respect, marine waters appear quite different from fresh waters as the latter are dominated by Trebouxiophyceae and Chlorophyceae (Fawley *et al*., 2005).

The few quantitative studies performed to date indicate that prasinophytes are more abundant in coastal systems than in the open ocean (Thomsen and Buck, 1998; Not *et al*., 2005) and that in the open ocean they are preferentially found near the deep chlorophyll maximum (Countway and Caron, 2006; Marie *et al*., 2006). The present study confirms these trends as prasinophytes were mainly recovered from the Morocco upwelling, the Strait of Gibraltar and the Strait of Sicily and were better represented in the top 50 m than in deeper samples. This was linked particularly to the presence of Mamiellales, and notably of a limited number of genera including *Micromonas* and *Ostreococcus*. *Micromonas* is recognized as a cosmopolitan species, having been detected in oceanic systems such as the Gulf of Mexico and the western North Pacific (Furuya and Maruno, 1983; Cottrell and Suttle, 1991), as well as in more coastal environments including in the Mediterranean Sea (Zingone *et al*., 1999; Massana *et al*., 2004b). Quantitative measurements using molecular probes have shown that *Micromonas* is more abundant near the coast than in the open ocean (Not *et al*., 2004; 2005), suggesting that this genus is probably best adapted to nutrient-rich waters. This is confirmed here as *Micromonas* was particularly dominant in surface waters of the Morocco upwelling and at

western coastal Mediterranean sites influenced by low salinity, low temperature Atlantic waters. As also reported by Marie and colleagues (2006), no *Micromonas* environmental sequences were recovered from the highly oligotrophic waters of the Ionian Sea. However, *Micromonas* sequences formed an important part of the genetic library from surface waters of the Strait of Gibraltar that are relatively depleted in nutrients, suggesting that other parameters, such as low temperature, could also influence *Micromonas* distribution. The recent report of a psychrophilic *Micromonas* ecotype, probably endemic to the Arctic Basin (Lovejoy *et al*., 2007), also suggests a role for temperature.

Since the first isolation of *Ostreococcus* from a coastal Mediterranean Sea lagoon (Courties *et al*., 1994), this genus has been cultured from many oceanic regions (Rodríguez *et al*., 2005). However, its contribution to picoeukaryote populations appears more sporadic than that of *Micromonas*, with low concentration (100 cells ml^{-1}) observed all year round in the English Channel (Not *et al*., 2004), contrasting with very high concentrations (10^5 cells ml^{-1}) occurring during short bloom periods in North Pacific or North Atlantic coastal ecosystems (O'Kelly *et al*., 2003; Countway and Caron, 2006). Zhu and colleagues (2005) reported *Ostreococcus* concentrations off the Spanish Mediterranean coast around 50 cells ml^{-1} between October 2001 and March 2002, and close to zero during the rest of the year. At the end of summer, Marie and colleagues (2006), estimated by Q-PCR low *Ostreococcus* concentrations (2.10^2 cells ml^{-1} on average) for the PROSOPE cruise, except in the Strait of Sicily where it reached 6.10^3 cells ml^{-1}. In the present study, *Ostreococcus* sequences were recovered mainly from the Strait of Sicily. Of note, the distributions of *Micromonas* and *Ostreococcus* deduced here from their relative abundance in genetic libraries present similar trends to those estimated by Q-PCR on the same samples (Marie *et al*., 2006).

While *Micromonas* and *Ostreococcus* environmental sequences were mainly recovered from relatively mesotrophic waters, *Mamiella*, *Dolichomastix*, *Crustomastix* and RCC 391 showed maximal contributions in oligotrophic surface waters from the Tyrrhenian or Ligurian Sea where *Micromonas* and *Ostreococcus* contributions were lowest. Such biogeographical partitioning may correspond to ecological differences, such as higher competitiveness of *Micromonas* and *Ostreococcus* in temperate coastal systems preventing the development of other Mamiellales. Two other Mamiellales genera, *Bathycoccus* and *Mantoniella*, had widespread distributions, environmental sequences being recovered throughout the whole Mediterranean transect including westward coastal locations.

Although *Dolichomastix* has been recorded by microscopy from coastal waters worldwide (Manton, 1977), previous records of this genus in the Mediterranean Sea were mostly from the Strait of Sicily and the oligotrophic surface waters of the Tyrrhenian Sea (Throndsen and Zingone, 1997). Our data confirm this distribution pattern. *Crustomastix* has also been isolated into culture from the Tyrrhenian Sea (Zingone *et al*., 2002). In contrast, *Mamiella* is reported here for the first time in the Mediterranean Sea, extending the known distribution of this genus previously recorded from England to Thailand (Moestrup, 1984), but also from North Pacific and Tasmanian coastal surface waters (Thomsen and Buck, 1998; LeRoi and Hallegraeff, 2006).

Prasinococcales were recovered from the Alboran to the Ionian Seas and therefore seem, as observed for the Mamiellales, to be able to adapt to a large range of conditions. This confirms previous observations as Prasinococcales strains have been isolated from both coastal and open waters in the western North Atlantic (Sieburth *et al*., 1999), the western North Pacific (Miyashita *et al*., 1993), as well as in the Arabian Sea (Fuller *et al*., 2006). Nevertheless, Prasinococcales appear to be present much more sporadically than the Mamiellales. For instance, Prasinococcales have been reported to be particularly abundant (34% of the Chloroplastida) periodically during autumn at an English Channel coastal site dominated by Mamiellales throughout the year (Not *et al*., 2004).

High-performance liquid chromatography pigment analyses have shown that prasinoxantin-containing picoeukaryotes (i.e. Mamiellales, Prasinococcales and Pycnococcaceae) constitute the main Chloroplastida in some oceanic and coastal ecosystems (Suzuki *et al*., 2002; Rodríguez *et al*., 2003). However, prasinoxanthin has been reported to be low or absent despite high concentrations of chlorophyll *b* (e.g. Obayashi *et al*., 2001; Mangoni *et al*., 2004), suggesting that other Chloroplastida lacking prasinoxanthin such as the Pyramimonadales, Chlorodendrales, Nephroselmidiaceae or members of clade VII (Latasa *et al*., 2004) can be sporadically important (Rodriguez *et al*., 2002; Hill *et al*., 2005). Three of these groups lacking prasinoxanthin were recovered here (Fig. 2). The Pyramimonadales showed maximal contributions in the Alboran and Algerian Basins, while the Chlorodendrales and clade VII were mainly recovered eastward in the Ionian, Tyrrhenian and Ligurian Seas. These orders occurred preferentially above 50 m, except for clade VII that was mainly recovered from deeper euphotic waters. Interestingly, the Chlorodendrales sequences from this study were all recovered from nutrient-depleted waters and were phylogenetically related to other sequences from oligotrophic environments, including from symbionts of radiolarians from the Sargasso Sea (Gast *et al*., 2000) and a sequence recovered from Blanes waters (Spain) in summer (Massana *et al*., 2004b; Fig. 2).

Prasinophyte ecotypic differentiation

Genetic differentiation into ecotypes may help understanding the widespread occurrence of a given taxa. The unicellular cyanobacterium *Prochlorococcus* provides one of the best examples of this concept with both high-light and low-light adapted ecotypes (Moore *et al*., 1998). In the case of *Micromonas*, the different clades observed in Mediterranean genetic libraries (clades 1–4, according to Worden, 2006) seem to have similar distributions (Fig. 6C), while clade 5 was not detected. The coexistence of clades 1–4, and the insignificant contribution of clade 5 observed at two coastal Mediterranean sites using clade-specific probes detected by FISH (E. Foulon, F. Not, F. Jalabert, T. Cariou, and N. Simon, submitted) supports our data. However, the dominant clade may vary on a seasonal scale at a given site (Foulon *et al*., submitted), suggesting that temperature or nutrient levels could play a role in *Micromonas* clade distribution.

In contrast, *Ostreococcus* clades seem to have more clearly distinct ecological niches as the two *Ostreococcus* clades recovered (clades A and B), never co-occurred in the present study. Clade A was present only in the Alboran Sea and the Algerian Basin (surface sample), while clade B was more ubiquitous as it was recovered from the Algerian Basin (deep euphotic waters), but also eastward in the central Mediterranean Sea. It is noteworthy that we failed, as in previous studies, to obtain environmental sequences from *Ostreococcus* clades C and D. This is probably due to the fact that culture strains from these two clades were isolated from quite particular environments (i.e. a coastal lagoon and a harbour respectively). *Ostreococcus* clade A seems to prefer surface waters influenced by low temperature, low salinity Atlantic waters, whereas clade B was present throughout the photic zone including deep euphotic waters (95 m in the Algerian Basin and 110 m in the Ionian Sea; Fig. 6C). This concords with the capacity of clade B to adapt to low light intensities as reported by Rodríguez and colleagues (2005). Similarly, *Crustomastix* clade C was mainly recovered from the deep euphotic waters of the Ionian and Ligurian Seas, while clades A and B were present only in surface waters (Fig. 6C), suggesting a possible clade partitioning according to depth.

Potential ecotypic differentiation was also observed in other prasinophyte orders including clade VII and IX (Fig. 6B). Clade VII was represented here by sequences belonging to the lineages A and B, for which no described species are available. While lineages A and B have been reported from 75 and 120 m in the central South Pacific (Moon-van der Staay *et al*., 2001), only lineage A was recovered from surface waters in the English Channel (Romari and Vaulot, 2004), suggesting that lineage B could be restricted to deep euphotic waters. Finally, the two lineages composing clade IX showed similar horizontal distributions with maximal contributions in the Ionian, Tyrrhenian and Ligurian Seas, i.e. in off-shore nutrient-depleted waters, a trend supported by the observation of similar sequences in the South Pacific gyre (X. Shi, pers. comm.), one of the most oligotrophic areas of the world oceans. Clade B was mostly present in surface waters (5 and 15 m), while clade A was recovered deeper (50 and 65 m), suggesting that the latter clade could be adapted to lower light intensity. However, this distribution could also be explained by others parameters like temperature (10°C lower at depth) or trophic level as the deep euphotic waters benefit from slightly higher nutrient levels. It is interesting to note that all of these examples of potential niche partitioning occur in the vertical dimension, as previously observed for *Ostreococcus* (Rodríguez *et al*., 2005).

Conclusion

In the present study, the use of a Chloroplastida-biased primer revealed many clades never before detected by cloning approaches, including the discovery of two new orders of prasinophytes with no cultured representatives. Although previous studies using general primers were able to detect several dominant prasinophyte taxa (*Micromonas*, *Ostreococcus*, *Bathycoccus*), our data suggest that these approaches certainly miss a large number of taxa present at lower concentrations and which could constitute a rare biosphere as suggested before for prokaryotic (Pedrós Alió, 2006; Sogin *et al*., 2006) or eukaryotic (Countway *et al*., 2005) microbial communities. In the future, new molecular approaches specifically targeting photosynthetic organisms, such as analysis of the plastid 16S rRNA gene (Fuller *et al*., 2006) or the physical sorting of photosynthetic populations by flow cytometry prior to analysis of genetic diversity (X. Shi, pers. comm.), should contribute to better assessment of this rare picoeukaryotic biosphere. The present study also provides novel information concerning the distribution of prasinophytes, in particular for taxa for which little knowledge was previously available. For instance, our data suggest that many prasinophyte genera could display niche-partitioning, in particular in the vertical dimension, similar to that observed for *Prochlorococcus* or *Ostreococcus*. This illustrates the power of using targeted primers for clone library construction to develop novel biogeography hypotheses.

Experimental procedures

Sampling and nucleic acid extraction

Seawater samples were collected from several stations along two transects in the Mediterranean Sea in September and October 1999 during the PROSOPE cruise aboard the N.O.

Thalassa (Fig. 1). For analysis of the molecular diversity of picoeukaryotic communities, 1.45–5 l water samples were retrieved using 12 l Niskin bottles fitted on a Rosette sampler equipped with conductivity, temperature and depth sensors. Seawater was pre-screened through a 200 μm mesh filter followed by a 3 μm pore-size polycarbonate filter (diameter 47 mm, Nuclepore, Whatman International, Maidstone, UK) in order to separate picoplankton from larger organisms. Cells were collected by filtration onto 0.45 μm pore-size polycarbonate filter (diameter 47 mm, Pall Supor-450, Pall, Ann Arbor, MI). The 0.45 μm filters were transferred into cryovials, covered with 3.5 ml of lysis buffer (0. 75 M sucrose, 50 mM Tris-HCl, pH 8), and frozen in liquid nitrogen. DNA was extracted as previously described (Marie *et al*., 2006). Ancillary data (nutrients, dissolved oxygen, Chl *a*, salinity and temperature) are available from the PROSOPE web site: http://www.obs-vlfr.fr/cd_rom_dmtt/pr_main.htm

PCR amplification and 18S rDNA genetic libraries

For construction of genetic libraries, one to three representative depths were selected per station (Table 1) based on preliminary analysis of the diversity of the picoeukaryotic community by DGGE and TTGE on a wider sample set (Marie *et al*., 2006). In particular, we aimed to select depths with widely different band patterns. The eukaryotic 18S rRNA gene was amplified by PCR using two different sets of primers. The almost complete 18S rRNA gene (about 1800 bp) was amplified using the general eukaryotic primers Euk328f and Euk329r (Romari and Vaulot, 2004). The second set of primers combined the Euk328f primer and the reverse primer CHLO02r (Zhu *et al*., 2005) targeting the Chloroplastida, and yielded a fragment of about 950 bp. The specificity of the CHLO02r primer was checked *in silico* using the Match Probe function of ARB (Ludwig *et al*., 2004). The ARB database used contained about 30000 18S rRNA gene sequences downloaded from the Silva web site (http://www.arb-silva.de/, Pruesse *et al*., 2007). Only sequences more than 1400 bp in length and containing no N (undetermined base) in the primer region (beginning at position 764 of *Escherichia coli*) were considered (23900 sequences). The taxonomic affiliation of each sequence was based on ARB analysis. Taxonomic groups were sorted based on the average number of computed mismatches (Table S1). The PCR mixture (25 μl final volume) contained about 10 ng of environmental DNA as template, 1 μM final concentration of each primer and 12.5 μl of HotStarTaq Master Mix (Qiagen, Courtaboeuf, France), a premixed solution containing a HotStarTaq DNA Polymerase, the PCR Buffer, $MgCl_2$ (1.5 mM final concentration), and deoxynucleoside triphosphates (dNTPs, 200 μM final concentration each). Polymerase chain reactions were carried out in an automated thermocycler (iCycler, Bio-Rad, Marne la Coquette, France) with the following conditions: an initial incubation step at 95°C for 15 min for the activation of the HotStarTaq DNA Polymerase, followed by 34 cycles with a denaturing step at 95°C for 1 min, an annealing step at 57°C for 1.5 min, and an extension step at 72°C for 1.5 min. These cycles were followed by a final extension step at 72°C for 10 min. Amplified products from four individual PCR reactions were pooled, fragment length was checked by electrophoresis on a 1% agarose gel and the appropriate band was extracted from the gel and then purified using the QiaEx II Gel Extraction Kit from Qiagen (Qiagen, Courtaboeuf, France). Purified PCR products were ligated into the pCR®2.1-TOPO® vector system and transformed into competent cells (*E. coli*) following the protocol provided by the manufacturer (Invitrogen, Carlsbad, CA, USA).

Screening of genetic libraries and sequencing

A total of 7200 clones were recovered from agar plates. The presence of inserts was checked by PCR. Between 79 and 397 positive clones per genetic library were sequenced using an ABI Prism 3100 (Applied Biosystems) (Table 1). A single sequencing reaction using the primer Euk328f was first performed for 3980 clones, resulting in 100–1000 bp sequences. Sequence processing is summarized in Fig. S3. Briefly, sequences smaller than 400 bp were discarded. The 3486 remaining sequences were first subjected to BLAST search against publicly available sequences. Eight hundred and thirty-one sequences affiliated to Chloroplastida based on BLAST were selected. These sequences were clustered into distinct OTUs with Clusterer (Klepac-Ceraj *et al*., 2006) using a similarity threshold of 99.5%. The algorithm is based on the nearest neighbour approach and adds a sequence to an OTU if there is at least one sequence that is within the selected similarity threshold. For each OTU, one sequence per library (461 sequences in total) was selected and re-sequenced using different primers, i.e. Euk328f, Euk1055r (Atkins *et al*., 2000), M13f or M13r from the pCR®2.1-TOPO® vector kit (Invitrogen), yielding several partial overlapping sequences for each clone. Assembly of sequences for each clone was conducted as follows: (i) vector sequences were removed using the Phred/Phrap software (Ewing and Green, 1998), (ii) overlapping sequences from the same clone were manually assembled to construct a consensus sequence, and (iii) remaining single-stranded fragments were trimmed away from the consensus sequence. The resulting consensus sequences were then aligned using the slow and iterative refinement method FFT-NS-i of the Mafft 5.8 software (Katoh *et al*., 2005). The alignment was corrected by hand using the BioEdit sequence editor (Hall, 1999). Using the BioEdit editor for the visualization of the alignment, we deduced the secondary structures by hand using previously published studies (see for example Lange *et al*., 1996). Poorly aligned sequences were eliminated. Chimeras were checked by BLAST analysis of different parts of the sequences and confronting alternative phylogenetic trees using 300 bp pieces from the 5′ and 3′ ends and from the middle part of the gene. Out of the 461 consensus sequences (831 clones), 38 chimeras (172 clones) were identified and removed from the analysis. The acquisition of high-quality consensus sequences removed the artefactual genetic variability often generated by single sequence reads and therefore induced redundancy in the consensus sequence data set (i.e. two partial sequences from the initial 461 clone set that appeared slightly different could turn out to be identical once high-quality consensus sequences had been obtained). In order to remove this redundancy, further clustering was performed on the 423 remaining consensus sequences with a threshold of 99.5% using the Clusterer software, yielding a data set of 174

consensus sequences representative of 659 clones. Among these 174 sequences, 19 came from universal eukaryotic genetic libraries and 155 from the libraries obtained with the CHLO02r primer.

Clone library coverage and diversity estimation

Phylotype redundancy and species richness were estimated using the program EstimateS (Colwell, 2006). The 174 high-quality sequences were grouped into OTUs at two different sequence identity levels: 98.5% and 99.5% using Clusterer. The choice of these two thresholds was based on preliminary phylogenetic analyses of the partial length sequences, and corresponds more or less to the genus and the species/infra-species levels respectively. Although it may seem high, a 99.5% threshold allows distinguishing of ecotypes for certain taxa, for example for *Ostreococcus*. The resulting contingency tables (lines: clones; rows: OTUs) were used as input for EstimateS to generate rarefaction curves ($S_{obs.}$) and the non-parametric richness estimators, Chao1 (Chao, 1984) and Ace (Chao and Lee, 1992) for each sequence identity threshold. EstimateS settings were: randomization without replacement and no estimator bias correction.

Phylogenetic analyses

Phylogenetic analyses were conducted using three different alignments. The first one was a general alignment including 47 representative partial 18S rDNA sequences (910 bp) from the different genetic libraries and a selection of publicly available sequences belonging to the Chloroplastida. The two other alignments were composed of sequences belonging, respectively, to the orders Mamiellales and Pyramimonadales. Poorly aligned and very variable regions of the alignments were automatically removed with Gblocks (Castresana, 2000) using the following parameters: allowing gaps in half position, minimum length of a block equal to five for the general analysis, and block equal to two for the Mamiellales and Pyramimonadales data sets. Each alignment was analysed by ML using Paup 4.0b10 (Swofford, 2002). Different nested models of DNA substitution and associated parameters were estimated using Modeltest (Posada and Crandall, 1998). The model selected was the GTR + G + I for each data set. Settings given by Modeltest were used to perform ML analysis. A heuristic search procedure using the tree bisection/reconnection branch swapping algorithm was performed to find the optimal ML tree topology. When there was discrepancy between ML topology and bootstrap analyses in NJ and MP, ML topology was favoured as it is considered more robust (Tateno *et al*., 1994) and always agreed with previous work based on full-length analyses. In the present study, new clusters were defined using the following criteria: (i) a cluster should contain a least two sequences; and (ii) a cluster is supported by ML topology as well as by NJ and MP bootstrap values higher than 75%.

Statistical analyses

Principal component analysis (PCA) was used to group samples according to physico-chemical variables (Fig. S1). Prasinophyte taxa were classified according to the present phylogenies and the matrices of the relative contribution of taxa in each sample were obtained for three different taxonomic levels (order, genus and species/infra-species level). Each matrix was subjected to correspondence analysis (CA) in order to establish relationships between taxonomic assemblages and sampling sites. Assuming the coherence of the results between the two ordinations (PCA and CA), physico-chemical factors influencing biogeography of the taxonomic assemblages were inferred. All statistics were computed with the Ginkgo software (Bouxin, 2005).

Nucleotide sequence accession numbers

The sequences have been deposited in the GenBank database under accession numbers EU143373–EU143546.

Acknowledgements

We wish to thank H. Claustre for coordinating the PROSOPE cruise and L. Garczareck, F. Partensky and D. Marie for sampling during the cruise. We are grateful to M. Perennou and N. Gloaguen for help with sequencing, to E. Corre for performing automatic BLAST analysis, to D. Davoult for help with statistics and to I. Probert for his careful correction of the English language. The work was supported by the programs PROSOPE (CNRS INSUE) and PICOCEAN (GIS Génomique). M. Viprey benefited from a doctoral fellowship from the French Ministry of Education and Research.

References

Atkins, M.S., Teske, A.P., and Anderson, O.R. (2000) A survey of flagellate diversity at four deep-sea hydrothermal vents in the Eastern Pacific Ocean using structural and molecular approaches. *J Euk Microbiol* **47:** 400–411.

Bass, D., and Cavalier-Smith, T. (2004) Phylum-specific environmental DNA analysis reveals remarkably high global biodiversity of Cercozoa (Protozoa). *Int J Syst Evol Microbiol* **54:** 2393–2404.

Berman, T., Walline, P.D., Schneller, A., Rothenberg, J., and Townsend, D.W. (1985) Secchi disk depth record: a claim for the eastern Mediterranean. *Limnol Oceanogr* **30:** 447–448.

Bouxin, G. (2005) Ginkgo, a multivariate analysis package. *J Veg Sci* **16:** 353–359.

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

Chao, A. (1984) Non-parametric estimation of the number of classes in population. *Scand J Statist* **11:** 265–270.

Chao, A., and Lee, S.M. (1992) Estimating the number of classes via sample coverage. *J Am Statist Assoc* **87:** 210–217.

Colwell, R.K. (2006) *Estimates: statistical estimation of species richness and shared species from samples*, Version 8. [WWW document]. URL http://purl.oclc.org/estimates

Corinaldesi, C., Danovaro, R., and Dell'Anno, A. (2005) Simultaneous recovery of extracellular and intracellular

DNA suitable for molecular studies from marine sediments. *Appl Environ Microbiol* **71:** 46–50.

Cottrell, M.T., and Suttle, C.A. (1991) Wide-spread occurence and clonal variation in viruses which cause lysis of a cosmopolitan, eukaryotic marine phytoplankter, *Micromonas pusilla*. *Mar Ecol Prog Ser* **78:** 1–9.

Countway, P.D., and Caron, D.A. (2006) Abundance and distribution of *Ostreococcus* sp. in the San Pedro Channel, California, as revealed by quantitative PCR. *Appl Environ Microbiol* **72:** 2496–2506.

Countway, P.D., Gast, R.J., Savai, P., and Caron, D.A. (2005) Protistan diversity estimates based on 18S rDNA from seawater incubations in the Western North Atlantic. *J Euk Microbiol* **52:** 95–106.

Courties, C., Vaquer, A., Trousselier, M., Lautier, J., Chrétiennot-Dinet, M.-J., Neveux, J., *et al.* (1994) Smallest eukaryotic organism. *Nature* **370:** 255.

Dawson, S.C., and Pace, N.R. (2002) Novel kingdom-level eukaryotic diversity in anoxic environments. *Proc Natl Acad Sci USA* **99:** 8324–8329.

Derelle, E., Ferraz, C., Rombauts, S., Rouzé, P., Worden, A.Z., Robbens, S., *et al.* (2006) Genome analysis of the smallest free-living eukaryote *Ostreococcus tauri* unveils many unique features. *Proc Natl Acad Sci USA* **103:** 11647–11652.

Díez, B., Pedrós-Alió, C., and Massana, R. (2001) Study of genetic diversity of eukaryotic picoplankton in different oceanic regions by small-subunit rRNA gene cloning and sequencing. *Appl Environ Microbiol* **67:** 2932–2941.

Ewing, B., and Green, P. (1998) Base-calling of automated sequencer traces using Phred. II. Error probabilities. *Genome Res* **8:** 186–194.

Fawley, M.W., Yun, Y., and Qin, M. (2000) Phylogenetic analyses of 18S rDNA sequences reveal a new coccoid lineage of the *Prasinophyceae* (*Chlorophyta*). *J Phycol* **36:** 387–393.

Fawley, M.W., Fawley, K.P., and Owen, H.A. (2005) Diversity and ecology of small coccoid green algae from Lake Itasca, Minnesota, USA, including *Meyerella planktonica*, gen. et sp nov. *Phycologia* **44:** 35–48.

Fuller, N.J., Campbell, C., Allen, D.J., Pitt, F.D., Zwirglmaier, K., Le Gall, F., *et al.* (2006) Analysis of photosynthetic picoeukaryote diversity at open ocean sites in the Arabian Sea using a PCR biased towards marine algal plastids. *Aquat Microb Ecol* **43:** 79–93.

Furuya, K., and Maruno, R. (1983) The structure of the phytoplankton community in the subsurface chlorophyll maxima in the western North Pacific Ocean. *J Plankton Res* **5:** 393–406.

Gascard, J.C., and Richez, C. (1985) Water masses and circulation in the Western Alboran Sea and in the Straits of Gibraltar. *Prog Oceanog* **15:** 175–216.

Gast, R.J., McDonnell, T.A., and Caron, D.A. (2000) srDNA-based taxonomic affinities of algal symbionts from a planktonic foraminifer and a solitary radiolarian. *J Phycol* **36:** 172–177.

Groisillier, A., Massana, R., Valentin, K., Vaulot, D., and Guillou, L. (2006) Genetic diversity and habitats of two enigmatic marine alveolate lineages. *Aquat Microb Ecol* **42:** 277–291.

Guillou, L., Eikrem, W., Chrétiennot-Dinet, M.J., Le Gall, F., Massana, R., Romari, K., *et al.* (2004) Diversity of picoplanktonic prasinophytes assessed by direct SSU rDNA sequencing of environmental samples and novel isolates retrieved from oceanic and coastal marine ecosystems. *Protist* **155:** 193–214.

Hall, T.A. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Ser* **41:** 95–98.

Hill, V., Cota, G., and Stockwell, D. (2005) Spring and summer phytoplankton communities in the Chukchi and Eastern Beaufort Seas. *Deep-Sea Res Part II* **52:** 3369–3385.

Johnson, P.W., and Sieburth, J.M. (1982) *In-situ* morphology and occurrence of eukaryotic phototrophs of bacterial size in the picoplankton of estuarine and oceanic waters. *J Phycol* **18:** 318–327.

Jones, H.L.J., Leadbeater, B.S.C., and Green, J.C. (1994) An ultrastructural study of *Marsupiomonas pelliculata* gen. et sp. nov., a new member of the *Pedinophyceae*. *Eur J Phycol* **29:** 171–181.

Katoh, K., Kuma, K., Toh, H., and Miyata, T. (2005) MAFFT version 5: improvement in accuracy of multiple sequence alignment. *Nucleic Acids Res* **33:** 511–518.

Klepac-Ceraj, V., Ceraj, I., and Polz, M.F. (2006) Clusterer: extendable java application for sequence grouping and cluster analyses. *Online J Bioinfo* **7:** 15–21.

Lange, M., Guillou, L., Vaulot, D., Simon, N., Amann, R.I., Ludwig, W., and Medlin, L.K. (1996) Identification of the class *Prymnesiophyceae* and the genus *Phaeocystis* with ribosomal RNA-targeted nucleic acid probes detected by flow cytometry. *J Phycol* **32:** 858–868.

Latasa, M., Scharek, R., Le Gall, F., and Guillou, L. (2004) Pigment suites and taxonomic groups in *Prasinophyceae*. *J Phycol* **40:** 1149–1155.

LeRoi, J.-M., and Hallegraeff, G. (2006) Scale-bearing nanoflagellates from southern Tasmanian coastal waters, Australia. II. Species of *Chrysophyceae* (*Chrysophyta*), *Prymnesiophyceae* (*Haptophyta*, excluding *Chrysochromulina*) and *Prasinophyceae* (*Chlorophyta*). *Bot Marina* **49:** 216–235.

Lovejoy, C., Massana, R., and Pedrós Alió, C. (2006) Diversity and distribution of marine microbial eukaryotes in the Artic Ocean and adjacent Seas. *Appl Environ Microbiol* **72:** 3085–3095.

Lovejoy, C., Vincent, W.F., Bonilla, S., Roy, S., Martineau, M.J., Terrado, R., *et al.* (2007) Distribution, phylogeny, and growth of cold-adapted picoprasinophytes in Artic seas. *J Phycol* **43:** 78–89.

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar, *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32:** 1363–1371.

Manca, B., Burca, M., Giorgetti, C., Coatanoan, C., Garcia, M.J., and Iona, A. (2004) Physical and biochemical averaged vertical profiles in the Mediterranean regions: an important tool to trace the climatology of water masses and to validate incoming data from operational oceanography. *J Mar Syst* **48:** 83–116.

Mangoni, O., Modigh, M., Conversano, F., Carrada, G.C., and Saggiomo, V. (2004) Effect of summer ice coverage on phytoplankton assemblages in the Ross Sea, Antarctica. *Deep-Sea Res Part I* **51:** 1601–1617.

Manton, I. (1977) *Dolichomastix* (Prasinophyceae) from Artic Canada, Alaska and South Africa: a new genus of flagellates with scally flagella. *Phycologia* **16:** 427–438.

Marie, D., Zhu, F., Balagué, V., Ras, J., and Vaulot, D. (2006) Eukaryotic picoplankton communities of the Mediterranean Sea in summer assessed by molecular approaches (DGGE, TTGE, QPCR). *FEMS Microbiol Ecol* **55:** 403–415.

Massana, R., Castresana, J., Balagué, V., Guillou, L., Romari, K., Groisillier, A., *et al.* (2004a) Phylogenetic and ecological analysis of novel marine stramenopiles. *Appl Environ Microbiol* **70:** 3528–3534.

Massana, R., Balagué, V., Guillou, L., and Pedrós Alió, C. (2004b) Picoeukaryotic diversity in an oligotrophic coastal site studied by molecular and culturing approaches. *FEMS Microbiol Ecol* **50:** 231–243.

Miyashita, H., Ikemoto, H., Kurano, N., Miyachi, S., and Chihara, M. (1993) *Prasinococcus capsulatus* gen. & sp. nov., a new marine coccoid prasinophyte. *J Gen Appl Microbiol* **39:** 571–582.

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.

Moon-van der Staay, S.Y., De Wachter, R., and Vaulot, D. (2001) Oceanic 18S rDNA sequences from picoplankton reveal unsuspected eukaryotic diversity. *Nature* **409:** 607–610.

Moore, L.R., Rocap, G., and Chisholm, S.W. (1998) Physiology and molecular phylogeny of coexisting *Prochlorococcus* ecotypes. *Nature* **393:** 464–467.

Moro, I., La Rocca, N., Dalla Valle, L., Moschin, E., Negrisolo, E., and Andreoli, C. (2002) *Pyramimonas australis* sp. nov. (*Prasinophyceae*, *Chlorophyta*) from Antarctica: fine structure and molecular phylogeny. *Eur J Phycol* **37:** 103–114.

Moutin, T., and Raimbault, P. (2002) Primary production, carbon export and nutrients availability in western and eastern Mediterranean Sea in early summer 1996 (MINOS cruise). *J Mar Syst* **33–34:** 273–288.

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

Not, F., Simon, N., Biegala, I.C., and Vaulot, D. (2002) Application of fluorescent *in situ* hybridization coupled with tyramide signal amplification (FISH-TSA) to assess eukaryotic picoplankton composition. *Aquat Microb Ecol* **28:** 157–166.

Not, F., Latasa, M., Marie, D., Cariou, T., Vaulot, D., and Simon, N. (2004) A single species *Micromonas pusilla* (*Prasinophyceae*), dominates the eukaryotic picoplankton in the Western English Channel. *Appl Environ Microbiol* **70:** 4064–4072.

Not, F., Massana, R., Latasa, M., Marie, D., Colson, C., Eikrem, W., *et al.* (2005) Late summer community composition and abundance of photosynthetic picoeukaryotes in Norwegian and Barents Seas. *Limnol Oceanogr* **50:** 1677–1686.

Not, F., Gausling, R., Azam, F., Heidelberg, J.F., and Worden, A.Z. (2007) Vertical distribution of picoeukaryotic diversity in the Sargasso Sea. *Environ Microbiol* **9:** 1233–1252.

O'Kelly, C.J., Sieracki, M.E., Thier, E.C., and Hobson, I.C. (2003) A transient bloom of *Ostreococcus* (*Chlorophyta*, *Prasinophyceae*) in west Neck Bay, Long Island, New York. *J Phycol* **39:** 850–854.

Obayashi, Y., Tanoue, E., Suzuki, K., Handa, N., Nojiri, Y., and Wong, C.S. (2001) Spatial and temporal variabilities of phytoplankton community structure in the northern North Pacific as determined by phytoplankton pigments. *Deep-Sea Res Part I* **48:** 439–469.

Partensky, F., Hess, W.R., and Vaulot, D. (1999) *Prochlorococcus*, a marine photosynthetic prokaryote of global significance. *Microbiol Mol Biol Rev* **63:** 106–127.

Pedrós Alió, C. (2006) Marine microbial diversity: can it be determined? *Trends Microbiol* **14:** 257–263.

Posada, D., and Crandall, K.A. (1998) Modeltest: testing the model of DNA substitution. *Bioinformatics* **14:** 817–818.

Pruesse, E., Quanst, C., Knittel, K., Fuchs, B., Ludwig, W., Peplies, J., and Glôckner, F.O. (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequences data compatible with ARB. *Nucleic Acids Res* **35:** 7188–7196.

Rodríguez, F., Varela, M., and Zapata, M. (2002) Phytoplankton assemblages in the Gerlache and Bransfield Straits (Antarctic Peninsula) determined by light microscopy and CHEMTAX analysis of HPLC pigment data. *Deep-Sea Res Part II* **49:** 723–747.

Rodríguez, F., Pazos, Y., Maneiro, J., and Zapata, M. (2003) Temporal variation in phytoplankton assemblages and pigment composition at a fixed station of the Ria of Pontevedra (NW Spain). *Est Coast Shelf Sci* **58:** 499–515.

Rodríguez, F., Derelle, E., Guillou, L., Le Gall, F., Vaulot, D., and Moreau, H. (2005) Ecotype diversity in the marine picoeukaryote *Ostreococcus* (*Chlorophyta*, *Prasinophyceae*). *Environ Microbiol* **7:** 853–859.

Romari, K., and Vaulot, D. (2004) Composition and temporal variability of picoeukaryote communities at a coastal site of the English Channel from 18S rDNA sequences. *Limnol Oceanogr* **49:** 784–798.

Sieburth, J.M., Keller, M.D., Johnson, P.W., and 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.

Simon, N., Campbell, L., Ornolfsdottir, E., Groben, R., Guillou, L., Lange, M., and Medlin, L.K. (2000) Oligonucleotide probes for the identification of three algal groups by dot blot and fluorescent whole-cell hybridization. *J Euk Microbiol* **47:** 76–84.

Slapeta, J., Lopez-Garcia, P., and Moreira, D. (2006) Global dispersal and ancient cryptic species in the smallest marine eukaryotes. *Mol Biol Evol* **23:** 23–29.

Sogin, M.L., Morrison, H.G., Huber, J.A., Welch, D.M., Huse, S.M., Neal, P.R., *et al.* (2006) Microbial diversity in the deep sea and the underexplored 'rare biosphere'. *Proc Natl Acad Sci USA* **103:** 12115–12120.

Stoeck, T., Hayward, B., Taylor, G.T., Varela, R., and Epstein, S.S. (2006) A Multiple PCR-primer approach to access the

microeukaryotic diversity in environmental samples. *Protist* **157:** 31–43.
Suda, S. (2004) Taxonomic characterization of *Pyramimonas aurea* sp. nov. (*Prasinophyceae*, *Chlorophyta*). *Phycologia* **43:** 682–692.
Suzuki, K., Minami, C., Liu, H., and Saino, T. (2002) Temporal and spatial patterns of chemotaxonomic algal pigments in the subarctic Pacific and the Bering Sea during the early summer of 1999. *Deep-Sea Res Part II* **49:** 5685–5704.
Swofford, D.L. (2002) *PAUP*. Phylogenetic Analysis Using Parsimony (*and Others Methods)*. Sunderland, MA, USA: Sinauer Associates.
Tateno, Y., Takezaki, N., and Nei, M. (1994) Relative efficiencies of the Maximum-Likelihood, Neighbor-Joining, and Maximum-Parsimony methods when substitution rate varies with site. *Mol Biol Evol* **11:** 261–277.
Thomsen, H.A., and Buck, K.R. (1998) Nanoflagellates of the central California waters: taxonomy, biogeography and abundance of primitive, green flagellates (*Pedinophyceae*, *Prasinophyceae*). *Deep-Sea Res Part II* **45:** 1687–1707.
Throndsen, J., and Zingone, A. (1997) *Dolichomastix tenuilepsis* sp. nov., a first insight into the microanatomy of the genus *Dolichomastix* (*Mamiellales*, *Prasinophyceae*, *Chlorophyta*). *Phycologia* **36:** 244–254.
Vaulot, D., Romari, K., and Not, F. (2002) Are autotrophs less diverse than heterotrophs in marine picoplankton? *Trends Microbiol* **10:** 266–267.
Vaulot, D., Le Gall, F., Marie, D., Guillou, L., and Partensky, F. (2004) The Roscoff Culture Collection (RCC): a collection dedicated to marine picoplankton. *Nova Hedwigia* **79:** 49–70.
Venter, J.C., Remington, K., Heidelberg, J., Halpern, A.L., Rusch, D., Eisen, J.A., *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304:** 66–74.
von Wintzingerode, F., Göbel, U.B., and Stackebrandt, E. (1997) Determination of microbial diversity in environmental samples: pitfalls of PCR-based rRNA analysis. *FEMS Microbiol Rev* **21:** 213–229.
Worden, A.Z. (2006) Picoeukaryote diversity in coastal waters of the Pacific Ocean. *Aquat Microb Ecol* **43:** 165–175.
Worden, A.Z., Nolan, J.K., and Palenik, B. (2004) Assessing the dynamics and ecology of marine picophytoplankton: the importance of the eukaryotic component. *Limnol Oceanogr* **49:** 168–179.
Zhu, F., Massana, R., Not, F., Marie, D., and Vaulot, D. (2005) Mapping of the picoeucaryotes in marine ecosystems with quantitative PCR of 18S rRNA gene. *FEMS Microbiol Ecol* **52:** 79–92.
Zingone, A., Sarno, D., and Forlani, G. (1999) Seasonal dynamics in the abundance of *Micromonas pusilla* (*Prasinophyceae*) and its viruses in the Gulf of Naples (Mediterranean Sea). *J Plankton Res* **21:** 2143–2159.
Zingone, A., Borra, M., Brunet, C., Forlani, G., Kooistra, W.H. C.F., and Procaccini, G. (2002) Phylogenetic position of *Crustomastix stigmatica* sp. nov. and *Dolichomastix tenuilepis* in relation to the *Mamiellales* (*Prasinophyceae*, *Chlorophyta*). *J Phycol* **38:** 1024–1039.
Zubkov, M.V., Sleigh, M.A., Tarran, G.A., Burkill, P.H., and Leakey, R.J.G. (1998) Picoplanktonic community structure on an Atlantic transect from 50 degrees N to 50 degrees S. *Deep-Sea Res Pt I* **45:** 1339–1355.

Supplementary material

The following supplementary material is available for this article online:

Table S1. Number of sequences that display between zero and six mismatch with the ChlO02r primer for each of the major taxonomic groups represented in the ARB 18S rDNA database. Only sequences longer than 1400 bp and without undetermined position (N) have been taken into account. 'Others' taxa include: Heterolobosea, Ichthyosporea, Haplosporida, Oxymonadida, Nucleariidae, Ancyromonas, Giardia, Haemosporida, Litostomatea, Microsporidia, Parabasilidea, Spumellaria and unknown taxa. Only four groups (Streptophyta, Chlorophyta, Cercozoa and Radiolaria) display on average less than two mismatches to the primer.

Fig. S1. Principal component analysis (PCA) of environmental parameters. The percentage of variability explained by each axis is indicated. Samples (square) and variables (arrows) are plotted against the first two axes. Samples are grouped according to their environmental characteristics. Samples are labelled as follows: Station.Depth. Stations M and D correspond to MIO and Dyfamed respectively. The sample collected at 30 m in the Morocco upwelling, characterized by very high nutrient concentrations and cold waters, was removed from the analysis as its influence would have grouped all the other samples together. Components I, II and III (component III, not shown) represent, respectively, 40%, 26% and 14.5% of the total variance and are mainly explained by nutrient levels (nitrate, nitrite, and silicate) and temperature, dissolved oxygen concentration and salinity, phosphate concentration respectively.

Fig. S2. Rarefaction curves for pooled Chloroplastida sequence data for the Ace diversity estimator (circles) and the number of observed species, S_{obs} (triangles) ($\pm$95% confidence interval). See text for details. Curves were constructed for two OTU sequence similarity cut off levels: 98.5% (black symbols) and 99.5% (white symbols).

Fig. S3. Schematics for sequence acquisition and analysis. SR and CS correspond to single read and consensus sequences respectively. Gr. 99.5% (or 98.5%) means of 'grouping using 99.5% (or 98.5%) identity'.

This material is available as part of the online article from http://www.blackwell-synergy.com

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

