

# Use of flow cytometric sorting to better assess the diversity of small photosynthetic eukaryotes in the English Channel

Dominique Marie, Xiao Li Shi, Fabienne Rigaut-Jalabert & Daniel Vaultot

Université Pierre et Marie Curie (Paris 06) et CNRS, UMR 7144, Station Biologique, Roscoff, France

**Correspondence:** Dominique Marie, Université Pierre et Marie Curie (Paris 06) et CNRS, UMR 7144, Station Biologique de Roscoff, Place G. Teissier, BP 74, 29682 Roscoff cedex, France. Tel.: +33 2 98 29 23 23; fax: +33 2 98 29 23 24; e-mail: marie@sb-roscoff.fr

Received 23 October 2009; revised 5 January 2010; accepted 6 January 2010.  
Final version published online 8 March 2010.

DOI:10.1111/j.1574-6941.2010.00842.x

Editor: Riks Laanbroek

## Keywords

flow cytometry; diversity; photosynthetic eukaryotes.

## Abstract

Small photosynthetic eukaryotes are key primary producers in marine waters. In recent years, their diversity has been studied by the analysis of 18S rRNA gene sequences directly amplified and cloned from filtered natural samples. However, these clone libraries are often dominated by nonphotosynthetic organisms and few sequences from autotrophs are recovered. In the present paper, we developed a new approach based on flow cytometry. Photosynthetic pico-, nano- and phycoerythrin-containing (PE-) eukaryotes from the coastal English Channel were sorted based on their size and pigment fluorescence. 18S rRNA gene libraries were constructed from the DNA of sorted cells. We addressed methodological issues linked to the relatively low concentration of these cells. This novel approach confirmed that, in the English Channel, pico-eukaryotes are dominated by three genera *Micromonas*, *Ostreococcus* and *Bathycoccus*, while PE-eukaryotes are mainly cryptophytes from clade 4. It also revealed that nano-eukaryotes are dominated by haptophytes with important contributions from small diatoms and Prasinophyceae. It should be emphasized that haptophytes were nearly absent from clone libraries constructed from filtered samples, which explains why they have been overlooked in previous studies. The new strategy should be very useful to conduct similar studies on other specific populations that can be discriminated by flow cytometry (e.g. red tide organisms or uncultivated protists).

## Introduction

In recent years, there has been a growing recognition of the importance of small-sized eukaryotes in both marine and freshwater environments. The abundance of pico-eukaryotes (defined as cells smaller than 2–3 µm) ranges from 1000 to 5000 cells mL<sup>-1</sup> in oceanic and coastal waters, respectively. Despite these relatively low concentrations compared with photosynthetic prokaryotes such as *Prochlorococcus* or *Synechococcus* (Partensky *et al.*, 1999), recent studies have pointed out that marine photosynthetic pico-eukaryotes are major primary producers (Worden & Not, 2008). Nano-eukaryotes (2–20 µm) are in general 10 times less abundant than pico-eukaryotes in the plankton (e.g. Moran, 2007), but many nanoplanktonic species such as *Emiliania huxleyi* can sporadically form dense blooms. Species of small marine photosynthetic eukaryotes belong to a variety of algal classes (Vaultot *et al.*, 2008), in particular within green (Prasinophyceae) and brown (Stramenopiles) algae for picoplankton, and within Haptophyta and Dinophyceae (dinoflagellates) for nanoplankton (Vaultot *et al.*, 2008).

Several strategies have been applied to analyze the diversity of small photosynthetic eukaryotes in marine systems. The culturing approach led to the isolation and description of many important species (Vaultot *et al.*, 2008), but is limited by medium selectivity and poor knowledge of actual growth requirements (Le Gall *et al.*, 2008). Molecular techniques, especially the cloning of the 18S rRNA gene from natural populations, have established that small-eukaryote assemblages are highly diverse in marine systems (Vaultot *et al.*, 2008).

One major limitation of current 18S rRNA gene studies is that environmental libraries generated from filtered samples with universal PCR primers are dominated by heterotrophic organisms (Vaultot *et al.*, 2002). For example, 18S clone libraries constructed from samples off Roscoff (Brittany, France) contain only few representative sequences from photosynthetic lineages such as the Prasinophyceae, with most sequences belonging to nonphotosynthetic lineages such as marine stramenopiles (MAST, see Massana *et al.*, 2004), Syndiniales (alveolate groups I and II, Guillou *et al.*, 2008), ciliates, Cercozoa, or choanoflagellates (Romari & Vaultot, 2004). This is surprising since microscopy counts

with probes detected by FISH suggest that in these waters an average on an annual basis of 85% of eukaryotic cells belong to the Chlorophyta with one Prasinophyceae sp., *Micromonas pusilla*, being dominant most of the year (Not et al., 2004).

Several approaches have recently been applied to target more specifically the photosynthetic fraction of small-eukaryote assemblages. For example, clone libraries have been constructed using 18S rRNA gene primers biased toward photosynthetic groups such as Chlorophyta (Viprey et al., 2008) or primers targeting the 16S rRNA plastid gene, only present in photosynthetic organisms (Fuller et al., 2006). These approaches have revealed that some important groups, such as Chrysophyceae (Fuller et al., 2006) had escaped detection. Even within classes thought to be well known, such as the Prasinophyceae, major novel clades have been detected (Viprey et al., 2008). However, these two approaches are somewhat biased because they rely on specific primers that can only be designed based on available sequences.

Flow cytometry has been used extensively in the last 25 years to characterize both auto- and heterotrophic plankton populations in natural samples. It allows discriminating and enumerating specific groups of cells (e.g. cyanobacteria) based on their size and fluorescence characteristics either due to natural cell pigments or staining, for example of DNA (Veldhuis & Kraay, 2000; Czechowska et al., 2008). However, the capacity of flow cytometry to sort specific subpopulations, which allows characterizing them subsequently using molecular techniques, has been little used in marine systems, and to the best of our knowledge only for prokaryotes (Wallner et al., 1997; Fuchs et al., 2005).

In the present paper, we develop a novel approach to assess the diversity of small photosynthetic eukaryotes by sorting specific populations of photosynthetic cells by flow cytometry, extracting DNA from the sorted cells, and constructing clone libraries with universal eukaryotic primers of the 18S rRNA gene. This approach was tested on samples from the English Channel collected during each season of the year and proved to be very efficient in minimizing the contribution of heterotrophic groups.

## Materials and methods

### Sample collection and preparation

Surface seawater samples were collected using a Niskin bottle, once each season (Table 1), at the SOMLIT-Astan site (48.46°N, 3.56°W) off Roscoff (Brittany, France). An extra sample was taken on August 30, 2007 to conduct tests. In order to sort enough cells in a reasonable period of time and better visualize the different cell populations, 3 L of seawater were concentrated to a volume of 30 mL by tangential flow filtration using a Vivaflow 200 cartridge equipped with a 100 000 MWCO RC membrane (Sartorius Biotechnologie SAS, France). Concentration factors varied between 50 and 80 times (Table 1). Cell recovery was lowest for *Synechococcus* cyanobacteria (49% recovery, mean value), and increased for larger cells (57%, 74% and 82% for pico-, nano- and PE-eukaryotes, respectively). Five milliliters of the concentrated sample were reserved for flow cytometry and prefiltered through a 50- $\mu$ m nylon mesh in order to avoid clogging of the flow cell. The remaining volume (not prefiltered and therefore possibly containing microplanktonic cells) was filtered using a syringe through a 0.22- $\mu$ m pore size Sterivex unit (Millipore, Billerica, MA). Seawater was completely removed from the Sterivex unit that was then filled with 2 mL of lysis buffer (0.75 M sucrose, 50 mM Tris-HCl, 40 mM EDTA), and stored at  $-80^{\circ}\text{C}$  until extraction.

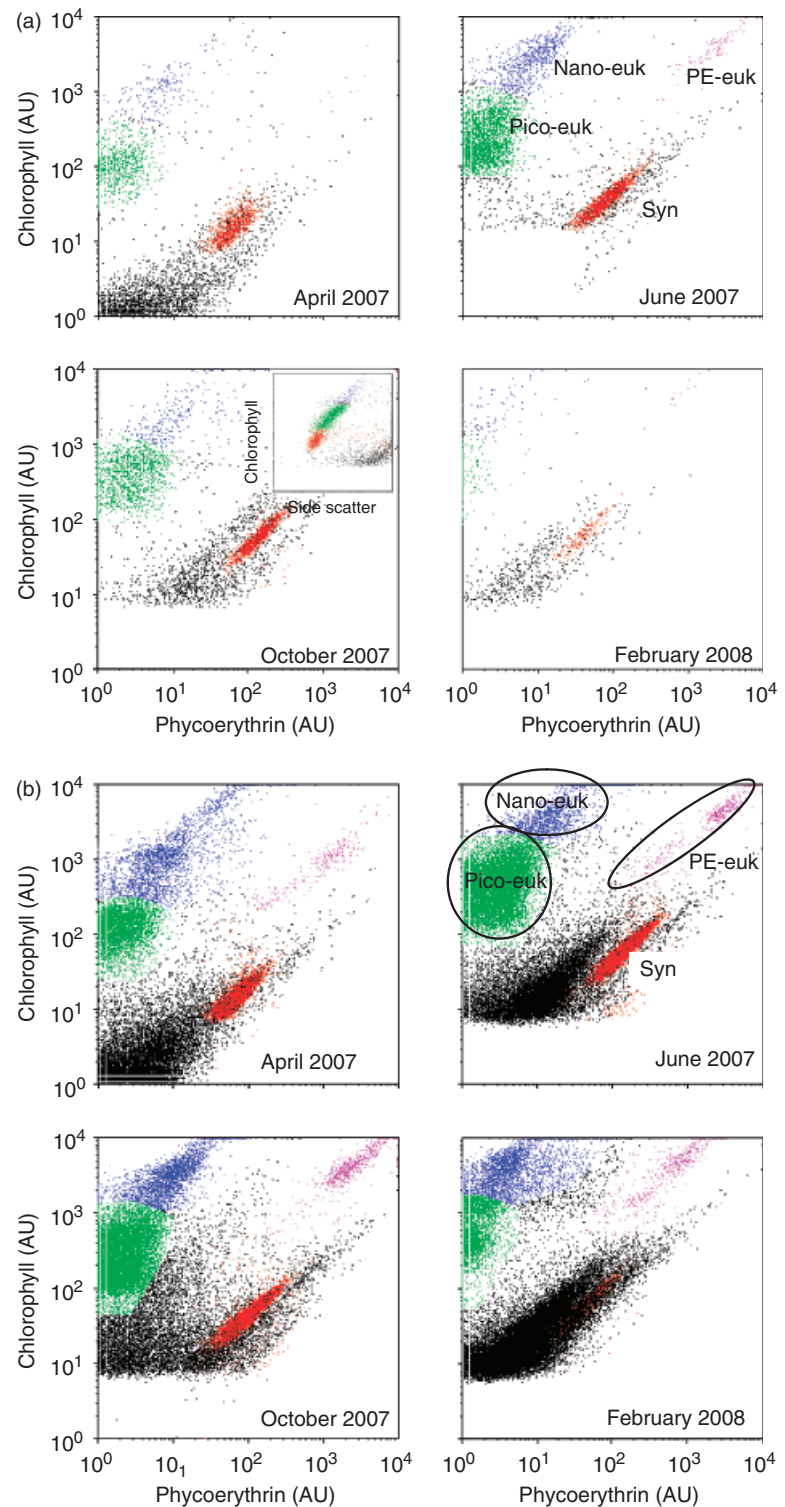
### Flow cytometry

Samples were analyzed as described previously (Marie et al., 2000) using a FACSAria flow cytometer (Becton Dickinson, San Jose, CA) equipped with a laser emitting at 488 nm and a 70- $\mu$ m nozzle. Emitted light was collected through the following set of filters: 488/10 band pass (BP) for side scatter, 576/26 BP for orange fluorescence, and 655 long pass for red fluorescence. Signal detection was triggered on the chlorophyll fluorescence. Samples were run for 2 min at a flow rate of 40  $\mu\text{L min}^{-1}$  to estimate cell abundance. Three

**Table 1.** Sampling dates, population abundance, percentages of recovery after tangential flow filtration, number of cells sorted by flow cytometry, and PCR primers used (see Table 2)

Date	<i>Synechococcus</i>		Pico-euk			Nano-euk			PE-euk			PCR primers
	Cells $\text{mL}^{-1}$	Recovery %	Cells $\text{mL}^{-1}$	Recovery %	Sorted	Cells $\text{mL}^{-1}$	Recovery %	Sorted	Cells $\text{mL}^{-1}$	Recovery %	Sorted	
April 11, 2007	5447	70	10 889	72	200 000	1064	113	50 000	55	70	5000	Euk328f/Euk329r
June 25, 2007	4410	44	18 677	40	205 000	1758	54	50 000	249	70	15 000	Euk328f/Euk329r and nested PCR Euk1Af/1492rE
October 04, 2007	3795	47	7924	60	250 000	516	60	50 000	60	111	15 000	373Cf/Euk329r
February 15, 2008	1582	35	3847	55	250 000	791	71	50 000	84	75	5500	373Cf/Euk329r

Note that some percentages of recovery are higher than 100, due to the difficulty in clearly finding the limits of the populations on nonconcentrated samples.



**Fig. 1.** Flow cytometric distributions obtained for Astan samples taken at four different seasons, before (a) and after (b) concentration by tangential filtration. Cytograms correspond to phycoerythrin orange fluorescence vs. chlorophyll red fluorescence, both in arbitrary units. The inset in (a) corresponds to the side scatter (a proxy of cell size) vs. chlorophyll fluorescence cytogram. Colored dots correspond to the limits of the different populations based on multiparameter gating: pico-eukaryotes (Pico-euk), nano-eukaryotes (Nano-euk), PE-eukaryotes (PE-euk), and *Synechococcus* (Syn) cyanobacteria. Ellipses emphasize the three eukaryote populations that were sorted from concentrated samples.

populations of pico-eukaryotes (Pico-euk), nano-eukaryotes (Nano-euk) and phycoerythrin-containing eukaryotes (PE-euk) were selected for sorting based on light scatter, orange phycoerythrin fluorescence, and red chlorophyll

fluorescence (Fig. 1). Between 5000 and 250 000 eukaryotic cells were sorted into Eppendorf tubes containing 180  $\mu$ L of lysis buffer (Tris-HCl, pH8; EDTA-Na2 2 mM; Triton X-100, 1.2%). Sorting was performed in high-purity sorting mode

at a frequency of 90 000 Hz and with a deflection voltage of 6000 V. For sorting, the flow rate was adjusted such that the analysis rate remained below 12 000 events per second.

In order to estimate the number of cells that have to be sorted for constructing 18S rRNA gene clone libraries and to determine whether the mode of cell collection influences DNA recovery, series of 10–200 000 pico-eukaryotes and 10–50 000 nano-eukaryotes from the Astan sample taken in August 2007 were sorted either in Eppendorf tubes containing 180 µL of lysis buffer or onto 0.2-µm Supor membrane filters (Pall Life Sciences, France). In the latter case, following sorting, filters were placed into cryovials containing 1 mL of lysis buffer. All samples were stored at –80 °C until extraction.

### DNA extraction

Samples were thawed at room temperature. Either 20 or 100 µL of lysozyme (20 mg mL<sup>-1</sup>) was added to populations sorted into Eppendorf tubes or onto filters, respectively. Incubation at 37 °C was performed for 30 min. One hundred and eighty microliters and 1 mL of lysis buffer AL from DNeasy Blood and Tissue Extraction Kit (Qiagen) was added to populations sorted into lysis buffer and onto filters, respectively. Proteinase K (50 µM final concentration) and 5 µL glycogen (5 mg mL<sup>-1</sup>), which helps to recover low amounts of DNA, were added, followed by incubation at 55 °C for 30 min. Proteinase K was inactivated by 10-min incubation at 70 °C. DNA was precipitated by addition of pure ethanol (30% final concentration). Samples were then transferred to Qiagen Kit columns, and washed following the manufacturer's instruction. Finally, purified DNA was eluted with 80 µL of sterile water.

The DNA of whole samples collected on Sterivex filters was extracted following the same procedure. Two milliliters of solution was recovered from the Sterivex unit and the different incubation steps were performed with 200 µL of lysozyme, 2 mL of lysis buffer AL, and 50 µM proteinase K, but without glycogen in contrast to the sorted populations. Finally, DNA was precipitated with 2 mL of ethanol and transferred to columns from the Qiagen Kit. Washing and elution of DNA was performed following the manufacturer's instruction.

### PCR amplification

A 30 µL PCR was performed using 10 µL of extracted DNA and Hot Star Taq Master Mix (Qiagen) for sorted populations from April and June and Go Taq (Promega) for all other samples. Universal primers Euk328f and Euk329r targeting the 18S rRNA gene were used for sorted populations from April and June, while 373Cf and Euk329r were used for the two other dates (Table 2). The change of primer set during the study was motivated by the fact that, for unknown reasons, we observed a higher cloning efficiency using the combination of 373Cf and Euk329r. Amplification conditions included 5 min at 95 °C to activate the Taq polymerase, followed by 10 cycles at 95 °C for 45 s, touchdown from 64 to 54 °C for 45 s and 72 °C for 75 s. Then, 25 cycles of 95 °C for 45 s, 57 °C for 45 s and 72 °C for 75 s were performed, followed by a final extension step for 10 min at 72 °C. The quality and amount of the PCR products were evaluated after migrating 10 µL of PCR reactions onto a 1% agarose gel and comparing it with 5 µL of Smart Ladder (Eurogentec).

For most populations from the sorting test, PCR products were not concentrated enough to be cloned and a second nested PCR with the internal primers Euk1Af and 1492rE was performed using 0.5 µL of the first PCR diluted 10-fold. Amplification conditions included 5 min at 95 °C followed by 35 cycles of 95 °C for 45 s, 45 °C for 45 s and 72 °C for 75 s with a final extension for 10 min at 72 °C.

### Cloning and sequencing

One microliter of fresh PCR product was cloned using the TA vector pCR2.1-TOPO (Invitrogen) following the manufacturer's protocol. Cloning reactions were dispersed on Petri dishes containing 30 mL of Luria–Bertani agar (Difco), 30 µL ampicillin (50 mg mL<sup>-1</sup>), and 70 µL of both X-Gal (40 mg mL<sup>-1</sup>) and IPTG (200 mg mL<sup>-1</sup>). Plates were incubated overnight at 37 °C and then transferred to 4 °C for approximately 2 h. White colonies were resuspended directly into 30 µL of PCR mix containing 0.25 µL Go Taq polymerase (Qiagen), 3 µL of reaction buffer, 1.5 µL of MgCl<sub>2</sub>, 1 µL of dNTP (Eurogentec), 0.25 µL of primers Euk1Af and 1492rE (10 pM) for sorted populations from April and June and 373Cf and Euk329r for the other samples, and 23.75 µL of sterile water. PCR

**Table 2.** Sequences of primers targeting the 18S rRNA gene used in this study

Primer name	Sequence from 5' to 3'	Position*	GC%	T <sub>m</sub> (°C)	Reference
Euk328f	ACC TGG TTG ATC CTG CCA G	1	58	51	Romari & Vaulot (2004)
Euk329r	TGA TCC TTC YGC AGG TTC AC	1747	53	50	Romari & Vaulot (2004)
Euk1Af	CTG GTT GAT CCT GCC AG	3	59	48	Sogin & Gunderson (1987)
1492rE	ACC TTG TTA CGR CTT	1724	43	37	Dawson & Pace (2002)
373Cf	GAT TCC GGA GAG GGA GCC TGA	361	62	55	Weekers et al. (1994)

\*Relative to *Ostreococcus tauri* 18S rRNA gene.

amplification was performed with one step of 5 min at 95 °C followed by 35 cycles of 95 °C for 45 s, 45 °C for 45 s and 72 °C for 75 s with a final extension for 10 min at 72 °C.

For each clone, the presence of a PCR product was evaluated after migration of 5 µL of the PCR reactions onto a 1% agarose gel by comparison with Smart Ladder. Selected PCR products were transferred to a sequencing reaction cleanup plate (Montage SEQ<sub>96</sub>, Millipore) and washed twice with 100 µL of sterile water.

PCR sequencing reactions were performed using 0.5 µL of the fluorescent Big Dye Terminator DNA sequencing kit V3.1 (Applied Biosystems, Foster City, CA), 0.75 µL of Buffer (400 mM Tris-HCl, 10 mM MgCl<sub>2</sub>, pH 9), 0.5 µL of 10 µM forward primer Euk1A for sorted populations from April and June and 373Cf for the other samples and 0.5–1 µL of the purified PCR product. Sterile water was used to bring the reaction volume to 5 µL. The PCR amplification involved an initial denaturation step at 96 °C for 5 min, followed by 30 cycles including 30 s of denaturation at 94 °C, 30 s of annealing at 56 °C and 2-min extension at 60 °C. A final extension of 7 min at 60 °C followed by cooling at 4 °C terminated the PCR program. The PCR product was then purified into 20 µL of injecting solution to eliminate the remaining fluorescent nucleotides and the sequencing was then performed using an ABI Prism 3100 (Applied Biosystems).

## Sequence analysis

Sequences were analyzed for close relatives using BLAST (<http://www.ncbi.nih.gov/BLAST/>, June 2008). Sequences were also aligned using the SILVA aligner (<http://www.arb-silva.de>, Pruesse *et al.*, 2007) and merged into an ARB (Ludwig *et al.*, 2004) database containing > 30 000 complete or partial SSU rDNA sequences from eukaryotes. Finally, sequences were analyzed with KEYDNATOOLS (<http://keydnatools.com/>), which provides taxonomic affiliation and chimera detection based on sequence motifs (Guillou *et al.*, 2008). Phylogenetic assignment of each sequence was finalized based on these three analyses. Sequences that presented > 98% identity to a given genus were assigned to this genus. Rarefaction curves and the Chao1 index, a statistical indicator of richness, were estimated with the FASTGROUPII software (<http://biome.sdsu.edu/fastgroup/>) using a sequence match similarity threshold of 80% (roughly equivalent to a 98% sequence identity threshold).

Sequences have been deposited to GenBank under accession numbers FJ431283–FJ432001.

## Results

### Flow cytometric analysis of photosynthetic plankton in the English Channel

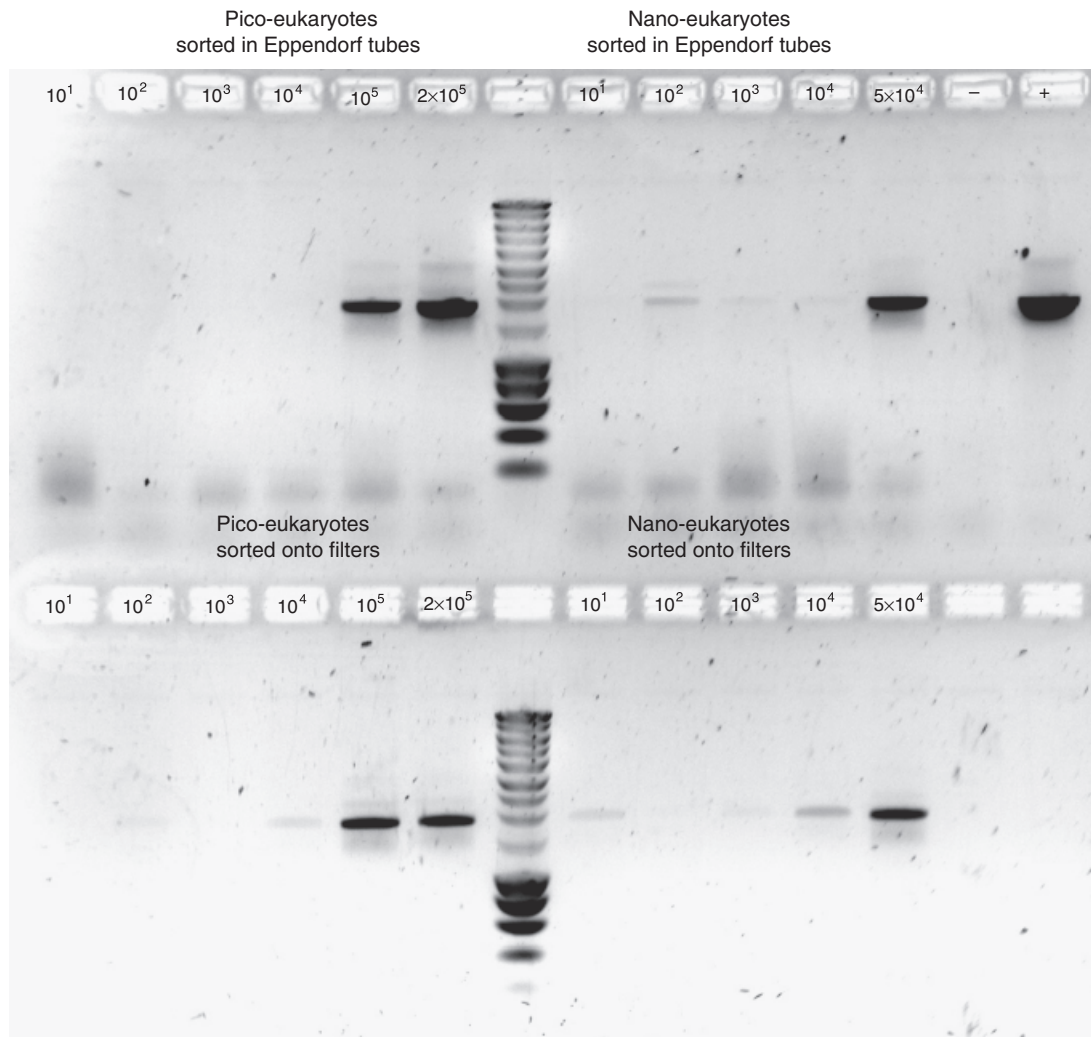
At least four populations of photosynthetic organisms could be detected by flow cytometry based on scatter and fluores-

cence signals (Fig. 1a) in English Channel coastal waters sampled off Roscoff. A population of *Synechococcus* was clearly identified by its small size and presence of both red and orange fluorescences, with cell concentrations ranging from 1600 cells mL<sup>-1</sup> in February up to 5500 cells mL<sup>-1</sup> in April (Table 1). A population of larger cells displaying both chlorophyll and phycoerythrin fluorescences, labeled PE-eukaryotes, was clearly seen in June with concentration of about 250 cells mL<sup>-1</sup>. Its flow cytometric signature was more difficult to detect for samples from October and February and seemed to be absent in April. The presence of this population was clearer in samples concentrated by tangential filtration with evidence of two groups of organisms that differed in their pigment content and scatter in all samples except that of October (Fig. 1b). The two other populations lacked orange phycoerythrin fluorescence and corresponded roughly to the pico- (< 2 µm) and nanophytoplankton (2–20 µm) size ranges (Sieburth *et al.*, 1978). Therefore, they are labeled in the rest of the article as pico- and nano-eukaryotes. These two populations reached maximum abundance in June with cell densities at about 19 000 and 1800 cells mL<sup>-1</sup>, respectively (Table 1).

### Methodological considerations

We tested the effect of the collection mode of sorted cells: either directly into Eppendorf tubes filled with lysis buffer or onto filters. We also tried to determine the number of cells necessary to obtain enough DNA to build reliable 18 rRNA gene clone libraries. These tests were achieved on pico- and nano-eukaryote populations sorted from a sample collected at the Astan Station in August 2007. No differences were observed between PCR products obtained on populations directly sorted into lysis buffer and those sorted onto filters (Fig. 2). Using Euk328f and Euk329r primers, large amounts of PCR product were obtained for 100 000 and 200 000 pico-eukaryotes and 50 000 nano-eukaryotes. For cell concentrations below these values, only a faint PCR product was observed. Therefore, a second nested PCR was performed before cloning for populations sorted into lysis buffer.

Clone library composition varied sharply as a function of the number of cells sorted. In particular, the percentage of sequences matching that of photosynthetic organisms increased with the number of sorted cells (Tables 3 and 4). For 10 sorted pico-eukaryote cells (Table 3, Supporting Information, Table S1), only two sequences of photosynthetic eukaryotes were obtained. However, they corresponded to cells (*Mantoniella*, *Chaetoceros*) with a size larger than a typical picoplankton. When 100 cells were sorted, all photosynthetic sequences belonged to the pico-eukaryotic genus *Ostreococcus*. With 1000 pico-eukaryotes sorted, *Ostreococcus* still dominated (66% of the photosynthetic sequences) with the rest corresponding to another



**Fig. 2.** PCR products obtained using the universal primers Euk328f and Euk329r on different numbers of pico-eukaryotes (10, 100, 1000, 10 000, 100 000, 200 000) and nano-eukaryotes (10, 100, 1000, 10 000, 50 000) sorted into Eppendorf tubes (top) and onto filters (bottom).

**Table 3.** Sequences obtained after cloning populations containing different numbers of sorted pico-eukaryotes (Astan sample from August 30, 2007)

Cells sorted	Sequences obtained	<i>Micromonas</i>	<i>Ostreococcus</i>	<i>Bathycoccus</i>	Other photosynthetic groups	Fungi	Metazoa	Others
10	5				2	2		1
100	20		16			4		
1000	17		10	5		2		
10 000	16	12	3			1		
100 000	20	10	8	1			1	
200 000	19	12	3			1	2	1

pico-eukaryotic species, *Bathycoccus prasinos*. For all samples with pico-eukaryote cells in excess of 10 000, clone composition was quite similar. Sequences matching *M. pusilla* became dominant with up to 80% of the photo-

synthetic sequences recovered, *Ostreococcus* was still present while *Bathycoccus* was much less represented or even absent. Sequences of nano-eukaryotes were dominated by diatoms (Table 4), especially *Chaetoceros socialis*, a small

**Table 4.** Sequences obtained after cloning populations containing different numbers of sorted nano-eukaryotes (Astan sample from August 30, 2007)

Cells sorted	Sequences obtained					
		Diatoms	Haptophyta	Fungi	Metazoa	Others
10	14	1		13		
100	11	3		6		2
1000	10	5			5	
10 000	14	4	1	1	3	5
50 000	10	4	1		3	2

chain-forming species (around 10 µm). When higher numbers of cells were sorted (10 000 and 50 000), some haptophyte sequences were also obtained (*Chrysochromulina*).

Sequences of nonphotosynthetic protists were also recovered in these populations including in particular Cercozoa in the nano-eukaryote population (Table S1). Fungal sequences were especially abundant in populations with low numbers of sorted cells (10 for pico-eukaryotes and 10–100 for nano-eukaryotes), but could also be found in populations with larger numbers of sorted cells. These sequences probably originated from laboratory contamination since some of them (Table S1) corresponded to common house molds such as *Penicillium* or to human-associated fungi such as *Trichosporon* (Pfaller & Diekema, 2004). Other sequences not originating from protists included some from metazoans and land plants. The former were affiliated to *Muggiaea atlantica*, a siphonophore whose gametes might be found within the picoplanktonic population, while the latter matched *Musa basjoo*, a banana tree commonly found in gardens along the coast of Roscoff.

### Seasonal study: 18S rRNA gene clone libraries from sorted populations

Three photosynthetic populations (pico-, nano- and PE-eukaryotes) were sorted by flow cytometry at each season (Table 1) and their 18S rRNA gene was amplified, cloned and sequenced. Using the software KEYDNATOOLS (Guillou *et al.*, 2008), we were able to detect that at least 19 sequences (6%) were likely chimeras (Table S3). All potential chimeras were manually checked by aligning the 5' and 3' ends of the sequences to their respective targets. In many cases, chimeras were formed by sequences from two very closely related taxa. For example, chimerical combinations of *Micromonas* and *Bathycoccus* were recovered from pico-eukaryotes sorted in April, and in June, the dominating *Ostreococcus* formed chimeras with *Micromonas* or *Bathycoccus*. For nano-eukaryotes sorted in June, two chimeras combined the two dominant genera *Chrysochromulina* and *Phaeocystis* (Table S3). For PE-eukaryotes, the four detected chimeras were combinations of cryptophyte nuclear and nucleomorph 18S rRNA genes. All these chimeras are not considered further in the rest of Results.

Among the 313 nonchimerical partial sequences obtained from sorted populations, 93% were most similar to those from photosynthetic organisms (Table 5 and Table S2). More than 88% of the sequences had at least 98% similarity with known sequences based on BLAST analysis. This proportion was higher among pico-eukaryotes and PE-eukaryotes (96% and 91%, respectively), but lower (79%) for nano-eukaryotes for which 8% of the sequences had < 96% of similarity with known organisms. Rarefaction curves and Chao1 estimators (Fig. 3) demonstrate that pico- and PE-eukaryotes were less diverse than nano-eukaryotes.

Among the 98 pico-eukaryote sequences, Prasinophyceae were dominant (Fig. 4) with 92% belonging to the three genera *Micromonas*, *Ostreococcus* and *Bathycoccus* (Mamiellales). If we exclude sequences of Metazoa and fungi, pico-eukaryote populations sorted in April, June and February only yielded prasinophyte sequences, while in October we also recovered four sequences of photosynthetic stramenopiles, and one cryptophyte (Table 5). Nano-eukaryotes were dominated by haptophytes (Fig. 4), especially the genera *Phaeocystis* and *Chrysochromulina*, but diatoms and nano-sized Prasinophyceae sp. were also important contributors (Fig. 4). In October, *Micromonas* sequences were abundant within the nano-eukaryotes (Table 5).

### Seasonal study: 18S rRNA gene clone libraries from filtered samples

For comparison with previous studies, 340 sequences were obtained from clone libraries constructed from filtered samples. Less than 4% of the sequences (12) were identified as chimeras. In contrast to chimerical sequences obtained from sorted populations that were composed of pieces from closely related phylogenetic groups, those from filtered samples resulted from the assemblage of widely different groups.

Alveolates, prasinophytes and stramenopiles contributed most to the 328 nonchimerical sequences recovered (42%, 20%, and 9% of the sequences, respectively, Table 5). Metazoa, Cercozoa, and cryptophytes represented approximately 8%, 6% and 4% of the clone libraries, respectively. Prymnesiophytes represented < 2% of the sequences and were not recovered in February. Four sequences of picobiliphytes were obtained in samples from April and February. Four sequences of *Telonemia* and two of *Radiolaria* were also obtained.

### Seasonal study: major groups

#### Chlorophyta

Sequences matching that of the three major genera *Micromonas*, *Ostreococcus* and *Bathycoccus* belonging to the Mamiellales (Prasinophyceae) were similarly distributed in sorted populations and filtered samples with a dominance of *Bathycoccus* in February, *Micromonas* in April, and

**Table 5.** Summary of phylogenetic assignments for sequences obtained at different times of the year for the three sorted populations and for filtered samples. 'All' corresponds to the total of the four seasonal samples

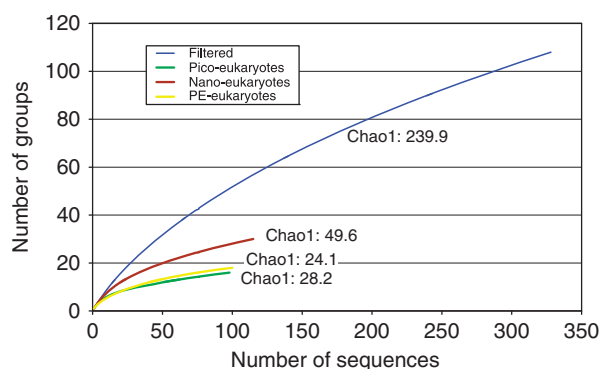
		Clone Library																																				
		RA070411S	RA070625S	RA071004S	RA080215S		RA070411B	RA070625B	RA071004B	RA080215B		RA070411C	RA070625C	RA071004C	RA080215C		RA070411T	RA070625T	RA071004T	RA080215T																		
		Pico-euks					Nano-euks					PE-euks					Filtered																					
Division	Class	Month	Apr	Jun	Oct	Feb	All	Apr	Jun	Oct	Feb	All	Apr	Jun	Oct	Feb	All	Apr	Jun	Oct	Feb	All	Apr	Jun	Oct	Feb	All	Total										
		Genus																																				
Photosynthetic																																						
Chlorophyta	Prasinophyceae	<i>Bathycoccus</i>	2	2	3	18	25																				4	6	1	5	16	41						
Chlorophyta	Prasinophyceae	<i>Micromonas</i>	19	1	2	12	34																				1	1	10	10	1	2	23	71				
Chlorophyta	Prasinophyceae	<i>Ostreococcus</i>	2	10	19		31																				2	2	1	25	1		27	60				
Chlorophyta	Prasinophyceae	Others			1		1	8	3	3		14																				1	1			2	17	
Cryptophyta	Cryptophyceae	<i>Geminigera</i>											4	13	10	5	32	2	1			2	5	37														
Cryptophyta	Cryptophyceae	<i>Teleaulax</i>											6	5	16	10	37	2	2	1	2	7	45															
Cryptophyta	Cryptophyceae	Others											5	2	2	12	21		1			1	22															
Picobiliphytes													1				1				3	4	5															
Haptophyta	Prymnesiophyceae	<i>Phaeocystis</i>											21	3	7	31		2	1	1		4	35															
Haptophyta	Prymnesiophyceae	<i>Chrysochromulina</i>						6	2	2	2	12		1				1				2	14															
Haptophyta	Prymnesiophyceae	Others											2	3	1	6							6															
Stramenopiles	Bacillariophyceae												5		4	13	22		6	9		11	26	49														
Stramenopiles	Bolidophyceae																		1			1	1															
Stramenopiles	Chrysophyceae																		1			1	2															
Stramenopiles	Raphidophyceae																						2															
Stramenopiles	Pelagophyceae	<i>Pelagomonas</i>																							1	1	1											
Stramenopiles	Phaeophyceae	<i>Sargassum</i>																							1	1	1											
Cercozoa	Chlorarachnio- phyceae																							1			1	1										
Alveolata	Dinophyceae							1				1		1		1		1	13	3	21	18	55	57														
Heterotrophic																																						
Alveolata	Ciliophora																		1		1	5		4	7	16	17											
Alveolata	Syndiniales Group I																		6	1	5	2	14	14														
Alveolata	Syndiniales Group II																		3	2	39	7	51	51														
Alveolata	Syndiniales Other																				2		2	2														
Cercozoa								3	1		2	6		6	8	1	4	19	25																			
Choanoflagellida																					1	1	1															
Fungi																								7														
Radiolaria	Acantharea																	1			2	3		2	2													
Stramenopiles	MAST and others																			1		2	2	6	3	3	2	14	17									
Telonemia																							5			5	5											
Unidentified																							1			1	1											
Multicellular																																						
Metazoa			1					1																				4	6	2	14	26	27					
Streptophyta	Embryophyta	<i>Musa</i>																								5	5											
		Sum	24	13	31	30	98	23	30	37	25	115	16	20	32	32	100	79	82	85	82	328	641															
		Chimera	7	6					13																				2	1	2	1	4	3	1	8	12	31



*Ostreococcus* in June and October (Table 5). Sequences of larger prasinophytes belonging to the genera *Pyramimonas* and *Mantoniella* were found among nano-eukaryotes except in February (Table S2). In October, one sequence matching the genus *Pycnococcus* was also found. Sequences of Prasinophyceae besides the three major Mamiellales genera were poorly represented in filtered samples with only one sequence of *Pyramimonas* in April and one of *Pycnococcus* in June.

### Cryptophyta

Although cryptophytes were poorly represented in filtered samples, they were very well recovered in sorted PE-eukaryotes (Table 5). Among the 104 partial sequences obtained



**Fig. 3.** Rarefaction curves of number of unique sequences recovered vs. number of clones sequenced, for sorted pico-eukaryotes, nano-eukaryotes, and PE-eukaryotes and for filtered samples. Libraries maintained at the four seasons were pooled together for each population. The Chao1 index is an estimate of the richness (total number of phylotypes) for each sample type. Rarefaction curves and Chao1 indices were computed with FASTGROUPII (Yu *et al.*, 2006), using a sequence match similarity threshold of 80% (roughly equivalent to a 98% sequence identity threshold).

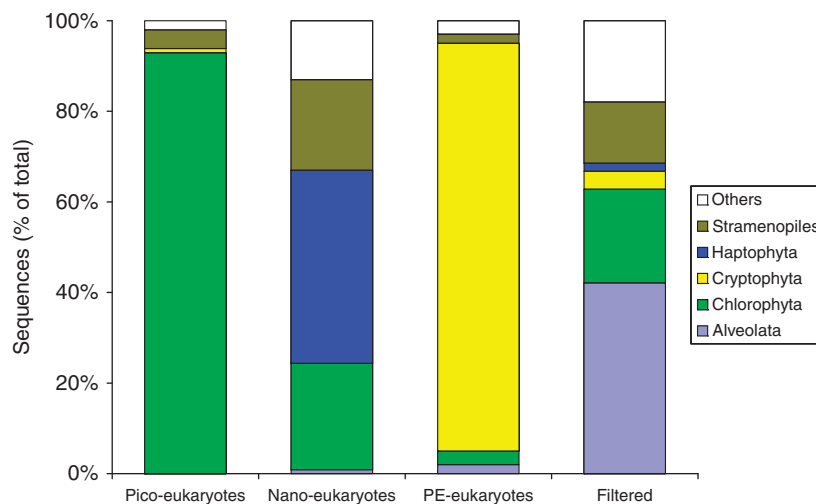
for cryptophytes (91 for sorted and 13 for filtered samples), 79% are affiliated to clade 4 (Hoef-Emden, 2008), mostly related to the genera *Geminigera* and *Teleaulax* (37% and 35%, respectively). *Plagioselmis*, also from clade 4, represented 9% of the sequences and was only recovered in sorted populations from October.

### Haptophyta

Sequences of haptophytes were poorly represented in clone libraries from filtered samples with a maximum of 4% in April and they were completely absent in February. In contrast, haptophytes constituted an important fraction of the sequences recovered for sorted nano-eukaryotes representing 26%, 83%, 21%, and 40% in April, June, October, and February, respectively. Many sequences were affiliated to two major genera; *Chrysochromulina* in April and *Phaeocystis* in June and February. Some sequences were related to uncultivated haptophytes.

### Stramenopiles

Sequences of photosynthetic stramenopiles were not recovered for nano-eukaryotes in June, while they represented 22%, 11%, and 52% of the sequences obtained in April, October and February, respectively. They were also absent from the October clone library from filtered samples. A wide variety of diatom genera, including *Thalassiosira*, *Minidiscus*, *Minutocellus*, *Guinardia*, *Chaetoceros*, and *Corethron* (Table S2) represented 87% of the sequences. Bolidophyceae, Chrysophyceae, Pelagophyceae, and Phaeophyceae were more sporadically found. Four photosynthetic stramenopiles (diatoms, Chrysophyceae, Raphidophyceae) sequences were found within the pico-eukaryote population sorted in October. Sequences of heterotrophic stramenopiles



**Fig. 4.** Overall phylogenetic composition of 18S clone libraries constructed from sorted populations and filtered samples.

(MAST, Massana *et al.*, 2004) were recovered in filtered samples (4% of all sequences), but also in nano- and PE-eukaryotes sorted in October.

### Cercozoa

Sequences of Cercozoa were retrieved in all clone libraries from filtered samples with a maximum contribution of 10% in June. Cercozoa contributed to 13% and 8% of the clones analyzed from nano-eukaryotes sorted in April and February, respectively, much less in June, and they were absent in October.

### Alveolata

Sequences of dinoflagellates (Dinophyceae) were recovered in all filtered samples making up from 16% to 25% of the sequences, except in June where they only represented < 4% of the sequences. At least nine genera were represented, including *Gymnodinium*, *Heterocapsa*, *Pheopolykrikos*, *Gyrodinium*, and *Karodinium*. In contrast, they were virtually absent from the sorted populations with only one sequence in the nano-eukaryotes and one in the PE-eukaryotes. Heterotrophic alveolates were major contributors in filtered samples, with three major groups: Ciliophora (5%), Syndiniales group I (4%) and group II (15%). With the exception of one ciliate sequence recovered in the PE-eukaryote population sorted in October, heterotrophic alveolates were completely excluded from clone libraries originating from sorted populations.

### Other protist sequences

A few sequences from the recently discovered picobiliphytes (Not *et al.*, 2007) were retrieved, one from the nano-eukaryotes sorted in June and four from filtered samples, one in April and three in February. Other sequences from heterotrophic protists (Radiolaria, choanoflagellates, Telonemia) were only obtained from filtered samples and completely excluded by sorting (Table 5).

### Metazoa, Fungi, and land plants

Sequences from nonprotist groups were found within clone libraries both from sorted populations and filtered samples. With the exception of one metazoan sequence within the pico-eukaryotes in April, Metazoa were only recovered from filtered samples, reaching a maximum of 15% of the sequences in February. Crustacea (15) and Polychaeta (nine) dominated as observed previously (Romari & Vaulot, 2004), probably originating from larval stages that are particularly abundant in coastal waters in spring. In contrast, fungal sequences were not recovered from filtered samples, but only in sorted popula-

tions, representing up to 8% of nano-eukaryote sequences in October. As in the case of the test sample, sequences of the banana tree, *M. basjoo*, were found (13% of sequences) in the October nano-eukaryote population.

## Discussion

### Diversity of photosynthetic eukaryotes

Our aim was to develop an approach that would allow a better assessment of the diversity of small photosynthetic eukaryotes from 18S rRNA gene clone libraries. The composition of clone libraries obtained from filtered samples during the course of the present work was very similar to that observed in a previous study at the same location (Romari & Vaulot, 2004). In contrast, sorting autotrophic subpopulations based on chlorophyll fluorescence (Fig. 1) allowed better targeting of photosynthetic groups (Fig. 4). Their overall contribution jumped from 54% for filtered samples up to 92% for the sum of the three sorted populations. For pico-eukaryotes, the share of photosynthetic sequences reached even 98% (Table 5). These photosynthetic populations are clearly less diverse than the whole eukaryotic microbial community as evidenced by the rarefaction curves and estimation of the Chao1 index (Fig. 3). Pico- and PE-eukaryote populations were 10 times less diverse than the total community and nano-eukaryote populations five times less. This supports an earlier claim that, indeed, photosynthetic pico-eukaryotes are less diverse than heterotrophic ones (Vaulot *et al.*, 2002). One advantage of sorting by flow cytometry lies in the reduction of the sequencing effort needed to assess the composition of the photosynthetic community. For example, when the PE-eukaryote concentration was very low in October (60 cells mL<sup>-1</sup>), only a single sequence was obtained from filtered samples, while a much more detailed view of this population was achieved following sorting.

The value of cell sorting is clearly illustrated by the fact that several sequences were only recovered in the sorted populations and not in the filtered samples. Sequences related to *Mantoniella* (Mamiellales) were only found in sorted nano-eukaryotes (Table S2) and one of these presented a 100% similarity to an environmental sequence recovered with Chlorophyta-biased primers from the Mediterranean Sea (Viprey *et al.*, 2008). In the same group, sequences related to *Pyramimonas* were found in the filtered sample only in April, but were present in sorted nano-eukaryote populations in April, June and October. Among Haptophyta, we detected novel sequences unrelated to any known genus in the October nano-eukaryotes (Table S2). Sequences related to *Chrysochromulina* were observed only in April and October in the filtered samples, but at all seasons in the sorted nano-eukaryotes.

Pico-eukaryote sequences were dominated by the three genera of *Micromonas*, *Ostreococcus*, and *Bathycoccus* (Mamiellales, Prasinophyceae) that are very typical of coastal waters (Vaulot *et al.*, 2008). If the four seasonal samples are grouped together, these three genera contributed almost equally. However, at a given time, either one (spring, summer) or two (fall, winter) genera appeared to dominate. Quite surprisingly in summer, *Ostreococcus* was the most abundant in clone libraries. In contrast, a previous quantitative study using FISH has shown that *Ostreococcus* was always low in abundance off Roscoff (at most 18% of Mamiellales, Not *et al.*, 2004). Other groups contributed very little to pico-eukaryotes and contaminants were quite low. Two identical stramenopile clones somewhat related to Raphidophyceae (precise assignment would require obtaining the full 18S rRNA gene) could belong to a yet undescribed class. The PE-eukaryote community was also quite simple, showing little seasonal variation with cryptophytes of clade 4 dominating, especially two closely related genera (*Geminigera* and *Teleaulax*). Cryptophytes are characteristic of coastal waters where their pigment alloxanthin is often detected (Breton *et al.*, 2000). Curiously, we did not recover any picobiliphyte sequences among the PE-eukaryotes despite the fact that sequences from this group were detected in the filtered samples from late winter and early spring as reported previously (Romari & Vaulot, 2004) and picobiliphytes have been shown to contain PE (Not *et al.*, 2007; Cuvelier *et al.*, 2008). This absence could be explained by the fact that the window chosen to sort PE-eukaryotes was inadequate for picobiliphytes, either because of a larger size (Cuvelier *et al.*, 2008) or a different pigmentation. Finally, nano-eukaryotes consisted mainly of haptophytes, diatoms, and Mamiellales (Prasinophyceae). Besides the bloom-forming genus *Phaeocystis*, which appears to be present throughout the year except in spring, and the highly diversified genus *Chrysochromulina*, some other haptophytes were only distantly related to known species. Among diatoms, genera containing small-sized species such as *Minutocellus* or *Minidiscus* were quite well represented. Diatoms were especially important in the nano-eukaryote population in winter, i.e. much before the diatom bloom that develops in Western Channel waters in late spring. Interestingly during the bloom (June), diatom sequences were only found in the filtered sample, but not in nano-eukaryotes, suggesting that they corresponded to large cells. These diatom sequences matched very closely to that of *Guinardia delicatula*, which is quite large (typically 50 µm long). *Guinardia delicatula* is one of the major blooming species off Roscoff (Sournia *et al.*, 1987) and was observed in a very high concentration in the Lugol's counts on that date (F. Rigaut-Jalabert & F. Jouenne, unpublished data). Sequences of the nanoplanktonic genus *Mantoniella* (3–5 µm) from the order Mamiellales (prasinophytes) were observed at all seasons except in winter.

## Methodological considerations

One of the major drawbacks of flow cytometric sorting is that, while purity is in general very high, the amount of material recovered is quite low and downstream analysis requires adaptation of existing protocols. We tried to solve this methodological bottleneck in several ways.

First, we preconcentrated the samples in order to better visualize the different subpopulations and to speed up sorting. As an example, sorting 5000 cells from a PE-eukaryote population at 84 cells mL<sup>-1</sup> as observed in February would require to run a 60-mL sample and would take about 12 h, while after preconcentration it took < 10 min. One potential problem with preconcentration is that it may induce differential cell loss. Indeed, the percentages of recovery of the different populations discriminated by flow cytometry after tangential flow filtration are lower for the small cells (*Synechococcus* and pico-eukaryotes) than for the larger cells (Table 1). However, the real extent of cell loss is very hard to evaluate since some rare and fragile taxa may be lost, but sequences from these taxa may not be recovered in unconcentrated samples because of their scarcity. Still the major groups are probably quite well recovered such as the Mamiellales or the cryptophytes from clade 4 that are found to be dominant in unconcentrated samples (Romari & Vaulot, 2004) and are very well recovered after tangential flow filtration (this work).

Second, we developed a cell collection and DNA extraction protocol that was efficient at recovering small quantities of DNA. In particular, cells were sorted directly into a lysis buffer to minimize cell loss. One problem stemming from the low number of collected cells is the potential for contamination, in particular by fungi often found on laboratory benches. For all practical purposes, it seems that sorted populations of 100 000 cells for pico-eukaryotes and 50 000 cells for nano-eukaryotes provide enough material for reliable clone library construction with minimal contamination and without relying on nested PCR.

If one is willing to use nested PCR, the number of sorted cells can be decreased by an order of magnitude. However, for populations below 1000 pico-eukaryotes or 100 nano-eukaryotes, contamination issues appear more drastic. More surprisingly, clone library composition seems to depend somewhat on the number of sorted cells even for quite large numbers. For example in our tests, *Micromonas* is only observed in pico-eukaryote populations for 10 000 sorted cells or more (Table 3). Moreover, *Bathycoccus* represent almost 30% of the sequences for 1000 sorted cells and disappears for larger numbers of sorted cells. Possible explanations are that for some DNA templates, once it reaches a certain concentration it can be better amplified, or alternatively that at low cell numbers (1000 and below), there is some stochastic effect either during sorting (rare

cells have a low chance to be sorted) or during PCR (low concentration templates have little chance to be amplified). In order to test this stochastic effect, it would be necessary to perform separate PCR and cloning reactions on replicate populations with low numbers of sorted cells (e.g. 100). Our data suggest that at least 10 000 cells need to be sorted in order to build clone libraries exempt of stochastic biases.

Another potential problem is cross-contamination between populations. For example in October, pico-eukaryote populations yielded sequences of photosynthetic stramenopiles and cryptophytes while the PE-eukaryotes contained noncryptophyte sequences. This may suggest that some technical problems during sorting in October led to cross-contamination between the different populations. Alternatively the presence of heterotrophic sequences among the PE-eukaryotes could be due to the recent ingestion of cryptophytes by these organisms, providing them with a fluorescent signal triggering sorting.

Sorting clearly allows a better assessment of the diversity of some groups that are virtually absent in clone libraries constructed from filtered samples. This is the case for Haptophyta. While their representative pigment 19' hexanoyl-oxyfucoxanthin is very abundant in marine waters, they are under-represented in clone libraries (Moon-van der Staay *et al.*, 2000). This under-representation could be linked to the fact that their 18S rRNA genes have a slightly higher GC% than for other groups. In our samples, the GC% of haptophytes was significantly higher (49.2% on average) in comparison with Chlorophyta (45.9%) and Stramenopiles (44.2%). In sorted populations, haptophyte templates have probably much less competition from other groups and can be readily amplified as demonstrated here.

The existence of chimeras among sequences recovered from environmental 18S clone libraries has received some attention lately (Berney *et al.*, 2004). The formation of such artefactual sequences often occurs within the last PCR cycles, when the concentration of primers decreases (Kanagawa, 2003). When working on sorted material, the amount of starting DNA is low and therefore the number of PCR cycles has to be increased (up to 40 here) to be able to clone the amplified material, increasing the potential for chimera formation. The use of the new software KEYDNATOOLS (Guilou *et al.*, 2008) proved to be very efficient to detect chimerical sequences (5% of all sequences from the seasonal study), even those occurring between very closely related organisms, such as the Mamiellales *Micromonas*, *Ostreococcus*, and *Bathycoccus* (Table S3) that are usually hard to spot. Interestingly, the highest percentages of chimerical sequences recovered were obtained for pico-eukaryotes, a population of quite low diversity. Chimeras were quite abundant for populations sorted in April and June (29% and 46% of the total sequences, respectively) for which the universal primers Euk328f/Euk329r were used. In contrast,

no chimera was observed for pico-eukaryotes sorted in October and February for which amplification was performed with the primer set 373Cf/Euk329r. For nano-eukaryotes, chimeras were only observed between closely related organisms (*Chrysochromulina* and *Phaeocystis*). These data suggest (1) that chimera formation could be higher for low diversity populations containing closely related organisms and (2) that primer sets may have an influence. It was however, surprising to observe chimeras between nuclear and nucleomorph 18S rRNA genes for cryptophytes since their sequences are phylogenetically quite different.

## Conclusions

Molecular analysis of flow cytometrically sorted populations appears to have many advantages over that of conventional filtered samples. It reduces the sequencing effort needed to target specific groups and provides access to deeper diversity (Shi *et al.*, 2009). The use of dyes, such as nucleic acid stains, would certainly help discriminating more specific populations, such as dinoflagellates, which have a large genome size compared with their cell size. Despite potential biases, linked for example to the necessary preconcentration, we anticipate that, coupled with whole genome amplification as recently demonstrated for cyanobacteria (Palenik *et al.*, 2009), our approach permits environmental metagenomic studies of marine eukaryotic microorganisms, a feat not yet possible when starting from conventional filtered samples, because the metagenomic signal is overwhelmed by bacterial sequences.

## Acknowledgements

This work has been funded in part by PICOFUNPAC (ANR Biodiversité 06-BDIV-013). X.L.S. benefited from fellowships from the Université Pierre et Marie Curie (Paris 6) and from the China Scholarship Council (CSC) managed by the Fondation Franco-Chinoise pour la Science et ses Applications (FFCSA). We are indebted to Peter von Dassow for helping us to improve the final version.

## References

- Berney C, Fahrni J & Pawlowski J (2004) How many novel eukaryotic 'kingdoms'? Pitfalls and limitations of environmental DNA surveys. *BMC Biol* **2**: 13.
- Breton E, Brunet C, Sautour B & Brylinski JM (2000) Annual variations of phytoplankton biomass in the Eastern English Channel: comparison by pigment signatures and microscopic counts. *J Plankton Res* **22**: 1423–1440.
- Cuvelier ML, Ortiz A, Kim E *et al.* (2008) Widespread distribution of a unique marine protistan lineage. *Environ Microbiol* **10**: 1621–1634.

- Czechowska K, Johnson DR & van der Meer JR (2008) Use of flow cytometric methods for single-cell analysis in environmental microbiology. *Curr Opin Microbiol* **11**: 205–212.
- Dawson SC & Pace NR (2002) Novel kingdom-level eukaryotic diversity in anoxic environments. *P Natl Acad Sci USA* **99**: 8324–8329.
- Fuchs BM, Woebken D, Zubkov MV, Burkill P & Amann R (2005) Molecular identification of picoplankton populations in contrasting waters of the Arabian Sea. *Aquat Microb Ecol* **39**: 145–157.
- Fuller NJ, Campbell C, Allen DJ, Pitt FD, Le Gall F, Vault D & Scanlan DJ (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.
- Guillou L, Viprey M, Chambouvet A *et al.* (2008) Widespread occurrence and genetic diversity of marine parasitoids belonging to Syndiniales (Alveolata). *Environ Microbiol* **10**: 3349–3365.
- Hoef-Emden K (2008) Molecular phylogeny of phycocyanin-containing cryptophytes: evolution of biliproteins and geographical distribution. *J Phycol* **44**: 985–993.
- Kanagawa T (2003) Bias and artifacts in multitemplate polymerase chain reactions (PCR). *J Biosci Bioeng* **96**: 317–323.
- Le Gall F, Rigaut-Jalabert F, Marie D, Garczarek L, Viprey M, Godet A & Vault D (2008) Picoplankton diversity in the South-East Pacific Ocean from cultures. *Biogeosciences* **5**: 203–214.
- Ludwig W, Strunk O, Westram R *et al.* (2004) ARB: a software environment for sequence data. *Nucleic Acids Res* **32**: 1363–1371.
- Marie D, Partensky F, Simon N, Guillou L & Vault D (2000) Flow cytometry analysis of marine picoplankton. In *Living Colors: Protocols in Flow Cytometry and Cell Sorting* (Diamond RA & DeMaggio S, eds), pp. 421–454. Springer Verlag, Berlin.
- Massana R, Castresana J, Balagué V *et al.* (2004) Phylogenetic and ecological analysis of novel marine stramenopiles. *Appl Environ Microb* **70**: 3528–3534.
- Moon-van der Staay SY, van der Staay GWM, Guillou L, Vault D, Claustre H & Medlin LK (2000) Abundance and diversity of prymnesiophytes in the picoplankton community from the equatorial Pacific Ocean inferred from 18S rDNA sequences. *Limnol Oceanogr* **45**: 98–109.
- Moran XAG (2007) Annual cycle of picophytoplankton photosynthesis and growth rates in a temperate coastal ecosystem: a major contribution to carbon fluxes. *Aquat Microb Ecol* **49**: 267–279.
- Not F, Latasa M, Marie D, Cariou T, Vault D & Simon N (2004) A single species *Micromonas pusilla* (Prasinophyceae) dominates the eukaryotic picoplankton in the western English channel. *Appl Environ Microb* **70**: 4064–4072.
- Not F, Valentin K, Romari K *et al.* (2007) Picobiliphytes, a new marine picoplanktonic algal group with unknown affinities to other eukaryotes. *Science* **315**: 252–254.
- Palenik B, Ren Q, Tai V & Paulsen IT (2009) Coastal *Synechococcus* metagenome reveals major roles for horizontal gene transfer and plasmids in population diversity. *Environ Microbiol* **11**: 349–359.
- Partensky F, Blanchot J & Vault D (1999) Differential distribution and ecology of *Prochlorococcus* and *Synechococcus* in oceanic waters: a review. *Marine Cyanobacteria. Bulletin de l'Institut Océanographique de Monaco. Numéro spécial 19* (Charpy L & Larkum AWD, eds), pp. 457–475. Musée Océanographique Monaco.
- Pfaller MA & Diekema DJ (2004) Rare and emerging opportunistic fungal pathogens: concern for resistance beyond *Candida albicans* and *Aspergillus fumigatus*. *J Clin Microbiol* **42**: 4419–4431.
- Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J & Glockner FO (2007) SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* **35**: 7188–7196.
- Romari K & Vault 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.
- Shi XL, Marie D, Jardillier L, Scanlan DJ & Vault D (2009) Groups without cultured representatives dominate eukaryotic picophytoplankton in the oligotrophic South East Pacific Ocean. *PLoS ONE* **4**: e7657.
- Sieburth JM, Smetacek V & Lenz J (1978) Pelagic ecosystem structure: heterotrophic compartments of the plankton and their relationship to plankton size fractions. *Limnol Oceanogr* **23**: 1256–1263.
- Sogin ML & Gunderson JH (1987) Structural diversity of eukaryotic small subunit ribosomal RNAs. Evolutionary implications. *Ann NY Acad Sci* **503**: 125–139.
- Sournia A, Birrien J-L, Douville J-L, Klein B & Viollier M (1987) A daily study of the diatom spring bloom at Roscoff (France) in 1985. I. The spring bloom within the annual cycle. *Estuar Coast Shelf S* **25**: 355–367.
- Vault D, Romari K & Not F (2002) Are autotrophs less diverse than heterotrophs in marine picoplankton? *Trends Microbiol* **10**: 266–267.
- Vault D, Eikrem W, Viprey M & Moreau H (2008) The diversity of small eukaryotic phytoplankton ( $\leq 3 \mu\text{m}$ ) in marine ecosystems. *FEMS Microbiol Rev* **32**: 795–820.
- Veldhuis MJW & Kraay GW (2000) Application of flow cytometry in marine phytoplankton research: current applications and future perspectives. *Sci Mar* **64**: 121–134.
- Viprey M, Guillou L, Ferréol M & Vault D (2008) Wide genetic diversity of picoplanktonic green algae (Chloroplastida) in the Mediterranean Sea uncovered by a phylum-biased PCR approach. *Environ Microbiol* **10**: 1804–1822.
- Wallner G, Fuchs B, Spring S, Beisker W & Amann R (1997) Flow sorting of microorganisms for molecular analysis. *Appl Environ Microb* **63**: 4223–4231.
- Weekers PH, Gast RJ, Fuerst PA & Byers TJ (1994) Sequence variations in small-subunit ribosomal RNAs of *Hartmannella*

*vermiformis* and their phylogenetic implications. *Mol Biol Evol* **11**: 684–690.

Worden AZ & Not F (2008) Ecology and diversity of picoeukaryotes. *Microbial Ecology of the Ocean* (Kirchman DL, ed), pp. 159–205. Wiley, New York.

Yu Y, Breitbart M, McNairnie P & Rohwer F (2006) FastGroupII: a web-based bioinformatics platform for analyses of large 16S rDNA libraries. *BMC Bioinformatics* **7**: 57.

## Supporting Information

Additional Supporting Information may be found in the online version of this article:

**Table S1.** Phylogenetic affiliation for sequences from samples used to test the effect of the number of cells sorted.

**Table S2.** Phylogenetic affiliation for sequences from the seasonal samples.

**Table S3.** List of chimerical sequences obtained for the seasonal samples.

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