FEMS MICROBIOLOGY ECOLOGY

RESEARCH ARTICLE

Whole-genome amplification (WGA) of marine photosynthetic eukaryote populations

Cécile Lepere[1], Mikihide Demura[2], Masanobu Kawachi[2], Sarah Romac[1], Ian Probert[1] & Daniel Vaulot[1]

[1]UPMC (Paris-06) and CNRS, UMR 7144, Station Biologique de Roscoff, BP74, Roscoff, France; and [2]National Institute for Environmental Studies, Ibaraki, Tsukuba, Japan

Correspondence: Present address: Cécile Lepere, Department of Biological Sciences, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK. Tel.: +44 2476 522697; fax: +44 2476 523568; e-mail: c.lepere@warwick.ac.uk

Received 11 June 2010; revised 8 November 2010; accepted 6 February 2011.

DOI:10.1111/j.1574-6941.2011.01072.x

Editor: Patricia Sobecky

Keywords
whole-genome amplification; photosynthetic microbial eukaryote; flow cytometry sorting; rRNA genes.

Abstract

Metagenomics approaches have been developing rapidly in marine sciences. However, the application of these approaches to marine eukaryotes, and in particular to the smallest ones, is challenging because marine microbial communities are dominated by prokaryotes. One way to circumvent this problem is to separate eukaryotic cells using techniques such as single-cell pipetting or flow cytometry sorting. However, the number of cells that can be recovered by such techniques remains low and genetic material needs to be amplified before metagenomic sequencing can be undertaken. In this methodological study, we tested the application of whole-genome amplification (WGA) to photosynthetic eukaryotes. We performed various optimization steps both on a mixture of known microalgal strains and on natural photosynthetic eukaryote populations sorted by flow cytometry. rRNA genes were used as markers for assessing the efficiency of different protocols. Our data indicate that WGA is suitable for the amplification of photosynthetic eukaryote genomes, but that biases are induced, reducing the diversity of the initial population. Nonetheless, this approach appears to be suitable for obtaining metagenomics data on microbial eukaryotic communities.

Introduction

Eukaryotic microorganisms, especially phytoplankton that is capable of carbon fixation, play important roles in oceanic waters. Analysis of phytoplankton communities to determine their distribution, diversity and specific role is fundamental to develop an understanding of how aquatic ecosystems function and evolve. Over the last 20 years, a number of studies have highlighted the major role of small eukaryotic phytoplankton ($< 3\,\mu m$) (Vaulot *et al*., 2008) in global carbon cycling in marine environments (Li, 1994; Liu *et al*., 2009; Jardillier *et al*., 2010), even though they are typically far less abundant than their prokaryotic counterparts (*Prochlorococcus* and *Synechococcus*). Despite their ecological importance, small eukaryotes have remained poorly described due to their size and to the lack of distinguishing morphological characteristics. In the last decade, the application of molecular approaches, especially the amplification, cloning and sequencing of the 18S rRNA genes in natural samples, has revealed the considerable diversity of small eukaryotic plankton and the existence of novel groups of sequences unrelated to cultured organisms. However, a major limitation of this type of approach is that environmental clone libraries generated with universal primers are typically dominated by heterotrophic organisms (Not *et al*., 2008). Thus, alternative approaches focusing on photosynthetic cells have been developed recently. These include studies targeting plastid genes (Fuller *et al*., 2006; Lepère *et al*., 2009), the use of specific primers for photosynthetic taxa (Viprey *et al*., 2008) and the construction of clone libraries from flow cytometry-sorted populations (Shi *et al*., 2009; Marie *et al*., 2010).

Genomics, i.e. the study of whole genomes, has been developing rapidly in marine sciences. Attention initially focused on marine prokaryotes such as *Prochlorococcus* (Rocap *et al*., 2003) because of the small size of their genome. More recently, the genomes of small microalgae such as the prasinophytes *Ostreococcus* and *Micromonas* or the diatom *Thalassiosira* have been deciphered (Armbrust *et al*., 2004; Derelle *et al*., 2006; Worden *et al*., 2009). Metagenomics, i.e. direct genomic sequencing of material sampled from the environment, allows the retrieval of genetic information on

© 2011 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

populations without cultivation (Wooley *et al.*, 2010). Metagenomic approaches have been used successfully to characterize prokaryotic communities in marine waters (Venter *et al.*, 2004). However, metagenomics is difficult to apply to eukaryotes because the filter-fractionated samples typically used are almost completely dominated by prokaryotic sequences (Massana *et al.*, 2008). Therefore, metagenomic analysis of eukaryotes requires the physical separation of eukaryotes from prokaryotes. Single-cell pipetting and flow cytometry sorting (e.g. Shi *et al.*, 2009) are two possible strategies to achieve this. However, these techniques provide very little material and due to the requirement for micrograms of DNA even for next-generation sequencing, preamplification is necessary.

In recent years, whole-genome amplification (WGA) of microbial populations based on multiple displacement amplification (MDA) has been developing (Binga *et al.*, 2008). On soil and sediment samples, MDA allows the generation of sufficient templates for 16S rRNA gene PCR and library construction (Gonzalez *et al.*, 2005; Abulencia *et al.*, 2006). Chen *et al.* (2008) showed that the combination of DNA stable isotope probing, WGA and metagenomics provided access to the genetic information of uncultivated methanotrophs. WGA has been applied successfully to amplify genetic material from a small number of cells or even from single cells (Zhang *et al.*, 2006; Rodrigue *et al.*, 2009), providing genetic data for uncultured organisms (Stepanauskas & Sieracki, 2007; Woyke *et al.*, 2009; Heywood *et al.*, 2010; Tripp *et al.*, 2010). Recently, Cuvelier *et al.* (2010) used flow cytometric sorting, followed by WGA to obtain genomic data on uncultured eukaryotic microorganisms.

In this methodological study, we optimized a WGA protocol to amplify the DNA of photosynthetic eukaryotes and successfully applied this protocol to samples obtained by flow cytometry sorting from the South-East Pacific Ocean.

Materials and methods

Cultures

Preliminary tests were performed on two cultures from the National Institute for Environmental Studies (NIES, Tsukuba, Japan) Microbial Culture Collection (http://mcc.nies.go.jp/): NIES-252 (*Nephroselmis astigmatica*) and NIES-1411 (*Micromonas pusilla*). Aliquots of 1, 10, 100 and 1000 cells were sorted by flow cytometry (EPICS Altra, Beckman Coulter) and immediately frozen at $-80\,^{\circ}C$.

A mix of 26 culture strains (cell size ranging from 2 to 100 μm) was prepared to simulate an environmental sample. Strains belonging to 16 classes (Table 1) were selected from the Roscoff Culture Collection (RCC, http://www.sb-roscoff.fr/Phyto/RCC/). The cell size and cell concentration of 400 mL cultures of each strain were quantified by flow cytometry (Cell Lab Quanta SC, Beckman Coulter). Subsamples of known volumes from each culture were mixed and the multistrain sample was diluted into sterile seawater (10 L final volume). The final concentrations of cells were calculated such that 50 mL of the culture mix would correspond to concentrations typically found in 10–15 L of seawater (the typical volume filtered for metagenomic analyses). For each culture, we computed the product of the final cell concentration by cell volume, which should be proportional to the number of rRNA gene copies per milliliter because rRNA gene copy number has been shown to be related to cell volume (Zhu *et al.*, 2005). Fifty milliliters of the mix was filtered onto 0.8 μm polycarbonate filters (47 mm diameter), flash frozen in liquid nitrogen and stored at $-80\,^{\circ}C$ until extraction.

Environmental samples

Sampling (Table 2) was performed in the surface layer and at the vicinity of the deep chlorophyll maximum at selected stations between October 26 and December 11, 2004 along a transect between the Marquesas Islands to Chile via Easter Island through the South-East Pacific Ocean during the BIOSOPE cruise (Claustre *et al.*, 2008). The region covered by this transect remains one the most sparsely sampled regions of the global ocean and corresponds to the most oligotrophic waters on Earth. This region is characterized by microbial communities with very low cell concentrations, particularly for photosynthetic picoeukaryotes, whose abundance is on average 600 cells mL^{-1} in the South Pacific Gyre. WGA optimization and reproducibility tests were also conducted on surface seawater samples collected at the SOMLIT-Astan site (48.461°N, 3.561°W) off Roscoff (Brittany, France). Seawater samples were collected using Niskin bottles mounted on a CTD frame. Samples were concentrated between 5- and 100-fold by tangential flow filtration using a 100 000 MWCO (Regenerated Cellulose – RC, ref. VF20C4) Vivaflow 200 cassette (Marie *et al.*, 2010). Concentrated samples were analyzed on board using a FACSAria flow cytometer (Becton Dickinson, San Jose, CA) equipped with a laser emitting at 488 nm and a 70-mm 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 (Marie *et al.*, 2010). The signal was triggered on the red fluorescence from chlorophyll. Photosynthetic eukaryotes were discriminated based on their side scatter and red fluorescence (see Shi *et al.*, 2009), and different populations were sorted in the 'purity' mode (Table 2). Cells were collected in Eppendorf tubes, and after centrifugation, the volume of the sorted samples was adjusted to 250 μL by

© 2011 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

Table 1. Amplification of a mixture of different eukaryotic microalgal strains to test for WGA biases

Strains	Class	Species or clade	Size (μm)	18S rRNA gene GC%	Initial concentration (× 1000 cells mL^{-1})	Final concentration (cells mL^{-1})	Concentration × cell volume	18S rRNA clones pre-WGA	18S rRNA clones post-WGA
RCC782	Bacillariophyceae	*Cylindrotheca closterium*	60 × 5	44.9	508.0	17 780	26670000		1
RCC1717	Bacillariophyceae	*Chaetoceros diadema*	10	44.9	258.0	9030	9030000	5	10
RCC91	Dinophyceae	*Scrippsiella trochoidea*	25	45.9	10.7	376	5878906.2		
RCC1453	Prymnesiophyceae	*Hyalolithus neolepis*	20	49.2	20.1	703	5628000		
RCC504	Eustigmatophyceae	*Nannochloropsis gaditana*	2.2	46.6	3486.0	469 269	4928950.1		2
RCC775	Bacillariophyceae	*Ditylum brightwellii*	100 × 20	47.4	3.3	117	4676000		4
Rsal	Cryptophyceae	*Rhodomonas salina*	5.7	46.0	409.5	14 332	2724742.6	9	1
RCC656	Prymnesiophyceae	*Chrysochromulina* sp.	3.5	49.0	1467.5	51 362	2336957.4		
RCC703	Bacillariophyceae	*Minutocellus* sp.	3.7	45.8	1156.5	40 477	2100584.5	3	3
RCC503	Pinguiophyceae	*Phaeomonas* sp.	2.5	47.4	3736.0	130 760	2018705.4	3	
RCC1537	Pavlovophyceae	*Pavlova lutheri*	3.3	48.4	1387.0	48 545	1728750.1	1	
RCC1563	Prasinophyceae	*Tetraselmis convolutae*	5.5	46.7	191.5	6702	1145818.3	6	4
RCC1216	Prymnesiophyceae	*Emiliania huxleyi*	3.5	51.2	673.0	23 555	984173.2		
RCC1491	Dinophyceae	*Pelagodinium beii*	6.5	45.4	91.5	3202	879486.6		
RCC475	Trebouxiophyceae	*Nannochloris* sp.	1.7	49.8	4642.5	162 487	855988.1	3	9
RCC287	Prasinophyceae	Clade VII	1.7	46.0	4197.5	146 912	760671.3		6
RCC2	Chlorophyceae	*Chlamydomonas reginae*	9.9	48.9	12.5	437	421938.2	1	
RCC21	Chrysophyceae	*Ochromonas* sp.	4.9	44.4	80.5	2817	323424.4	3	
TW15	Prymnesiophyceae	*Scyphosphaera apsteinii*	20	48.7	0.6	23	181160		
RCC1512	Dinophyceae	*Thoracosphaera heimii*	9.1	46.3	5.0	175	131007.3		
RCC1082	Dictyochophyceae	*Pseudochattonella verruculosa*	6.4	48.9	12.0	420	108048.9		
RCC239	Bolidophyceae	*Bolidomonas mediterranea*	1.7	44.6	296.0	10 360	50898.7		
RCC365	Cercozoa	*Partenskyella glossopodia*	2.3	49.8	110.0	3832	44239.2		
							Total number of clones	34	40
							Number of strains recovered	9	9

The table shows the measured cell size, GC% (18S rRNA gene, average for genus), initial mean cell concentration, final concentration per milliliter in the 10 L mixture, concentration multiplied by cell volume providing a proxy of rRNA copy number in the final mixture, 18S rRNA gene sequence numbers obtained in clone libraries constructed from DNA before (pre-WGA) and after WGA (post-WGA).

Table 2. BIOSOPE sample locations, photosynthetic picoeukaryote (PPE) abundances, size range of the sorted population (pico for picoplankton, micro for microplankton) and number of sorted cells

Sample code	Station	Longitude (°W)	Latitude (°S)	Trophic status	Depth (m)	Sorted PPE populations	Number of sorted cells
T19	STB1	134.10	11.74	Meso	25	Pico	103 074
T41	STB7	120.38	22.05	Oligo	40	Pico	125 000
T39	STB7	120.38	22.05	Oligo	175	Pico	106 000
T52	GYR2	114.00.	25.58	Oligo	5	Micro	540
T60	STB11	107.29	27.77	Oligo	0	Pico	171 440
T65	STB12	104.31	28.54	Oligo	40	Pico	123 000
T105	EGY2	91.45	31.84	Oligo	5	Micro	9100
T142	UPW1	73.37	34.00	Eutro	5	Pico	104 000
T149	UPW3	73.28	33.84	Eutro	30	Pico	233 000
T173	UPX2	72.47	34.63	Eutro	5	Pico	370 000

Meso, oligo and eutro are related to mesotrophic, oligotrophic and eutrophic regions of the BIOSOPE transect.

© 2011 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

adding filtered seawater. Samples were frozen in liquid nitrogen.

DNA extraction and WGA

For preliminary tests, NIES cultures were used directly for WGA without prior DNA extraction. For the RCC mix, filters were crushed (6 knocks s^{-1} for 1 min; FreezerMill 6700, Fisher Scientific, France). Approximately 1 g of material was obtained per filter. DNA was extracted using the Nucleospin RNAII kit (Macherey-Nagel, Hoerdt, France) and quantified using a Nanodrop ND-1000 Spectrophotometer (Labtech International, France). Extract quality was checked on an agarose gel (1.5%). DNA from the sorted environmental populations was extracted using a DNeasy blood and tissue kit (Qiagen, Courtaboeuf, France), as recommended by the manufacturer (see Shi *et al.*, 2009 for details).

WGA was carried out using the REPLI-g Mini kit (Qiagen) following the manufacturer's protocol. Lysis and neutralization buffers were, however, modified according to Gonzalez *et al.* (2005). All samples were treated with the same protocol, although DNA templates did not require a lysis step. Briefly, 1 μL of cells (corresponding to 500 cells) or 1 μL of DNA (corresponding to 3–5 ng of DNA) in 2.5 μL of phosphate-buffered saline was chemically lysed with the addition of 3.5 μL of an alkaline solution (400 mM KOH, 100 mM DTT, 10 mM EDTA) and incubated on ice for 10 min. Lysed samples were neutralized with 3.5 μL of neutralization buffer (2 mL 1 M HCl, 3 mL 1 M Tris-HCl). The product was used as the template in the WGA reaction. Reactions were carried out in 50 μL volumes. Reaction buffer (29 μL), water (9.5 μL) and 1 μL of Phi29 DNA polymerase were added to 10.5 μL of template and incubated at 30 °C for 16 h. A final incubation at 65 °C for 5 min inactivated the Phi29 DNA polymerase. Some samples were also subjected to a second round of amplification in order to obtain more DNA. Two microliters of the initial amplification reaction was used for the second round using the same protocol. Five microliters of the amplified product was run on an agarose gel (1%) in order to estimate the amplification efficiency. WGA is highly susceptible to contamination. Purity of the reagents is crucial and the level of care is similar to that needed for PCR reactions with low template quantities. We systematically used dedicated pipettes and applied standard methods to create work areas and instruments free of DNA contamination (in particular, the use of an UV hood). Appropriate blank controls (sterile water) were included for each experiment. In some cases, 10–15 reactions were performed and then pooled together. One microliter was used as a template for PCR/cloning. Amplicons were purified and concentrated using a Microcon YM-100 column (Millipore, Molsheim, France) or by ethanol precipitation.

Quantification of genomic DNA after WGA

After WGA, amplified products were visualized by agarose gel (1%) electrophoresis to assess whether the reaction was successful. In some cases, products obtained after WGA were analyzed by pulsed-field gel electrophoresis (PFGE) using the LEADER DR-II (Bio-Rad) system. The electrophoresis was conducted with a 1% agarose gel in TBE 0.5 ×, at 200 V for 20 h, with initial and final pulse parameters of 0.5 and 1.5 s, respectively. The use of a High Range DNA Ladder (Fermentas Life Sciences) allowed the evaluation of fragment sizes. DNA was stained with ethidium bromide (final concentration 0.5 μg mL^{-1}) for 10 min.

DNA was also quantified in the final reaction volume with Quanti-iT™ PicoGreen dsDNA (Invitrogen, Carlsbad, CA), a sensitive fluorescent stain suitable for quantifying double-stranded DNA that excludes nucleotides and single-stranded nucleic acids from the signal. The stain was used according to the manufacturer's instructions and DNA was quantified with a Tecan microplate reader (Tecan, Männedorf, Switzerland) using the MAGELLAN 5 software. Amplification levels were estimated by the ratio of DNA concentrations after and before WGA for each sample.

PCR reactions, cloning and sequencing

The full (or nearly full)-length 18S rRNA gene was PCR amplified using the eukaryotic primers Euk 328f and Euk 329r (Moon-van der Staay *et al.*, 2001) or 63f (5′-ACG-CTT-GTC-TCA-AAG-ATT-A-3′) and 1818r (5′-ACG-GAA-ACC-TTG-TTA-CGA-3′) (designed by M.K.). The PCR mixture (30 μL final volume) contained 1 μL of the amplicon with 0.5 μM final concentration of each primer and 15 μL HotStar Taq® *Plus* Master Mix (Qiagen). PCR reactions were performed as described previously (Viprey *et al.*, 2008) with an initial incubation step at 95 °C for 5 min for activation of the HotStar Taq® *Plus* DNA Polymerase. For samples T142 and T149, the general bacterial primers 8f (Martinez-Murcia *et al.*, 1995) and 1492r (Lane, 1991) were also used to amplify the 16S rRNA gene. PCR reactions were performed using the following program: initial denaturation at 95 °C for 5 min, 30 standard cycles of denaturation at 95 °C for 1 min, annealing at 55 °C for 1 min, extension at 72 °C for 1 min and a final extension at 72 °C for 5 min.

PCR products were cloned into pCR®2.1-TOPO® vectors and transformed into *Escherichia coli* competent cells following the manufacturer's instructions (Invitrogen). Sequencing reactions were performed with purified PCR products using Big Dye Terminator V3.1 (Applied Biosystems, Foster city, CA) and the primer Euk528f (Romari & Vaulot, 2004) for the 18S rRNA gene and the primer 8f for the 16S rRNA gene and run on an ABI prism 3100 sequencer (Applied Biosystems). Partial sequences were compared with those available in public databases with the NCBI BLAST

© 2011 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

web application (http://www.ncbi.nih.gov/BLAST/). Partial 18S rRNA gene sequences obtained from sorted BIOSOPE samples before and after WGA were clustered into distinct operation taxonomic units (OTUs) with the CD-HIT software (http://www.bioinformatics.org/cd-hit/) based on a 98% similarity threshold consistent with previous work (Romari & Vaulot, 2004; Shi *et al.*, 2009). No chimeras were detected among 18S rRNA gene sequences when using the Ribosomal Database project II program CHECK_CHIMERA (http://rdp.cme.msu.edu/). Partial 18S rRNA gene sequences were aligned with related sequences from public databases using the global alignment with free end gaps from GENEIOUS 4.8 software (http://www.geneious.com/, Biomatters Ltd, NZ). Alignments were analyzed by neighbor joining using GENEIOUS. Bootstrap values were estimated from 1000 replicates. The new sequences reported in this paper have been submitted to GenBank under the following accession numbers: HM474420–HM474786.

Results

Optimization of the WGA protocol

Amplification conditions

The commercial kit tested, REPLI-g Mini from Qiagen, is MDA based. Two denaturation protocols were tested in order to evaluate their influence on the size and yield of amplified DNA based on gel analysis. Chemical lysis produced, after amplification, high-molecular-weight fragments between 20 and 50 kbp (Fig. 1), whereas thermal denaturation generated lower fragment sizes. In addition, chemical lysis generated larger amounts of DNA. Typical yields were 800–1200 ng for each reaction and the success rate was on average 85% for approximately 120 reactions.

Starting material

The quantity of initial material is an important parameter. In some cases, it may be necessary to start from a very small number of cells, or even a single cell, in particular to obtain metagenomes from organisms not yet available in culture (Rodrigue *et al.*, 2009; Woyke *et al.*, 2009). In other cases, it may be relevant to start from a larger pool of cells, in order to assess the genetic diversity within an environmental population for example. We quantified the minimal quantity of small photosynthetic eukaryote cells required for efficient amplification. Tests were conducted on flow cytometry-sorted cells (1000, 100, 10 and 1) from *Micromonas* and *Nephroselmis*. The 18S rRNA gene was amplified after WGA by PCR using universal primers. The presence of a product was verified on an agarose gel and the product was sequenced. With one cell as the starting material, either 18S rRNA gene amplification was unsuccessful or the sequence obtained did not correspond to the original culture, but to a contaminant, in general a fungus. From 10, 100 and 1000 cells, however, 18S rRNA gene amplification was successful and the sequences obtained matched that of the initial culture. In subsequent experiments, all WGA reactions were undertaken from 400 to 500 cells or 3 to 5 ng of DNA.

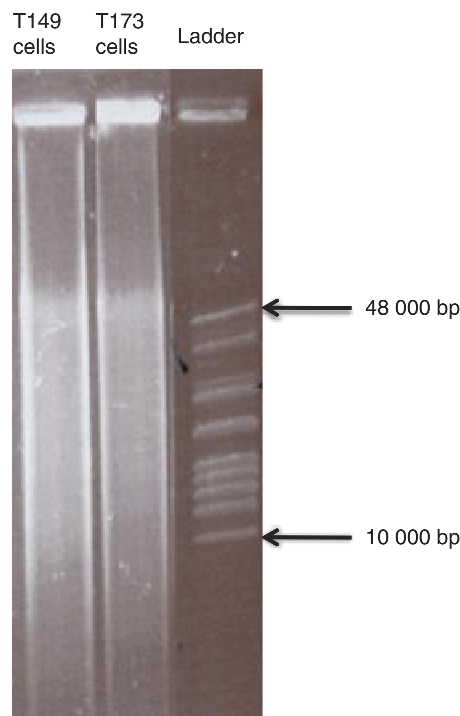

Fig. 1. PFGE analysis of 20 μL of the WGA product from two sorted South-East Pacific Ocean samples (T149 and T173).

Reproducibility

In order to test reproducibility, 10 WGA reactions were performed simultaneously on the same sample using the same amplification conditions. Amplification appeared to occur somewhat randomly (Fig. 2). Out of the 10 reactions, three did not yield any positive amplification, four reactions yielded a weak signal, while three reactions showed strong amplification.

Estimation of biases in amplification induced by WGA

Amplification of a mix of cultures

In order to obtain a reliable representation of an environmental sample, WGA must evenly amplify all of the genomes present. To determine the extent of biases induced by WGA, a laboratory mix of 26 eukaryotic culture strains was prepared. We compared the composition of 18S rRNA gene

© 2011 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

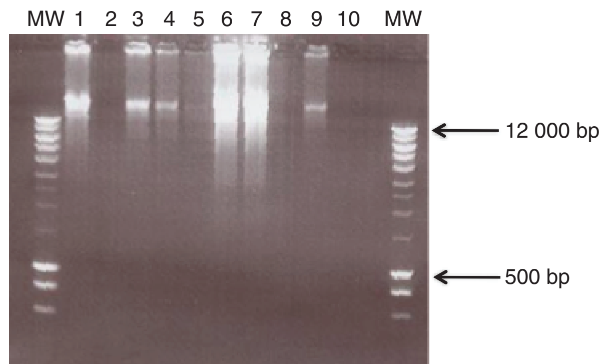

Fig. 2. Agarose gel electrophoresis of 5 μL of 10 replicate WGA products. WGA reactions were performed under the same conditions on photosynthetic picoeukaryote cells sorted by flow cytometry from an English Channel sample (SOMLIT-Astan site). MW, molecular weight marker.

clone libraries built from DNA obtained before (34 clones) and after (40 clones) WGA of the culture mix (Table 1). Of the nine taxa recovered in the 18S rRNA gene library constructed before WGA, only five were retrieved after WGA. However, four additional taxa were recovered after WGA, such that the total number of taxa obtained before and after WGA was equivalent. Only strains with high rRNA gene copies in the mix (as estimated from the product of cell volume and cell concentration; see Table 1 and Materials and methods) were represented in the 18S rRNA gene libraries. Strains with very low copy number (*Scyphosphaera*, *Thoracosphaera*, *Pseudochattonella*, *Bolidomonas*, *Partenskyella*) were absent both before and after WGA. This was the case for representatives of the Dinophyceae and the Prymnesiophyceae even when cells were abundant in the mix (e.g. *Scrippsiella trochoidea*). Four strains (two diatoms and two green algae) had comparable 18S rRNA gene representation before and after WGA. The cryptophyte *Rhodomonas*, the pinguiophyte *Phaeomonas*, the pavlovophyte *Pavlova*, the chlorophyte *Chlamydomonas* and the chrysophyte *Ochromonas* were evidently not well amplified because either fewer or no clones were recovered after WGA. In contrast, *Nannochloropsis* (Eustigmatophyceae), the diatoms *Cylindrotheca* and *Ditylum*, and especially the clade VII prasinophyte strain showed the opposite trend as they were present after WGA, but not recovered before.

Amplification from extracted DNA or directly from cells

One sorted sample from the South-East Pacific Ocean (T149, Table 2) was amplified by WGA either directly from cells or from extracted DNA. In the 18S rRNA gene clone library constructed from DNA before WGA, photosynthetic eukaryotes were represented by Prasinophyceae (*Bathycoccus*, *Ostreococcus* and *Micromonas*) and Chrysophyceae.

Table 3. 18S rRNA gene clone library composition (total number of sequences) constructed pre- and post-WGA from cells and DNA of the South-East Pacific Ocean sorted sample T149

		T149 (DNA)		T149 (cells)
Class	Genus or order	Pre-WGA	Post-WGA	Post-WGA
Prasinophyceae	*Bathycoccus*	11	29	7
Prasinophyceae	*Micromonas*	9		1
Prasinophyceae	*Ostreococcus*	1		
Chrysophyceae		4		24
Acantharea			1	
Dinophyceae	Syndiniales	6		

Syndiniales, which are likely to be heterotrophic parasites (Chambouvet *et al*., 2008), were also present (Table 3). After WGA from DNA, the 18S rRNA gene clone library was dominated by *Bathycoccus* (29 sequences out of 30), while after WGA from cells, the clone library was more diversified, containing Prasinophyceae (seven sequences of *Bathycoccus* and one sequence of *Micromonas*) and Chrysophyceae (24 sequences). In both cases, however, Syndiniales were not recovered after WGA (Table 3).

Second round of WGA

In some cases, for example when starting from a single cell, the quantity of DNA generated by a single round of WGA may not be sufficient for metagenomic sequencing (Rodrigue *et al*., 2009). The effect of a second round of amplification was tested on two sorted environmental samples (T149 and T173, Table 4). The quantity of DNA obtained after the second round was similar to that obtained for the first round of WGA. For sample T149, no clear difference was observed in the composition of the 18S rRNA gene clone libraries between the two rounds, *Bathycoccus* and Chrysophyceae being present in both cases. For sample T173, *Nannochloris* disappeared after the first round of WGA, whereas sequences of Chrysophyceae, which could not be detected before WGA or after the first round, were present after the second round.

Application of an optimized WGA protocol to natural populations from the South-East Pacific Ocean

Nine different DNA samples from sorted photosynthetic eukaryotes from the South-East Pacific Ocean were used to test the finalized protocol (Table 5). Most of the sequences in clone libraries constructed after WGA using bacterial primers for samples T142 and T149 were affiliated to *Proteobacteria* (data not shown). 18S rRNA gene clone libraries were constructed in order to compare diversity before and after WGA (at least for the dominant sequence types,

© 2011 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

Table 4. 18S rRNA gene clone library composition (total number of sequences) constructed from two South-East Pacific Ocean sorted samples (T149 and T173) after the first and second round of WGA

		T149 (cells)		T173 (DNA)		
Class	Genus	First round of WGA	Second round of WGA	before WGA	First round of WGA	Second round of WGA
Prasinophyceae	*Bathycoccus*	7	8	6	3	1
Prasinophyceae	*Micromonas*	1		6		
Prasinophyceae	*Ostreococcus*			1		
Trebouxiophyceae	*Nannochloris*				6	
Chrysophyceae		24	20			7

because the diversity in the clone libraries was not saturated). Partial 18S rRNA gene sequences were clustered into distinct OTUs based on a 98% similarity threshold (Table 5 and Fig. 3). Libraries from two samples corresponding to microplankton populations (T52 and T105) were dominated by Dinophyceae sequences and a large fraction of OTUs were similar before and after WGA (Table 5). For samples corresponding to smaller size plankton, WGA induced a change in diversity in some cases (Table 5). The phylogenetic tree of sequences obtained before and after WGA from samples T142 and T149 shows a clear decrease of diversity, with only *Bathycoccus* sequences recovered after WGA (Fig. 3). In samples T19 and T41, Prasinophyceae from clades VII and/or IX were present before and after WGA. In these two samples, the number of OTUs was lower in the post-WGA libraries and few OTUs were common between the two conditions. Clone library dominance changed before and after WGA in T39 from Syndiniales to Chrysophyceae (present before WGA, but not dominant), in T60 from Chrysophyceae to Bolidophyceae and in T65 from Prasinophyceae clade IX/Chrysophyceae to Prymnesiophyceae (*Phaeocystis*). In T39 and T65, the diversity was reduced after WGA and in T60 and T65 samples, the dominant group after WGA was not observed before WGA. Interestingly, WGA seems not to amplify (or only weakly) the heterotrophic Syndiniales initially present in samples T41, T39 and T65.

Discussion

In order to characterize natural communities of microorganisms by molecular approaches, it is often necessary to physically separate and concentrate specific groups of cells. This is especially true for eukaryotic microorganisms that constitute a minor part, at least in terms of abundance, of the community. Because separation techniques often result in small yields, WGA is a potentially promising approach for amplifying the genetic signal. The Phi29 polymerase used for WGA has exceptional strand displacement and the highest processivity reported for any DNA polymerase in the absence of cellular multisubunit complexes. This enzyme also exhibits an exonuclease activity that enables proofreading and has been shown to amplify DNA of up to 70 kb (Blanco *et al*., 1989). The latter property is particularly interesting for metagenomics (e.g. fosmid library construction, Chen *et al*., 2008). To date, WGA has been mostly applied to prokaryotes (Stepanauskas & Sieracki, 2007; Binga *et al*., 2008). In the present study, we used the commercial REPLI-g Mini kit after the modification of the denaturation buffer (Gonzalez *et al*., 2005). Amplification was successful in all tested samples containing photosynthetic eukaryotes. We were able to amplify either directly from cells sorted by flow cytometry (from as little as 10 cells) or from DNA extracted either from sorted cells or from cultures. DNA fragments obtained were of a high molecular weight (up to 50 kb). The yield of amplified DNA, quantified by PicoGreen, was lower than the manufacturer's claim (10 μg per reaction), ranging from 0.8 to 1.2 μg in a 50 μL final volume corresponding to a 115–170-fold amplification. This is coherent with the 50–160-fold amplification obtained for bacteria with the REPLI-g Mini kit (Bouzid *et al*., 2009; Woyke *et al*., 2009).

WGA has been shown to induce biases during amplification of the original DNA template. These biases are potentially due to several factors, such as the number of cells used in the reaction (Arakaki *et al*., 2010), GC%, chromosome length and the presence of repeat regions (Pinard *et al*., 2006). These biases may be more or less critical depending on the aim of the study. For the study of community structure starting from a large number of cells, for example sorted by flow cytometry, any bias in amplification of the gene of interest will make the data obtained after WGA difficult to interpret. In contrast, for obtaining metagenomic data, biases may be much less critical. For example, Rodrigue *et al*. (2009) were able to reconstruct the genome of individual *Prochlorococcus* cells, despite very large random variations in genome coverage following WGA. In the present study, we investigated the effect of certain parameters on amplification biases for samples containing a range of genotypes using the 18S rRNA gene as a marker. Such a marker can indicate whether genotypes are over- or under-amplified, but it cannot reveal uneven amplification

© 2011 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

Table 5. 18S rRNA gene clone library composition constructed on pre- and post-WGA of nine environmental sorted samples from the South-East Pacific Ocean

	Sample	T19		T39		T41		T60		T65		T142		T149		T52		T105	
	Sorted population	Pico		Pico		Pico		Pico		Pico		Pico		Pico		Micro		Micro	
Class	Genus or order	Pre-WGA	Post-WGA	Pre-WGA	Post-WGA	Pre-WGA	Post-WGA	Pre-WGA	Post-WGA	Pre-WGA	Post-WGA	Pre-WGA	Post-WGA	Pre-WGA	Post-WGA	Pre-WGA	Post-WGA	Pre-WGA	Post-WGA
Prasinophyceae	*Bathycoccus*			2									31	11	29				
Prasinophyceae	*Micromonas*											9		9			1		1
Prasinophyceae	*Ostreococcus*											10		1					
Prasinophyceae	Clade VII	17	25		1					4									
Prasinophyceae	Clade IX	2	6	3		18	30	2		13									
Trebouxiophyceae	*Nannochloris*		5																
Chrysophyceae				4	23		1	14	4	9				5					
Prymnesiophyceae	*Chrysochromulina*							2	1	4	2								
Prymnesiophyceae	*Phaeocystis*										26								
Acantharea															1				
Bolidophyceae	*Bolidomonas*								22										
Bacillariophyceae		1																	
Pelagophyceae	*Pelagomonas*			1	4														
Cryptophyceae				1															
Bicosoecida				2															
Dictyochophyceae								1											
Dinophyceae																5	20	9	9
Dinophyceae	Syndiniales			26	1	3				3	1			6					
	Number of OTUs (98%)	8	4	28	22	10	7	12	13	17	3	5	5	19	2	3	4	7	6
	Common OTUs	2		7		1		5		1		0		1		2		5	
	Unique OTUs	6	2	21	15	9	6	7	8	16	2	5	5	18	1	1	2	2	1

Numbers of OTUs (98% of similarity) for each sample pre- and post-WGA, numbers of common OTUs obtained in both conditions, and number of unique OTUs for each condition. Pico and micro correspond to the size range of the sorted population.

© 2011 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

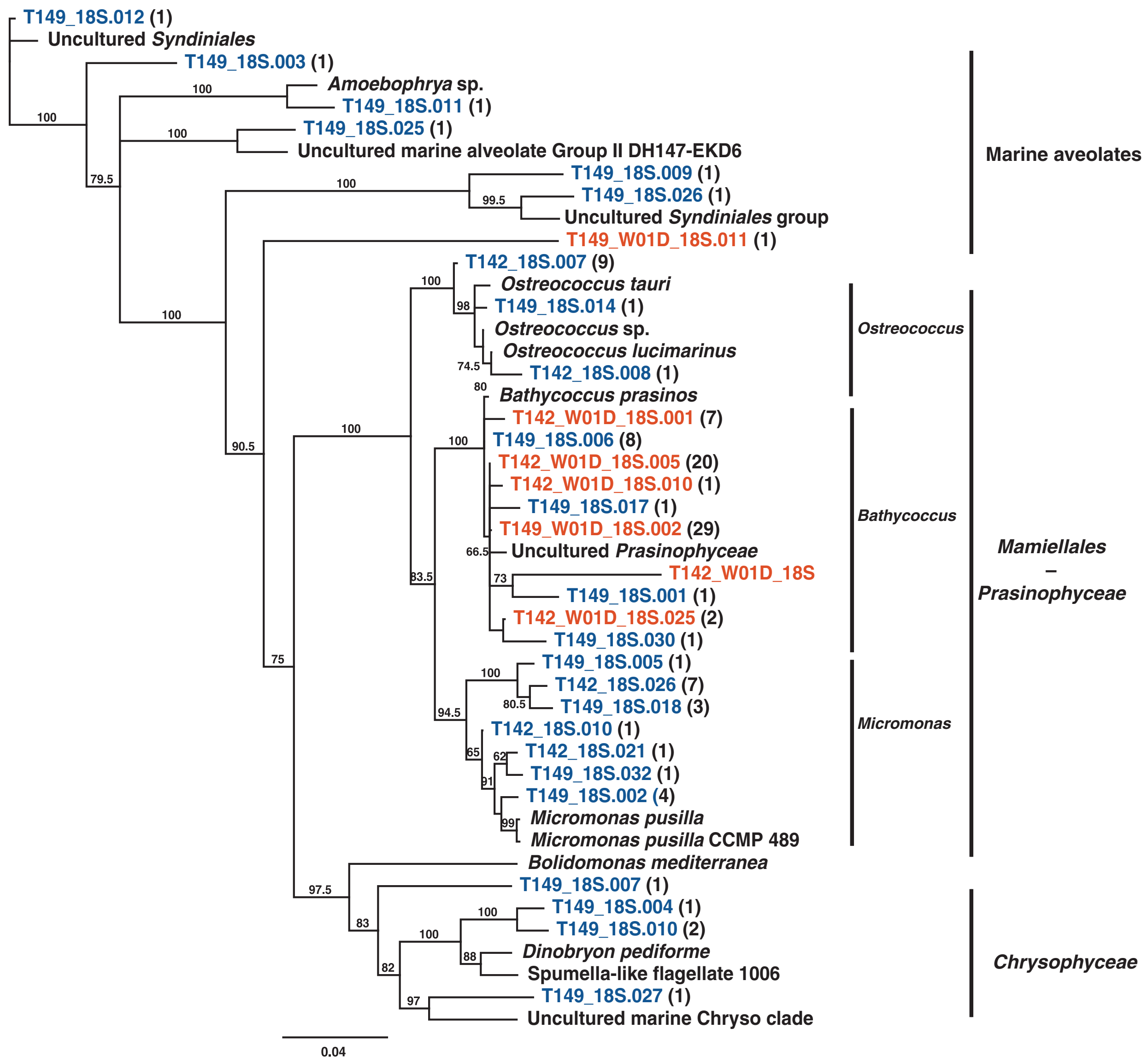

Fig. 3. Phylogenetic tree of 18S rRNA genes from South-East Pacific Ocean environmental samples T142 and T149. OTUs (98% similarity) and the number of clones associated (in parentheses) obtained from nonamplified DNA material are highlighted in blue. OTUs obtained from WGA-amplified DNA appear in red. Bootstrap values correspond to the neighbor-joining method (1000 replicates, values $> 60\%$ shown).

across the genome of a given genotype (as in Rodrigue *et al.*, 2009). Most of our tests of amplification biases were performed on DNA extracted from photosynthetic eukaryotes, either from a culture mix or from natural samples sorted by flow cytometry. In most cases, we observed significant differences in the composition of clone libraries before and after WGA.

The culture mix offers the advantage over natural samples of knowing precisely the community composition and hence being able to estimate initial gene copy numbers. Although the number of taxa recovered from the culture mix was identical before and after WGA, the composition was quite different (Table 1), as observed previously for soil bacteria (Abulencia *et al.*, 2006). It should be noted that even before WGA, clone library composition in fact poorly reflected the initial mixture composition in terms of relative rRNA gene abundance. In particular, some taxa that were abundant in the mix were never recovered in clone libraries, either before or after WGA. This was the case in particular for the Prymnesiophyceae, whose rRNA gene is known to be poorly amplified by universal primers when mixed with other groups (Potvin & Lovejoy, 2009; Marie *et al.*, 2010). This bias could be due to the slightly higher GC% of the Prymnesiophyceae 18S rRNA gene (Table 1), which may also be unfavorable for WGA (Pinard *et al.*, 2006). However, this explanation does not hold for the 18S rRNA gene of

© 2011 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

dinoflagellates, the GC% of which falls in the same range as for other strains recovered in clone libraries. It is noteworthy that taxa that appeared in clone libraries after WGA (*Cylindrotheca*, *Nannochloropsis*, *Ditylum*) were more abundant in the initial mixture (at least in terms of estimated copies of the rRNA gene) than those present in clone libraries constructed before WGA.

In contrast to the culture mix, the initial community composition of sorted samples from the South-East Pacific Ocean was not known. For these samples, clone libraries on microplankton populations had almost the same composition before and after WGA according to the clustering analysis (T52 and T105, Table 5). In picoplankton samples, however, either some taxa not present before WGA dominated clone libraries after WGA (*Bolidomonas* in T60, *Phaeocystis* in T65, *Bathycoccus* in T142, Table 5 and Fig. 3) or, when the same group was retrieved, OTUs were different before and after WGA (T19 and T41). Syndiniales (*Dinophyceae*), which are heterotrophic parasites of autotrophic plankton species, in particular dinoflagellates (Chambouvet *et al*., 2008), were probably a minor component of the sorted populations because only photosynthetic organisms were selected based on chlorophyll fluorescence. In some cases, however, they were very prominent in clone libraries before WGA, but, interestingly, they were eliminated after WGA.

It is difficult to assess why some samples exhibit biases in contrast to others. For example, *Phaeocystis*, being a Prymnesiophyceae with a relatively high GC% of the 18S rRNA gene (50.1%), should not have been favored by WGA in sample T65. Although some random factors may be implicated (Rodrigue *et al*., 2009), in the case of the natural samples, we started from DNA extracted from a population of several hundred (microplankton) to several hundred thousand (picoplankton) cells and several (at least 10 reactions when possible) independent WGA reactions were pooled together in order to minimize random effects. On the positive side, it should be noted that we did not detect any chimeras in our clone libraries.

Our study helps to make some basic recommendations for WGA amplification of photosynthetic eukaryotes. Great care should be used to avoid contamination: in our experience, working under a UV-equipped PCR hood resolved most problems. Also, it may be preferable when possible to start from cells rather than from extracted DNA, because this may decrease biases. Other amplification kits may be less prone to biases (F. Humily, pers. commun.). In any case, biases can easily be checked by constructing clone libraries from one or several genes before and after WGA as shown in this study, or by other more rapid fingerprinting techniques such as terminal restriction fragment length polymorphism (F. Humily, pers. commun.). Despite these biases, WGA of uncultured microorganisms may be the only way to obtain valuable metagenomic data on these organisms, as demonstrated recently by Cuvelier *et al*. (2010), who successfully used WGA to obtain genomic data on small uncultured prymnesiophytes sorted by flow cytometry from subtropical North Atlantic waters.

Acknowledgements

This research was supported by the following programs: PICOFUNPAC (ANR biodiversité 06-BDIV-013), JST-CNRS PhytoMetagene, CPER Souchothèque de Bretagne (RCC), the ASSEMBLE EU FP7 Infrastructure project (227799) and the EU BioMarKs project (Biodiversity of Marine Eukaryotes). We thank D. Marie for flow cytometry analysis and sorting, X.L. Shi for providing DNA extracts, S. Mazard for her precious help in bioinformatics and F. Humily for his help with PFGE analysis.

References

Abulencia CB, Wyborski DL, Garcia JA *et al.* (2006) Environmental whole genome amplification to access microbial population in contaminated sediments. *Appl Environ Microb* **72**: 3291–3301.

Arakaki A, Shibusawa M, Hosokawa M & Matsunaga T (2010) Genomic DNA preparation from a single species of uncultured magnetotactic bacterium by multiple displacement amplification. *Appl Environ Microb* **76**: 1480–1485.

Armbrust EV, Berges JA, Bowler C *et al.* (2004) The genome of the diatom *Thalassiosira pseudonana*: ecology, evolution, and metabolism. *Science* **306**: 79–86.

Binga EK, Lasken RS & Neufeld JD (2008) Something from (almost) nothing: the impact of multiple displacement amplification on microbial ecology. *ISME J* **2**: 233–241.

Blanco L, Bernad A, Lazaro JM, Martin G, Garmendia C & Salas M (1989) Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. Symmetrical mode of DNA replication. *J Biol Chem* **264**: 8935–8940.

Bouzid M, Heavens D, Elwin K *et al.* (2009) Whole genome amplification (WGA) for archiving and genotyping of clinical isolates of *Cryptosporidium* species. *Parasitology* **137**: 27–36.

Chambouvet A, Morin P, Marie D & Guillou L (2008) Control of toxic marine dinoflagellate blooms by serial parasitic killers. *Science* **322**: 1254–1257.

Chen Y, Dumont MG, Neufeld JD *et al.* (2008) Revealing the uncultivated majority: combining DNA stable-isotope probing, multiple displacement amplification and metagenomic analyses of uncultivated *Methylocystis* in acidic peatlands. *Environ Microbiol* **10**: 2609–2622.

Claustre H, Sciandra A & Vaulot D (2008) Introduction to the special section bio-optical and biogeochemical conditions in the South East Pacific in late 2004: the BIOSOPE program. *Biogeosciences* **5**: 679–691.

© 2011 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved

Cuvelier ML, Allen AE, Monier A *et al.* (2010) Targeted metagenomics and ecology of globally important uncultured eukaryotic phytoplankton. *P Natl Acad Sci USA* **107**: 14679–14684.

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

Fuller N, Campbell C, Allen DJ *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.

Gonzalez J, Portillo MJ & Saiz-Jimenez C (2005) Multiple displacement amplification as a pre-polymerase chain reaction (pre-PCR) to process difficult to amplify samples and low copy number sequences from natural environments. *Environ Microbiol* **7**: 1024–1028.

Heywood JL, Sieracki ME, Bellows W, Poulton NJ & Stepanauskas R (2010) Capturing diversity of marine heterotrophic protists: one cell at a time. *ISME J*, DOI: 10.1038/ismej.2010.155.

Jardillier L, Zubkov MV, Pearman J & Scanlan DJ (2010) Significant CO_2 fixation by small prymnesiophytes in the subtropical and tropical northeast Atlantic Ocean. *ISME J* **4**: 1180–1192.

Lane DJ (1991) 16S/23S rRNA sequencing. *Nucleic Acid Techniques in Bacterial Systematic* (Stackebrandt E & Goodfellow M, eds), pp. 115–117. John Wiley, New York.

Lepère C, Vaulot D & Scanlan DJ (2009) Photosynthetic picoeukaryote community structure in the South East Pacific Ocean encompassing the most oligotrophic waters on Earth. *Environ Microbiol* **11**: 3105–3117.

Li WKW (1994) Primary production of prochlorophytes, cyanobacteria and eucaryotic ultraphytoplankton: measurements from flow cytometric sorting. *Limnol Oceanogr* **39**: 169–175.

Liu H, Probert I, Uitz J, Claustre H, Aris-Brosou FM, Not F & De Vargas C (2009) Extreme diversity in non-calcifying haptophytes explains a major pigment paradox in open oceans. *P Natl Acad Sci USA* **106**: 12803–12808.

Marie D, Shi X, Rigaut-Jalabert F & Vaulot D (2010) Diversity of small photosynthetic eukaryotes in the English Channel from samples sorted by flow cytometry. *FEMS Microbiol Ecol* **72**: 165–178.

Martinez-Murcia AJ, Acinas SG & Rodriguez-Valera F (1995) Evaluation of prokaryotic diversity by restrictase digestion of 16S rDNA directly amplified from hypersaline environments. *FEMS Microbiol Ecol* **17**: 247–256.

Massana R, Karniol B, Pommier T, Bodaker I & Bej O (2008) Metagenomic retrieval of a ribosomal DNA repeat array from an uncultured marine alveolate. *Environ Microbiol* **10**: 1335–1343.

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

Not F, Latasa M, Scharek R *et al.* (2008) Phytoplankton diversity across the Indian Ocean: a focus on the picoplanktonic size fraction. *Deep-Sea Res Pt I* **55**: 1456–1473.

Pinard R, Winter A, Sarkis GJ *et al.* (2006) Assessment of whole genome amplification-induced bias through high-throughput, massively parallel whole genome sequencing. *BMC Genomics* **7**: 216–222.

Potvin M & Lovejoy C (2009) PCR-based diversity estimates of artificial and environmental 18S rRNA gene libraries. *J Eukaryot Microbiol* **56**: 174–181.

Rocap G, Larimer FW, Lamerdin J *et al.* (2003) Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424**: 1042–1047.

Rodrigue S, Malmstrom RR, Berlin AM, Birren BW, Henn MR & Chisholm SW (2009) Whole genome amplification and *de novo* assembly of single bacterial cells. *PLoS One* **4**: e686.

Romari K & 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.

Shi X, Marie D, Jardillier L, Scanlan DJ & Vaulot D (2009) Groups without cultured representatives dominate eukaryotic picophytoplankton in the oligotrophic South East Pacific Ocean. *PLoS One* **4**: 7657.

Stepanauskas R & Sieracki ME (2007) Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. *P Natl Acad Sci USA* **104**: 9052–9057.

Tripp HJ, Bench SR, Turk KA *et al.* (2010) Metabolic streamlining in an open-ocean nitrogen-fixing cyanobacterium. *Nature* **464**: 90–94.

Vaulot D, Eikrem W, Viprey M & Moreau H (2008) The diversity of small eukaryotic phytoplankton ($\leq$ 3 μm) in marine ecosystems. *FEMS Microbiol Rev* **32**: 795–820.

Venter JC, Remington K, Heidelberg JF *et al.* (2004) Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66–74.

Viprey M, Guillou L, Ferréol M & Vaulot D (2008) Wide genetic diversity of picoplanktonic green algae (Chloroplastidia) in the Mediterranean Sea uncovered by a phylum-biased PCR approach. *Environ Microbiol* **10**: 1804–1822.

Wooley JC, Godzik A & Friedberg I (2010) A primer on metagenomics. *PLoS Comput Biol* **6**: e1000667.

Worden AZ, Lee JH, Mock T *et al.* (2009) Green evolution and dynamic adaptations revealed by genomes of the marine picoeukaryotes *Micromonas*. *Science* **324**: 268–272.

Woyke T, Xie G, Copeland A *et al.* (2009) Assembling the marine metagenome, one cell at a time. *PLoS One* **4**: e5299.

Zhang K, Martiny AC, Reppas NB, Barry KW, Malek J, Chisholm SW & Church GM (2006) Sequencing genomes from single cells by polymerase cloning. *Nat Biotechnol* **24**: 680–686.

Zhu F, Massana R, Not F, Marie D & Vaulot D (2005) Mapping of picoeucaryotes in marine ecosystems with quantitative PCR of the 18S rRNA gene. *FEMS Microbiol Ecol* **52**: 79–92.

© 2011 Federation of European Microbiological Societies
Published by Blackwell Publishing Ltd. All rights reserved