

## ORIGINAL ARTICLE

# Improved group-specific PCR primers for denaturing gradient gel electrophoresis analysis of the genetic diversity of complex microbial communities

Martin Mühling<sup>1,3</sup>, John Woolven-Allen<sup>1,3</sup>, J Colin Murrell<sup>2</sup> and Ian Joint<sup>1</sup>

<sup>1</sup>*Plymouth Marine Laboratory, The Hoe, Plymouth, UK and* <sup>2</sup>*Department of Biological Sciences, University of Warwick, Coventry, UK*

Phylum- and class-specific PCR primers were tested for the production of clone libraries and for denaturing gradient gel electrophoresis (DGGE) analysis of complex bacterial communities. Primers were designed to specifically amplify 16S rRNA gene fragments of the phyla *Bacteroidetes*, *Planctomycetes* and *Firmicutes*, of three classes of the phylum *Proteobacteria*, the *Alphaproteobacteria*, *Betaproteobacteria* and *Gammaproteobacteria*, and of the *Cyanobacteria* (including chloroplast 16S rRNA genes). The specificity of the seven primer pairs was tested by producing clone libraries from environmental DNA samples from mesotrophic (Norwegian coastal) and oligotrophic (Northern Atlantic Gyre) environments. Five of the seven primer pairs specifically amplified target 16S rRNA gene sequences. Exceptions were the *Betaproteobacteria*- and *Firmicutes*-specific primers, which were relatively successful with coastal water mesocosm samples but less so with the Northern Atlantic Gyre sample. Phylogenetic analysis of sequences from the *Gammaproteobacteria* clone library revealed that the coastal sample yielded a number of clones that clustered within clades that belong to the oligotrophic marine *Gammaproteobacteria* (OMG) group, indicating that this group is not confined exclusively to the oligotrophic environment. Comparison of the bacterial diversity of the environmental DNA sample from the coastal and the open ocean using a two- or three-step nested PCR-DGGE process revealed significant differences in the bacterial communities. The application of the group-specific primers provides a higher resolution genetic fingerprinting approach than existing DGGE primer sets.

*The ISME Journal* (2008) 2, 379–392; doi:10.1038/ismej.2007.97; published online 13 March 2008

**Subject Category:** microbial population and community ecology

**Keywords:** 16S rRNA gene; DGGE; clone libraries; genetic diversity; bacteria; group-specific PCR primers

## Introduction

The study of biodiversity of marine microbial communities has become commonplace in microbial ecology, with the 16S ribosomal RNA (rRNA) gene being the most frequently used phylogenetic marker (Amann *et al.*, 1995; Pernthaler and Amann, 2005). This marker gene has revealed the great genetic diversity of bacteria that is now assembled in the Ribosomal Database Project-II (RDP-II; <http://rdp.cme.msu.edu/>). This currently stores approximately 418 000 partial 16S rRNA gene sequences, though it is likely that there is some

degree of redundancy and sequence anomalies within the database (Ashelford *et al.*, 2005). This genetic diversity also means that it is very unlikely that the whole range of diversity can be detected in a single sample, especially where communities are complex. Attempts to estimate the total bacterial community diversity have used partial analysis of the total community (that is clone library screening) combined with theoretical models. This approach, however, does not reveal the identity of the less abundant components of the assemblage. Increasing the number of clones per clone library has been successful in detecting novel bacterial clades (for example, Chouari *et al.*, 2005) or comparing different marine environments in terms of bacterial community composition (for example, Rappé *et al.*, 2000; Zaballos *et al.*, 2006). However, despite the decreasing costs for nucleotide sequencing, the success of this approach is still limited because of the huge scale of bacterial diversity — perhaps as

Correspondence: M Mühling, Plymouth Marine Laboratory, Prospect Place, The Hoe, Plymouth PL1 3DH, UK.  
E-mail: mmuh@pml.ac.uk

<sup>3</sup>These authors contributed equally to this work.

Received 19 July 2007; revised 9 October 2007; accepted 10 October 2007; published online 13 March 2008

many as  $2 \times 10^6$  different species in the oceans (Curtis *et al.*, 2002). The recent arrival of pyrosequencing of 16S rRNA tags (Sogin *et al.*, 2006) may represent an alternative because of the much lower cost per sequence; but pyrosequencing still does not allow analyses and comparison of the bacterial diversity in different environments on a routine basis.

Alternative approaches, such as denaturing gradient gel electrophoresis (DGGE), are routinely used to determine diversity because they avoid large-scale sequencing efforts. However, these are also likely to detect only a small subfraction of the total diversity. The use of bacterial PCR primers is likely to miss minor fractions of the microbial community because most of the PCR product will be composed of the more abundant species. Faint DNA bands on DGGE gels are unlikely to be detected or their identity determined. To overcome this limitation and to detect less abundant sequence clones, Holben *et al.* (2004) fractionated 16S rRNA gene sequences from a microbial community according to their G + C content before DGGE analysis. However, with the possible exception of the high G + C-containing *Actinobacteria*, this method has limited application and does not separate bacteria by phylogeny. Combining bromodeoxyuridine immunocapture and DGGE has been proposed to separate the DNA of the actively growing bacteria from the rest of the environmental DNA (Hamasaki *et al.*, 2007). However, this results in the analysis of subgroups of bacteria that are not defined on phylogenetic criteria, and does not allow screening of the whole range of genetic bacterial diversity.

An alternative approach is the application of group-specific oligonucleotide probes. This has mainly been used in combination with fluorescence *in situ* hybridization (FISH). FISH has been a very effective technique but, in contrast to the analysis of clone libraries produced from PCR fragments, FISH cannot reveal unknown bacterial groups. It only detects and quantifies those bacteria that the probes were designed to detect. Group-specific PCR primers have already been used for the analysis of complex microbial communities. For example, Mehling *et al.* (1995) designed and used PCR primers for the 16S rRNA gene sequences of *Streptomyces*. Driven by the biotechnological potential of compounds and activities found in *Actinobacteria* in general, Stach *et al.* (2003) developed new primers to detect a wide range of *Actinobacteria* in terrestrial and marine environments. Cyanobacterial and chloroplast 16S rRNA gene fragments can also be amplified specifically (Nübel *et al.*, 1997). PCR primers have also been designed to amplify more defined groups of bacteria, such as those of the genus *Pseudomonas* (for example, Widmer *et al.*, 1998). Many studies on pathogens have focused on well-defined species or strains of pathogens in the environment (for example, Miyamoto *et al.*, 1997). Other microbial ecology

studies have involved PCR primers specific for functional groups, such as the ammonia-oxidizing bacteria (McCaig *et al.*, 1994; Freitag and Prosser, 2003; Calvo *et al.*, 2004) and sulphate-reducing bacteria (Stubner, 2004; Dar *et al.*, 2005). The study by Dar *et al.* (2005) provides also a convincing case that group-specific primers can be used in combination with DGGE to detect less abundant bacterial groups in a complex community. The novelty of that approach was to use group-specific primers in a PCR followed by reamplification of the PCR product with nested, universal bacterial DGGE primer. This overcomes the difficulty in designing novel DGGE primers and establishing the optimal conditions for those primers (Dar *et al.*, 2005).

As 16S rRNA gene sequence databases continue to grow, group-specific PCR primers must be continually re-evaluated for their specificity and range of sequence matches (Baker *et al.*, 2003). Furthermore, new and more sophisticated software packages, such as PRIMROSE (Ashelford *et al.*, 2002) and ARB (Ludwig *et al.*, 2004), can help researchers to design 16S rRNA gene probes. These software packages use different sequence databases and each has its specific strengths. For example, as well as differences in the algorithm used for searching for priming sites, it is possible to use PRIMROSE to design degenerate probes, a function that is not available in ARB. ARB, however, allows the search for probes to be based on a greater number of specific parameters.

To analyse the total bacterial community during changing environmental conditions and seasons, we have developed group-specific primers for seven different taxonomic groups. Apart from expensive large-scale sequencing of clone libraries, DGGE remains one of the few fingerprinting techniques currently available that allows analysis of the whole microbial community and identification, by sequence analysis of DGGE bands, of specific members of that community. Therefore, we used a similar DGGE approach to that of Dar *et al.* (2005), that is reamplification of the bacterial group-specific PCR products with a nested or semi-nested universal DGGE-PCR primer set. We tested these primers for their specificity and applicability by sequence analysis of clone libraries and comparative DGGE analysis of total environmental DNA from a coastal nutrient-rich and an oligotrophic open ocean environment.

## Materials and methods

### PCR primer design

PCR primers were designed using the ARB (Ludwig *et al.*, 2004) and PRIMROSE (Ashelford *et al.*, 2002) programs; the latter is part of the TOOLKIT software package (<http://www.cf.ac.uk/biosi/research/biosoft>). The name given to each primer consists of a short name indicating the target group, followed by

a number, representing the position of the first base of the primer within the *Escherichia coli* 16S rRNA gene sequence and 'f' or 'r' indicating whether the primer is the forward or reverse primer, respectively. The sequences of the primers and their specificity are summarized in Table 1. The number of target and non-target matches of the primers was tested *in silico* using the PROBE MATCH function within the RDP-II database (<http://rdp.cme.msu.edu>) and the ROSE function of TOOLKIT which identifies the number of sequences, among the 15 104 *Bacterial* sequences in its database, to which the primer sequence is homologous.

A nested *Bacteria* primer pair was selected for each group-specific primer pair to be used in a nested re-PCR (Table 2). Only one nested *Bacteria* primer was available in the cases of the *Betaproteobacteria*-, the *Bacteroidetes*- and the *Cyanobacteria*/chloroplast-specific primers. Therefore, one of the group-specific primers was paired with the *Bacteria* primer for a semi-nested re-PCR (Table 2), and a GC clamp was attached to the 5'-position of one of the two primers (indicated in the primer name with '-GC'; Table 2).

### Sampling

Samples from two contrasting environments were used in this study — a mesotrophic coastal and an oligotrophic open ocean environment. The coastal sample was collected in May 2003 from a 20-m<sup>3</sup> mesocosm after 12 days of incubation during the 'Pelagic Ecosystem CO<sub>2</sub> Enrichment Study' (PeECE; <http://peece.ifm-geomar.de/>) at the EU Large-Scale-Facility, University of Bergen, Norway, located at Espeland in the Raunefjord (60.27°N 5.22°E), 20 km south of Bergen. During the experiment, nitrate, phosphate and silicate were added at typical winter concentrations to stimulate a phytoplankton bloom. The open ocean, oligotrophic sample was obtained from 15 m depth in the Northern Atlantic Gyre (35°N 20°W) collected during the Atlantic Meridional Transect (AMT-15) cruise in September 2004 (Robinson *et al.*, 2006).

Samples (4 l) from the mesocosm and samples (7 l) from the Northern Atlantic Gyre, respectively, were filtered through Sterivex cartridges with 0.2 µm pore size filter membrane (Millipore, Watford, UK). The cartridges were stored at -70 °C until analysis. Total environmental DNA was isolated from the cartridges using the method described by Jameson *et al.* (2007). This involved lysozyme treatment with 1.5 ml of lysis buffer (200 mM Tris-HCl, pH 8.0; 1.5 mg lysozyme; 125 mM EDTA), incubated at 37 °C on a rotary shaker (10 r.p.m.) for 30 min, with a 90° turn of the cartridge every 6 min to ensure coverage of the whole inner surface area of the cartridge. After addition of 0.25 ml of proteinase K lysis buffer (to a final concentration of 1.25% w/v SDS, 300 µg proteinase K, 70 mM NaCl), the cartridge was incubated at 37 °C for 15 min followed by another

60 min at 55 °C, again with regular rotation of the cartridge. Subsequently, the lysis mix was transferred to microcentrifuge tubes and extracted twice with one volume of chloroform/isopropanol (24:1). The DNA was precipitated (1 vol isopropanol; 0.4 vol 7.5 M ammonium acetate), and the pellet was washed with 70% (v/v) ethanol. The DNA pellet was resuspended in sterile, double-distilled water.

### PCR and preparation of 16S rRNA gene clone libraries

Clone libraries were prepared from PCR-amplified 16S rRNA gene fragments for each of the group-specific primer pairs. A nested PCR approach was required for three of the primer pairs to obtain sufficient PCR product for subsequent cloning (see Table 1). In these cases, aliquots of the PCR products obtained with the *Bacteria* primer pair 9bfm/1512uR were used as templates for a reamplification with a nested group-specific primer set. A typical PCR reaction was carried out in 25-µl volumes and contained 2 mmol l<sup>-1</sup> MgCl<sub>2</sub>, 200 µmol l<sup>-1</sup> dNTPs, 2.5 U of *Taq* DNA polymerase (Promega, Southampton, UK) and 50 nmol l<sup>-1</sup> of each of the primers. The temperatures for primer annealing were dependent on the primer pair used (summarized in Table 1). They were determined using the equation  $T_m = (4 \times GC) + (2 \times AT)$ . A range of annealing temperatures (ATs) similar to the calculated  $T_m$  were then tested in a gradient PCR to determine empirically the temperature that resulted in the most specific PCR product. All PCRs used the same basic cycle protocol except for the AT: following an initial denaturation step of 4 min at 96 °C, 30 PCR cycles were performed (96 °C for 60 s, AT for 60 s, 74 °C for 60 s) followed by a final extension step at 74 °C for 10 min. The extension time was increased to 90 s for primer pair 9bfm/1512uR. In all cases, the PCR yielded only specific products, that is single bands as judged by electrophoresis of the PCR products on agarose gels.

Aliquots of the products from PCRs or re-PCRs were cloned into a TA vector using the pGEM-T Easy Vector System I cloning kit (Promega). Twenty or 50 clones were picked from each of the group-specific clone libraries prepared from the mesocosm and the Northern Atlantic Gyre samples, respectively, and the 16S rRNA gene fragments were reamplified using vector primers (M13). The PCR products were used for the sequencing reactions.

### Sequence analysis

PCR products were treated with ExoSapIT (Amersham Biosciences, Little Chalfont, UK) according to the manufacturer's instructions and used directly for sequence analysis. Nucleotide sequencing was performed using the BigDye Terminator v3.1 cycle sequencing kit (ABI). An M13 primer was used in the cycle sequencing reaction. Sequences were analysed on an ABI 3100 automatic sequencer.

**Table 1** Summary of group-specific 16S rRNA gene PCR primers, their specificity towards taxonomic groups as revealed by *in silico* analysis and the annealing temperatures used in the PCR reactions

Primer <sup>a</sup>	Target group	Sequence (5' → 3')	<i>Escherichia coli</i> position	Includes variable regions	Identical matches within target group <sup>b</sup>	Percentage matches within the target group (%) <sup>c</sup>	Matches outside target group <sup>b</sup>	AT group-specific PCR for clone libraries (°C)	16S rRNA gene fragment used for 'Probe Match' <sup>d</sup>	Reference
Alf28f <sup>e</sup>	Alphaproteobacteria	ARCGAACGCTGGCGGCA	28–44	V1–V4	891	81.2 (83.6)	13	69	1–100	Ashelford <i>et al.</i> (2002) <sup>f</sup>
Alf684r <sup>e</sup>	Alphaproteobacteria	TACGAATTTYACCTCTACA	684–702		1359	88.5 (89.3)	242		650–750	This study
Beta359f	Betaproteobacteria	GGGGAATTTTGGACAATGGG	359–378	V3–V4	851	93.0 (88.8)	10	63	300–400	Ashelford <i>et al.</i> (2002) <sup>f</sup>
Beta682r	Betaproteobacteria	ACGCATTTCACTGCTACACG	682–701		701	86.4 (82.4)	6		650–750	Ashelford <i>et al.</i> (2002) <sup>f</sup>
Gamma395f	Gammaproteobacteria	CMATGCCCGGTGTGTGAA	395–412	V3–V5	1412	52.8 (59.1)	203	54	350–450	This study
Gamma871r	Gammaproteobacteria	ACTCCCCAGGCGGTCDACTTA	871–891		1579	64.8 (62.1)	50		850–950	This study
CFB555f	Bacteroidetes	CCGGAWTYATTGGGTTTAAAGGG	555–577	V4–V5	549	83.9 (85.0)	2	61	500–600	This study
CFB968r	Bacteroidetes	GGTAAGGTTCCCTCGCGTA	968–985		573	90.5 (92.2)	305		900–1000	This study
CYA361f	Cyanobacteria, chloroplasts	GGAATTTTCCGCAATGGG	361–378	V3–V4	434	91.0 (92.2)	16	59	300–400	This study
CYA785r	Cyanobacteria, chloroplasts	GACTACWGGGGTATCTAATCC	785–805		345	88.2 (87.8)	42		750–850	This study
Plancto352f <sup>e</sup>	Planctomycetes	GGCTGCAGTCGAGRATCT	350–367	V3–V5	209	84.0 (87.4)	140	68	300–400	This study
Plancto920r <sup>e</sup>	Planctomycetes	TGTGTGAGCCCCCGTCAA	920–937		103	98.1 (87.1)	6		900–1000	This study
Firm350f <sup>e</sup>	Firmicutes	GGCAGCAGTRGGGAATCTTC	350–369	V3–V4	1140	24.9 (37.7)	7	57	300–400	This study
Firm814r <sup>e</sup>	Firmicutes	ACACYTAGYACTCATCGTTT	814–833		1087	25.1 (30.2)	17		750–850	This study
9bfm	Bacteria	GAGTTTGATYHTGGCTCAG	9–27	V1–V9	3101	77.7 (86.4)	1	52	1–100	This study
1512uR	Universal (bacteria and archaea)	ACGGHTACCTTGTTACGACTT	1492–1512		3284	78.5 (80.0)		52	1450–1542	Weisburg <i>et al.</i> (1991)

Abbreviations: AT, annealing temperature; FISH, fluorescence *in situ* hybridization.

<sup>a</sup>The number in the primer name indicates the starting position of the primer sequence within the *E. coli* 16S rRNA gene sequence.

<sup>b</sup>Information based on analysis using the ROSE function within PRIMROSE; percentage of positive hits depends on number of sequences in the database with sufficient information; there is less information available for the termini compared to the centre of the 16S rRNA gene.

<sup>c</sup>Numbers in brackets indicate percentage matches within target group as obtained from using the online tool 'Probe Match' within the Ribosomal Database Project-II (RDP-II) database.

<sup>d</sup>Numbers indicate *E. coli* position; the longest 16S rRNA gene sequences in the RDP are 1542 bases.

<sup>e</sup>The template for the PCR was the PCR product obtained with primers 9bfm/1512uR.

<sup>f</sup>These primers were suggested for use as FISH probes by Ashelford *et al.* (2002), but have not yet been tested.

**Table 2** Nested PCR approach with DGGE primers

Primers used for group-specific PCR	Primers used for re-PCR for DGGE <sup>a</sup>	AT (semi-) nested PCR (°C)	Denaturing gradient used for DGGE (%) <sup>b</sup>
Alf28f/Alf684r	341f-GC/518r	56	40–60
Beta359f/Beta682r	518f-GC/Beta682r	60	40–55
Gamma395f/ Gamma871r	518f-GC/785r	56	40–60
CFB555f/CFB968r	CFB555f-GC/907r	64	40–60
CYA361f/CYA785r	518f-GC/CYA785r	56	40–55
Plancto352f/ Plancto920r	518f-GC/907r	60	40–60
Firm350f/Firm814r	518f-GC/785r	56	40–60
9bfm/1512uR <sup>c</sup>	341f-GC/518r	56	40–60

Abbreviations: AT, annealing temperature; DGGE, denaturing gradient gel electrophoresis.

PCR products produced with the group-specific primers were reamplified using a set of bacterial or universal DGGE primers.

<sup>a</sup>Nucleotide sequences of the primers, which have been published previously, and the GC clamp are: 341f, CCTACGGGAGGCAGCAG (Muyzer *et al.*, 1993); 518f, CCAGCAGCCGCGGTAAT (Muyzer *et al.*, 1993); 518r, ATTACCGCGGCTGCTGG (Muyzer *et al.*, 1993); 785r, CTACCAGGG TATCTAATCC (Lee *et al.*, 1993); 907r, CCGTCAATTCMTTGTGATTT (Muyzer *et al.*, 1998); GC clamp, CGCCCGCCGCGCGCGCGGG CGGGCGGGGACGGGGGG (Muyzer *et al.*, 1993); Please note, the sequence of the bacterial primer 785r (*E. coli* position 785–803) matches 81.2% of all 16S rRNA gene sequences in the ROSE database, but only 1% of the *Cyanobacteria*/chloroplast 16S rRNA gene sequences.

<sup>b</sup>100% denaturant contains 7 M urea and 40% (v/v) formamide.

<sup>c</sup>The approach taken for DGGE analysis of the domain bacteria involved reamplification also.

Generally, only one strand of the DNA fragments was sequenced. This proved to be sufficient for the taxonomic identification of the cloned 16S rRNA gene fragments using the BLAST search function within the NCBI database. Both DNA strands were sequenced in those cases where sequences could not be assigned easily to taxonomic groups or where the quality of the nucleotide sequence could not be determined unambiguously.

#### Group-specific PCR-DGGE analysis

Denaturing gradient gel electrophoresis (DGGE) was based on Muyzer *et al.* (1998) and Dar *et al.* (2005). In essence, group-specific PCR products were used as template in a re-PCR with a nested *Bacteria* DGGE-PCR primer pair (Table 2). Semi-nested re-PCRs were used in the case of the *Cyanobacteria*/chloroplasts, the *Bacteroidetes* and the *Betaproteobacteria* (Table 2). Only one bacterial primer was available within the 16S rRNA gene fragment that was amplified by the primers specific for these phylogenetic groups. PCR conditions and cycle protocol for the nested PCRs with the DGGE primer pairs were the same as those used for the PCRs with the group-specific primer pairs, except for the ATs which are shown in Table 2.

DGGE of the PCR products was performed on a 8% (w/v) polyacrylamide gel with urea and formamide as denaturants. The denaturing gradients varied with the DGGE primer pair used for the PCR but were generally between 40% and 60% (Table 2). Electrophoresis was performed in 1 × Tris-acetate EDTA buffer at 60 °C at constant voltage of 60 V for 18 h. Subsequently, gels were stained in 1 × SYBR Gold nucleic acid gel stain (Molecular Probes) for 45 min and rinsed in distilled water prior to image analysis on a Syngene GelDoc station. Individual

DGGE bands were cut out from the gel and incubated overnight at 4 °C in 30 µl H<sub>2</sub>O. An aliquot (ca 5 µl) was used in a PCR with the same primer set used for DGGE (Table 2; but without the GC clamp attached to one primer) to reamplify the insert. The nucleotide sequences of these 16S rRNA gene fragments were determined as described above.

#### Phylogenetic analysis

Sequences from the clone library prepared with the *Gammaproteobacteria*-specific primers were first compared to sequences stored in GenBank using the BLAST algorithm. Subsequently, the sequences from the clone libraries, and sequences with high similarity as identified by the BLAST searches, were imported into the ARB software program (<http://www.mikro.biologie.tu-muenchen.de/pub/ARB>) and aligned to other *Gammaproteobacteria* 16S rRNA gene sequences using the automated alignment tool within ARB (Ludwig *et al.*, 2004). The alignment was further corrected 'by eye', taking the secondary structure prediction of the ARB program into account. Phylogenetic analyses were based on the alignment of the 16S rRNA gene sequences from the mesocosm and Northern Atlantic Gyre clone libraries with representative sequences from the sequence databases. Calculation of the phylogenetic trees was based on these sequence alignments using the neighbour-joining method with Jukes–Cantor corrections, as well as the maximum likelihood and parsimony algorithms. Analyses were carried out using a maximum frequency filter. For each of the phylogenetic analyses in this study, the grouping of strains and environmental clones within the different clusters of the tree was identical for all of the above three phylogenetic methods for calculating trees (maximum likelihood, maximum parsimony,

neighbour-joining). However, those branching points within a tree that were not supported by each of the three algorithms were collapsed within the neighbour-joining tree using a strict consensus rule until the branching was supported in all three analyses. The neighbour-joining tree was chosen for depicting the phylogenetic relationship of the 16S rRNA gene clones and strains, and bootstrap values were calculated from 100 trees using the neighbour-joining method.

#### Nucleotide sequence accession numbers

The sequence data of 16S rRNA gene fragments have been submitted to the EMBL database with accession numbers AM706671-707020 (Northern Atlantic Gyre clone library), AM706537-AM706670 (mesocosm clone library), AM747394-AM747468 (sequences of DGGE bands).

## Results and discussion

#### *In silico* analysis of the group-specific PCR primers

Blackwood *et al.* (2005) used ARB to develop five group-specific primers including primers for four of the groups of bacteria investigated in this study. Ashelford *et al.* (2002) compared in detail the PRIMROSE and ARB programs and concluded that in many cases it was possible to identify better oligonucleotide probes (judged by *in silico* analysis) using PRIMROSE rather than ARB. We based the development of our primers on the independent use of both software packages to ensure the best design. The theoretical specificities of all primers were tested with the ROSE program of the TOOLKIT software package, as well as the PROBE MATCH function within the RDP database (<http://rdp.cme.msu.edu/>). The sequences of the primers and their specificity are summarized in Table 1. For

the purpose of comparison (see below under 'Group-specific clone libraries'), we preferred the use of the ROSE program of the TOOLKIT software package, since this uses exactly the same sequences as a basis for the comparison. In contrast, results of sequence searches in the current version of the PROBE MATCH function within the RDP database depend on the search window set for the priming site, which is different for different primer locations along the 16S rRNA gene.

PCR primers were developed for the amplification of 16S rRNA gene fragments that provide valuable additions to existing primers. *In silico* analyses indicate that they have generally a higher number of exact matches to the 16S rRNA gene sequences from members of the target group of bacteria for which they were designed, while their specificity is generally similar to that of published primers (Table 3).

However, some of the most suitable primers were identical to the FISH probes suggested by Ashelford *et al.* (2002) — Alf28f, Beta359f and Beta682r. Other probes suggested by Ashelford *et al.* (2002) or previously published primers (for example, Nübel *et al.*, 1997; Blackwood *et al.*, 2005) exploit similar priming sites to those used in our study but were, for example, of different length (for example, Alf684r, CFB555 and CYA785r) and degeneracy (for example, CYA361f). Blackwood *et al.* (2005) and Blümel *et al.* (2007) paired their group-specific primers with either the universal primer 1392r or the *Bacteria* primers Eub338 or 27f. In contrast, we designed two group-specific primers (that is a primer pair) for each taxonomic group to increase specificity of the PCR, or to ensure that a more group-specific primer of the primer pair would compensate for lower specificity of the other primer (for example, CFB555f/CFP968r compared to 27F/Cyt1020R of Blümel *et al.*, 2007).

**Table 3** Comparison of 16S rRNA gene group-specific primers developed in this study with those used in previous studies by Blackwood *et al.* (2005) and Nübel *et al.* (1997)

Primer pairs	<i>Escherichia coli</i> position	Percentage of matches within the target group	Matches outside target group
Alf684r	684–702	88.5	242
<b>ADF681F</b>	<b>682–698</b>	<b>87.3</b>	<b>47</b>
Beta682r	682–701	86.4	6
<b>Beta680F</b>	<b>680–694</b>	<b>85.7</b>	<b>5</b>
Plancto920r	920–937	47.5	6
<b>Plancto930R</b>	<b>931–947</b>	<b>43.1</b>	<b>3</b>
Firm350f	350–369	24.9	7
<b>BLS342F</b>	<b>352–369</b>	<b>21.1</b>	<b>0</b>
CFB968r	968–985	90.5	16
<b>Cyt1020R</b>	<b>978–995</b>	<b>8.5</b>	<b>0</b>
CYA361f	361–378	91.0	16
<b>CYA359F</b>	<b>359–378</b>	<b>83.4</b>	<b>73</b>
CYA785r	785–805	88.2	42
<b>CYA781R<sup>a</sup></b>	<b>781–805</b>	<b>87.2</b>	<b>40</b>

The *in silico* analysis was carried out using the PRIMROSE program (Ashelford *et al.*, 2002). The primers used by Blackwood *et al.* (2005) and Nübel *et al.* (1997) are indicated in bold.

<sup>a</sup>Comparison is based on the mixture of primers CYA781R(a) and CYA781R(b) (Nübel *et al.*, 1997).

### A new bacterial forward primer used for nested reamplification of PCR products

The group-specific primer pairs were generally used to amplify 16S rRNA gene fragments directly from environmental DNA. However, primer pairs for the *Alphaproteobacteria*, the *Planctomycetes* and the *Firmicutes* resulted in only low yield when environmental DNA was used as template. Therefore, a nested PCR approach was designed to overcome this problem. Obviously, the forward primer used in the first PCR had to be upstream of the *Alphaproteobacteria*-specific forward primers (Alf28f), which were located close to the 5'-end of the 16S rRNA gene. Concerns have been raised for some time when the bacterial forward primers 8f and 27F are used because of limited amplification efficiency and potential mismatches with newly discovered strains or environmental 16S rRNA gene sequences (for example, Marchesi *et al.*, 1998). These primers were designed when the sequence databases consisted only of a few thousand clones (for example, see under <http://rdp.cme.msu.edu/misc/history.jsp>). Testing *Bacteria* primers for specificity revealed that the sequences were homologous to only 56.5% (8f; Hicks *et al.*, 1992) or 72.9% (27F; Giovannoni *et al.*, 1996) of the over 15 000 sequences within the PRIMROSE database. The sequence of the universal (bacteria and archaea) reverse primer 1512uR (Weisburg *et al.*, 1991) was homologous to 78.5% of the sequences tested (Table 1).

We decided to use reverse primer 1512uR but modify the forward primer to account for sequence differences at this priming site. *In silico* analysis showed that the sequence of the new *Bacteria* primer 9bfm (Table 1) is homologous to 77.7% of the bacterial sequences in the PRIMROSE database but also to one *Archaeon* (*Methanobrevibacter* sp strain MB-9; accession AB017514). However, an archaeal 16S rRNA gene sequence was never detected, either in any of the libraries screened in this study (see below) or in any other analyses in our laboratory. Although there is only a small increase (5%) in the number of sequences to which the sequence of primer 9bfm is homologous, this is likely to lead to a far higher amplification rate. The greater degeneracy of the primer is likely to result in the successful amplification of more novel and yet unknown genotypes since many of the sequences that are not homologous to primer 9bfm will differ only in one base (as tested using the OLIGO CHECK function within the TOOLKIT software package). Furthermore, this one-base difference is not within the last three or four 3'-end bases, which are important for the polymerase to start extending the DNA strand during PCR.

However, despite the high percentage of exact matches to bacterial 16S rRNA gene sequences, primers may still be biased against certain groups of bacteria while matching the 16S rRNA gene sequence of most strains of other phylogenetic groups. Therefore, primers should always be tested with

respect to the specific bacterial group of interest prior to use.

In contrast to group-specific primers (see below), the application of these general *Bacteria* primers aims to amplify the whole genetic diversity of the domain. Therefore, use of less stringent conditions in the PCR reactions, for example a lower AT than the melting temperature of the primer, will allow for mismatches during the annealing, thus overcoming some of the above-mentioned biases. This further increases the genetic range of bacterial 16S rRNA gene fragments that are amplified in a PCR. In practice, it was found that, despite the increased degeneracy of primer 9bfm and the lower than optimal AT used in our PCRs (52 °C instead of  $\geq 54$  °C  $T_m$ ), all PCRs resulted in specific amplification of 16S rRNA gene fragments as judged by sequence analyses of clone libraries prepared with these primers (unpublished data). It should be noted that none of the PCRs with this primer pair in our laboratory has yet led to the amplification of nonspecific products.

### Group-specific 16S rRNA gene clone libraries

The newly developed primers were tested with environmental DNA samples from contrasting environments—a mesocosm experiment in a cold, coastal surface environment and a subtropical oligotrophic environment of the Northern Atlantic Gyre. As mentioned above, the following comparisons concerning specificity of the primers are based on the program ROSE of the TOOLKIT software package, though the specificity of the primers when compared using the PROBE MATCH function within the RDP database are also provided (Table 1).

*Alphaproteobacteria*. The oligonucleotides designed to specifically amplify the *Alphaproteobacteria* were identical or similar (Alf28f and Alf684r, respectively) to those suggested for use as molecular probes by Ashelford *et al.* (2002), and in one case (Alf684r; Table 3) similar to a primer proposed by Blackwood *et al.* (2005). The primer sequence of reverse primer Alf684r was homologous to that of a relatively large number (242) of bacteria outside the *Alphaproteobacteria*, mainly *Fusobacteria* and bacteria belonging to the orders *Desulfovibrionales*, *Desulfobacterales* and *Desulfuromonadales* of the *Deltaproteobacteria*. In contrast, the sequence of forward primer Alf28f was homologous only to 13 16S rRNA gene sequences outside the target group, equally distributed among the *Gammaproteobacteria*, *Deltaproteobacteria* and the *Verrucomicrobia*. Due to the high specificity of the forward primer and the high AT, the application of both primers as a primer pair in a PCR resulted in the amplification of 16S rRNA gene fragments from organisms that all belonged to the target group (Table 4). This primer pair therefore demonstrated the advantage of both primers of a primer pair being biased towards a

particular bacterial group of interest, rather than one group-specific primer combined with a second, *Bacteria* or universal 16S rRNA gene primer.

Most of the sequences that were detected in either of the two clone libraries belong to members of the Roseobacter group and the genus *Sphingomonas*, but several were also members of the genera *Rhodobium* and *Brucella*. However, the rest of the sequences belonged to a wide range of different groups of the *Alphaproteobacteria* (Supplementary Table S1). This demonstrates the value of the group-specific primer approach in detecting specifically a wide genetic diversity within each taxonomic group.

*Betaproteobacteria*. Ashelford *et al.* (2002) also suggested a number of oligonucleotides specific for the *Betaproteobacteria*. Our independent search led us to two primers (Beta359f, Beta682r; Table 1), which were identical to two of those suggested by Ashelford *et al.* (2002). Reverse primer Beta682r exploits the same discriminatory sequence stretch as primer Beta680F used by Blackwood *et al.* (2005) as a forward primer, although it is different in length and precise sequence position (Table 3). All sequences analysed from the mesocosm sample were from *Betaproteobacteria* (Table 4). In contrast, despite the high specificity of the primers, only one-third of the sequences from the Northern Atlantic Gyre clone library were from *Betaproteobacteria* (Table 4). Interestingly, three of the clones detected in the clone libraries from the Northern Atlantic Gyre show sequence similarities (at 92–95% sequence similarity) to 16S rRNA gene sequences of *Burkholderia* (Supplementary Table S1)—also prevalent in the Sargasso Sea data set (Venter *et al.*, 2004).

To test whether this low yield of positive hits could be improved using a nested reamplification-PCR approach, we also prepared and screened a second clone library prepared using a two-step

nested PCR approach. This used an aliquot of the PCR product obtained with primers 9bfm/1512uR as template for nested reamplification with primers Beta359f/682r in a second PCR. Again, only about one in four of the 25 sequences screened from this clone library were from *Betaproteobacteria*. The reason for this low yield of positive hits from this oligotrophic environment is not known. We believe that the use of primer Beta682r generally results in the group-specific amplification of 16S rRNA gene fragments as the last 3'-end (G) base of the primer is highly specific for the vast majority of *Betaproteobacteria* sequences in the ROSE database. This is supported by the fact that the clone libraries prepared from the coastal sample proved to be composed entirely of target sequences (Table 4). However, the low specificity at the 3'-end of primers Beta359f (most 16S rRNA gene sequences have three guanosines at *E. coli* positions 376–378), combined with the potentially low abundance of *Betaproteobacteria* in the Northern Atlantic Gyre sample may have led to the low percentage of target hits in this clone library (Table 4).

*Gammaproteobacteria*. As in the case of the *Alphaproteobacteria*, one of the primers (Gamma359f) appears to have a high number (203) of matches outside the target group (Table 1). However, more than 160 of these sequence hits are due to a large number of sequences for a small number (five) of particular bacterial species. For example, 45 non-target hits are due to homologous sequences within the 16S rRNA gene of *Acetobacter*, *Gluconobacter* and *Gluconacetobacter*. Furthermore, 118 hits are due to the *Ralstonia* (54) and *Burkholderia* (64) sequence hits. The reason for the high number of sequences obtained from these bacteria in the databases is because they are important pathogens. Given this, and the higher specificity of the reverse primer Gamma871r (Table 1), it is not surprising that the specificity found in both clone libraries is 100% (Table 4). This is further confirmed by the fact that the application of the *Betaproteobacteria*-specific primers demonstrated the presence of sequences with 92–95% sequence similarity to 16S rRNA gene sequences of *Burkholderia* as the next nearest isolated species (Supplementary Table S1), but no such sequences were detected within the clone libraries produced with the *Gammaproteobacteria*-specific primers (Supplementary Table S1).

Many of the sequences obtained from the mesocosm sample have a high sequence similarity to 16S rRNA gene sequences that fall within phylogenetic clades that belong to the oligotrophic marine *Gammaproteobacteria* (OMG) group (Figure 1; Supplementary Table S1). This group of *Gammaproteobacteria* was introduced by Cho and Giovannoni (2004) to indicate that all isolates were able to grow only in low-nutrient (oligotrophic) media. Phylogenetic analysis of their 16S rRNA gene sequences further confirmed that they form independent

**Table 4** Results from the BLAST searches with the sequences of the 16S rRNA gene fragments obtained from the screening of the two clone libraries by sequence analysis

Primers	Percentage of positive hits within mesocosm library (out of 20 clones tested)	Percentage of positive hits within Northern Atlantic Gyre library (out of 50 clones tested)
Alf28f/Alf684r	100	100
Beta359f/Beta682r	100	30
Gamma395f/	100	100
Gamma871r		
CFB555f/CFB968r	100	100
CYA361f/CYA785r	100	100
Plancto352f/	100	100
Plancto920r		
Firm350f/Firm814r	70	0



phylogenetic clades, which together comprise the OMG group of *Gammaproteobacteria* (Cho and Giovannoni, 2004). The fact that *Gammaproteobacteria* 16S rRNA gene sequences from the mesocosm (Figure 1) cluster within two (SAR92 and OM182) of the five OMG clades identified by Cho and Giovannoni (2004) may indicate that this phylogenetic clade represents a group within the *Bacteria* that is genetically diverse and occurs also in coastal nutrient-rich environments. In fact, Stingl *et al.* (2007) state for the SAR92 clade, into which several of the 16S rRNA gene sequences from the mesocosm sample cluster, that 'the peak of abundance correlates with the relatively high nutrient concentrations found in an upwelling region off the Oregon coast. In the lower nutrient regions farther off the coast, the abundance of the SAR92 was low, close to the limit of detection'. A possible explanation for this difference may be that under laboratory conditions, SAR92 isolates grow only under low-nutrient conditions while in their natural environment they thrive at higher nutrient concentrations such as provided by coastal environments. Alternatively, the phylotypes that we detected in the mesocosm samples may be physiologically different from those isolated by Cho and Giovannoni (2004).

It should be noted in this context that the phylogenetic analyses carried out in this study were rigorous. These included independent analyses based on the three main algorithms (maximum likelihood, neighbour-joining, parsimony; individual trees not shown) followed by comparison of the resulting trees. The phylogenetic tree shown in Figure 1 is a strict consensus of the three trees that was achieved by collapsing the tree on those branching points where there were differences between the three different trees. Also, the sets of sequences that form the individual clades and the branching within these clades were identical in all of the three different trees (individual trees not shown), thus confirming the robustness of the consensus tree (Figure 1).

The success of this *Gammaproteobacteria*-specific primer pair is demonstrated by the high specificity achieved by primers Gamma395f/871r and the resulting discovery of 16S rRNA gene sequences in a nutrient-rich environment that appear to belong to the OMG group of *Gammaproteobacteria*. In addition, a relatively small sample (70 sequences) from the open ocean and the mesocosm clone libraries revealed sequences that cluster in a wide range of *Gammaproteobacteria* clades (Figure 1). In this context, it will be interesting to analyse a larger number of clones per clone library than has been possible in this study, and from a variety of different geographic origins and seasons; a more extensive screening programme has the potential to discover novel phylogenetic clades of *Gammaproteobacteria*.

*Bacteroidetes*. This phylum was previously referred to as 'Cytophaga-Flavobacterium-Bacteroides' or

'CFB', hence the primer names. Ashelford *et al.* (2002) suggested an oligonucleotide (CFB560) as a FISH probe that exploits a similar discriminatory position within the 16S rRNA gene sequence to that of our forward primer CFB555f, which is located five bases further downstream. Primer CFB555f is highly specific with only two non-target hits (Table 1), while its sequence is identical to 84% of the *Bacteroidetes* sequences in the PRIMROSE database. In contrast, the reverse primer CFB968r has a relatively high number (305) of non-target hits, but the sequence is identical to over 90% of those in the database. However, 298 of the 305 non-target hits are due to identical sequences within the 16S rRNA gene of three members (*Borrelia*, *Spirochaeta* and *Treponema*) of the phylum *Spirochaeta*, and these pathogens are unlikely to be sufficiently abundant in the marine environment to be amplified by PCR. This, and the high specificity of the forward primer, seems to be responsible for the fact that all of the sequences in the clone libraries produced using this primer pair derived from members of the *Bacteroidetes*.

*Planctomycetes*. *Planctomycetes* species in general, and marine members of this phylum in particular, are one of the least studied groups of bacteria, mainly due to the lack of successful laboratory cultivation. Only members of two (*Planctomyces* and *Pirellula*) of the four recognized genera have been grown in culture. The Sargasso Sea shotgun sequencing project (Venter *et al.*, 2004) revealed only a small number of molecular marker genes identifying the genomic DNA clone as deriving from the *Planctomycetes*. FISH studies (Glöckner *et al.*, 1999; Brinkmeyer *et al.*, 2003) and 16S rRNA gene clone library-based approaches (Brinkmeyer *et al.*, 2003) have indicated low abundance (up to 3%) of *Planctomycetes* in a variety of environments. However, these studies may have underestimated *Planctomycetes* genetic diversity and abundance since the 16S rRNA gene sequences of the *Planctomycetes* may contain mismatches to several commonly used bacteria primers (Vergin *et al.*, 1998). We were interested to develop new PCR primers that specifically amplify the 16S rRNA gene fragments of the *Planctomycetes* with the potential to discover novel sequences and possibly phylogenetic lineages.

As with *Alphaproteobacteria*, *Gammaproteobacteria* and the *Bacteroidetes*, one primer (Plancto-920r) was highly specific for the target group, while the sequence of the other primer (forward primer Plancto350f) was homologous to 140 non-target sequences (Table 1). However, 130 of these 140 non-target hits belonged to the genera *Chlamydia* and *Chlamydophila*, which are pathogens that have not yet been detected and are unlikely to occur in the marine environment. The application of these oligonucleotides as a primer pair therefore led to specific amplification of members of the *Planctomycetes* only (Table 4).

*Cyanobacteria and chloroplasts of eukaryotic algae (oxygenic photoautotrophs)*. The primers developed here exploited similar priming sites to those developed by Nübel *et al.* (1997) but differ in length and degeneracy (Tables 1 and 3); in particular, forward primer CYA361f matched a larger number of target sequences and had fewer non-target hits (Tables 1 and 3). The screening of the 16S rRNA gene clone libraries demonstrated that the primers were very selective and they specifically amplified only members of the target group. The two environment studies showed contrasting results. The sample from the mesocosm experiment was dominated by 16S rRNA gene fragments of chloroplasts (Supplementary Table S1) and had a high abundance of eukaryotic photoautotrophs (data not shown), but 96% of the sequences from the Northern Atlantic Gyre sample had high similarity with *Prochlorococcus* isolates or environmental sequence clones. *Prochlorococcus* is known to dominate the oligotrophic regions of the oceans due to its temperature range, high surface to volume ratio and an efficient nutrient uptake mechanism (Scanlan and West, 2002), and would not be expected to occur in Norwegian waters—as indicated by the results obtained here.

*Firmicutes*. We were intrigued by the fact that the vast majority of marine bacteria detected routinely in 16S rRNA gene clone libraries are derived from Gram-negative bacteria, although a considerable part (up to 0.05% by weight of clones) of the bacterial metagenome in the Sargasso Sea were derived from the *Firmicutes* (Venter *et al.*, 2004). Results from the screening of PCR-amplified 16S rRNA gene fragment clone libraries generally contained only a small fraction of 16S rRNA gene sequences from Gram-positive bacteria (for example, Zaballos *et al.*, 2006; Martín-Cuadrado *et al.*, 2007), suggesting that Gram-negative bacteria far outnumber Gram-positive bacteria in the sea. However, this large difference in the abundance of these two groups may be exacerbated by using a molecular approach. For example, the considerably thicker peptidoglycan layer in the *Firmicutes* and the higher GC content of the *Actinobacteria* are likely to result in lower genomic DNA yield after DNA isolation from environmental samples

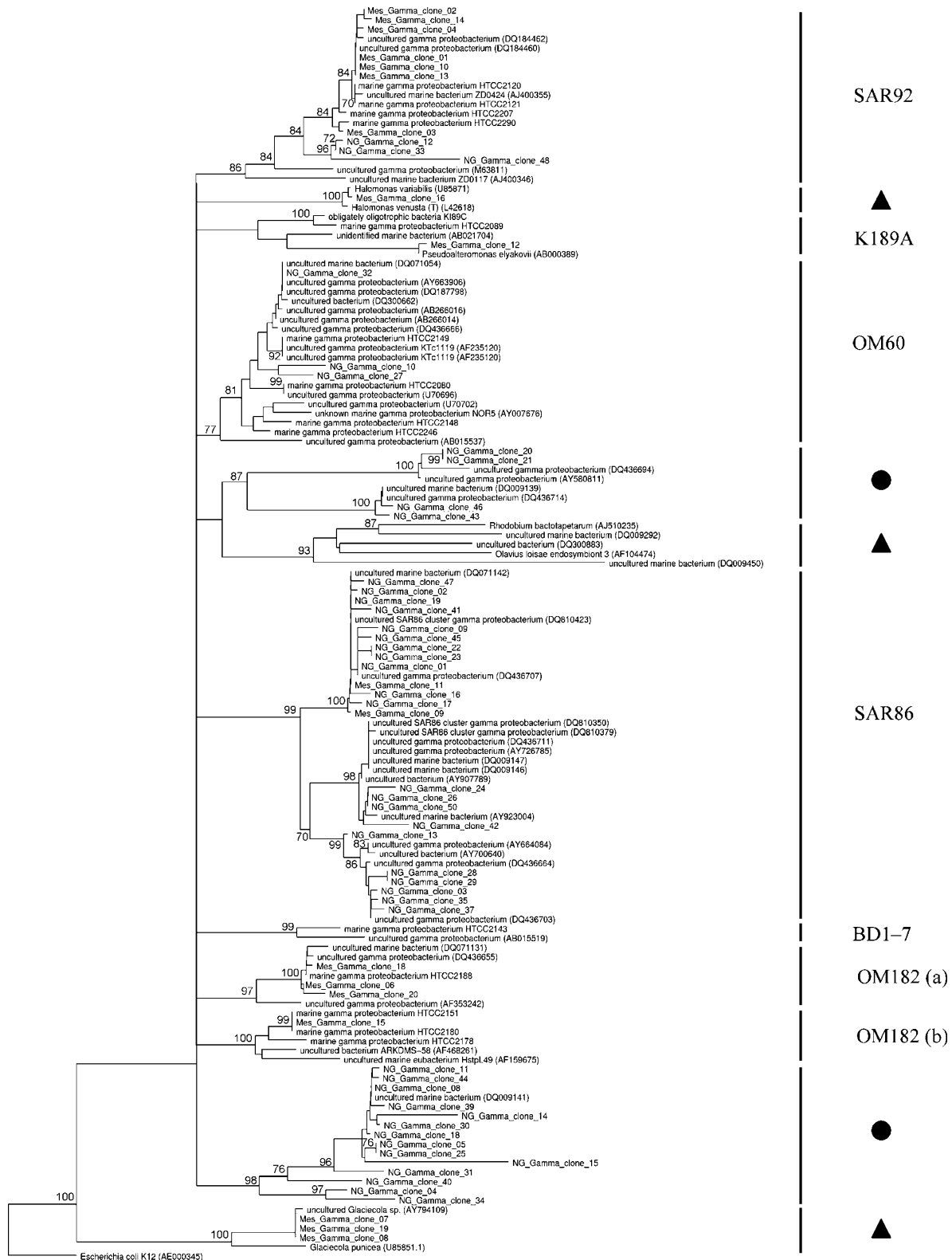
and lower yields during PCR amplification, respectively.

To investigate specifically the diversity of Gram-positive bacteria in the oceans, we designed specific primers for this group. *Actinobacteria*-specific primers are already available and proved to be successful for marine samples (Stach *et al.*, 2003). Interestingly, all our attempts to provide improved PCR primers for the *Actinobacteria* resulted in exactly the same primers as those suggested by Stach *et al.* (2003) or Blackwood *et al.* (2005). Due to the specificity and extensive testing by Stach *et al.* (2003) and Blackwood *et al.* (2005), we decided to concentrate further efforts on the *Firmicutes* group. This group consists of three classes with 33 families of bacteria (Garrity *et al.*, 2003), so we attempted to design primers that amplify 16S rRNA gene sequences from the whole genetic range of this phylum, thus avoiding the need for multiple PCRs for the screening of the many different *phylogenetic* groups. Given the huge genetic range among the *Firmicutes* and the low number of matches of the primers to 16S rRNA gene sequences outside the target group (Table 1), it is not surprising that the sequences of the primers are identical only to about 25% of the 2267 *Firmicutes* sequences within the PRIMROSE database. However, despite the high specificity of the primers (Table 1), 6 of the 20 sequences of the clone library prepared from the mesocosm DNA sample using these primers were from *Cyanobacteria* (two) or *Alphaproteobacteria* (four). Four of the seven non-target hits of forward primer, Firm350f, were to 16S rRNA gene sequences from chloroplasts and two were to *Alphaproteobacteria*. The sequence of the reverse primer, Firm814r, is homologous to 17 non-target hits, 14 of which are members of the *Geobacter* group within the *Delta*-*proteobacteria* and only one from an *Alphaproteobacterium* (*Ehrlichia*), but none to chloroplasts or *Cyanobacteria*. The clone library from the oligotrophic Northern Atlantic Gyre sample consisted entirely of cyanobacterial (*Prochlorococcus*) 16S rRNA gene fragments. The precise reason for this is unknown. We believe that the bias of the last 3'-end (C) base of primer Firm350f for a wide range of the *Firmicutes* sequences in the ROSE database was sufficient to amplify 16S rRNA gene fragments mainly from the target group in the case of the

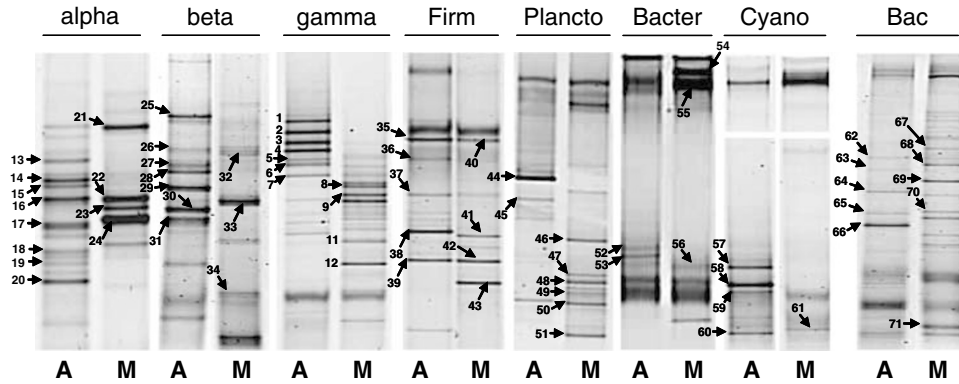
**Figure 1** Phylogenetic analysis of representative 16S rRNA gene nucleotide sequences from the *Gammaproteobacteria* clone libraries prepared from samples from mesocosm (coastal) and Northern Atlantic Gyre (open ocean) samples and representative sequences from the NCBI and ARB sequence database. The tree was calculated from a nucleotide alignment of 16S rRNA gene fragments (356 bases) using the neighbour-joining method within ARB, with Jukes–Cantor corrections and a maximum frequency filter (Ludwig *et al.*, 2004). *Escherichia coli* (accession J01859) was used as outgroup. The confidence of branch points was determined by three separate analyses (maximum likelihood, neighbour-joining, maximum parsimony), with multifurcations indicating branch points that were collapsed using a strict consensus rule until supported in all three analyses. Values of 100 bootstrap replicates (calculated using the neighbour-joining method) are given as numbers at branching points, but those <70 are omitted. Clade OM182 is split into two on this tree (a and b) as the two subclades grouped differently in one of the three methods used, and therefore the branching was collapsed to represent the most stringent consensus branching. The strains that are grouped together in the two OM182 clades are also grouped together into subclades in the phylogenetic analysis of Cho and Giovannoni (2004). Filled triangles indicate clades containing formally described species, and filled circles indicate clades or subclades for which no formally described species is available.

mesocosm sample. However, in the North Atlantic Gyre sample, the low specificity at the 3'-end of primer Firm814r for their target groups (that is most 16S rRNA gene sequences have three adenines

at *E. coli* positions 814–816) combined with the potential low abundance of bacteria of the target group may have led to the absence of any target hits in the Firmicutes clone libraries (Table 4). This



0.10



**Figure 2** Denaturing gradient gel electrophoresis (DGGE) analyses of 16S rRNA gene fragments amplified from DNA samples from the Northern Atlantic Gyre and the mesocosm experiments. Details on the primers used for the two- or three-step nested PCR approach are outlined in the text. The sequence identity of the numbered bands is given in the Supplementary data. The abbreviations indicate the target group of the corresponding primers: alpha, *Alphaproteobacteria*; beta, *Betaproteobacteria*; gamma, *Gammaproteobacteria*; Firm, *Firmicutes*; Plancto, *Planctomycetes*; Bacter, *Bacteroidetes*; Cyano, *Cyanobacteria*; Bac, *Bacteria*.

explanation is supported by the fact that the last three 3'-end bases of forward primer Firm350f are also present in all *Cyanobacteria*/chloroplast 16S rRNA gene sequences in the ROSE database, and all of the unspecific sequences derived from the cyanobacterium *Prochlorococcus*, which was the dominant autophototroph in this sample ( $6.4 \times 10^4$  cells ml<sup>-1</sup>; Jameson *et al.*, 2007).

#### DGGE analysis to detect differences in assemblage diversity between a coastal and an open ocean environment

As outlined above, DGGE with group-specific PCR primers is a very powerful method for the detailed study of microbial assemblages. All of the group-specific primers described in this study were originally designed as DGGE primers, that is a GC clamp was attached to one of the primers of each pair. However, the primers did not provide reproducible DGGE patterns and we decided to adapt the PCR-DGGE approach of Dar *et al.* (2005). This requires a nested PCR with *Bacteria* primers (one primer with a GC clamp attached to it) subsequent to the PCR with the group-specific primers, thus resulting in a two-step nested or, in the cases of the *Alphaproteobacteria*, the *Planctomycetes* and the *Firmicutes*, a three-step nested PCR-DGGE approach. Tests of the two-step nested PCR approach showed that low yield of the PCRs that sometimes occurred with the specific primers was not a problem, since the nested reamplification with the bacterial DGGE primer sets resulted in an amount of PCR product sufficient for subsequent DGGE analysis.

This nested PCR approach was successful with all primer pairs when applied to environmental DNA samples from both environments tested. In the case of the *Betaproteobacteria*, *Bacteroidetes* and the *Cyanobacteria*/chloroplast, *Bacteria* primers are not available for nested PCRs. Therefore, a

semi-nested approach was used in the second PCR, with one *Bacteria* primer and one that was specific for the group (Table 2). The results of the DGGE analysis are summarized in Figure 2. In all cases, there were detectable differences in the microbial assemblages between the Northern Atlantic Gyre and the mesocosm sample. However, the separation of the PCR fragments appeared to be better in those cases where a nested PCR was used as compared to a semi-nested PCR (that is for the *Betaproteobacteria*, the *Bacteroidetes* and the *Cyanobacteria*/chloroplast; see above). This was particularly the case with the *Bacteroidetes* and the *Cyanobacteria*/chloroplast-specific primer pair (Figure 2). In the latter case, it may therefore be advisable to use the primer pair developed by Nübel *et al.* (1997) for DGGE analysis rather than the one used here, but to consider primers CYA361f/CYA785r for the preparation of clone libraries.

The identification of the DGGE bands by sequence analysis showed that the PCR fragments mostly belonged to the bacterial group that the primers specifically targeted. An exception was the DGGE analysis of *Betaproteobacteria*. As with the results from clone library screening, a large fraction (60%) of the bands on the DGGE gel from the Northern Atlantic Gyre DNA sample were from prokaryotes other than the *Betaproteobacteria*. In contrast to this, all of the bands on the *Firmicutes* DGGE gel that were identified by sequence analysis derived from strains belonging to the target group (Supplementary Table S2). However, this may only be an artefact due to the sequence analysis of only a relatively small number of DGGE 'bands'. Increasing this number may again reveal non-target 16S rRNA gene fragments.

This research has shown the clear benefits of using group-specific PCR primers. Much more information is obtainable on bacterial diversity than with general bacteria primers as it is possible to target specific bacterial groups for more detailed

investigation of diversity, either by screening clone libraries or by using these primers in association with the well-tried methodology of DGGE; there is greater discrimination on individual gels and more bands are visible and can be sampled for sequencing. Although these primers have been designed for use in studies of the bacterial diversity in marine aquatic environments, they are also likely to be useful for the analysis of the bacterial diversity in other environments.

## Acknowledgements

This work is a core research activity of the Plymouth Marine Laboratory, a Collaborative Centre of the Natural Environment Research Council. Additional financial support came from the EU project 'Microbial Marine Communities Diversity: from Culture to Function' (MIRACLE) (EVK3-CT-2002-00087) and from a Natural Environment Research Council (NERC)-funded studentship (NER/S/A/2004/12171) allocated to JAW-A. We are grateful to Dr Hendrik Schäfer (Warwick University, UK) for useful discussion during this project. We also thank Ellie Jameson (PML) for the gift of the environmental DNA from the Northern Atlantic Gyre collected during the Atlantic Meridional Transect (AMT-15) cruise. The DNA sample from the mesocosm was collected during the 'Pelagic Ecosystem CO<sub>2</sub> Enrichment Study' (PeECE; <http://peece.ifm-geomar.de/>) organized by Professor Ulf Riebesell (IFM-GEOMAR, Kiel, Germany) and carried out at the EU Large-Scale-Facility, University of Bergen, Norway.

## References

- Amann RI, Ludwig W, Schleifer K-H. (1995). Phylogenetic identification of individual microbial cells without cultivation. *Microbiol Rev* **59**: 143–169.
- Ashelford KE, Chuzhanova NA, Fry JC, Jones AJ, Weightman AJ. (2005). At least 1 in 20 16S rRNA sequence records currently held in public repositories is estimated to contain substantial anomalies. *Appl Environ Microbiol* **71**: 7724–7736.
- Ashelford KE, Weightman AJ, Fry JC. (2002). PRIMROSE: a computer program for generating and estimating the phylogenetic range of 16S rRNA oligonucleotide probes and primers in conjunction with the RDP-II database. *Nucleic Acids Res* **30**: 3481–3489.
- Baker GC, Smith JJ, Cowan DA. (2003). Review and re-analysis of domain-specific 16S primers. *J Microbiol Methods* **55**: 541–555.
- Blackwood CB, Oaks A, Buyer JS. (2005). Phylum- and class-specific PCR primers for general microbial community analysis. *Appl Environ Microbiol* **71**: 6193–6198.
- Blümel M, Süling J, Imhoff JF. (2007). Depth-specific distribution of Bacteroidetes in the oligotrophic Eastern Mediterranean Sea. *Aquat Microb Ecol* **46**: 209–224.
- Brinkmeyer R, Knittel K, Jürgens J, Weyland H, Amann R, Helmke E. (2003). Diversity and structure of bacterial communities in Arctic and Antarctic pack ice. *Appl Environ Microbiol* **69**: 6610–6619.
- Calvo L, Vila X, Abella CA, Garcia-Gil LJ. (2004). Use of ammonia-oxidizing bacterial-specific phylogenetic probe Nso1225 as a primer for fingerprint analysis of ammonia-oxidizer communities. *Appl Microbiol Biotechnol* **63**: 715–721.
- Cho JC, Giovannoni SJ. (2004). Cultivation and growth characteristics of a diverse group of oligotrophic marine gammaproteobacteria. *Appl Environ Microbiol* **70**: 432–440.
- Chouari R, Le Paslier D, Daegelen P, Ginestet P, Weissenbach J, Sghir A. (2005). Novel predominant archaeal and bacterial groups revealed by molecular analysis of an anaerobic sludge digester. *Environ Microbiol* **7**: 1104–1115.
- Curtis TP, Sloan WT, Scannell JW. (2002). Estimating prokaryotic diversity and its limits. *Proc Natl Acad Sci USA* **99**: 10494–10499.
- Dar SA, Kuenen JG, Muyzer G. (2005). Nested PCR-denaturing gradient gel electrophoresis approach to determine the diversity of sulfate-reducing bacteria in complex microbial communities. *Appl Environ Microbiol* **71**: 2325–2330.
- Freitag TE, Prosser JI. (2003). Community structure of ammonia-oxidizing bacteria within anoxic marine sediments. *Appl Environ Microbiol* **69**: 1359–1371.
- Garrity GM, Bell JA, Lilburn TG. (2003). *Taxonomic Outline of the Prokaryotes. Bergey's Manual of Systematic Bacteriology*, 2nd edn. Release 4.0. Springer-Verlag: New York. 395 pp.
- Giovannoni SJ, Rappé MS, Vergin KL, Adair NL. (1996). 16S rRNA genes reveal stratified open ocean bacterioplankton populations related to the green non-sulfur bacteria. *Proc Natl Acad Sci USA* **93**: 7979–7984.
- Glöckner FO, Fuchs BM, Amann R. (1999). Bacterioplankton compositions of lakes and oceans: a first comparison based on fluorescence *in situ* hybridization. *Appl Environ Microbiol* **65**: 3721–3726.
- Hamasaki K, Taniguchi A, Tada Y, Long RA, Azam F. (2007). Actively growing bacteria in the inland sea of Japan, identified by combined bromodeoxyuridine immunocapture and denaturing gradient gel electrophoresis. *Appl Environ Microbiol* **73**: 2787–2798.
- Hicks RE, Amann RI, Stahl DA. (1992). Dual staining of natural bacterioplankton with 4',6-diamidino-2-phenylindole and fluorescent oligonucleotide probes targeting kingdom-level 16S rRNA sequences. *Appl Environ Microbiol* **58**: 2158–2163.
- Holben WE, Feris KP, Kettunen A, Apajalahti JHA. (2004). GC fractionation enhances microbial community diversity assessment and detection of minority populations of bacteria by denaturing gradient gel electrophoresis. *Appl Environ Microbiol* **70**: 2263–2270.
- Jameson E, Joint I, Mann NH, Mühling M. (2007). Application of a novel *rpoC1*-RFLP approach reveals that marine *Prochlorococcus* populations in the Atlantic gyres are composed of greater microdiversity than previously described. *Microb Ecol* (in print). Published online, doi:10.1007/s00248-007-9259-5.
- Lee SH, Malone C, Kemp PF. (1993). Use of 16S rRNA-targeted fluorescent probes to increase signal strength and measure cellular RNA from natural planktonic bacteria. *Mar Ecol Prog Ser* **101**: 193–201.
- Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar B *et al*. (2004). ARB: a software environment for sequence data. *Nucleic Acid Res* **32**: 1363–1371.

- Marchesi JR, Sato T, Weightman AJ, Martin TA, Fry JC, Hiom SJ *et al.* (1998). Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. *Appl Environ Microbiol* **64**: 795–799.
- Martín-Cuadrado A-B, López-García P, Alba J-C, Moreira D, Monticelli L, Strittmatter A *et al.* (2007). Metagenomics of the deep Mediterranean, a warm bathypelagic habitat. *PLoS ONE* **2**: e914. Published online doi:10.1371/journal.pone.0000914.
- McCaig AE, Embley TM, Prosser JL (1994). Molecular analysis of enrichment cultures of marine ammonia oxidisers. *FEMS Microbiol Lett* **120**: 363–367.
- Mehling A, Wehmeier UF, Piepersberg W. (1995). Nucleotide sequences of streptomycete 16S ribosomal DNA—towards a specific identification system for streptomycetes using PCR. *Microbiology* **141**: 2139–2147.
- Miyamoto H, Yamamoto H, Arima K, Fujii J, Maruta K, Izu K *et al.* (1997). Development of a new semi-nested PCR method for detection of *Legionella* species and its application to surveillance of legionellae in hospital cooling tower water. *Appl Environ Microbiol* **63**: 2489–2494.
- Muyzer G, Brinkhoff T, Nübel U, Santegoeds C, Schäfer H, Waver C. (1998). Denaturing gradient gel electrophoresis (DGGE) in microbial ecology. In: Akkermans ADL, van Elsas JD, de Bruijn FJ (eds). *Molecular Microbial Ecology Manual*. Kluwer Academic Publishers: Dordrecht, The Netherlands. pp 1–27.
- Muyzer G, De Waal EC, Uitterlinden AG. (1993). Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl Environ Microbiol* **59**: 695–700.
- Nübel U, Garcia-Pichel F, Muyzer G. (1997). PCR primers to amplify 16S rRNA genes from cyanobacteria. *Appl Environ Microbiol* **63**: 3327–3332.
- Pernthaler J, Amann R. (2005). Fate of heterotrophic microbes in pelagic habitats: focus on populations. *Microbiol Mol Biol Rev* **69**: 440–461.
- Rappé MS, Vergin K, Giovannoni SJ. (2000). Phylogenetic comparisons of a coastal bacterioplankton community with its counterparts in open ocean and freshwater systems. *FEMS Microbiol Ecol* **33**: 219–232.
- Robinson C, Poulton AJ, Holligan PM, Baker AR, Foster G, Gist N *et al.* (2006). The Atlantic Meridional Transect Programme (AMT): a contextual view 1995–2005. *Deep Sea Res II* **53**: 1485–1515.
- Scanlan DJ, West NJ. (2002). Molecular ecology of the marine cyanobacterial genera *Prochlorococcus* and *Synechococcus*. *FEMS Microbiol Ecol* **40**: 1–12.
- Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR *et al.* (2006). Microbial diversity in the deep sea and the underexplored ‘rare biosphere’. *Proc Natl Acad Sci USA* **103**: 12115–12120.
- Stach JEM, Maldonado LA, Ward AC, Goodfellow M, Bull AT. (2003). New primers for the class Actinobacteria: application to marine and terrestrial environments. *Environ Microbiol* **5**: 828–841.
- Stingl U, Desiderio RA, Cho J-C, Vergin KL, Giovannoni SJ. (2007). The SAR92 clade: an abundant coastal clade of culturable marine bacteria possessing proteorhodopsin. *Appl Environ Microbiol* **73**: 2290–2296.
- Stubner S. (2004). Quantification of Gram-negative sulphate-reducing bacteria in rice field soil by 16S rRNA gene-targeted real-time PCR. *J Microbiol Methods* **57**: 219–230.
- Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA *et al.* (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66–74.
- Vergin KL, Urbach E, Stein JL, DeLong EF, Lanoil BD, Giovannoni SJ. (1998). Screening of a fosmid library of marine environmental genomic DNA fragments reveals four clones related to members of the order Planctomycetales. *Appl Environ Microbiol* **64**: 3075–3078.
- Weisburg WG, Barns SM, Pelletier DA, Lane DJ. (1991). 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol* **173**: 697–703.
- Widmer F, Seidler RJ, Gillevet PM, Watrud LS, Di Giovanni GD. (1998). A highly selective PCR protocol for detecting 16S rRNA genes of the genus *Pseudomonas* (sensu stricto) in environmental samples. *Appl Environ Microbiol* **64**: 2545–2553.
- Zaballos M, Lopez-Lopez A, Ovreas L, Bartual SG, D’Auria G, Alba JC *et al.* (2006). Comparison of prokaryotic diversity at offshore oceanic locations reveals a different microbiota in the Mediterranean Sea. *FEMS Microbiol Ecol* **56**: 389–405.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)