The Journal of Microbiology (2011) Vol. 49, No. 5, pp. 729-740
DOI 10.1007/s12275-011-1205-z

Comparative Approach to Capture Bacterial Diversity of Coastal Waters[§]

Hyunsoo Na[1,2], Ok-Sun Kim[2,3]*, Seok-Hwan Yoon[2], Yunmin Kim[2], and Jongsik Chun[2]

[1]Center for Geomicrobiology, University of Aarhus, Ny Munkegade, Bld. 1535, 8000 Århus C, Denmark
[2]School of Biological Sciences, Seoul National University, Seoul 151-742, Republic of Korea
[3]Division of Life Sciences, Korea Polar Research Institute, Incheon 406-840, Republic of Korea

(Received April 20, 2011 / Accepted May 27, 2011)

Despite the revolutionary advancements in DNA sequencing technology and cultivation techniques, few studies have been done to directly compare these methods. In this study, a 16S rRNA gene-based, integrative approach combining culture-independent techniques with culture-dependent methods was taken to investigate the bacterial community structure of coastal seawater collected from the Yellow Sea, Korea. For culture-independent studies, we used the latest model pyrosequencer, Roche/454 Genome Sequencer FLX Titanium. Pyrosequencing captured a total of 52 phyla including 27 candidate divisions from the water column, whereas the traditional cloning approach captured only 15 phyla including 2 candidate divisions. In addition, of 878 genera retrieved, 92.1% of the sequences were unique to pyrosequencing. For culture-dependent analysis, plate culturing, plate washing, enrichment, and high-throughput culturing (HTC) methods were applied. Phylogenetic analysis showed that the plate-washing clones formed a cluster devoid of any previously cultured representatives within the family *Rhodobacteraceae*. One HTC isolate (SF293) fell into the OM182 clade, which was not recovered by other culturing methods described here. By directly comparing the sequences obtained from cultures with those from culture-independent work, we found that only 33% of the culture sequences were identical to those from clone libraries and pyrosequences. This study presents a detailed comparison of common molecular and cultivation techniques available in microbial ecology. As different methods yielded different coverage, we suggest choosing the approach after carefully examining the scientific questions being asked.

Keywords: diversity, pyrosequencing, Roche/454 GS-FLX titanium, cloning, high-throughput culturing (HTC), DGGE, plate-washing PCR (PWPCR)

In this era of rapid development of high-throughput sequencing technology, microbial diversity is estimated at increasingly higher resolution. Microbial community studies have benefited greatly from the advent of so-called 'next generation sequencing', revealing that the rare biosphere, i.e. the relatively low-abundance taxa which usually account for most of the observed phylogenetic diversity, is considerably larger and more diverse than previously appreciated by traditional Sanger sequencing and cultivation methods (Curtis *et al.*, 2002; Pedrós-Alió, 2006).

However, it was reported that an intrinsic error of pyrosequencing associated with homopolymer error could propagate spurious phylotypes resulting in overestimation of microbial diversity (Huse *et al.*, 2007; Kunin *et al.*, 2010). Unlike genome sequencing projects where sequencing errors can be compensated by sequencing depth and assembly, microbial diversity studies cannot take advantage of highly redundant assemblies and are therefore prone to misinterpretation by incorrect base calling. Recent studies attempt to minimize this problem by applying different strategies. Kunin and colleagues (2010) used a stringent quality filter system with a minimum of 0.2% error and low clustering thresholds (no greater than 97%) to avoid overestimation. Quince and colleagues (2009) developed a program called 'PyroNoise' which infers true sequences from amplicons by reducing sequencing noise (Quince *et al.*, 2009). Sun and colleagues demonstrated that their new program ESPRIT (Sun *et al.*, 2009) using pairwise alignments estimates microbial species richness more accurately than commonly used multiple sequence aligners such as CLUSTAL W (Thompson *et al.*, 1994), MUSCLE (Edgar, 2004), or NAST (DeSantis *et al.*, 2006).

Despite the exhaustive sequencing efforts and development of new algorithms to minimize errors, questions have been raised as to whether we have captured the full spectrum of microbial diversity. Donachie and colleagues re-emphasized the importance of culturing microorganisms, arguing that microbial diversity would be best described when metagenomic techniques went parallel with culture libraries (Donachie *et al.*, 2007). The study cited 7 different studies where they had recorded less than a 30% overlap of the sequences detected by the two methods contemporaneously applied to the same samples. Cheng and Foght (2007) reported a similar study where they observed the increased bacterial diversity by combining culture-based studies with molecular analyses.

Taken together, it seems that microbial diversity would be better described by not relying on a single approach that might harbor intrinsic errors or introduce biases. To circumvent this problem, we applied and evaluated several approaches, combining a wide range of molecular and culturing techniques to characterize the bacterial community of the Yellow

* For correspondence. E-mail: oskim@kopri.re.kr; Tel.: +82-32-260-6315; Fax: +82-32-260-6301
§ Supplemental material for this article may be found at http://www.springer.com/content/120956

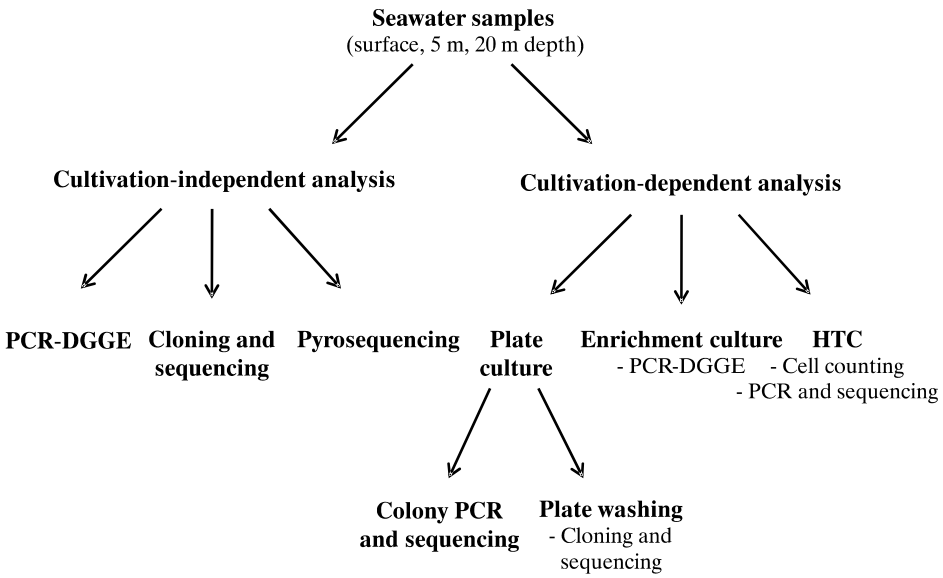

Fig. 1. Schematic overview of the study.

Sea, as shown in the diagram, Fig. 1. For improved phylogenetic and taxonomic analyses of sequences, EzTaxon-e (http://www.eztazon-e.org), the recently developed phylogenetic tree-based database and the software CLcommunity (http://www.chunlab.com) were applied. We believe that our study will provide microbiologists a chance to compare results using an integrative approach when they try to analyze microbial diversity.

Materials and Methods

Sample collection

Seawater was collected at high tide on March 19th, 2008, from the surface, 5 m depth, and 20 m depth from the Yellow Sea (N 37° 38′ 24.8″, E 126° 23′ 40″), off the coast of Ganghwa Island, Korea. Samples were immediately shipped to the laboratory in ice boxes and stored in the dark at 4°C until further processing.

Preparation of media

Seawater samples were pre-filtered with Whatman filter paper to remove soil particles, and then filtered through a 0.22-μm-pore-diameter membrane and immediately autoclaved. After sterilization, the seawater was kept at ambient temperature overnight to cool and restore pH. A total of 8 oligotrophic media were prepared with different amendments, which were commonly adjusted to the final concentration of 0.001% (w/v). Detailed contents per liter of sterilized seawater of each medium are as follows. YAF medium: 0.001 g of yeast extract, 0.01 g of sodium acetate (anhydrous), and 0.001 g of sodium formate. Fish extract medium: 0.3 ml of fish extract made from ground mullets and breams, 50 g of each in 1.5 L of seawater. Shellfish extract (Sh) medium: 0.6 ml of shellfish extract made from ground mussels and ark shells, 25 g of each, in 1.5 L of seawater. Mollusk extract (ML) medium: 0.3 ml of mollusk extract made from ground octopuses, 25 g of each in 1.5 L of seawater. Zooplankton extract (ZP) medium: 10 ml of zooplankton extract. Seaweed extract (SW) medium: 0.6 ml of seaweed extract made from brown and green algae, 50 g of each in 1.5 L of seawater. Tidal flat slurry extract (TFS) medium: 1 ml of tidal flat slurry extract, containing 10 g of tidal flat sediment (sampled during low tide at the sampling site) mixed in 1 L of seawater. Seafood extract mixture (Mix) medium: 0.3 ml of fish extract, 0.6 ml of shellfish extract, 0.3 ml of mollusk extract, 10 ml of zooplankton and 0.3 ml of seaweed extract. All extracts were prepared by grinding the material in water and autoclaving. After centrifuging at 4,000×g for 10 min, the supernatants were filtered to yield the final form of extracts. Zooplankton and algae were obtained from the sampling site, and other materials were purchased from the local markets near the sampling site. The media described above were intended to closely resemble natural seawater, except the YAF medium which served as a standard oligotrophic medium amended with chemicals commonly found in a laboratory. All of the media had 0.1 μM KH_2PO_4 and vitamin mix at a 10^{-5} dilution of stock solution (DSM 141). For agar plate cultures, agar (Difco) was added at the concentration of 1.5% (w/v).

Inoculation, cultivation conditions, and screening

Seawater was serially diluted by 10 fold. For plate cultures, 100 μl from each dilution was spread on agar plates of the eight different types of media mentioned above. The inoculated plates were incubated at 16°C for at least three weeks. For enrichment cultures, 1 ml from each dilution was inoculated into the conical tube containing 9 ml of the media. The inoculated tubes were incubated at 16°C for at least 2 months. For high-throughput culturing (HTC), seawater was diluted to either 10 or 30 cells/ml and dispensed into 48-well microtiter plates. After incubation in the dark at 16°C for at least 2 months, 150 μl of samples drawn from each well was fixed with formalin, and stained with 4'6-diamidino-2-phenylindole (DAPI) as previously described (Porter and Feig, 1980). After staining, the samples were filtered on a 0.22-μm-pore-diameter membrane and screened for growth as previously described (Connon and Giovannoni, 2002).

DNA extraction and PCR

For culture-dependent analyses, DNeasy Blood & Tissue kit (QIAGEN, USA) was used to extract genomic DNA from cultures, following the

manufacturer's guidelines. For culture-independent molecular analyses, the seawater sample (500 ml) was filtered through a 0.22-μm-pore-diameter membrane. Using a sterile scalpel, the membrane was cut into four pieces and one of them was sliced further into fine pieces for DNA extraction. Genomic DNA was extracted using the UltraClean® Soil DNA Isolation kit (MO BIO Laboratories, Inc, USA) according to the manufacturer's instructions.

PCR amplification of bacterial 16S rRNA gene sequences was performed with the GeneAmp 9700 PCR system (PE Applied Biosystems, USA). The conditions for PCR reactions were as follows: a total volume of 50 μl, 20 pmol of each primer, 2 U of Taq polymerase (TaKaRa, Japan), 2/25 volume of dNTP mix, and 1/10 volume of 10× buffer provided with the enzyme. The primers generally used in this study were Bact27F (5′-AGAGTTTGATCMTGGCTCAG-3′) and Univ1492R (5′-GGYTACCTTGTTACGACTT-3′) as previously described (Chun and Goodfellow, 1995). After initial denaturation for 5 min at 94°C, amplification reactions were performed with 30 cycles of denaturation (1 min, 94°C), annealing (1 min, 55°C), and extension (1 min, 72°C). After that, the final extension was carried out for 7 min, at 72°C. For HTC, the annealing temperature was lowered to 50°C and 40 amplification cycles were applied.

PCR-DGGE

For DGGE analysis, nested PCR was additionally performed to amplify the V3 region of 16S rRNA gene fragments after the first PCR reaction. The primers (Dgge338f with GC clamp, and Dgge518r) and amplification conditions were as previously described (Ampe *et al*., 1999). DGGE was performed with the D-code system from Bio-Rad Laboratories. The PCR products were loaded onto 8% (w/v) polyacrylamide gels, 1 mm thick, in 1× TAE buffer. The denaturant gradient ranged from 30 to 55%. 100% denaturant [7 M urea and 40% (v/v) formamide] was added to the bottom of the gel to prevent the sample from diffusing out of the gel. Electrophoresis was performed either for 14 h at 100 V or 8 h at 200 V. The temperature was set at 60°C. As standards, a mixture of DNAs was used which were obtained from a pre-performed DGGE with pure cultures. Gels were stained with SYBR Gold for 1 h. Images of the gels were obtained by using a Gel Doc 2000 Gel Documentation system and Quantity One software (Bio-Rad Laboratories). DGGE band patterns were analyzed by GelCompar II software (Applied Maths). Similarity matrices, based on band presence-absence, were produced by the Dice coefficient, from which dendrograms were constructed by UPGMA (Griffiths *et al.*, 2000; Kim, 2007). After DGGE, the bands were excised directly from the gel with a sterile scalpel, reamplified, and purified. For reamplification, the primers used were the same ones as in DGGE, without the GC clamp. The PCR was performed as described previously for nested PCR. After purification, cloning was carried out as described below.

Clone libraries

Cloning was carried out using pGEM-T Easy vector (Promega) following the manufacturer's instructions before transformation into HIT competent cells (DH5α, RBC). Transformed cells were spread onto Luria-Bertani agar plates containing 80 μg of X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) per ml, 0.5 mM of IPTG (isopropyl-β-D-thiogalactopyranoside), and 100 μg of ampicillin per ml. White colonies were picked using sterile toothpicks and transferred into PCR tubes. Colony PCR was performed with Bact27F and Univ1492R primers.

Sequencing and phylogenetic analysis of culture sequences

The sequences were determined on an automated DNA sequencer (ABI Gene Scan 3100, USA). Bact27F and Univ1492R primers were used for sequencing of clone libraries. For DGGE clone sequencing, Dgge340F primer without GC clamp was used.

The forward and reverse sequence fragments were assembled using the ContigExpress program. All of the sequences were checked for possible chimeras using the ChimeraCheck function of the Ribosomal Database Project (Cole *et al.*, 2003). Identification was carried out using EzTaxon (available at http://www.eztaxon.org) (Chun *et al.*, 2007) and Ribosomal Database Project (RDP) servers. For analysis of plate-washing clones, the calculation of sequence similarity among sequences was carried out using the jPHYDIT program (Jeon *et al.*, 2005). Sequences showing a similarity value greater than 97% were designated as identical clones, and only one of them was selected to be analyzed as a full sequence. Phylogenetic trees were constructed by the maximum parsimony method, using the software ARB. Sequences were imported into the Silva SSU database (SSURef_100_SILVA_04_10_08_opt) supplied on the ARB-homepage (http://www.arb-silva.de).

Pyrosequencing and data analysis

The extracted genomic DNA was amplified using primers targeting the V1 to V3 hypervariable regions of the bacterial 16S rRNA gene (V1-9F: 5′-X-AC-GAGTTTGATCMTGGCTCAG-3′ and V3-541R: 5′-X-AC-WTTACCGCGGCTGCTGG-3′ where X denotes a 5 nucleotide long barcode specifically designed for each depth (surface, 5 m and 20 m), followed by a common linker AC. The primer design and PCR amplification were carried out as previously described (Chun *et al.*, 2010). The DNA sequencing was performed by Macrogen Inc. (Korea) using the 454 GS FLX Titanium Sequencing System (Roche).

Sequencing reads from different depths were separated by unique barcodes. The barcode, linker, and PCR primers were removed from both ends of the reads. Then the reads were subjected to a filtering process as previously described (Chun *et al.*, 2010).

Each pyrosequencing read was taxonomically assigned using a newly developed database EzTaxon-e database (http://www.eztaxon-e.org), which contains a complete hierarchical taxonomic structure (from phylum to species) of the representative phylotypes of both cultured and uncultured entries in GenBank belonging to the domain *Bacteria*. We consider the database to be suitable for pyrosequencing analysis as it provides taxonomic information for uncultured bacterial 16S rRNA gene sequences, which usually account for a large portion of the microbial diversity and might represent microorganisms playing important ecological roles in nature. Naming convention of tentative taxa generated from uncultured sequences are described on the website, and also previously reported by Chun and colleagues (2010). The details of the database and related bioinformatic software are in preparation for publication.

For statistical analyses including clustering sequences and calculating diversity indices, the programs CD-HIT (Li and Godzik, 2006) and mothur (Schloss *et al.*, 2009) were used.

Nucleotide sequence accession numbers

The sequences of clones and isolates obtained in this study have been deposited under accession no. GU061004 to GU061323 in GenBank.

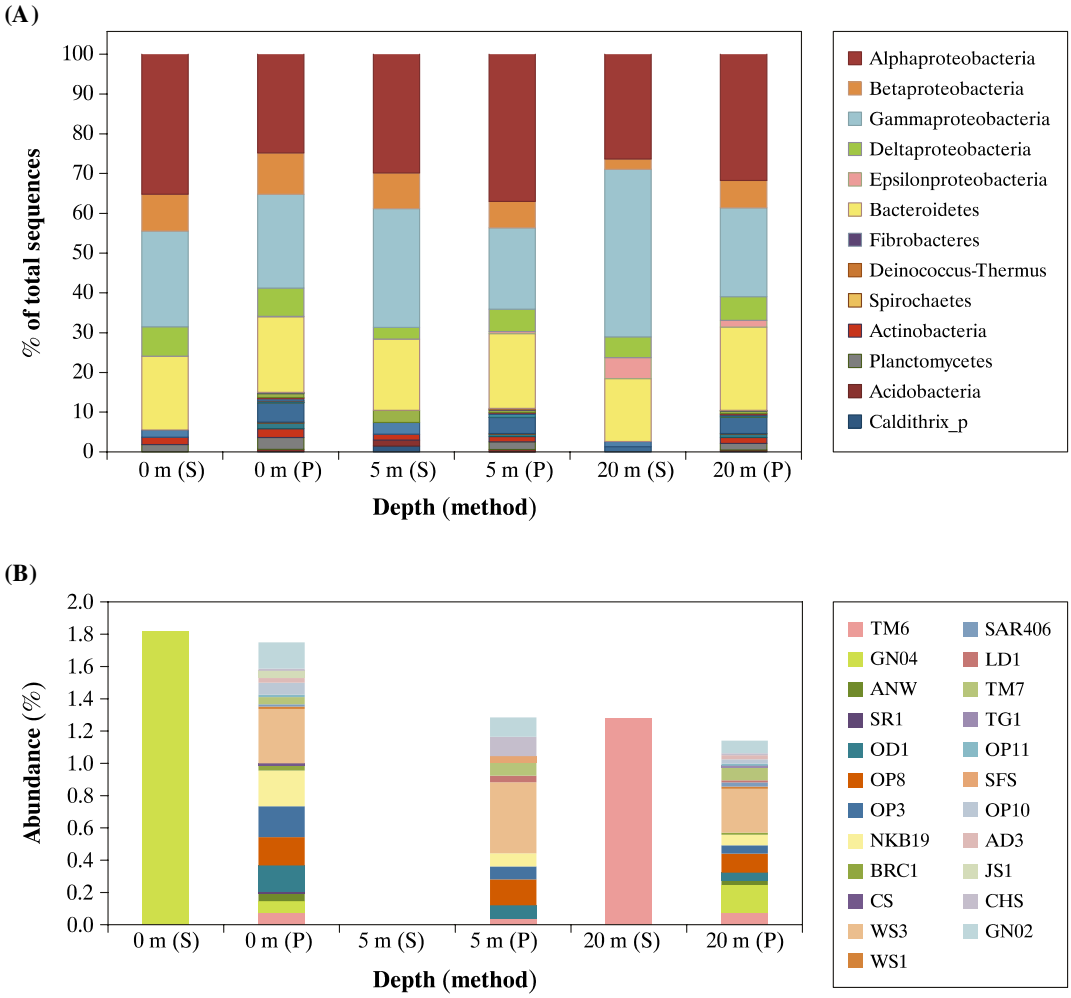

Fig. 2. Phylogenetic classification of bacterial 16S rRNA gene sequences from different depths of the Yellow Sea, analyzed by pyrosequencing (denoted as P) and the Sanger (denoted as S) method. (A) Comparison of known phyla detected by both methods. The phylum *Proteobacteria* is shown at class level resolution. Only major phyla are shown in the figure legend. (B) Candidate divisions detected by pyrosequencing and Sanger method.

Results and Discussion

Culture-independent analysis

A total of 17,012 reads with an average length of 434 bp was retrieved from pyrosequencing of PCR amplicons (6807, 2490, and 7715 reads from the surface, 5 m and 20 m depth, respectively). From clone libraries, a total of 200 reads were obtained (55, 67, and 78 from the surface, 5 m and 20 m depth, respectively). The two methods revealed slightly different community structures, presumably because we obtained almost 100 times more sequences from pyrosequencing as compared to the cloning approach. The different community structures produced by pyrosequencing were analyzed in more detail, comparing them with those from the Sanger method. For well-recognized phyla, pyrosequencing, and Sanger datasets were overall in good agreement, detecting *Proteobacteria* (66.7% and 76.0% from each method) and *Bacteroidetes* (19.4 and 17.0% from each) as major constituents of the whole water column bacterial community (Fig. 2A). This type of congruity between the datasets from the two methods has been reported previously (Edwards *et al.*, 2006), where they obtained similar taxonomic profiles of 16S rRNA from the deep mine samples. However, a slightly different tendency in relative abundance was observed in the two datasets. For instance, *Epsilonproteobacteria* was only detected at 20 m depth (5.1%) by Sanger, whereas it was shown to be present at all depths by pyrosequencing (0.1% at surface, 0.5% at 5 m, 1.7% at 20 m depth). *Chloroflexi* turned out to be the third most abundant (4.3%) phylum across all depths by pyrosequencing. This could have been overlooked if we had only applied the cloning approach in the analysis as *Chloroflexi* was only present at 20 m depth (1.3%) in the Sanger dataset (data not shown). Also, unclassified proteobacterial sequences were only detected by pyrosequencing. On the other hand, *Gammaproteobacteria* and *Verrucomicrobia* were relatively overrepresented at all depths in the Sanger dataset.

With candidate divisions, pyrosequencing showed much greater detection capability. The pyrosequencing dataset con-

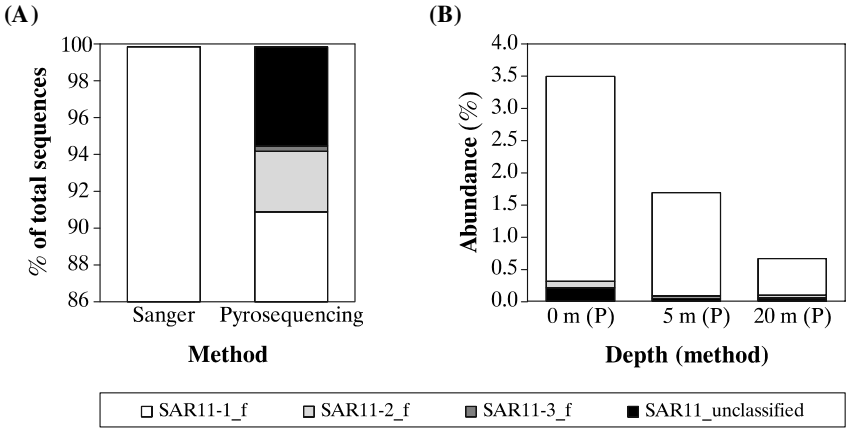

Fig. 3. Detection of SAR11 subclusters. (A) Comparison of Sanger and pyrosequencing datasets. (B) Vertical distribution of SAR11 subclusters in the water column, revealed by pyrosequencing.

tained 52 candidate divisions in total, while the Sanger method detected only one candidate phylum (GN4 from surface and TM6 from 20 m depth) or none (Fig. 2B). The candidate phylum TM6 only occurred at 20 m depth with the cloning approach, while it appeared at all depths with pyrosequencing. Likewise, pyrosequencing captured a number of other candidate phyla such as OD1, OP3, OP8, NKB19, WS3, and TM7 that were absent in the Sanger dataset at all depths. In particular, WS3 accounted for more than half (55%) of the candidate phyla in the pyrosequencing dataset while the Sanger dataset contained none. This is consistent with the previous study of bacterial community by Kim and colleagues (2008), where they found WS3 to be the most abundant candidate phylum in the tidal flat sediment geographically close to our sampling site. The candidate phylum WS3 has been reported to be present in a variety of niches such as mud volcano (Kormas *et al.*, 2008), marine sponge (Webster *et al.*, 2010), sulfur-rich spring (Elshahed *et al.*, 2007), anaerobic sludge digester (Chouari *et al.*, 2005), and rice roots (Derakshani *et al.*, 2001). With pyrosequencing, the percentage of candidate divisions was shown to decrease with depth from 1.7% at the surface to 1.2% at 5 m depth, and to 1.2% at 20 m depth (Fig. 2B). For both validated and candidate phyla, it was evident that pyrosequencing and Sanger datasets yielded different results. Going down to the order level, the SAR11 clade, well-known by its prevalence in seawater DNA (Morris *et al.*, 2002), was detected by both methods, but more subclusters, based on the groupings established previously (Morris *et al.*, 2005), were detected with pyrosequencing (Fig. 3A). SAR11 subgroups 1, 2, and 3 were detected by pyrosequencing, whereas only subgroup 1 was detected by the Sanger method. This subgroup 1, designated as SAR11-1_f, accounted for 2.5% of the whole water column community. Furthermore, pyrosequencing retrieved 18 unclassified sequences within the SAR11 clade. Specifically, 6 of the unclassified sequences showed 87-90% similarity values to the sequence of *Candidatus Pelagibacter ubique*, HTCC 1062 (CP000084), the first cultivated member of the SAR11 clade (Giovannoni *et al.*, 2005). With pyrosequencing, SAR11 showed a decreasing abundance with depth (Fig. 3B).

In the following analysis, we compared the results at a higher resolution of family and genus levels as shown in Fig. 4. For the analysis, the datasets from 3 depths were combined for each sequencing method. A total of 577 families were detected when the similarity cutoff value was set at 90%, and 878 genera were detected with cutoff value of 94%, according to the criteria of taxonomic ranks in EzTaxon-e DB. Surprisingly, the two methods shared very few taxa, 8.3% at the family

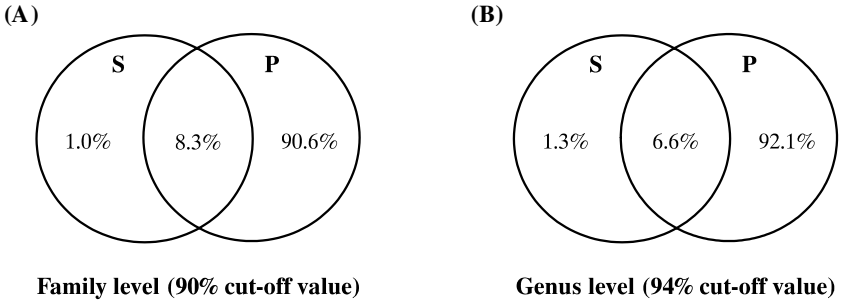

Fig. 4. Distribution of shared and unique bacterial taxa for the Sanger (denoted as S) and pyrosequencing (denoted as P) methods, at the family level (A), and genus level (B). Analyses were performed across all datasets for the 3 depths.

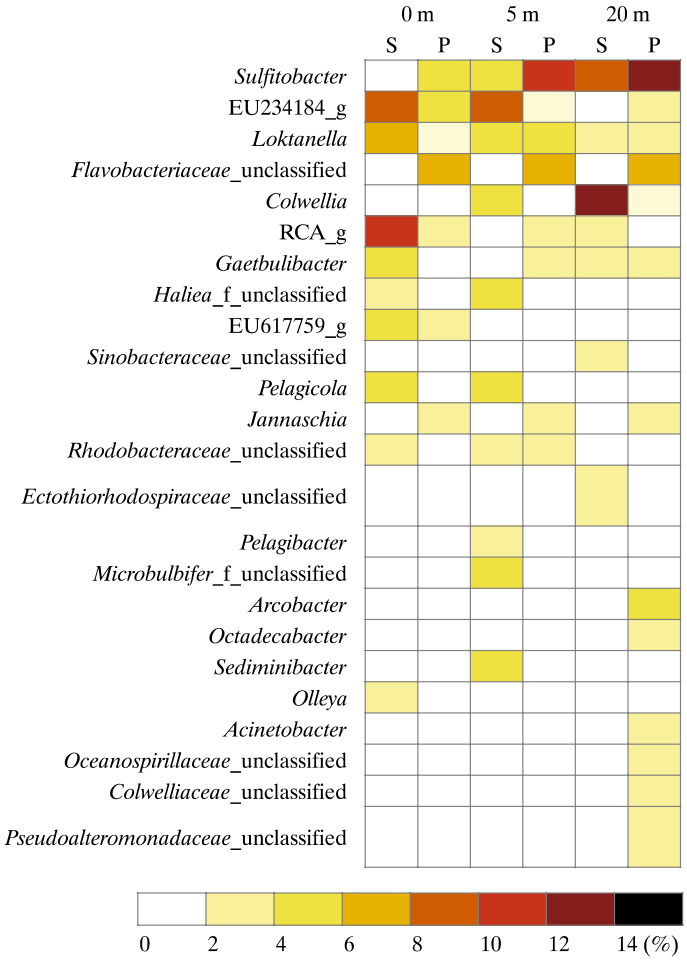

Fig. 5. Heat plot showing relative abundance of each bacterial genus depending on depth and the sequencing method used. Twenty-five genera having more than 2% abundance for at least one of the columns were selected, and their abundances were depicted in the plot. The color code indicates relative abundance, ranging from light yellow (low abundance) via orange and red to black (high abundance); white indicates that no sequence was assigned to the respective genus.

level and 6.6% at the genus level. At the family level, a large percentage of sequences (90.6%) was unique for pyrosequencing (P), and only 1.0% was unique for the Sanger method (S). A similar pattern was observed at the genus level.

The results at the genus level were analyzed at a fine scale with the heat plot (Fig. 5). For *Sulfitobacter* and EU234184_g, data from the two methods were in good agreement, showing either increase or decrease with depth. Overall, pyrosequencing showed a greater detection capacity for minor members, which comprise less than 2% of the community. It was also confirmed that the sequences unique for pyrosequencing from Fig. 4 came from this portion of minor constituents.

Given that pyrosequencing produces a greater number of reads, detection of more phylotypes by this method was expected. This was also supported by rarefaction analysis when we compared the number of OTUs detected with the same number of sequences sampled for the two methods (data not shown). Such a small difference could be explained by the overlapping region of the 16S rRNA gene covered by the two primer sets. The rarefaction curve generated from pyrosequencing data indicated that further sampling might retrieve additional taxa (Supplementary data Fig. 1). However, the diversity with depth could not be explained in a simple manner as rarefaction and diversity indices gave different results (Supplementary data Table 1).

Taken together, the taxonomic profiles generated by pyrosequencing and the Sanger method were in good agreement, with pyrosequencing being capable of detecting lower-abundance organisms. These results are consistent with the previous report (Schellenberg *et al.*, 2009), where they found that the pyrosequencing dataset included virtually the entire Sanger dataset, along with an additional level of microbial richness. It is speculated that pyrosequencing can detect minor members of the community by avoiding the biases resulting from the cloning step. In addition, since the GS FLX Titanium platform generates sequence reads around 450 bp, spanning the most informative regions of the 16S rRNA gene (Sundquist *et al.*, 2007), we believe this contributes to overcoming the major limitations of short read length, thereby providing a reliable description of a microbial community. However, the

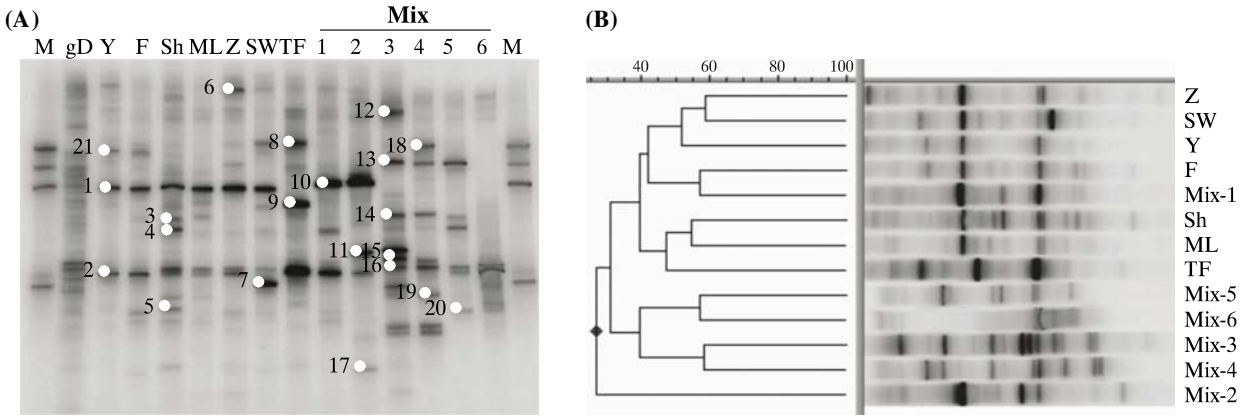

Fig. 6. (A) DGGE profiles of V3 domain fragments of bacterial 16S rRNA genes amplified from cultures enriched from surface seawater. (B) Dendrogram obtained by UPGMA clustering of DGGE patterns. M, marker; gD, genomic DNA extracts; Y, yeast extract-acetate-formate (YAF) medium; F, Fish extract media; Sh, Shellfish extract medium; ML, Mollusk extract medium; Z, Zooplankton extract medium, SW, Seaweed extract medium; TF, Tidal flat slurry extract medium; Mix, mixture of seafood extract.

part detected only by the Sanger method might still indicate that the non-overlapping region in the different primer sets plays a role in selective detection.

Culture-dependent analysis

To investigate how well the existing culturing methods capture the bacterial diversity of the Yellow Sea, enrichment, plate culturing and dilution-to-extinction (HTC) methods were evaluated using their detection capacity for the previously uncultured members of the community.

Influence of media composition and inoculum dilution on microbial communities

First, the effects of media types and the amount of inoculum

Table 1. Identification of the DGGE bands in Fig. 6A from enrichment cultures by 16S rRNA gene sequences. Lanes from 10 to 15, designated as M1 to M6 (Mix media) are shown. Circles indicate the presence and position of each DGGE band

Band No.	Clone	M1	M2	M3	M4	M5	M6	Closest sequence	Acc. No.	Similarity (%)
Band 2	DB2	o	o	o	o	o	o	*Tropicibacter naphthalenivorans*	AB302370	99
Band 10	DB1	o	o					*Cycloclasticus pugetii*	U12624	98
Band 11	DB11		o	o				*Cycloclasticus pugetii*	U12624	96
Band 13	DB13-1			o	o	o		*Tropicibacter naphthalenivorans*	AB302370	98
	DB13-2			o	o	o		Uncultured organism clone MAT-CR-H4-A07	EU245200	98
	DB13-3			o	o	o		*Tenacibaculum aiptasiae*	EF416572	96
	DB13-4			o	o	o		*Novosphingobium indicum*	EF549586	100
Band 14	DB14-1			o	o	o		*Roseovarius tolerans*	Y11551	100
	DB14-2			o	o	o		*Sulfitobacter marinus*	DQ683726	99
	DB14-3			o	o	o		*Novosphingobium indicum*	EF549586	97
	DB14-4			o	o	o		*Tropicibacter naphthalenivorans*	AB302370	99
	DB14-5			o	o	o		*Jannaschia pohangensis*	DQ643999	98
	DB14-6			o	o	o		*Flavobacteriaceae* bacterium TJD809	DQ993346	97
Band 15	DB15-1			o				Uncultured alpha proteobacterium clone CB11G01	EF471465	100
	DB15-2			o				*Cycloclasticus pugetii*	U12624	97
	DB15-3			o				Uncultured bacterium clone B78-57	EU287021	99
	DB15-4			o				*Tenacibaculum aiptasiae*	EF416572	97
	DB15-5			o				*Tropicibacter* naphthalenivorans	AB302370	100
Band 17	DB17		o					*Cycloclasticus pugetii*	U12624	96
Band 18	DB18-1				o			*Tropicibacter naphthalenivorans*	AB302370	99
	DB18-2				o			Uncultured bacterium clone 1C226568	EU799016	100
	DB18-3				o			*Thalassobacter arenae*	EU342372	99
Band 19	DB19				o			*Roseovarius pacificus*	DQ120726	100
Band 20	DB20-1					o	o	*Novosphingobium indicum*	EF549586	98
	DB20-2					o	o	*Gaetbulibacter marinus*	EF108219	97
	DB20-3					o	o	*Antarctobacter heliothermus*	Y11552	99

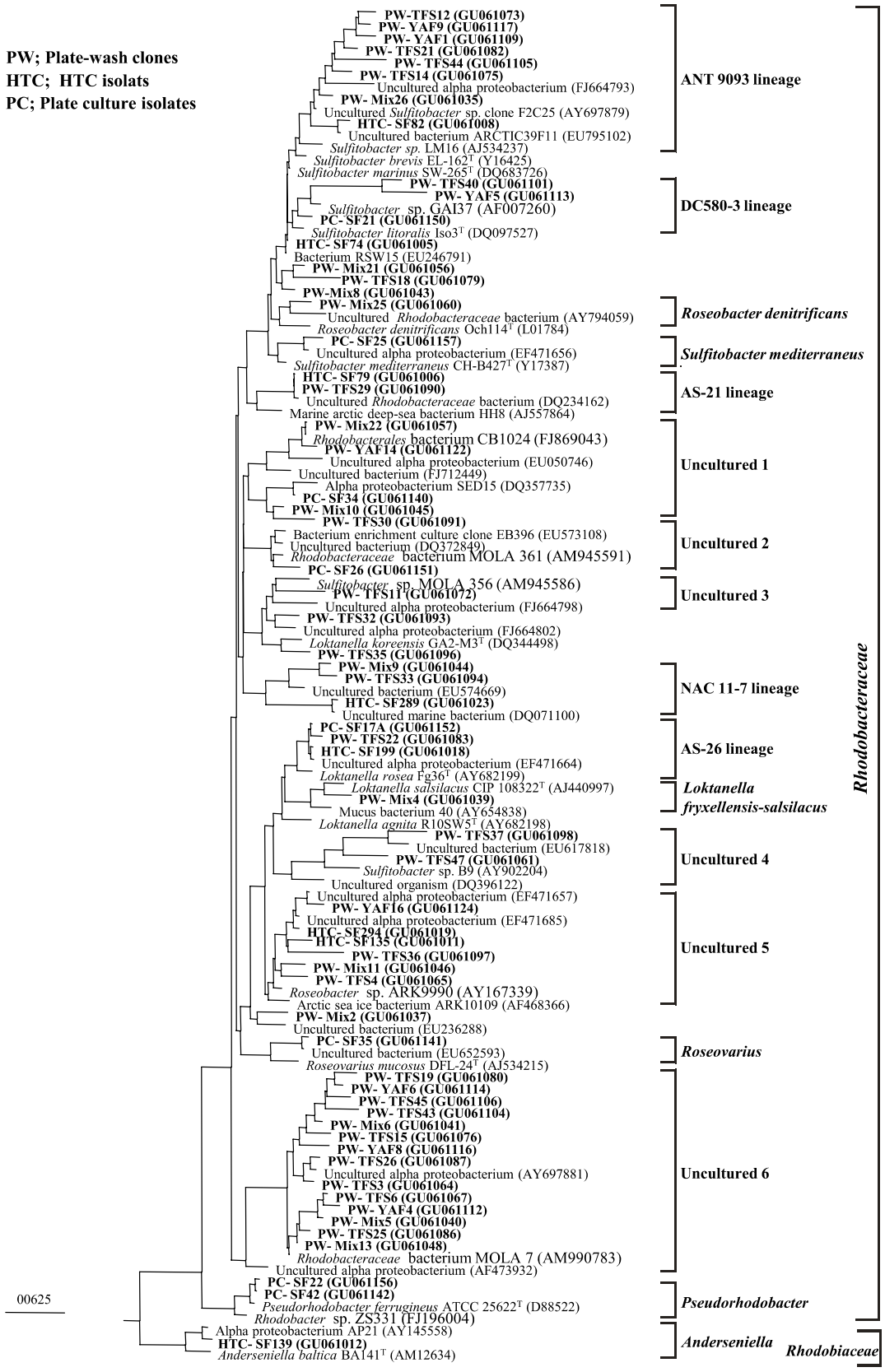

Fig. 7. Phylogenetic tree showing the relationships 16S rRNA gene sequences of clones and isolates belonging to *Alphaproteobacteria*. Sequences from this study are in bold. PW for Plate-wash clones, HTC for HTC isolates and PC for Plate culture isolates.

on the microbial community structure were tested. Enrichment cultures were subjected to DGGE analysis, and the band patterns varied with media types and inoculum dilution rates (Fig. 6A), indicating changes in relative abundances of enriched bacterial phylotypes. Clone libraries were constructed for the bands of interest to examine their taxonomical profiles, and the results are shown in Table 1.

At the 1/10 dilution (lanes 2 to 9), band 1 and 2 were most prominent. Band 2 was common to all lanes showing the strongest intensity in TFS medium, and was closely related to the Alphaproteobacterium, *Tropicibacter naphthalenivorans* (Table 1). Band 1, affiliated to the Gammaproteobacterium, *Cycloclasticus pugetii*, was predominant across all media types except TFS. It might imply that the difference in enriched phylotypes of bacteria in TFS was due to unknown components in the medium. Interestingly, the bacteria correlated with band 1 and 2 (*Cycloclasticus pugetii* and *Tropicibacter naphthalenivorans*) have been shown to degrade aromatic hydrocarbons in previous studies (Dyksterhouse *et al.*, 1995; Harwati *et al.*, 2009).

With respect to the different media types, lane 4 was characterized by the greatest number of bands, indicating that the Sh medium supported the growth of relatively diverse groups of bacteria. In addition to *Alpha-* and *Gammaproteobacteria*, *Beta-* and *Deltaproteobacteria* were also detected in the culture. The ZP medium supported the growth of *Deltaproteobacteria* and *Flavobacteria* (band 6).

In the Mix medium with varying dilution rates (lane 10 to 15), lane 12 with the dilution factor of 10^3 was characterized by the largest number of bands, suggesting the most diverse community structure. Among the bands detected in lane 12, bands 13 and 15 included sequences (DB13-2, 15-1, and 15-3) closely correlated with previously uncultured bacterial clones (Table 1). The clone DB18-2 matched with the uncultured bacterial clone 1C226568 at 100% similarity, which was known to affiliate with the *Roseobacter* clade DC5-80-3 lineage (Shaw *et al.*, 2008). Clones DB13-3, 14-6, and 20-2 belonged to the class *Flavobacteria* and became detectable in relatively highly diluted cultures (lanes 12-15). It is possible that these bacteria were only able to survive when they were exempted from competition with others in the initial inoculum. On the other hand, band 10 (normalized to band 1) became non-detectable below the dilution factor of 10^3. We speculate that the bacteria outcompeted others in the cultures representing lanes 10 and 11, but could not survive at more diluted conditions due to a possible requirement for unknown nutrients or chemicals in the seawater inoculum. Indeed, all the bands had their range of appearance in the DGGE gel, implying a preference for a certain level of competition or crowdedness of the bacteria.

Dendrogram analysis of the DGGE band patterns divided the lanes into two main clusters (Fig. 6B). The dendrogram reflected the media types and dilution rates in its pattern by grouping the same dilution rate (1/10) cultures in one cluster and the Mix medium cultures in the other cluster. However, the clustering patterns did not always appear to directly reflect the media types and the inoculum dilution rate, as observed with the cultures inoculated with seawater of 5 m and 20 m depth (data not shown).

Sequence and phylogenetic analysis of plate-wash clones, plate and HTC isolates

Pure culture isolates were obtained from solid agar plates and HTC cultures. After all the visible colonies were removed from the agar plates, the plates were flushed with sterile seawater and DNA was extracted for PCR. This plate-wash PCR (PWPCR) (Stevenson *et al.*, 2004) was carried out to evaluate and complement the plate culturing method under the assumption that PWPCR can detect any colonies invisible to the naked eye. Out of 8 different kinds of media, YAF, TFS, and MIX were selected for PWPCR to represent varying types of amendments from chemicals, sediment, and *in situ* biomaterials, respectively.

The three culturing methods recovered different taxa from the inoculum (Supplementary data Fig. 2A). HTC sequences fell into *Alphaproteobacteria* and *Gammaproteobacteria*. Solid agar plate sequences retrieved *Bacteroidetes* and *Actinobacteria* in addition to *Alpha* and *Gammaproteobacteria*. For PWPCR, *Alpha*, *Gammaproteobacteria*, and *Bacteroidetes* were obtained from TFS and MIX media, whereas only *Alphaproteobacteria* were recovered from YAF medium. We assume that this might be due to the limitations of manufactured chemicals to supply all the nutrients necessary for growth of certain bacteria most likely including as-yet-uncultured, fastidious bacteria. Overall, all culturing methods agreed that *Alphaproteobacteria* were the most abundant (more than 70%) members of the community.

Among *Alphaproteobacteria*, the family *Rhodobacteraceae* comprised the majority (more than 95% for each culturing method) of the sequences. In the family, *Sulfitobacter*, *Jannaschia*, *Loktanella*, and *Pelagicola* were detected by all three methods, with *Sulfitobacter* being most abundant, i.e., more than one third of the community for each method (Supplementary data Fig. 2B). Clone TFS-47 detected by PWPCR was 95% similar to the RCA cluster, which is composed exclusively of previously uncultured phylotypes. The RCA cluster is known to be widely distributed and abundant in temperate to polar oceans (Selje *et al.*, 2004).

Phylogenetic analysis was performed with the sequences belonging to *Rhodobacteraceae*, as shown in Fig. 7. According to the SILVA 100 database, five major lineages were found within the *Roseobacter* clade. Five plate-wash (PW) clones (TFS-12, 14, 21, 44, YAF-1, 9, MIX-26) and one HTC isolate (SF82) fell into the ANT 9093 lineage, which includes members from diverse environments such as polar sea ice, sediments, and sponges (Buchan *et al.*, 2005). One PW clone (TFS-29) and one HTC isolate (SF79) belonged to the AS-21 lineage. The AS-26 cluster included three sequences obtained from each method, one plate isolate (SF17A), PW clone (TFS-22), and HTC isolate (SF199). Many members of the AS-21 and AS-26 lineages are known to be derived from coastal seawater or sediment. Two PW clones and one plate isolate were affiliated with the DC5-80-3 cluster, for which the global distribution was systematically determined (Selje *et al.*, 2004). The members of the DC5-80-3 clusters were found in temperate and polar oceans, but have not been detected yet in tropical or subtropical seawater samples. Two PW clones (MIX-9 and TFS-33) and one HTC isolate (SF289) fell into the NAC 11-7 lineage. Previous studies suggest that several members of this lineage are often associated with algae and algal blooms (Gonzalez *et al.*, 2000; Zubkov *et al.*, 2002;

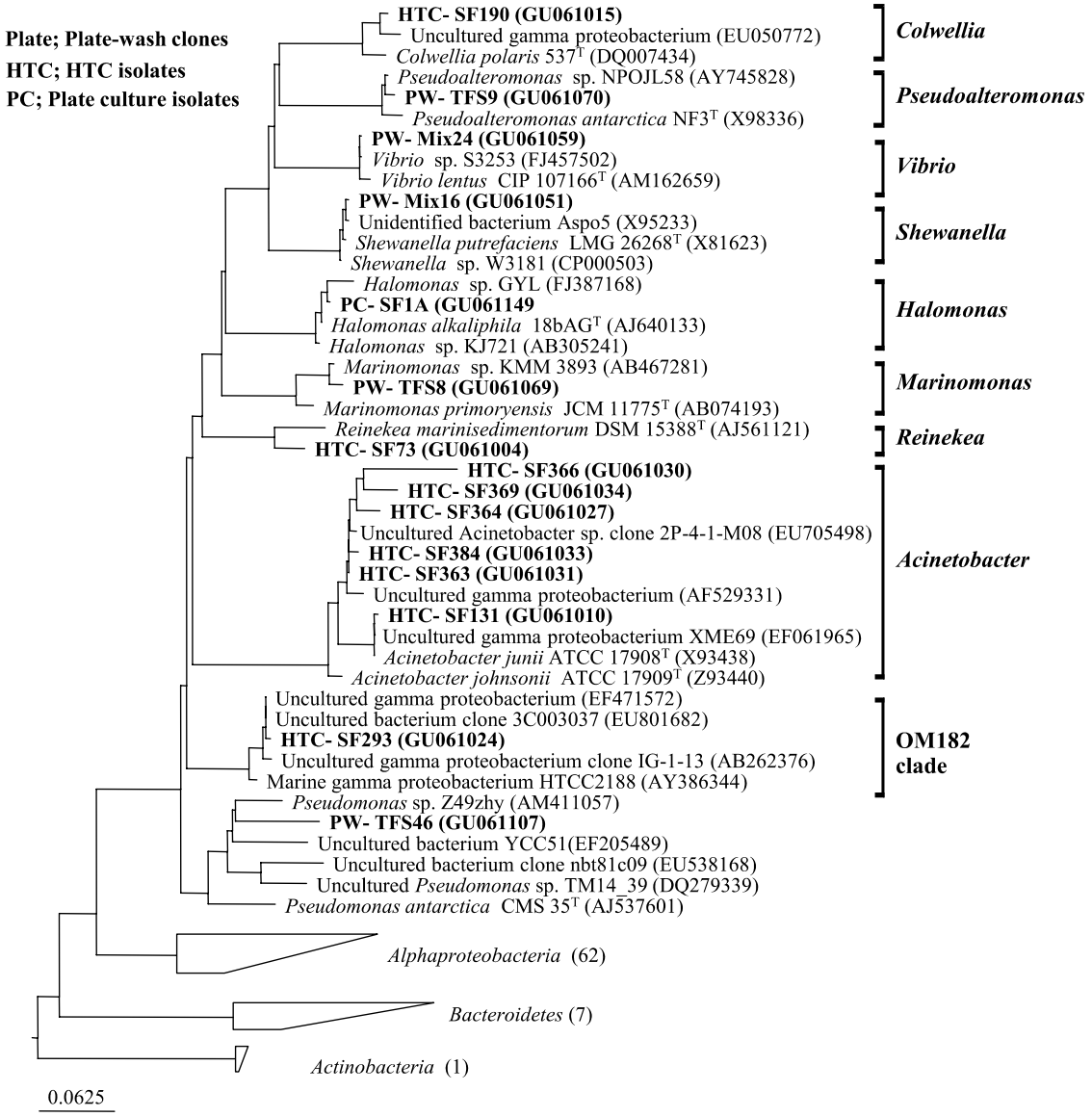

Fig. 8. Phylogenetic tree showing the relationships 16S rRNA gene sequences of clones and isolates belonging to *Gammaproteobacteria*. Sequences from this study are in bold. PW for Plate-wash clones, HTC for HTC isolates and PC for Plate culture isolates.

O'Sullivan *et al.*, 2004; Suzuki *et al.*, 2004). There were two isolates obtained from coastal seawater on oligotrophic media (Wagner-Dobler and Biebl, 2006). The HTC isolate (SF289) we found in this cluster was cultured in the zooplankton-amended (ZP) medium.

Most notably, there were six clusters found that were characterized only by uncultured, environmental clones, named 'Uncultured 1 to 6'. Especially, the cluster 'Uncultured 6' was only represented by a group of PW clones, deprived of any plate culture or HTC sequences. This outcome suggests that PWPCR can be used to detect sequences that are not retrieved by simply picking visible colonies from solid agar plates.

In *Gammaproteobacteria*, one HTC isolate from ZP medium (SF293) was found to be associated with the OM 182 clade (Fig. 8). This clade is one of the five clades belonging to the OMG (oligotrophic marine *Gammaproteobacteria*) group, characterized by sequences derived from exclusively marine environments. Our isolate SF293 was also monophyletic with another HTC isolate (AY386344) from a previous study (Cho and Giovannoni, 2004).

The culture-dependent methods used in this study were evaluated in terms of detecting previously uncultured sequences. For the analysis, the culture sequences were subjected to BLAST analysis, and 100 top-most similar sequences were examined for each query sequence. On average, it turned out that 58% and 48% of them from plate culturing and HTC, respectively, were uncultured sequences (Supplementary data Fig. 3).

The HTC method is well-known for its success in cultivating members of the SAR11 clade, which attracted much interest

for its overwhelming prevalence in surface bacterioplankton communities (Morris *et al.*, 2002). However, in application, the method required time-consuming procedures and an enormous amount of labor compared to its productivity. For instance, in this study, a total of 2304 wells were screened one month after inoculation, and approximately a quarter of them (633) were positive for growth. Among the 233 wells analyzed for identification, 42 wells turned out to be mixed cultures and 53 wells did not yield visible bands with the PCR conditions applied. After filtering the sequences with bad quality, a total of 31 HTC isolates were eventually obtained.

Comparison of culture-dependent and culture-independent data

Finally, we directly compared the sequences obtained from both culture-independent and dependent studies. The culture sequences were BLASTed against the database composed of clone library and pyrosequences obtained from this study. It turned out that only 33% of the culture sequences (26 out of 79) were identical to the sequences in the database. With the dissimilarity cutoff value of 3%, 27% (21 out of 79) of the culture sequences were less than 97% similar. We found that despite the relatively greater number of sequences generated by the culture-independent method, there were still sequences detected only by culturing methods. This indicates that any single approach may not be completely reliable, and that the integrative approach should be taken to describe microbial diversity. To reduce the gap of the results from the molecular and culture-based methods, improved culturability of microorganisms is necessary along with improvement of DNA sequencing technology. In the present study, we believe that culturability was maximized by using natural substances in the media.

Conclusions

In the present study, bacterial community structure of the Yellow Sea was characterized in an integrative way featuring successful application of the recent DNA sequencing technology, Roche/454 FLX Titanium. Interestingly, the traditional clone libraries and culturing methods also detected sequences uncaptured by pyrosequencing. Based on the direct comparison of these culture-dependent and independent techniques, we recommend an integrative approach combining different molecular and culturing methods in characterizing microbial diversity. These findings will serve as a basis for future studies on microbial diversity in nature.

Acknowledgements

This work was supported by the National Research Foundation of Korea Grant (#2011-0016498) funded by the Korea Government (MEST).

References

Ampe, F., N. ben Omar, C. Moizan, C. Wacher, and J.P. Guyot. 1999. Polyphasic study of the spatial distribution of microorganisms in Mexican pozol, a fermented maiz dough, demostrates the need for cultivation-independent methods to investigate traditional fermentations. *Appl. Environ. Microbiol.* 65, 5464-5473.

Buchan, A., J.M. Gonzalez, and M.A. Moran. 2005. Overview of the marine roseobacter lineage. *Appl. Environ. Microbiol.* 71, 5665-5677.

Cheng, S.M. and J.M. Foght. 2007. Cultivation-independent and -dependent characterization of bacteria resident beneath John Evans Glacier. *FEMS Microbiol. Ecol.* 59, 318-330.

Cho, J.C. and S.J. Giovannoni. 2004. Cultivation and growth characteristics of a diverse group of oligotrophic marine Gammaproteobacteria. *Appl. Environ. Microbiol.* 70, 432-440.

Chouari, R., D. Le Paslier, C. Dauga, P. Daegelen, J. Weissenbach, and A. Sghir. 2005. Novel major bacterial candidate division within a municipal anaerobic sludge digester. *Appl. Environ. Microbiol.* 71, 2145-2153.

Chun, J. and M. Goodfellow. 1995. A phylogenetic analysis of the genus *Nocardia* with 16S rRNA gene sequences. *Int. J. Syst. Bacteriol.* 45, 240-245.

Chun, J., K.Y. Kim, J.H. Lee, and Y. Choi. 2010. The analysis of oral microbial communities of wild-type and toll-like receptor 2-deficient mice using a 454 GS FLX Titanium pyrosequencer. *BMC Microbiol.* 10, 101.

Chun, J., J.H. Lee, Y. Jung, M. Kim, S. Kim, B.K. Kim, and Y.W. Lim. 2007. EzTaxon: a web-based tool for the identification of prokaryotes based on 16S ribosomal RNA gene sequences. *Int. J. Syst. Evol. Microbiol.* 57, 2259-2261.

Cole, J.R., B. Chai, T.L. Marsh, R.J. Farris, Q. Wang, S.A. Kulam, S. Chandra, and *et al*. 2003. The Ribosomal Database Project (RDP-II): previewing a new autoaligner that allows regular updates and the new prokaryotic taxonomy. *Nucleic Acids Res.* 31, 442-443.

Connon, S.A. and S.J. Giovannoni. 2002. High-throughput methods for culturing microorganisms in very-low-nutrient media yield diverse new marine isolates. *Appl. Environ. Microbiol.* 68, 3878-3885.

Curtis, T.P., W.T. Sloan, and J.W. Scannell. 2002. Estimating prokaryotic diversity and its limits. *Proc. Natl. Acad. Sci. USA* 99, 10494-10499.

Derakshani, M., T. Lukow, and W. Liesack. 2001. Novel bacterial lineages at the (sub)division level as detected by signature nucleotide-targeted recovery of 16S rRNA genes from bulk soil and rice roots of flooded rice microcosms. *Appl. Environ. Microbiol.* 67, 623-631.

DeSantis, T.Z., Jr., P. Hugenholtz, K. Keller, E.L. Brodie, N. Larsen, Y.M. Piceno, R. Phan, and G.L. Andersen. 2006. NAST: a multiple sequence alignment server for comparative analysis of 16S rRNA genes. *Nucleic Acids Res.* 34, W394-399.

Donachie, S.P., J.S. Foster, and M.V. Brown. 2007. Culture clash: challenging the dogma of microbial diversity. *ISME J.* 1, 97-99.

Dyksterhouse, S.E., J.P. Gray, R.P. Herwig, J.C. Lara, and J.T. Staley. 1995. *Cycloclasticus pugetii* gen. nov., sp. nov., an aromatic hydrocarbon-degrading bacterium from marine sediments. *Int. J. Syst. Bacteriol.* 45, 116-123.

Edgar, R.C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792-1797.

Edwards, R.A., B. Rodriguez-Brito, L. Wegley, M. Haynes, M. Breitbart, D.M. Peterson, M.O. Saar, S. Alexander, E.C. Alexander Jr., and F. Rohwer. 2006. Using pyrosequencing to shed light on deep mine microbial ecology. *BMC Genomics* 7, 57.

Elshahed, M.S., N.H. Youssef, Q. Luo, F.Z. Najar, B.A. Roe, T.M. Sisk, S.I. Buhring, K.U. Hinrichs, and L.R. Krumholz. 2007. Phylogenetic and metabolic diversity of Planctomycetes from anaerobic, sulfide- and sulfur-rich Zodletone Spring, Oklahoma. *Appl. Environ. Microbiol.* 73, 4707-4716.

Giovannoni, S.J., L. Bibbs, J.C. Cho, M.D. Stapels, R. Desiderio, K.L. Vergin, M.S. Rappe, and *et al*. 2005. Proteorhodopsin in the ubiquitous marine bacterium SAR11. *Nature* 438, 82-85.

Gonzalez, J.M., R. Simo, R. Massana, J.S. Covert, E.O. Casamayor, C. Pedrós-Alió, and M.A. Moran. 2000. Bacterial community structure associated with a dimethylsulfoniopropionate-producing North Atlantic algal bloom. *Appl. Environ. Microbiol.* 66, 4237-

4246.

Griffiths, R.I., A.S. Whiteley, A.G. O'Donnell, and M.J. Bailey. 2000. Rapid method for coextraction of DNA and RNA from natural environments for analysis of ribosomal DNA- and rRNA-based microbial community composition. *Appl. Environ. Microbiol.* 66, 5488-5491.

Harwati, T.U., Y. Kasai, Y. Kodama, D. Susilaningsih, and K. Watanabe. 2009. *Tropicibacter naphthalenivorans* gen. nov., sp. nov., a polycyclic aromatic hydrocarbon-degrading bacterium isolated from Semarang Port in Indonesia. *Int. J. Syst. Evol. Microbiol.* 59, 392-396.

Huse, S.M., J.A. Huber, H.G. Morrison, M.L. Sogin, and D.M. Welch. 2007. Accuracy and quality of massively parallel DNA pyrosequencing. *Genome Biol.* 8, R143.

Jeon, Y.S., H. Chung, S. Park, I. Hur, J.H. Lee, and J. Chun. 2005. jPHYDIT: a JAVA-based integrated environment for molecular phylogeny of ribosomal RNA sequences. *Bioinformatics* 21, 3171-3173.

Kim, B.S. 2007. Molecular ecology of prokaryotic community in Korean tidal flat. Seoul National University, Seoul.

Kim, B.S., B.K. Kim, J.H. Lee, M. Kim, Y.W. Lim, and J. Chun. 2008. Rapid phylogenetic dissection of prokaryotic community structure in tidal flat using pyrosequencing. *J. Microbiol.* 46, 357-363.

Kormas, K.A., A. Meziti, A. Dahlmann, G.J. De Lange, and V. Lykousis. 2008. Characterization of methanogenic and prokaryotic assemblages based on *mcrA* and 16S rRNA gene diversity in sediments of the Kazan mud volcano (Mediterranean Sea). *Geobiology* 6, 450-460.

Kunin, V., A. Engelbrektson, H. Ochman, and P. Hugenholtz. 2010. Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ. Microbiol.* 12, 118-123.

Li, W. and A. Godzik. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22, 1658-1659.

Morris, R.M., M.S. Rappe, S.A. Connon, K.L. Vergin, W.A. Siebold, C.A. Carlson, and S.J. Giovannoni. 2002. SAR11 clade dominates ocean surface bacterioplankton communities. *Nature* 420, 806-810.

Morris, R.M., K.L. Vergin, J.C. Cho, S.J. Giovannoni, M.S. Rappe, and C.A. Carlson. 2005. Temporal and spatial response of bacterioplankton lineages to annual convective overturn at the Bermuda Atlantic time-series study. *Limnol. Oceanogr.* 50, 1687-1696.

O'Sullivan, L.A., K.E. Fuller, E.M. Thomas, C.M. Turley, J.C. Fry, and A.J. Weightman. 2004. Distribution and culturability of the uncultivated 'AGG58 cluster' of the Bacteroidetes phylum in aquatic environments. *FEMS Microbiol. Ecol.* 47, 359-370.

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

Porter, K.G. and Y.S. Feig. 1980. The use of DAPI for identifying and counting aquatic microflora. *Limnol. Oceanogr.* 25, 943-948.

Quince, C., A. Lanzen, T.P. Curtis, R.J. Davenport, N. Hall, I.M. Head, L.F. Read, and W.T. Sloan. 2009. Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat. Methods* 6, 639-641.

Schellenberg, J., M.G. Links, J.E. Hill, T.J. Dumonceaux, G.A. Peters, S. Tyler, T.B. Ball, A. Severini, and F.A. Plummer. 2009. Pyrosequencing of the chaperonin-60 universal target as a tool for determining microbial community composition. *Appl. Environ. Microbiol.* 75, 2889-2898.

Schloss, P.D., S.L. Westcott, T. Ryabin, J.R. Hall, M. Hartmann, E.B. Hollister, R.A. Lesniewski, and *et al*. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75, 7537-7541.

Selje, N., M. Simon, and T. Brinkhoff. 2004. A newly discovered *Roseobacter* cluster in temperate and polar oceans. *Nature* 427, 445-448.

Shaw, A.K., A.L. Halpern, K. Beeson, B. Tran, J.C. Venter, and J.B. Martiny. 2008. It's all relative: ranking the diversity of aquatic bacterial communities. *Environ. Microbiol.* 10, 2200-2210.

Stevenson, B.S., S.A. Eichorst, J.T. Wertz, T.M. Schmidt, and J.A. Breznak. 2004. New strategies for cultivation and detection of previously uncultured microbes. *Appl. Environ. Microbiol.* 70, 4748-4755.

Sun, Y., Y. Cai, L. Liu, F. Yu, M.L. Farrell, W. McKendree, and W. Farmerie. 2009. ESPRIT: estimating species richness using large collections of 16S rRNA pyrosequences. *Nucleic Acids Res.* 37, e76.

Sundquist, A., S. Bigdeli, R. Jalili, M.L. Druzin, S. Waller, K.M. Pullen, Y.Y. El-Sayed, M.M. Taslimi, S. Batzoglou, and M. Ronaghi. 2007. Bacterial flora-typing with deep, targeted, chip-based pyrosequencing. *BMC Microbiol.* 7, 108.

Suzuki, M.T., C.M. Preston, O. Beja, J.R. de la Torre, G.F. Steward, and E.F. DeLong. 2004. Phylogenetic screening of ribosomal RNA gene-containing clones in Bacterial Artificial Chromosome (BAC) libraries from different depths in Monterey Bay. *Microbiol. Ecol.* 48, 473-488.

Thompson, J.D., D.G. Higgins, and T.J. Gibson. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673-4680.

Wagner-Dobler, I. and H. Biebl. 2006. Environmental biology of the marine Roseobacter lineage. *Annu. Rev. Microbiol.* 60, 255-280.

Webster, N.S., R.E. Cobb, R. Soo, S.L. Anthony, C.N. Battershill, S. Whalan, and E. Evans-Illidge. 2010. Bacterial community dynamics in the marine sponge *Rhopaloeides odorabile* under *in situ* and *ex situ* cultivation. *Mar. Biotechnol.* 13, 296-304.

Zubkov, M.V., B.M. Fuchs, S.D. Archer, R.P. Kiene, R. Amann, and P.H. Burkill. 2002. Rapid turnover of dissolved DMS and DMSP by defined bacterioplankton communities in the stratified euphotic zone of the North Sea. *Deep-Sea Res. II.* 49, 3017-3038.