Acta Oceanol. Sin., 2017, Vol. 36, No. 10, P. 106–114
DOI: 10.1007/s13131-017-1055-6
http://www.hyxb.org.cn
E-mail: hyxbe@263.net

Development of a 16S rRNA gene-based microarray for the detection of marine bacterioplankton community

ZHAO Wei[1], WANG Jingjing[1], LIANG Yajie[1], HUANG Zhiyong[1]*

[1] Tianjin Key Laboratory of Industrial Biological Systems and Process Engineering, Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China

Received 26 April 2016; accepted 6 December 2016

Abstract

A better understanding of bacterioplankton community shifts following change in marine environments is critical to predict the marine ecosystem function. In order to get a snapshot of the microbial taxonomy profiling of a wide range marine area, a quick, convenient and low cost method would be favorable. In this study, we developed a 16S rRNA gene-based microarray using ARB software, which contained 447 probes targeting 160 families of marine bacteria. The specificity, sensitivity and quantitative capability of this microarray were assessed by single cloned 16S rRNA genes. The reliability of this microarray was tested by eight environmental samples. The results showed that the microarray was specific, only 1.16% false results were detected in five single-clone hybridization tests. The microarray could detect DNA samples as few as 1 ng/μL and the signal intensity could reflect the relative abundance of the bacteria in the range of 1 ng/μL to 100 ng/μL of DNA concentration. Hybridization with environmental samples showed that it can discriminate bacterioplankton communities by sites and time. High throughput sequencing results from the eight samples confirmed the hybridization results. It indicated that this developed microarray could be used as a convenient tool to monitor the bacterioplankton community in marine environment.

Key words: microarray, bacterioplankton community, 16S rRNA gene, marine environment

Citation: Zhao Wei, Wang Jingjing, Liang Yajie, Huang Zhiyong. 2017. Development of a 16S rRNA gene-based microarray for the detection of marine bacterioplankton community. Acta Oceanologica Sinica, 36(10): 106–114, doi: 10.1007/s13131-017-1055-6

1 Introduction

Microorganisms play an important role in marine ecosystem, which has been regarded as the key engine of global geochemical processes, including carbon, nitrogen, sulfur cycles, etc. (Fuhrman, 2009). They are also probably more diverse than any other organisms, so it is easy to see why the structure of microbial communities, that is, the different kinds of organisms and their abundances, is so important to the way in which ecosystems function (Fuhrman, 2009). As reported, a strong correlation exists between marine microorganisms and marine pollution or ecological disasters (Rusch et al., 2003; Kim et al., 2008; Liu et al., 2008; Mahmud et al., 2008; Wang et al., 2010; Chen et al., 2010; Xiao et al., 2010). The microbial community structure and composition could reflect the physical and chemical properties of water and sediments (Ramaiah et al., 1996; Nakashima et al., 2006; Fuhrman, 2009). Therefore, it is necessary to detect the microorganisms in marine environment to evaluate the marine ecological system function, which may further provide reference for the improvement of marine eco-environmental security.

In the past few years, 16S rRNA gene-based oligonucleotide microarray or microchip or PhyloChip have been successfully used in environmental studies based on the advantages, which include: (1) it is a high-throughput format for the parallel detection of bacteria in environmental samples (Bodrossy et al., 2003); (2) it benefits from large databases for probe design and data interpretation (Sanguin et al., 2006a); (3) probe design can be quickly and easily completed by ARB software at various taxonomic levels (Ludwig et al., 2004); (4) it is a superior tool for directly identifying the bacteria present in a sample and for semi-quantitative community analysis. Until now various formats of microarrays had been developed and evaluated for bacterial detection and microbial community analysis in complex environments (Wu et al., 2001; Zhou and Thompson, 2003). Loy et al. (2002) developed a glass-based microarray containing 132 SSU rRNA-targeted oligonucleotide probes, representing all recognized groups of sulfate-reducing prokaryotes. They used this microarray to determine the diversity of sulfate-reducing prokaryotes in two different environments. The results were a consistent with those obtained using the conventional molecular methods, which suggested that microarray hybridization is a powerful tool for analyzing community structure. Liles et al. (2010) designed an oligonucleotide microarray using probe sequences based upon the 16S rRNA genes recovered from members of the bacterial division *Acidobacteria*, which successfully revealed a lineage-specific distribution in a soil clay fraction. Besides, Kyselková et al. (2008) developed a 16S rRNA-based microarray that targets key *Actinomycetes*, which enabled discrimination between *Actinomycete* communities from three forest soil samples, and the cloning and sequencing of one soil sample confirmed the microarray results. Moreover, Sanguin et al. (2008) developed a prototype 16S rRNA-based taxonomic microarray for *Alphaproteobacteria*. These studies mostly focused on a certain

Foundation item: The Science and Technology Support Program of Tianjin Municipal under contract Nos 15ZCZDSF00620 and 11ZCKFSY07800; the Open Fund for Priority Discipline of Zhejiang Province under contract No. xkzsc10.
*Corresponding author, E-mail: huang_zy@tib.cas.cn

kind of microorganism, which could not apply to environmental studies of abundant microorganisms. A high-density oligonucleotide microarray containing 500 000 probes developed by Brodie et al. (2006) could detect samples from different environments, such as heavily contaminated soil, rhizosphere microbiome (Mendes et al., 2011), contaminated marine environment (Hazen et al., 2010), which revealed the great applied potential of the "big PhyloChip". As we can see, the "big PhyloChip" could be widely used in many kinds of environments, benefitting from its high density and wide species coverage probes. However, for some specific environments where the main common species are well known, there may be no need to examine with the super "big Phylochip", since hundreds of thousands of probes may produce redundant data which leads to time-consuming analysis. Additionally, too many probes leads to an expensive construction of the chip. In this study we mainly targeted the monitoring of planktonic bacteria, which was reported critical to predict ecosystem function in the aquatic ecosystems (Comte and del Giorgio, 2011; Xiong et al., 2014), and the designation of the oligonucleotide probes was also mainly included the marine bacteria taxa, in which the 16S rRNA sequences were collected through the literature search and NCBI website with source of marine water or related environments.

In our study, a microchip targeting the detection of bacterioplankton was constructed, which contained most of the marine bacterioplankton, with a suitable size for the application of it. The probes were designed with ARB software. Single cloned 16S rRNA gene was used to validate the probe specificity. A series of different concentrations of *Rhodobacteraceae* 16S rRNA gene PCR products were used to test the sensitivity and quantitative capability of the microchip. Eight seawater samples from different sources were used to evaluate the reliability of this chip, which were then tested by an independent pyrosequencing analysis to validate the hybridization results.

2 Materials and methods

2.1 *Oligonucleotide probe design*

Oligonucleotide probes were designed using the ARB software package and the ssu_jan04_corr_opt. arb dataset of sequences (http://www.arb-home.de/). Additional marine bacteria sequences deposited in GenBank were also included. The parameters of the Probe Design function were set to the following values: length of output=100; length of probe=30, 31 and 32 nt; G+C content (%)=40–60, position=1–10 000 in *E. coli*, max. non-group hits=0, min group=10%–50%. The best probes were selected from the Probe Results window, and their predicted melting temperature T_m (according to the nearest neighbor method) were calculated using Oligo7 (Molecular Biology insights, West Cascade, CO) with default settings. Probes were chosen preferably with a same melting temperature (T_m), no hairpin and no stable homoduplex. Then, the specificity of probes was checked in BLAST. Probes having higher similarity (>90%) with non-target group were removed from the probe set. When these requirements could not be met, probes with suboptimal conditions were accepted in order to provide adequate phylogenetic coverage.

2.2 *Control probes*

Several control probes are included in the probe set. PC (ACT CCT ACG GGA GGC AGC AGT GGG GAA TAT) as positive control is a conservative part targeting the V3 region of 16S rRNA gene. NC (TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT) used as negative controls which corresponded to sequences not amplified during the PCR. CK (TTT TTT TTT TTT TTT TTT TTT TTT TTT TTT-HEX) was used as printing control, of which the HEX has fluorescence signal either there is hybridization or not, which could guide position.

2.3 *Microarray manufacturing and processing*

Probes were synthesized (Sangon, Shanghai, China) with a 5′C6-NH2 group for covalent attachment onto aldehyde slides AL (CapitalBio, Beijing, China). The 5′ end of each oligonucleotide probe was tailed with ten dTTP molecules (T-spacer) to increase the on-chip accessibility of spotted probes to target DNA (Loy et al., 2002b). Probes were spotted onto slide with a Smart Arrayer 48 (CapitalBio, Beijing, China), which was conducted by CapitalBio Corporation. There are four blocks in one microarray. The basic probe pattern on each block consisted of four spots of the positive control probes, six spots of the negative probes, four spots for the probe HEX (used as landmarks for image analysis) and one spot for each feature probe. The basic pattern was replicated three times on a microarray, leading thus to three spots per feature probe per microarray.

2.4 *Single clones and environmental samples*

The single clones used to validate the microarray are listed in Table 1. The 16S rRNA gene of the clones was sequenced. All the sequences have been submitted to NCBI database under accession numbers KT764021-KT764026. Seawater sample 22A and 151A were taken from the East China Sea in summer of 2013. Seawater sample P203, P602, Y1P404, Y1MB601, PE104 and HSP104 were sampled from a shrimp pond located in Zhanqi, Ningbo, China (29°32′N, 121°31′E) (Xiong et al., 2014), among which P203 and P602 were sampled in 2012 and the remaining were sampled in 2014. Five hundred milliliters of each seawater sample were filtered through the filtration membrane with a diameter of 0.22 μm. The filter membrane was used for DNA extraction with a Power Water DNA Isolation Kit (MO BIO Laboratories, lnc.) according to manufacturer's instructions.

2.5 *Polymerase chain reaction amplification of 16S rRNA gene and labelling*

The 16S rRNA gene of the six clones and eight seawater

Table 1. 16S rRNA gene clones used in the hybridization

Clones	Similar organisms	Identity /%	Group	Accession number	Environmental source
A	*Nautella* sp. BBD-FLK-1d	99	*Rhodobacteraceae*	KT764022	Sediments from the East China Sea
B	Uncultured *Comamonadaceae* bacterium clone DS140	97	*Comamonadaceae*	KT764023	East China Sea
C	*Francisella philomiragia*	99	*Francisellaceae*	KT764024	East China Sea
D	Uncultured *Haliscomenobacter* sp.	99	*Saprospiraceae*	KT764025	Sediments from the East China Sea
E	*Candidatus Aquiluna rubra*	98	*Microbacteriaceae*	KT764021	East China Sea
F	*Aureimarina marisflavi* strain IMCC3054	99	*Flavobacteriaceae*	KT764026	East China Sea

samples were amplified by primers 27F (5′-AGA GTT TGA TCM TGG CTC AG-3′) and 1492R (5′-GGY TAC CTT GTT ACG ACT T-3′) (Tanner et al., 1999). The PCR reaction mixtures (25 μL) contained 12.5 μL 2×Green PCR Master Mix (Thermo Fisher Scientific, Shanghai, China), 1 μL each primer (2.5 μmol/L), 1 μL seawater genomic DNA or 16S rRNA gene (~20 ng) and 9.5 μL ddH_2O. Thermal cycling was carried out with a denaturation step of 94°C for 3 min, 30 cycles with 1 min denaturation at 94°C, 1 min annealing at 55°C, 1 min 20 seconds elongation at 72°C. The 5 μL PCR products were denatured with 3 μL random primers Cy3-NNN NNN NNN (Sangon Biotech Co., Ltd, Shanghai, China) and 9 μL ddH_2O at 95°C for 3 min, and then 3 min on ice. Add 2.5 μL 5×Klenow buffer, 2.5 μL NTP (2.5 mmol/L) and 1 μL Klenow enzyme (NEB, Beijing, China) in the mixture, incubated at 37°C for 90 min, and then 70°C for 10 min (Schenk et al., 2000).

2.6 *Microarray hybridization*

Labelled 16S rRNA gene PCR pruducts of Clones A, B, C, D, E were hybridized with the array separately, then a series of different concentrations (1 ng/μL, 10 ng/μL, 100 ng/μL and 1 000 ng/μL) of 16S rRNA gene PCR products from Clone A were hybridized to test the sensitivity and quantitative capability of the microchip. Moreover, equally mixed labelled 16S rRNA gene PCR products of Clones A, B, C, D, F were hybridized with the array, and following were labelled 16S rRNA gene amplicons of the eight seawater samples. The hybridization protocols were as following: labelled DNA (15 μL) was mixed with 5 μL of hybridization buffer (5×SSC, 0.02% SDS, 5% formamide), and denatured at 95°C for 5 min, placed on ice immediately. The slides were placed in a hybridization chamber (CapitalBio, Beijing, China) and the chamber preheated to 60°C. The mixture was transferred onto the slide, covered with a Hybri-slip (CapitalBio, Beijing, China). Hybridization was conducted overnight at 60°C, then the slides were immediately immersed into 2×SSC containing 0.03% SDS at 60°C to remove the Hybri-slips and then transferred the slide to a new container, washed by shaking at 60°C for 5 min, and successively washed by shaking at 60°C for 5 min in fresh 0.2×SSC and then for 5 min in 95% ethyl alcohol at room temperature (Sanguin et al., 2006b). Finally, the slides were quickly dried by centrifugation with a mini slide centrifuge (ChipMate PMC-082, Tomy, Japan) at the default set values.

2.7 *Scanning and array normalization*

The slides were scanned at 10 μm resolution with a NimbleGen MS200 scanner (Roche NimbleGen) at 532 nm wavelength for Cy3, and images were analyzed with the GenePixPro 6.0 software (Axon, Union City, CA, USA). Spot quality was visually checked and spots of poor quality (presence of dust, spot morphology and saturation) had been excluded before further analysis. The signal intensity of each probe was calculated by the following formula: signal intensity=Log_2(signal mean-background intensity). Spots with a signal-to-noise ratio (SNR) (SNR=(signal mean-background intensity)/(background standard deviation)) greater than 2.0 were thought to be meaningful (Zhao et al., 2014; Liu et al., 2015). The value of the meaningful line was equal to the signal intensity of the spot with the SNR equalling 2. Additionally, probes present only once in three replicates were removed.

2.8 *Pyrosequencing analysis of the eight marine samples*

An aliquot (50 ng) of DNA from eight samples (22A, 151A, Y1MB601, Y1P404, PE104, HSP104, P203 and P602) was used as the template for amplification. The V1–V3 hyper variable regions of bacterial 16S rRNA genes were amplified using the primer set 27F and 519R (Zhang et al., 2014). This region furnished nearly the same resolution as that of the nearly full length sequence (Kim et al., 2011). Each sample was amplified in triplicate with a unique barcode primer. PCR products for each sample were combined and purified with a PCR fragment purification kit (TaKaRa Biotech, Japan). An equimolar amount of PCR products (assuming that the same size of amplicons had a similar molar mass) for each sample were combined in a single tube and run on a Roche FLX 454 pyrosequencing machine (Roche Diagnostics Corporation, Branford, CT, USA), producing reads from the forward direction 27F with the barcode. Sequencing reads were quality filtered and chimera checked using the Quantitative Insights into Microbial Ecology (QIIMEv1.5.0) workflow (Caporaso et al., 2010).

2.9 *Statistic analysis*

The probe sequences and 16S rRNA gene sequences used for testing were aligned through MEGA and the number of mismatches between them was counted. Analysis for the nonspecific probes was based on the in silico prediction of the secondary structure formed between the probe and the test duplex. The formation of stable hairpin structures was estimated using the DINAMELT server (http://www.bioinfo.rpi.edu/applications/hybrid/; Markham and Zuker, 2005) in the presence of 1 mol/L Na^+ in DNA mode at 60°C for the Two-state melting (hybridization) algorithm. Principal Coordinate Analysis (PCoA) for the eight samples was conducted by the PAST software (Hammer et al., 2001). The microarray data used for the statistics of the bacterial composition of the eight environmental samples was taken from the hybridization data above the meaningful line.

3 Results

3.1 *Probe design*

The microchip contained 447 probes targeting 160 families. Probes were designed to discriminate at family level. In all, 23 phyla were included in this microchip. In detail, 28 families were stemmed from *Actinobacteria*; 13 families were came out of *Bacteroidetes*; 18 families were belong to *Firmicutes*; 23 families were originated from *Alphaproteobacteria*; 9 families were from *Betaproteobacteria*; 31 came from *Gammaproteobacteria*; 10 were from *Deltaproteobacteria*; The other 28 families belonged to *Acidobacteria* (1), *Aquificae* (1), *Chlamydiae* (1), *Chlorobi* (1), *Chloroflexi* (3), *Deferribacteres* (2), *Deinococcus-Thermus* (3), *Fibrobacteres* (1), *Fusobacteria* (1), *Gemmatimonadetes* (1), *Lentisphaerae* (2), *Planctomycetes* (2), *Spirochaetes* (2), *Tenericutes* (1), *Verrucomicrobia* (3) and *Epsilonproteobacteria* (2). A probe of mitochondria representing the eukaryotic microorganism was also included.

In order to have homogenous hybridization conditions, in the beginning, the probe design parameters were set within limits (see Materials and Methods). In some cases these requirements could not be met and so probes with suboptimal conditions were accepted in order to provide adequate phylogenetic coverage. The length of probes ranged from 28 bp to 34 bp, the melting temperature ranged from 56.3°C to 76.6°C. The GC content were mostly between 40% and 60%, and only 2% of all the probes were lower than 40% and 8.7% of all were higher than 60%. The hairpin ΔG were mostly above –4 kcal/mol (Supplementary Table S1).

3.2 *Hybridization results of the single clone*

A set of single, cloned 16S rRNA genes were hybridized with

the microarray to evaluate the specificity of the probes. Each of the expected positive features produced hybridization signal intensity significantly greater than background signal intensity, indicating the presence of *Rhodobacteraceae* (Clone A targeted probes 20140620-173 (15.85), 20140620-174 (13.85) and 20140620-175 (14.14), values in the bracket are the corresponding signal intensity of the probes) (Fig. 1a), *Comamonadaceae* (Clone B targeted probes 20140620-120 (15.24) and 20140620-121 (15.93)) (Fig. 1b), *Francisellaceae* (Clone C targeted probes 20140620-130 (15.46), 20140620-131 (13.99) and 20140620-132 (13.96)) (Fig. 1c), *Saprospiraceae* (Clone D targeted probes 20140825-5 (15.59) and 20140620-52 (13.33)) (Fig. 1d) and *Microbacteriaceae* (Clone E targeted probes 20140825-1 (12.22) and 20140825-2 (12.48)) (Fig. 1e).

During the hybridization, some unexpected probes also produced significant hybridization signal, which accounted for 1.16% among the five single clone hybridizations (i.e., 5 targets × 447 probes). The most commonly present unexpected probe was 20140620-144, which was detected every time in the five single clone hybridization. Probe 20140620-168 and 20140825-10 ap-

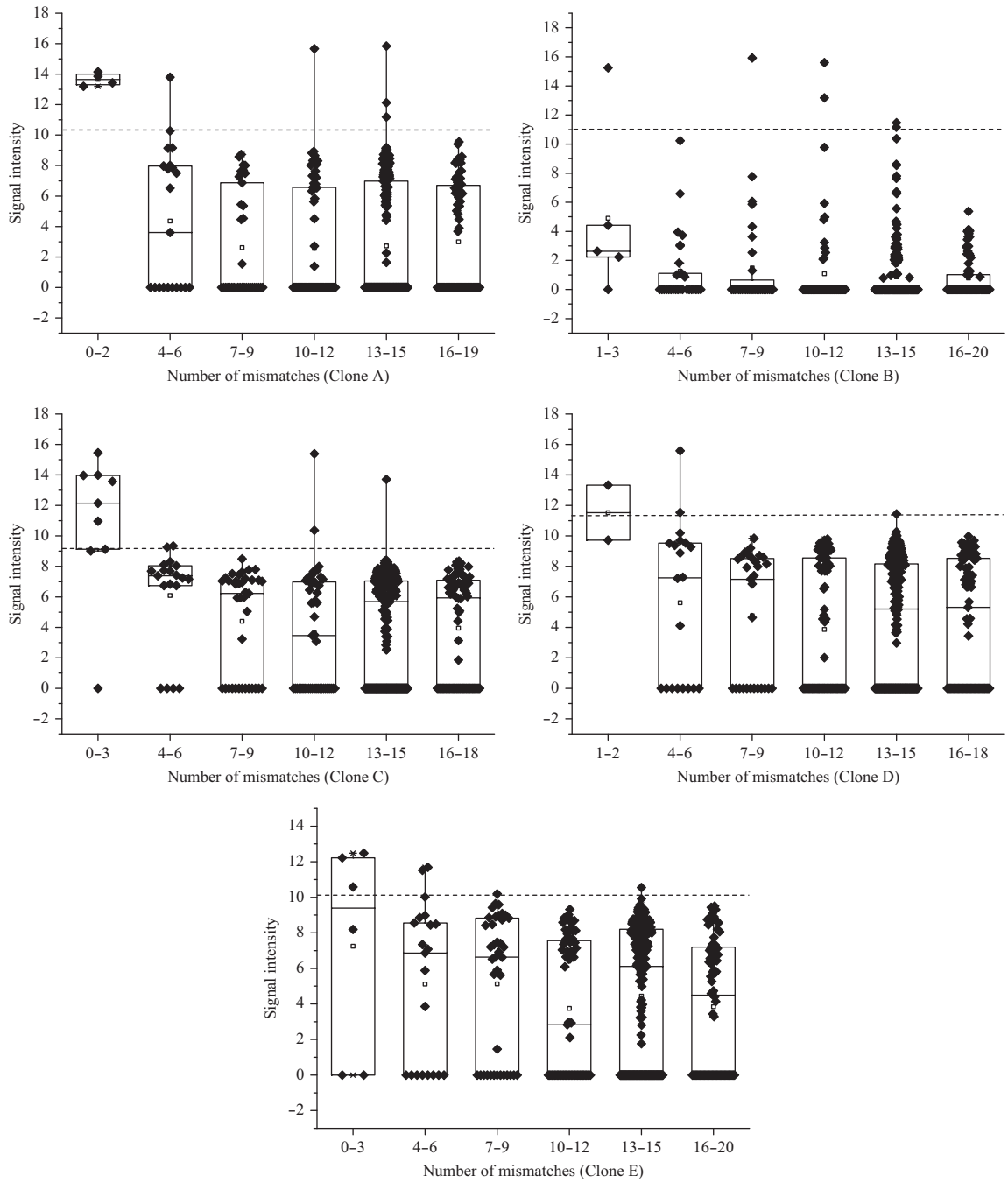

Fig. 1. Box plot of normalized signal intensity in probe-target combinations categorized according to number of mismatches. Five single, cloned 16S rRNA genes from *Rhodobacteraceae* (A), *Comamonadaceae* (B), *Francisellaceae* (C), *Saprospiraceae* (D) and *Microbacteriaceae* (E) were used as template to hybridize with the microarray. For each clone, the hybridization signal intensity of each probe on the microarray was categorized according to the number of mismatches between the probe and the test 16S rRNA gene. The dash line indicates the position of meaningful signal value of each clone's hybridization.

peared more than twice. The statistic of mismatch number between these unexpected probes and the tested rRNA sequence showed a low identity. The in silico prediction of the secondary structures between the nonspecific probes and target duplexes showed that the nonspecific hybridization was mainly due to two reasons: first, the higher content of continuous GC base composition; second, opportunistic hybridization led to stable duplex and high signal intensity (Supplementary Figs S1–S4) (Kyselková et al., 2008).

Besides, the distribution of the signal value to the mismatch number showed that the probes with a mismatch number during 0 to 3 had relatively high signal value and the expected probes were mostly located in this interval. The unexpected probes had the signal intensity scattered in the region with the mismatch number above 3.

3.3 *Hybridization results with different gradient concentrations of Rhodobacteraceae 16S rRNA gene*

Single, cloned *Rhodobacteraceae* 16S rRNA genes under gradient concentrations were hybridized with the microchip. Of the three features of this family, two features (20140620-175, 20140620-174) had significant hybridization signal. The results showed that when the concentration was 100 ng/μL the highest signal intensity was achieved (Fig. 2). The signal intensity could be detected when the rRNA genes was 1 ng/μL, which indicated that the detection limit of the microarray was 1 ng/μL. When the concentration was 1 000 ng/μL the signal value was saturated. When the rRNA gene concentration increased from 1 ng/μL to 100 ng/μL the signal intensities of the two spots were also increased (Fig. 2), which revealed the quantitative capability of the microchip.

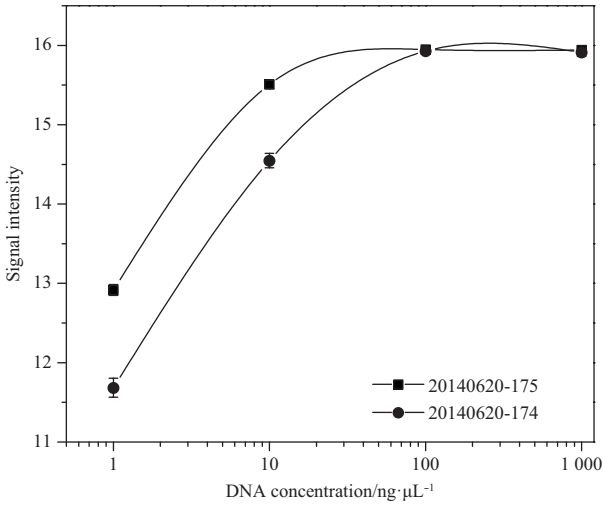

Fig. 2. Sensitivity and quantitative capability test of the microarray. The single, cloned 16S rRNA gene from *Rhodobacteraceae* was used as template to hybridize with the microarray under the template concentration of 1 ng/μL, 10 ng/μL, 100 ng/μL and 1 000 ng/μL. Two probes (20140620-175 and 20140620-174) designed for the detection of *Rhodobacteraceae* had positive signal. The normalized signal intensity of the two probes under the tested condition was correlated with the concentration of the template.

3.4 *Hybridization results with the mixed clones*

Mixed clones (Clones A, B, C, D, and F) were hybridized with the microchip (Fig. 3). Each of the expected positive features produced hybridization signal intensity significantly greater than the meaningful value, indicating the presence of *Comamonadaceae* (Clone B targeted probes 20140620-121 (15.94) and 20140620-120 (15.34)), *Francisellaceae* (Clone C targeted probes 20140620-131 (15.94), 20140620-130 (15.94) and 20140620-132 (15.17)), *Rhodobacteraceae* (Clone A targeted probes 20140620-173 (15.94), 20140620-175 (15.10) and 20140620-174 (14.98)), *Flavobacteriaceae* (Clone F targeted probes 20140620-39 (14.31)) and *Saprospiraceae* (Clone D targeted probes 20140825-5 (14.58) and 20140825-4 (13.23)). All the five clones were detected by the microarray and there still were unexpected probes: some (6 probes) are the same with the former single clone hybridization, others (11 probes) may be caused by the new added Clone F.

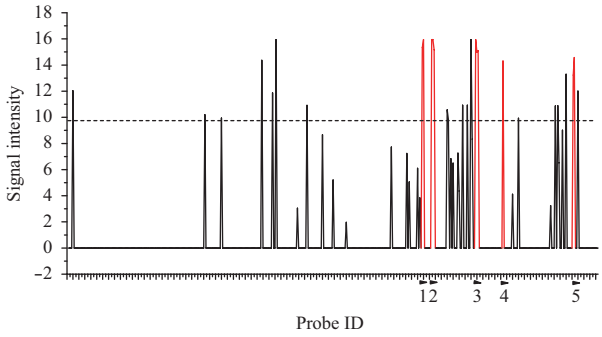

Fig. 3. Hybridization results of the developed marine bacterial microarray with clone mixtures. Five clones from *Rhodobacteraceae*, *Francisellaceae*, *Comamonadaceae*, *Saprospiraceae*, and *Flavobacteriaceae* were used. Above the dash line is the signal value of probes that had a meaningful SNR value when hybridized with the microarray chip. The lines with red color are the probes hybridized with the targeted clone sequence. The position of triangle with number 1 locates probes 20140620-121 and 20140620-120, which are targeted at the *Comamonadaceae* duplex; the position of triangle with number 2 locates probes 20140620-131, 20140620-130 and 20140620-132, which are targeted at the *Francisellaceae* duplex; the position of triangle with number 3 locates probes 20140620-173, 20140620-175 and 20140620-174, which are targeted at the *Rhodobacteraceae* duplex; the position of triangle with number 4 locates probe 20140620-39, which are targeted at the *Flavobacteriaceae* duplex; the position of triangle with number 5 locates probes 20140825-5 and 20140825-4, which are targeted at *Saprospiraceae* duplex.

3.5 *Hybridization results of seawater samples*

Eight water samples from different sources were assessed by the microarray. Hybridization results showed that the dominant phyla or classes across the eight samples were *Alphaproteobacteria* (averaged relative abundance, 31.8%), *Gammaproteobacteria* (25.7%), *Actinobacteria* (12.0%), *Bacteroidetes* (7.3%), *Betaproteobacteria* (6.1%) and *Deltaproteobacteria* (5.6%), representing more than 80% of the total bacterial amount (Fig. 4). Based on the detected bacteria at the family level across the samples, principal component analysis (PCoA) (Fig. 5) showed that Samples 22A and 151A were separated from the other six samples on the first axis (explaining 41.55% of variability) and Samples P602 and P203 were separated from the others on the second axis (explaining 18.88% of variability). On the whole, Samples 22A and 151A clustered together, Samples P203 and P602 clustered together, and Samples PE104, HSP104, Y1P404 and Y1MB601 clustered together.

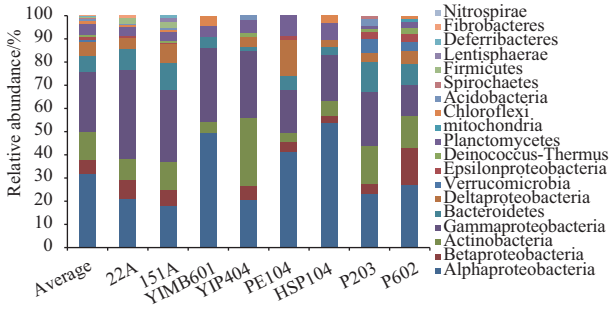

Fig. 4. Bacterial composition based on the hybridization data, and relative abundances of bacterial phyla in seawater samples and the average across the samples. Relative abundances are based on the proportional of those probe signal intensity that could be classified at the phylum level, with the exception that the predominant phyla of Proteobacteria were grouped at the class level.

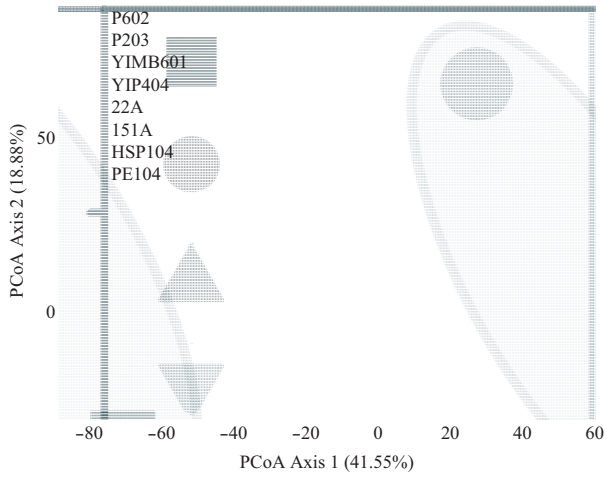

Fig. 5. PCoA analysis of the eight sea water samples based on the hybridization data. Principal coordinate analysis (PCoA) derived from the Bray-Curtis distances between seawater samples according to the probe signal intensity that was classified on the family level. The percentage of variation explained by each axis is shown. The samples in ellipse had relatively similar bacterial community structure. Samples 22A and 151A were from East China Sea, Samples P203 and P602 were from shrimp pond of 2012, and Samples PE104, HSP104, Y1P404 and Y1MB601 were from shrimp pond of 2014.

3.6 *Pyrosequencing results of seawater samples*

The pyrosequencing results of the eight samples showed that the dominant phyla or classes across the samples were *Alphaproteobacteria* (averaged relative abundance, 25.9%), *Gammaproteobacteria* (21.2%), *Actinobacteria* (18.3%) and *Bacteroidetes* (16.8%), representing more than 80% of the bacterial reads (Fig. 6). The PCoA analysis with pyrosequencing data were also based on the family level. The eight samples were divided into three groups (Fig. 7): Samples 22A and 151A were separated from Samples P602 and P203, and Samples PE104, HSP104, Y1P404 and Y1MB601 on the first axis (explaining 39.52% of variability) and the second axis (explaining 26.14% of variability).

4 Discussion

The bacterioplankton communities are extremely sensitive and reactive to subtle environmental changes, which indicated that the microbial diversity could be an indicator to evaluate the environmental conditions (Shan and Li, 2007; Or et al., 2012). In this study, we developed a microarray aimed at detecting bacterioplankton in marine and freshwater environment. The massively parallel nature of microarray hybridizations, combined with a large database of 16S rRNA gene sequences, enables rapid identification of bacterial phylotypes (Liles et al., 2010).

Compared to the labor- and resource-intensive efforts to clone and sequence a representative number of clones from a 16S rRNA gene clone library, phylogenetic microarrays can provide an efficient readout of the phylogenetic diversity present in an environmental sample (Liles et al., 2010). Compared to the high throughput sequencing method, the microarray could provide direct identification of the bacterioplankton present in seawater sample (Kyselková et al., 2008).

In many studies using phylogenetic microarrays, the probes would be validated against a bacterial target (Loy et al., 2002, 2005; Kyselková et al., 2008). In our study, five clones isolated from environmental seawater sample were used, but this is not the case for all, and for the latter the targets will need to be identified through hybridization analysis of multiple samples. Furthermore, the very large number of unique oligonucleotide probes

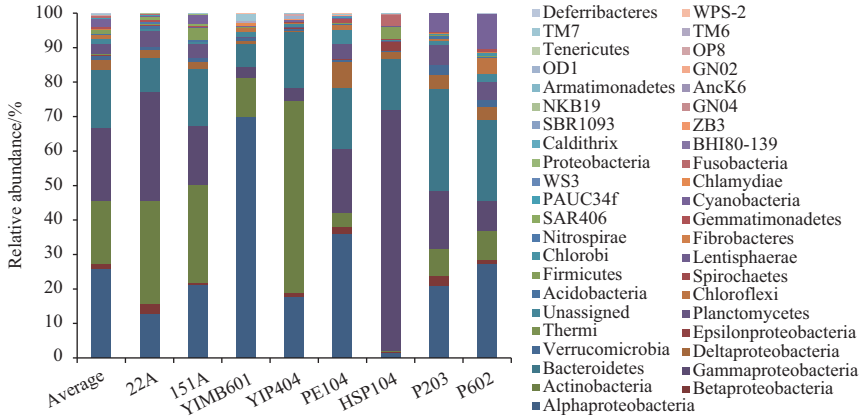

Fig. 6. Bacterial composition based on the sequencing data, and relative abundances of bacterial phyla in seawater samples and the average across the samples. Relative abundances are based on the proportional of those DNA sequences that could be classified at the phylum level, with the exception that the predominant phyla of Proteobacteria were grouped at the class level.

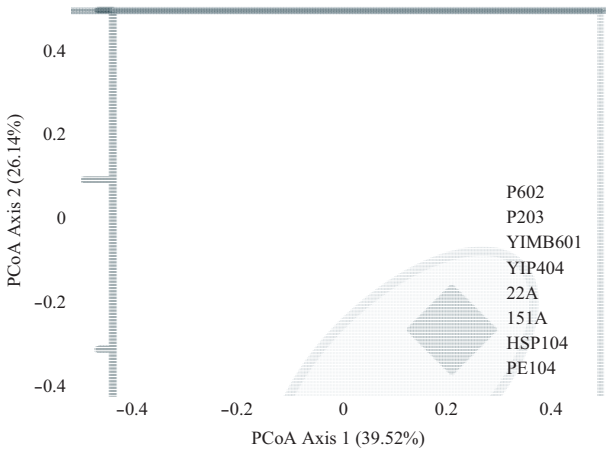

Fig. 7. PCoA analysis of the eight sea water samples based on the sequencing data. Principal coordinate analysis (PCoA) derived from the Bray-Curtis distances between seawater samples according to the DNA sequences that were classified on the family level. The percentage of variation explained by each axis is shown. The samples in ellipse had relatively similar bacterial community structure. The sources of these eight samples had been described in Fig. 5.

identified from an ARB database phylogenetic analysis precluded individual assessment of each probe prior to its inclusion on the microarray. Each probe selected from the ARB output was blast online and probes having higher similarity (>90%) with non-target group were removed from the probe set. The statistics of mismatches between the probes and the tested rRNA gene sequence showed that the target probes mostly had the less than three mismatches, which was consistent with the selection principle that the identity of the probes between different families should not be higher than 90%. Inevitably, nonspecific probes appeared during the single or mixed clone hybridization, since the account of the false probes was very low, it did not significantly affect the overall mean hybridization results.

For the single or mixed clone hybridization, probes designed for *Saprospiraceae*, *Microbacteriaceae* or *Flavobacteriaceae* were not all positively hybridized. Only one or two of three features had positive signals. This was because the design of the probes were according to the "multiple probe concept" (Loy et al., 2003), for which three probes assigned to separate signature sites for one target family were constructed to avoid the detecting deficiency for the multiple 16S ribosomal sequences. In this study, we used 16S rRNA products of clones isolated from the seawater samples, which may be affected by the sequence amplification bias. However, all the clones detecting experiments showed a good specificity of the microarray.

Evaluation of the sensitivity and quantitative ability of the microarray with increasing concentrations of 16S rRNA gene produced almost linear increased signal intensity, but variations in the signal among different probes cannot be generally interpreted as variations in quantity, signal intensity could reflect the abundance variation between different microarray set but not for the same array set. Due to potential kinetic differences between individual probes, the mean probe hybridization signals cannot be taken to be indicative of the quantitative abundance of any phylogenetic group present in an environmental sample (Liles et al., 2010). In the following analysis, relative amount was used for the comparison of bacterial abundance between different samples with the hybridization data of the environmental samples.

In this study we paid more attention on the microarray in detecting environmental samples. To assess the applicability of the developed microarray in environmental studies, eight environmental samples were analyzed with both microarray and pyrosequencing technique. When we analyze the diversity and abundance of the bacteria with the microarray data, the probes with signal intensity under the meaningful line were wiped off, and the PCoA analysis for hybridization results and pyrosequencing data were both based on the family level with the ratios of each family to all the bacteria abundance. The dominant bacterial phyla between the eight samples detected by microarray and pyrosequencing method were almost consistent. The hybridization results showed that samples from the East China Sea, shrimp pond of 2012 and shrimp pond of 2014 could be separated out, which indicated that this microchip could discriminate samples by sample sites and time. PCoA performed with pyrosequencing data showed the same tendency among the eight samples, which further illustrated the reliability of the microarray in application of environmental samples detection. Although in PCoA results with microarray data, YIP404 was slightly off the other three samples which were all sourced from pond in 2014, the overall trend was the same with pyrosequencing results and the difference between the two methods could be permitted according to Field et al. (2010). In their study, microbial composition was analyzed by PhyloChip and clone library and the results showed that there existed difference between the two methods, evenly the method used to identify the community appeared to influence the clustering of the data more heavily than the environment variable (Field et al., 2010). In our study, although some difference exists between this two methods, the overall trend was consistent, which indicated that the microarray may be closer to the result of pyrosequencing method in analyzing microbial community diversity than the clone library. This is in accordance with our design concept to capture a snapshot of the bacterioplankton community variation under various environmental implication. Admittedly, the high-throughput sequencing has its obvious advantages with high sensitivity, since the greater sampling depth provides detection of low-relative-abundance microorganisms in the environment (Roesch et al., 2007; Huse et al., 2008; Andersson et al., 2010). But restricted by the high cost and complex data analysis, it was not suitable for the general application, especially for the smaller research groups (Kircher and Kelso, 2010). Relatively, the microchip hybridization result was intuitive, easily to be analyzed and it had mature microarray data analysis theory system (Tong et al., 2006; Grant et al., 2007; Shiu and Borevitz, 2008), which could provide powerful support for data analysis.

Our collective understanding of marine bacterioplankton will be enhanced both through recent culture or unculture-based genomic studies (Hofer, 2013), and through a better understanding of their diversity following temporal and spatial variations. From the above, it appears that this developed microarray could be applied to detect the bacterioplankton in marine environment, moreover, the marine archaea will be considered to include in this chip for a more comprehensive detection for the estimates of marine microbial diversity. Maybe it is not the most exact, but coarse scanning could also reflect the overall trend, which may provide a beforehand evaluation of the marine microbial ecology and help to locate the main research area.

5 Conclusions

We developed a 16S rRNA gene-based microarray for the ap-

plication of bacterioplankton community detection. Total 447 probes were designed with ARB software, which targeted 160 families and 23 phyla. Hybridization results with the five single, cloned 16S rRNA genes showed that the microarray was specific, with only 1.16% false results. Hybridization with increasing concentrations of 16S rRNA gene qualified the sensitivity and quantitative ability of the microchip, which showed that the microarray could detect DNA samples as few as 1 ng/μL and the signal intensity could reflect the relative abundance of the bacteria in the range of 1 ng/μL to 100 ng/μL of DNA concentration. The results of environmental samples hybridization showed that the microchip can discriminate bacterioplankton communities by sites and time. The pyrosequencing data of the environmental samples showed consistent microbial community structure with the hybridization results, which confirmed the application of this developed microarray. This study showed that the 16S rRNA-based taxonomic microarray is a promising tool to monitor bacterioplankton community changes in marine water environments.

References

Andersson A F, Riemann L, Bertilsson S. 2010. Pyrosequencing reveals contrasting seasonal dynamics of taxa within baltic sea bacterioplankton communities. ISME J, 4(2): 171–181

Bodrossy L, Stralis-Pavese N, Murrell J C, et al. 2003. Development and validation of a diagnostic microbial microarray for methanotrophs. Environ Microbiol, 5(7): 566–582

Brodie E L, DeSantis T Z, Joyner D C, et al. 2006. Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation. Appl Environ Microbiol, 72(9): 6288–6298

Caporaso J G, Kuczynski J, Stombaugh J, et al. 2010. Qiime allows analysis of high-throughput community sequencing data. Nat Methods, 7(5): 335–336

Chen Mingxia, Li Heyang, Li Gang, et al. 2010. The study on the diversity and distribution of vibrios and the correlation between them and their ambient environmental factors in shenzhen coastal waters. Haiyang Xuebao (in Chinese), 32(5): 117–126

Comte J, del Giorgio P A. 2011. Composition influences the pathway but not the outcome of the metabolic response of bacterioplankton to resource shifts. PLoS One, 6(9): e25266

Field E K, D'Imperio S, Miller A R, et al. 2010. Application of molecular techniques to elucidate the influence of cellulosic waste on the bacterial community structure at a simulated low-level-radioactive-waste site. Appl Environ Microbiol, 76(10): 3106–3115

Fuhrman J A. 2009. Microbial community structure and its functional implications. Nature, 459(7244): 193–199

Grant G R, Manduchi E, Stoeckert C J Jr. 2007. Analysis and management of microarray gene expression data. In: Ausubel F M, ed. Current Protocols in Molecular Biology. Hoboken, NJ: John Wiley & Sons, Inc

Hammer Ø, Harper D A T, Ryan P D. 2001. Past: paleontological statistics software package for education and data analysis. Palaeontol Electron, 4(1): 1–9

Hazen T C, Dubinsky E A, DeSantis T Z, et al. 2010. Deep-sea oil plume enriches indigenous oil-degrading bacteria. Science, 330(6001): 204–208

Hofer U. 2013. Environmental microbiology: exploring diversity with single-cell genomics. Nat Rev Microbiol, 11(9): 598–599

Huse S M, Dethlefsen L, Huber J A, et al. 2008. Exploring microbial diversity and taxonomy using SSU rRNA hypervariable tag sequencing. PLoS Genet, 4(11): e1000255

Kim D, Kim J F, Yim J H, et al. 2008. Red to red-the marine bacterium hahella chejuensis and its product prodigiosin for mitigation of harmful algal blooms. J Microbiol Biotechnol, 18(10): 1621–1629

Kim M, Morrison M, Yu Zhongtang. 2011. Evaluation of different partial 16S rRNA gene sequence regions for phylogenetic analysis of microbiomes. J Microbiol Methods, 84(1): 81–87

Kircher M, Kelso J. 2010. High-throughput DNA sequencing-concepts and limitations. BioEssays, 32(6): 524–536

Kyselková M, Kopecký J, Felföldi T, et al. 2008. Development of a 16S rRNA gene-based prototype microarray for the detection of selected actinomycetes genera. Antonie van Leeuwenhoek, 94(3): 439–453

Liles M R, Turkmen O, Manske B F, et al. 2010. A phylogenetic microarray targeting 16S rRNA genes from the bacterial division *Acidobacteria* reveals a lineage-specific distribution in a soil clay fraction. Soil Biol Biochem, 42(5): 739–747

Liu Jiqing, Lewitus A J, Kempton J W, et al. 2008. The association of algicidal bacteria and raphidophyte blooms in south carolina brackish detention ponds. Harmful Algae, 7(2): 184–193

Liu Shanshan, Wang Feng, Xue Kai, et al. 2015. The interactive effects of soil transplant into colder regions and cropping on soil microbiology and biogeochemistry. Environ Microbiol, 17(3): 566–576

Loy A, Horn M, Wagner M. 2003. Probebase: an online resource for rrna-targeted oligonucleotide probes. Nucleic Acids Res, 31(1): 514–516

Loy A, Lehner A, Lee N, et al. 2002. Oligonucleotide microarray for 16s rrna gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. Appl Environ Microbiol, 68(10): 5064–5081

Loy A, Schulz C, Lücker S, et al. 2005. 16S rRNA gene-based oligonucleotide microarray for environmental monitoring of the betaproteobacterial order "rhodocyclales". Appl Environ Microbiol, 71(3): 1373–1386

Ludwig W, Strunk O, Westram R, et al. 2004. Arb: a software environment for sequence data. Nucleic Acids Res, 32(4): 1363–1371

Mahmud Z H, Neogi S B, Kassu A, et al. 2008. Occurrence, seasonality and genetic diversity of *Vibrio vulnificus* in coastal seaweeds and water along the Kii Channel, Japan. FEMS Microbiol Ecol, 64(2): 209–218

Markham N R, Zuker M. 2005. DINAMelt web server for nucleic acid melting prediction. Nucleic Acids Res, 33(S2): W577–W581

Mendes R, Kruijt M, de Bruijn I, et al. 2011. Deciphering the rhizosphere microbiome for disease-suppressive bacteria. Science, 332(6033): 1097–1100

Nakashima T, Miyazaki Y, Matsuyama Y, et al. 2006. Producing mechanism of an algicidal compound against red tide phytoplankton in a marine bacterium γ-*proteobacterium*. Appl Microbiol Biotechnol, 73(3): 684–690

Or A, Shtrasler L, Gophna U. 2012. Fine-scale temporal dynamics of a fragmented lotic microbial ecosystem. Sci Rep, 2: 207

Ramaiah N, Raghukumar C, Sheelu G, et al. 1996. Bacterial abundance, communities and heterotrophic activities in the coastal waters off Tamil Nadu. Indian J Mar Sci, 25: 234–239

Roesch L F W, Fulthorpe R R, Riva A, et al. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. ISME J, 1(4): 283–290

Rusch A, Huettel M, Reimers C E, et al. 2003. Activity and distribution of bacterial populations in middle atlantic bight shelf sands. FEMS Microbiol Ecol, 44(1): 89–100

Sanguin H, Herrera A, Oger-Desfeux C, et al. 2006a. Development and validation of a prototype 16s rrna-based taxonomic microarray for *Alphaproteobacteria*. Environ Microbiol, 8(2): 289–307

Sanguin H, Kroneisen L, Gazengel K, et al. 2008. Development of a 16S rRNA microarray approach for the monitoring of rhizosphere *Pseudomonas* populations associated with the decline of take-all disease of wheat. Soil Biol Biochem, 40(5): 1028–1039

Sanguin H, Remenant B, Dechesne A, et al. 2006b. Potential of a 16S rRNA-based taxonomic microarray for analyzing the rhizosphere effects of maize on *Agrobacterium* spp. And bacterial communities. Appl Environ Microbiol, 72(6): 4302–4312

Schenk P M, Kazan K, Wilson I, et al. 2000. Coordinated plant defense responses in *Arabidopsis* revealed by microarray analysis. Proc Natl Acad Sci U S A, 97(21): 11655–11660

Shan Yu, Li Zhengyan. 2007. An index system for ecological monitor-

ing of the marine environment. Trans Oceanol Limnol (in Chinese), (2): 52–56

Shiu S H, Borevitz J O. 2008. The next generation of microarray research: applications in evolutionary and ecological genomics. Heredity, 100(2): 141–149

Tanner M A, Shoskes D, Shahed A, et al. 1999. Prevalence of corynebacterial 16S rRNA sequences in patients with bacterial and "nonbacterial" prostatitis. J Clin Microbiol, 37(6): 1863–1870

Tong Weida, Lucas A B, Shippy R, et al. 2006. Evaluation of external RNA controls for the assessment of microarray performance. Nat Biotechnol, 24(9): 1132–1139

Wang Xin, Li Zhijiang, Zheng Tianling. 2010. Distribution and activity of marine bacterioplankton at frequent hab area of east china sea. Environ Sci (in Chinese), 31(2): 287–295

Wu Liyou, Thompson D K, Li Guangshan, et al. 2001. Development and evaluation of functional gene arrays for detection of selected genes in the environment. Appl Environ Microbiol, 67(12): 5780–5790

Xiao Hui, Tang Xuexi, Qiao Xudong, et al. 2010. The abundance of benthic bacteria in tianjin nearshore waters in the bohai bay and its correlation with environmental factors. Period Ocean Univ China (in Chinese), 40(6): 87–90, 160

Xiong Jinbo, Zhu Jianlin, Wang Kai, et al. 2014. The temporal scaling of bacterioplankton composition: High turnover and predictability during shrimp cultivation. Microb Ecol, 67(2): 256–264

Zhang Demin, Wang Xin, Xiong Jinbo, et al. 2014. Bacterioplankton assemblages as biological indicators of shrimp health status. Ecol Indic, 38: 218–224

Zhao Mengxin, Xue Kai, Wang Feng, et al. 2014. Microbial mediation of biogeochemical cycles revealed by simulation of global changes with soil transplant and cropping. ISME J, 8(10): 2045–2055

Zhou Jizhong, Thompson D. 2003. Microarrays: applications in environmental microbiolgy. In: Bitton G, ed. Encyclopedia of Environmental Microbiology. New York: John Wiley & Sons, Inc

Supplementary information:

Table S1. List of probes and their targets.

Fig. S1. In *silico* prediction of secondary structures between selected probe 20140620-173 (a) / probe 20140825-10 (b) / probe 20130902-243 (c) / probe 20130902-265 (d) / probe 20140620-144 (e) 20140825-9 (f) / probe 20140825-5 (g) and *Rhodobacteraceae* duplexes. A and B represent the mole fraction of the two molecules (probe and species). Free energy (dG) and enthalpy (dH) are in kcal/mol; entropy is in e.u. (cal/mol/K). The free energy, enthalpy and entropy changes associated with the transition from 'hybridized at temperature T′ to random coil are denoted by ΔG, ΔH and ΔS, respectively. They are related by the equation $\Delta G = \Delta H - T\Delta S$ and $\Delta S = 1000\times(\Delta H - \Delta G)/T$. Both ΔG and ΔH are computed using published nearest neighbor coefficients. And dG was equal to ΔG, dH was equal to ΔH. The computation details were referenced from Markham NR's publishment which was mentioned in the main body.

Fig. S2. In *silico* prediction of secondary structures between selected probe 20140620-168 (a) / probe 20140620-173 (b) / probe 20140620-144 (c) and *Comamonadaceae* duplexes.

Fig. S3. In *silico* prediction of secondary structures between selected probe 20130902-392 / probe 20140620-144 (b) and *Microbacteriaceae* duplexes.

Fig. S4. In *silico* prediction of secondary structures between selected probe 20140825-5 (a) / probe 20140620-144 (b) / probe 20130902-36 (c) / probe 20140620-186 (d) / probe 20130902-28 (e) and *Microbacteriaceae* duplexes.

The supplementary information is available online at www.hyxb.org.cn/aosen/ch/index.aspx. The supplementary information is published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.