Article

Comprehensive Analysis of Prokaryotes in Environmental Water Using DNA Microarray Analysis and Whole Genome Amplification

Takeshi Akama [1], Akira Kawashima [1], Kazunari Tanigawa [1], Moyuru Hayashi [1], Yuko Ishido [1], Yuqian Luo [1], Akihisa Hata [2], Noboru Fujitani [2], Norihisa Ishii [3] and Koichi Suzuki [1,*]

[1] Laboratory of Molecular Diagnostics, Department of Mycobacteriology, Leprosy Research Center, National Institute of Infectious Diseases, Tokyo 189-0002, Japan; E-Mails: akama@umich.edu (T.A.); akirak@jichi.ac.jp (A.K.); tanigawa@pharm.teikyo-u.ac.jp (K.T.); mhayashi@rs.kagu.tus.ac.jp (M.H.); yishido@nih.go.jp (Y.I.); yuqianluo31@gmail.com (Y.L)

[2] Department of Medical Risk Management, Faculty of Risk Management, Chiba Institute of Science, Faculty of Risk and Crisis Management, Choshi, Chiba 288-0025, Japan; E-Mails: ahata@cis.ac.jp (A.H.); nfujitani@cis.ac.jp (N.F.)

[3] Leprosy Research Center, National Institute of Infectious Diseases, Tokyo 189-0002, Japan; E-Mail: norishii@nih.go.jp

* Author to whom correspondence should be addressed; E-Mail: koichis@nih.go.jp; Tel.: +81-42-391-8211; Fax: +81-42-394-9092.

Abstract: The microflora in environmental water consists of a high density and diversity of bacterial species that form the foundation of the water ecosystem. Because the majority of these species cannot be cultured *in vitro*, a different approach is needed to identify prokaryotes in environmental water. A novel DNA microarray was developed as a simplified detection protocol. Multiple DNA probes were designed against each of the 97,927 sequences in the DNA Data Bank of Japan and mounted on a glass chip in duplicate. Evaluation of the microarray was performed using the DNA extracted from one liter of environmental water samples collected from seven sites in Japan. The extracted DNA was uniformly amplified using whole genome amplification (WGA), labeled with Cy3-conjugated 16S rRNA specific primers and hybridized to the microarray. The microarray successfully identified soil bacteria and environment-specific bacteria clusters.

The DNA microarray described herein can be a useful tool in evaluating the diversity of prokaryotes and assessing environmental changes such as global warming.

Keywords: metagenome; microarray; 16S rRNA; WGA; pathogenic bacteria

1. Introduction

Diverse prokaryotic populations, including bacteria and archaea, co-exist in the microflora of various environments such as soil, air and water. Microbes in environmental water constitute 50%–90% of ocean biomass, with 3×10^4 species per liter at a density of 10^4 to 10^7 cells per milliliter [1,2]. Therefore, the microflora in environmental water forms the basis of the food chain, and changes in microflora biomass or composition influence the entire water ecosystem.

Composition of the environmental microflora is dependent on various conditions such as temperature, salinity, organic or inorganic nutrients, and human activities [1,3]. *In vitro* culture and isolation of each organism were employed in an attempt to determine the microbial composition of environmental samples. However, it became evident that most environmental bacteria cannot be cultured [4]. As an alternative, genomic DNA was extracted from samples of marine water, fresh water, and soil for the analysis of DNA sequences. The analysis of marine water in the Sargasso Sea found 148 novel phylotypes and 1.2 million genes among 1.6 Gbp of DNA sequence [5]. These metagenomic data were integrated with those of a global ocean sampling expedition that collected samples at 44 sites, from the Atlantic Ocean near the Nova Scotia peninsula in Canada to the Pacific Ocean near French Polynesia via the Gulf of Panama [6]. The 6.3 Gbp of sequence demonstrated taxonomic, subspecies, genetic, and geographical diversity. Comparative metagenomic analysis of samples collected from marine, hyper saline, fresh, and coral reef water uncovered differences between the metabolic profiles in each environment [7].

Taxonomic analysis of metagenomes using DNA microarrays is simpler and more reproducible than sequencing, which is still an expensive and complicated procedure [8]. However, previous 16S rRNA arrays covered only 800 to 9,000 sequences, which were not enough to analyze microflora in environmental metagenomic DNA samples [9].

Using the DNA Data Bank of Japan (DDBJ), which has collected >140,000 16S rRNA sequences, we designed probes for a DNA microarray to detect prokaryotic species. We then applied this array to the metagenomic analysis of microflora from water samples around Japan. Individual pathogenic bacteria were detected, as well as specific prokaryotic clusters that might be associated with the temperature of the water that was sampled.

2. Results and Discussion

2.1. Sampling of Environmental Water

According to information supplied by the Japan Meteorological Agency [10] the Black Current flows from the south of Japan along the Pacific side of the archipelago and partially into the Sea of Japan (Figure 1a). The frigid Oyashio Current flows from north of Japan (Hokkaido) along the Pacific

coast and runs into the Black Current. These two currents create the basis of warm and cold sea surface temperatures around Japan (Figure 1b). Environmental water was collected from seven sites in Japan (Figure 1c). The temperature, pH and amounts of DNA recovered from 1 L of water samples before whole genome amplification (WGA) are shown in Table 1.

Figure 1. Ocean flow, marine temperature, and sampling sites in Japan. (**a**) Mean sea surface current (direction and knots) for 1–10 August 2009; (**b**) Monthly mean sea surface temperature (°C) in August 2009 obtained from the Japan Meteorological Agency [11]; (**c**) Sampling sites and associated sample numbers (see Table 1).
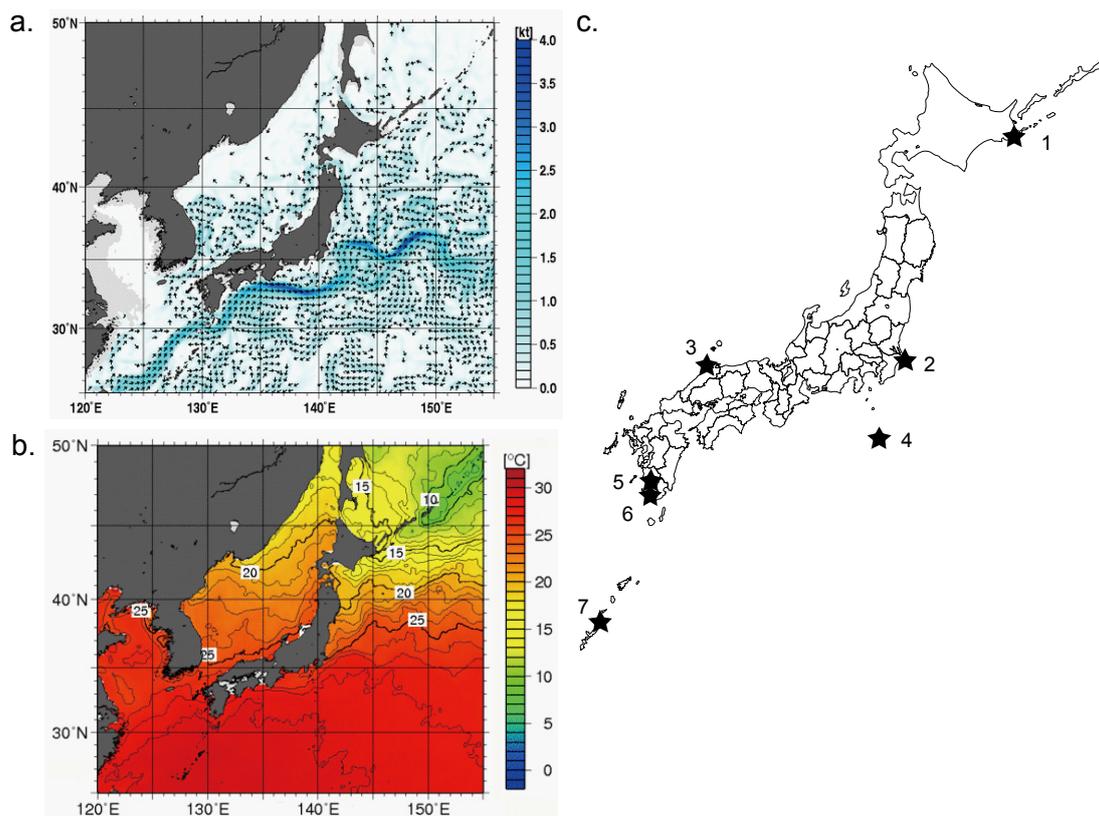
**Table 1.** Sampling sites, environmental data and recovered total DNA.

No.	Sampling site	Salinity	Sampling	Temperature (°C)	pH	DNA recovery (μg)
1	Nemuro	Marine	July 28	13.5	7.9	1.10
2	Tonegawa river	Estuarine water	May 23	21.0	8.0	5.60
3	Shimanekaka	Marine	August 18	27.0	8.2	6.20
4	Hachijo	Marine	August 3	28.0	8.3	0.12
5	Kagoshima	Marine	September 6	29.9	8.2	0.24
6	Ikeda lake	Fresh water	August 3	28.8	8.3	0.96
7	Yoron	Marine	September 10	29.5	8.0	0.66

2.2. Taxonomic Composition in Each Water Sample

DNA extracted from water samples collected from seven sites around Japan was amplified using WGA (typically 10 to 100-fold amplification was achieved) and subjected to DNA microarray analysis. Since signals from most probes were almost negative, probes that showed top 1,000 signals were subjected for taxonomical analysis (see Section 3.5 Data Analysis of DNA Microarray).

Phyla distribution was somewhat similar among samples and that of original probe composition (Figure 2 and Table S1), suggesting that this array detected each phylum without having significant difference in affinity and sensitivity of each probe spotted on the array. Sample 1 obtained from northern part of Japan with much lower water temperature (Table 1) showed no significant difference in the taxonomical distribution (Figure 2, Table S1 and Figure S1), suggesting that the local fluctuation of surface temperature may not be the most decisive factor to affect microbiota, and that the most bacterial flora around Japan islands are essentially similar.

Figure 2. Microarray data classified at the phylum level. The percentage of each phylum comprising the microarray probes and the top 1,000 target sequences from each samples 1–7. The actual percentage in each phylum including those for classified as "others" were shown in Table S1. $p = 10^{-8}$ as evaluated by the proposed algorithm [12].
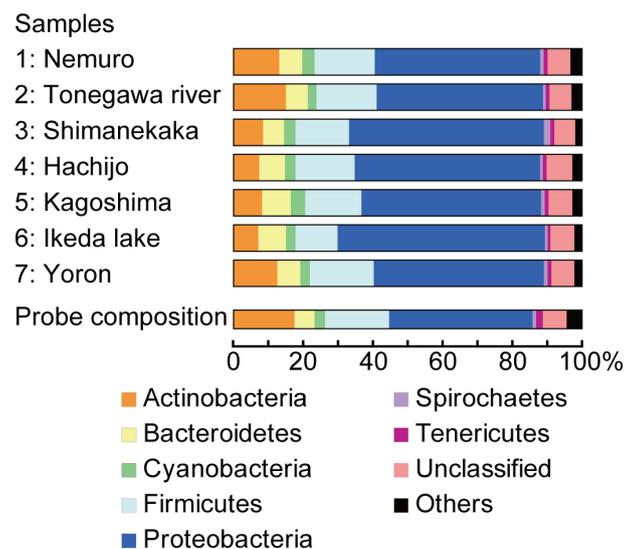
Nevertheless, small differences in distribution as shown in the levels of phylum (Figure 2), class, order, family, genus and species (Figure S1) also suggest that some of the prokaryotes are unevenly distributed around Japan. Indeed, sample 6, the only lake sample, showed a unique pattern compared to others (Figure 2 and Figure S1). These results also suggest that although the composition of most bacteria in the environment is similar, this microarray could detect small differences in the prokaryotic species in environmental water. However, it is still difficult to demonstrate clear differences in the proportion of water prokaryotes at the phylum level or at the any level of taxonomy; therefore, we proceeded to perform further identification of some specific bacteria and deeper taxonomy analysis by cluster analysis.

2.3. Detection of Pathogenic Bacteria

We examined whether DNA from pathogenic bacteria can be detected in the top 1,000 signals generated in each sample (see Section 3.5 Data Analysis of DNA Microarray). Five species were detected in the seven samples (Table 2). *Aeromonas sobria* [13], *Citrobacter freundii* [14], and *Clostridium perfringens* [15] were detected in multiple samples, while *Mycobacterium gordonae* [16] was only found in sample 7 and *Mycobacterium marinum* [17] in sample 2. These data indicate a possibility that this microarray can be used to screen pathogenic bacteria in the environment, although further confirmation is needed.

**Table 2.** Pathogenic bacteria detected by the DNA microarray.

Species	Sample number						
	1	**2**	**3**	**4**	**5**	**6**	**7**
Aeromonas sobria	−	+	+	+	+	+	+
Citrobacter freundii	−	−	+	+	+	+	+
Clostridium perfringens	+	+	+	+	+	+	+
Mycobacterium gordonae	−	−	−	−	−	−	+
Mycobacterium marinum	−	+	−	−	−	−	−

*2.4. Detection of Site-Specific Clusters and Species*

To identify habitat-specific patterns of microflora at the species level, hierarchical clustering analysis was performed for the top 200 sequences found in each sample as described in the Experimental Section. As a result, seven clusters were identified that were specific to each sample (Figure 3a). In cluster 2 (Tonegawa River), bacteria were detected that are not indigenous to the environment but are isolated from domestic animals or humans, such as *Mycoplasma conjunctivae* HRC/583, *Actinomyces hyovaginalis*, and *Campylobacter hominis*, which is consistent with the fact that there are many cattle farms upstream of this river (Figure 3b). In contrast, sequences in cluster 1 (Nemuro, northern part of Japan) included some identified at high latitude, such as *Pseudoalteromonas* sp. BSi20301 and BSi20493 and *Marinomonas* sp. BSi20412 detected in Arctic Sea samples [18] (Figure 3c). Nested PCR analysis confirmed that *Marinomonas* sp. BSi20412 was detected only in sample 1 (Figure 4).

In contrast, sequences coding for thermophilic bacteria, such as Alpha proteobacterium HTA473 isolated from the Mariana Trench [19], or *Thermus* sp. HR13 isolated from a hot spring [20] were found only in cluster 7 (Yoron, the southern part of Japan) (Figure 3d). A detailed listing of species that generated strong signals in clusters 3–6 are shown in Figure 5, and the most frequently detected species in each sample are shown in Table S2. These data indicate the possibility that some of the prokaryote species can be geo-located using DNA microarrays and cluster analysis, and that the process might depend on the marine current, temperature, and human activities.

Figure 3. Cluster analysis of sequences detected on the microarray. Sequences with the top 200 signal intensities in each sample were selected. The rank of the signal intensity in each microarray was changed to the reciprocal number, analyzed with Cluster 3.0, and visualized using Treeview 1.0. (**a**) The clusters that exhibited specific, strong signals only in one of the sampling sites 1 to 7 are marked as clusters 1 to 7. The detail of clusters 2 (**b**), 1 (**c**) and 7 (**d**) are shown with species names. Species indicated by arrows are the characteristic types in each cluster.
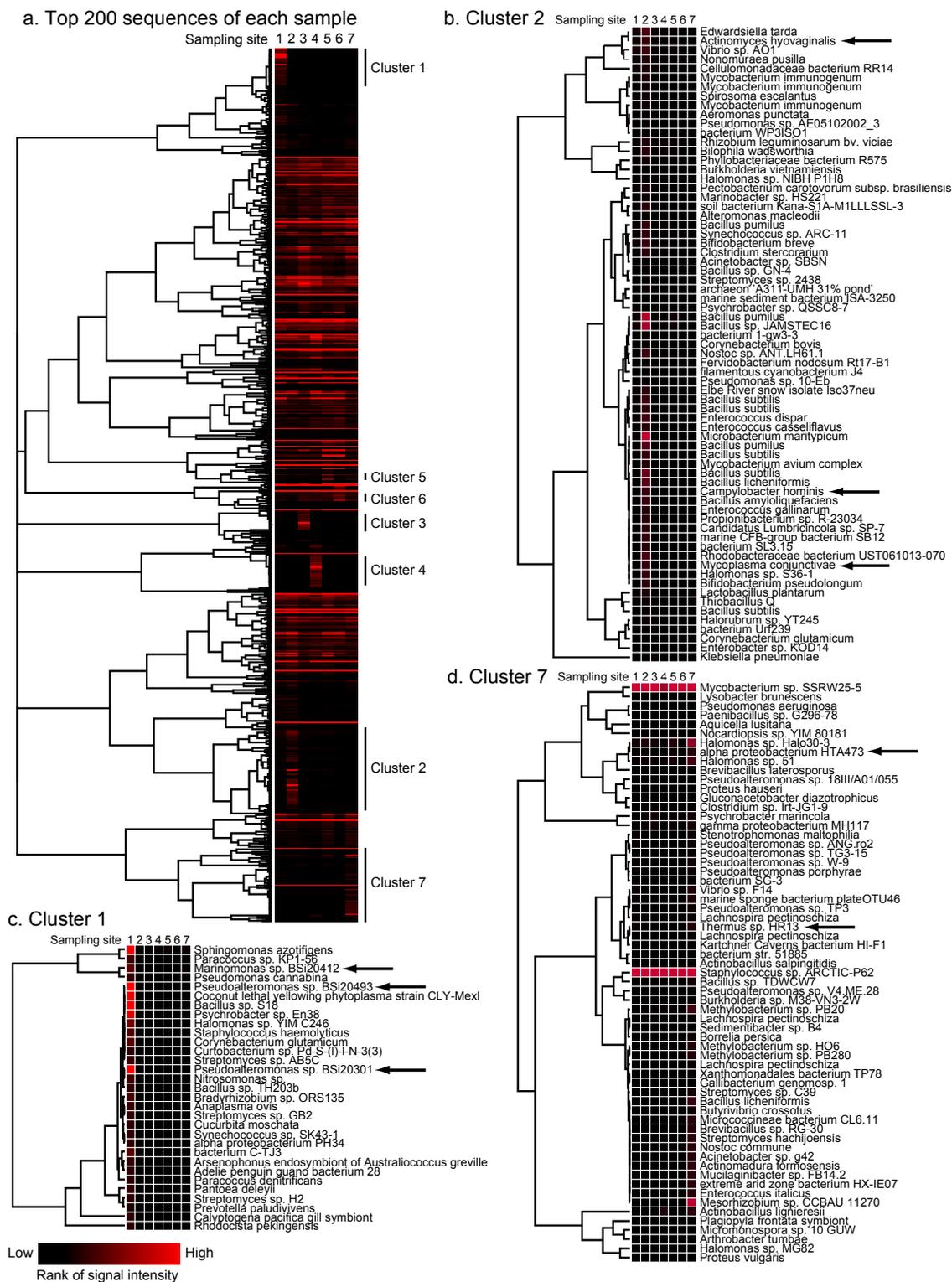
Figure 4. Nested PCR to detect a specific product of *Marinomonas* sp. BSi20412. 16S rRNA sequences of *Marinomonas* sp. BSi20412, which was identified only in sample number 1, was aligned with those of *Escherichia coli*, and *Bacillus subtilis* to detect species-specific sequences. The primer set for the first nested PCR reaction was designed based on unique *Marinomonas* sp. BSi20412 sequences. The second primer set was designed within the first primer set. The specificity of these primer sets was examined using the same template DNA that was applied to the microarrays. All of the samples were examined with *E. coli* genomic DNA as a positive control and distilled water as a negative control (N.C.). Pan-prokaryotic primer set was used to demonstrate the existence of prokaryotes DNA in the samples.
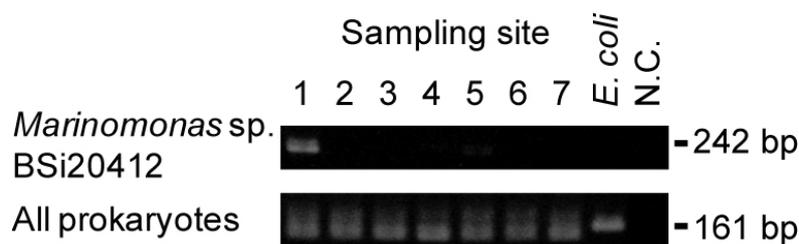
Figure 5. Clusters that had strong signals only in particular samples. The detail of cluster 3 (**a**), 4 (**b**), 5 (**c**), and 6 (**d**) identified in Figure 3a was shown with the name of species.
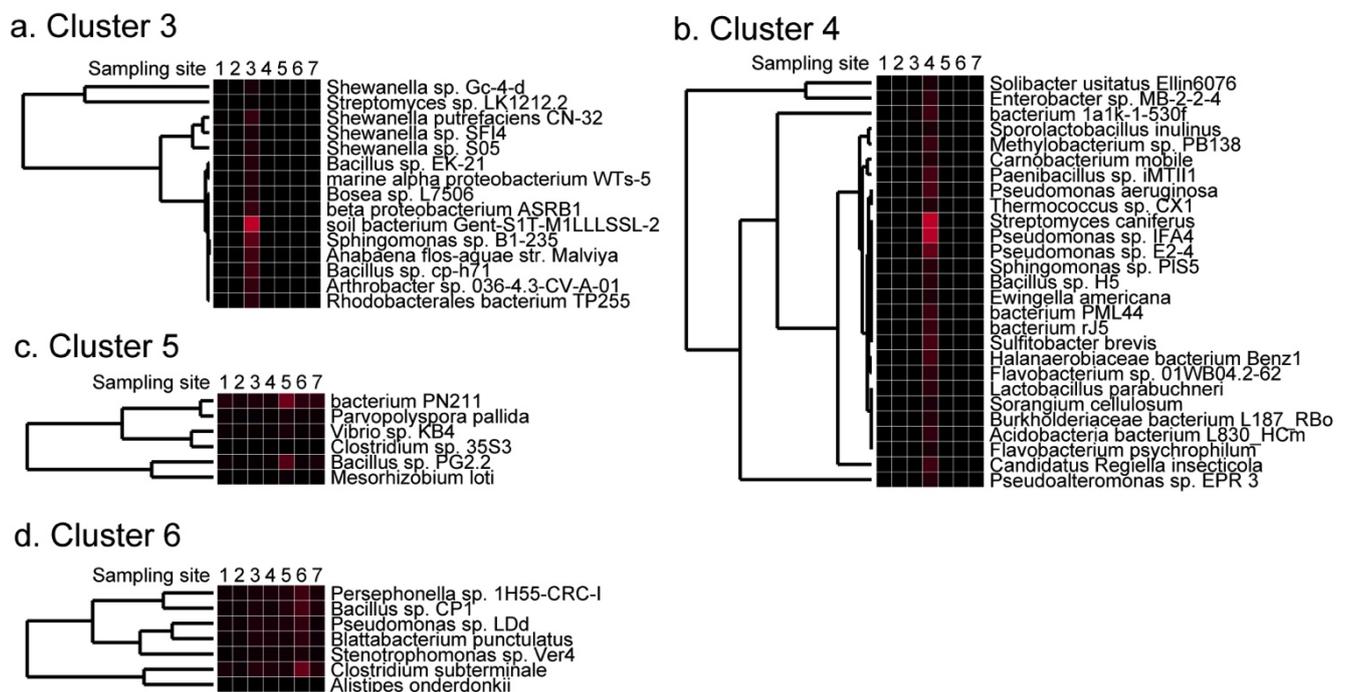
2.5. Discussion of Results

We designed a novel DNA microarray that can theoretically be specific to approximately 64,000 prokaryotic species in the database, and evaluated it by a comprehensive analysis of microflora in the environmental water around Japan. Previous metagenomic analyses of water in Japan only studied specific conditions such as subsurface or geothermal water in gold mines [21,22] or deep sea hydrothermal sites [23], or only searched for specific taxonomic groups [24]. We collected environmental water from

north to south of Japan, and used cluster analysis to reveal several unique groups of species among the samples. A cluster specific to sample 2 contained three pathogenic bacteria. It is possible that these bacteria are derived from sewage or cattle farms, because the collection site was downstream of the Tonegawa River, which has a large amount of livestock production in its basin [25].

The microarray also detected potentially "pathogenic" bacteria; however, this method cannot immediately evaluate if they were actually pathogenic or even if they were viable. For example, *Campylobacter jejuni*, *Escherichia coli*, and *Klebsiella pneumoniae* can exist in a viable but non-culturable (VNC) state in the environment [26]. The pathogenic bacteria identified in this study (Table 2) are known indigenous bacteria that cause waterborne diseases [27,28]. Therefore, data from this DNA microarray might be able to provide information that can be related back to the number of such pathogens and risk of infection. It will also be useful for cases of *Clostridium perfringens* [29,30], in which it will be suspected that most contamination originates from human activities like sewage or livestock production facilities [31]. Any relationship between these pathogenic bacteria and diseases or food poisoning around their sampling sites would have to be independently evaluated to verify actual pathogenicity.

Among the metagenomic analyses of water microflora, previous studies have reported the correlation of its composition with temperature [26,32]. The cluster that had unique signals in sample 7 (Yoron) contained bacteria that are found in warmer environments. Together with *Vibrio cholerae*, the bacteria in that specific cluster may be useful in evaluating changes in microflora caused by the influence of warmer marine temperatures. This hypothesis needs to be confirmed by studying more materials obtained from multiple sampling points having similar environmental temperature. Specific signals were also detected in sample 1 (Nemuro), which was collected from a lower temperature environment. Three of the species in this sample (*Pseudoalteromonas* sp. BSi20301 and BSi20493, and *Marinomonas* sp. BSi20412) were detected in a study of Arctic Sea samples from Canada [18]. Therefore, the decreases of signal intensities against these probes may indicate the progress of marine warming. Moreover, nested PCR revealed the presence of *Marinomonas* sp. BSi20412, which is an easier assay than microarray analysis. Prokaryotes that belong to cluster 1 may be analyzed simultaneously for a more precise index of global warming. Thus, it is possible that this microarray can distinguish prokaryotic species that are applicable to the analysis of environmental microflora, with results comparable to those obtained by pyrosequencing [33]. Determination of the bacterial index that reflects environmental changes will provide a tool to assess global warming. In our ongoing studies we are analyzing more samples from the same and different sampling points for several years. Using these approaches it may be possible to find specific index species that represent particular environments, and/or to illustrate the characteristics of each environment in an easily recognizable ways, such as by self-organization maps.

3. Experimental

3.1. DNA Microarray Design

Microarray probes were designed from 142,860 sequences of 16S rRNA released on 24 January 2009 by the DDBJ database [34]. In order to generate the most specific probes to each sequence, 24

nucleotides (nts) of each sequence were first searched at 1-bp intervals as an initial pool of candidate probes, and then applied filters based on homology against all other sequences in the 16S rRNA database [35]. In the filtering process, scores of each 24-mer oligo were calculated for the average 15-mer frequency in the genome, the oligo frequency in the database, and uniqueness. The 15-mer frequency score was used to avoid probes in repetitive regions while the oligo frequency score was used to identify the number of exact matches within the database. The uniqueness information was reduced to a Boolean value based on the weighted mismatch score. To be considered unique, a 24-mer oligo must have a weighted mismatch score greater than 10. The *Tm*, base pair composition, and self-complemetarity were also part of the probe selection parameters.

Finally, one to three different 29-mer probes were designed for each DNA sequence in the database. As a result, most probes were located within 5' variable region. Eventually, a total of 258,697 specific probes against 97,927 sequences were designed. It was not possible to design specific probes for the rest of 44,933 16S rRNA sequences in the database, since all their candidate probes had failed the filtering criteria. The 97,927 sequences included 72,233 sequences in which highly stringent probes were designed, and other 25,694 sequences with rather loose probes.

The sequences to which only loose probes were made were also included in the array, since the definition of the stringency was calculated based on the sequence in the database that covers only a few percent of the bacterial species in the nature. Therefore, DNA microarray may not determine specific species because there are many unidentified species that may have homologous sequences. However, it will be useful to represent gross characteristics and changes of particular environment. Designed sequence-specific probes with NimbleGen sample tracking control probes (Roche) and random probes were mounted to the array chip in duplicate.

Duplicate probes were distributed in a random fashion on the array chip and utilized for the evaluation for accuracy (see Section 3.5 Data Analysis of DNA Microarray). This prokaryotic microarray was theoretically expected to identify about 64,000 bacterial and archaeal species in the database. Data from probes designed in the same species were treated as independent data.

3.2. DNA Extraction from Environmental Water

One liter of environmental water was passed through a 0.22-µm filter and DNA was extracted from the filter using the Ultraclean Water DNA Kit (Mo Bio Laboratories Inc., Carlsbad, CA, USA) according to the manufacturer's guidelines. Only clear water with no visible turbidity was sampled. Briefly, the membrane was cut into small pieces and vortexed with 5 g of microbeads and 4 mL of bead solution for 30 s in a 15-mL sterile test tube. The addition of 0.5 mL of solution WD1 was followed by another 30-s vortex. The tube was then set horizontally and further vortexed for 10 min. After centrifugation at 2,500 ×*g* for 1 min, the supernatant was transferred to a new tube, mixed with 0.6 mL of solution WD2, and incubated at 4 °C for 5 min. After centrifugation at 2,500 ×*g* for 4 min, the supernatant was transferred to a new tube and mixed with 8 mL of solution WD3. This mixture was loaded into a spin filter tube and centrifuged at 2,500 ×*g* for 2 min. After discarding the flow through, 3 mL of solution WD4 was added, and the spin filter was centrifuged at 2,500 ×*g* for 3 min. The flow through was discarded before the filter was centrifuged at 2,500 ×*g* for 5 min. After transferring the

spin filter to a new tube, DNA was recovered by adding 3 mL of solution WD5 and centrifuging at 2,500 ×*g* for 2 min, followed by ethanol precipitation.

3.3. Whole Genome Amplification (WGA)

Recovered metagenomic DNA was uniformly amplified using WGA in order to perform fluorescent labeling and for further PCR analysis. Although non-uniform amplification is noted in some WGA applications, we used Genomeplex Whole Genome Amplification Kit (Sigma, St. Lois, MO, USA) based on the reported comparative study [36]. We confirmed in our previous study that the results of this WGA gave superior detection of trace amounts of DNA in subsequent PCR detection [37]. Briefly, 10-µL of DNA was treated with 1 µL fragmentation buffer at 95 °C for 4 min and chilled on ice. Library preparation buffer and library stabilization buffers (1 µL of each) were added to the samples. The solution was incubated at 95 °C for 2 min and chilled on ice. After the addition of 1 µL of Library Preparation Enzyme, the DNA was sequentially incubated at 16 °C for 20 min, 24 °C for 20 min, 37 °C for 20 min, and 75 °C for 5 min. The solution was amplified with 7.5 µL of Amplification Master Mix, 47.5 µL of distilled water, and 5 µL (12.5 units) of Jumpstart Taq DNA polymerase under the following conditions: initial denaturation at 95 °C for 3 min, 14 cycles of 94 °C for 15 sec and 65 °C for 5 min.

3.4. Fluorescent Labeling, Hybridization and Analysis of DNA Microarray

The 16S rRNA sequences in the sample were fluorescently labeled using 10-mer primers designed against conserved prokaryotic sequences [38–40]. Four Cy3-labeled reverse primers and three unlabeled forward primers were used simultaneously to label sample DNA for hybridization (Table S3). The primers were located in the 5' variable region of the 16S rRNA gene, where most of the array probe sequences were located (Figure S2). The labeling and hybridization protocols were modifications of the method used by Akama *et al*. [41]. Cy3 labeling reactions were performed with the NimbleGen One-color DNA Labeling Kit (Roche NimbleGen, Inc. Madison, WI, USA): 1 µg DNA was incubated for 10 min at 98 °C with 1 OD unit of 16S rRNA-specific primers. A random nonamer was used as the control primer. The addition of 8 mmol of dNTPs and 100 U of Klenow fragment (New England Biolabs, Ipswich, MA, USA) was followed by incubation at 37 °C for 2 h. The reaction was stopped by adding 0.1 volume of 0.5 M EDTA, and the labeled DNA was precipitated with isopropanol and resuspended in NimbleGen Hybridization Buffer (Roche NimbleGen). The Cy3-labeled samples were denatured at 95 °C for 5 min and hybridized to the arrays in a NimbleGen Hybridization System (Roche NimbleGen) for 18 h at 42 °C. The arrays were washed using the NimbleGen Wash Buffer Kit (Roche NimbleGen), dried by centrifugation, and scanned at a 5-µm resolution using a GenePix 4000B scanner (Molecular Devices, Sunnyvale, CA, USA) and NIMBLESCAN 2.5 software (Roche NimbleGen) to obtain fluorescence intensities. Accuracy of hybridization was assured by internal controls spotted in duplicates, and reproducibility of array results was confirmed in preliminary studies by analyzing the same sample twice.

3.5. Data Analysis of DNA Microarray

Raw signal intensities were first corrected based on background subtraction methods proposed by Xie *et al*. [42], by which most signals were shifted close to zero (Figure S3). Remaining noise was further evaluated by setting several threshold levels based on *p*-values. MAS5.0 background method from Affymetrix [43] was also applied. Data from duplicate probes were treated as a pair of independent data sets and such data sets were compared to verify the reproducibility of the array. The original and corrected data were assessed by histogram, box plot and MA plot (Figure S4).

Then, mean corrected signal intensities were calculated from the duplicate data. Because the signal intensities from each sample did not fall into a Gaussian distribution, and the number of probes that showed strong signals in each samples were about 1,000 (Figure S5), we decided to analyze these top 1,000 sequences in each sample for taxonomic classification and detection of pathogenic bacteria. Reliability of the data was evaluated by comparing the data from duplicate probes in each sample to create heat maps and perform clustering for phylum, class, order, family, genus and species, and to calculate *p*-values based on probability distribution (data not shown).

A total of 743 non-redundant sequences that were included in the top 200 sequences in each of the seven samples were subjected to further hierarchical cluster analysis. The rank of the signal intensity in each microarray was changed to the reciprocal number and analyzed with Cluster 3.0 [44]. Briefly, in the beginning of the hierarchical cluster analysis each element (data) was in a cluster of its own. The clusters were then sequentially combined into larger clusters, until all elements end up being in the same cluster. At each step, the two clusters separated by the shortest distance were combined. The definition of "shortest distance" was calculated using Lance-Williams algorithms (Ward's method). Java Treeview 1.0 [45] was then used for tree visualization of the microarray data.

3.6. Nested PCR

The 16S rRNA sequence of *Marinomonas* sp. BSi20412 (GenBank Accession No. DQ537503) was aligned with those of *Escherichia coli* (GenBank Accession No. J01859) and *Bacillus subtilis* (GenBank Accession No. AY219900) to identify a *Marinomonas*-specific sequence. In this region, the following PCR primer sets were designed for nested PCR: 5'-TTCAGGGGTGAGGAAGGGCGTTTG-3' (initial forward); 5'-CTCACAGTTCCCGAAGGCACTCCA-3' (initial reverse); 5'-ACCCAGATGTCTTGACGTTAGCCC-3' (second forward); 5'-GCATCTCTACCGGATTCTGTGGA-3' (second reverse) to amplify a 242-bp fragment. Touchdown PCR was performed with the external primers and the same template DNA that was used for microarray analysis. The PCR product was diluted and used as the template for the second round of nested PCR as described previously [37]. The primer set that could amplify all of the prokaryotic 16S rRNA sequences was used as a control: 5'-CCTACGGGAGGCAGCAG-3' (forward); 5'-ATTACCGCGGCTGCTGG-3' (reverse) to amplify a 161-bp fragment [46].

4. Conclusions

In conclusion, we have developed a 16S rRNA microarray that allows comprehensive detection of prokaryotes in a simple and cost-effective manner, making it possible to analyze clusters that include

both known and novel index species. This DNA microarray will be a useful tool in the detection of pathogenic and temperature-sensitive bacteria in environmental samples.

Supplementary Materials

Supplementary materials can be accessed at: http://www.mdpi.com/2076-0817/02/1/591/s1.

Acknowledgments

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Whitman, W.B.; Coleman, D.C.; Wiebe, W.J. Prokaryotes: The unseen majority. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6578–6583.
2. Amaral-Zettler, L.A.; Mccliment, E.A.; Ducklow, H.W.; Huse, S.M. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS One* **2009**, *4*, e6372.
3. Sogin, M.L.; Morrison, H.G.; Huber, J.A.; Mark, W.D.; Huse, S.M.; Neal, P.R.; Arrieta, J.M.; Herndl, G.J. Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 12115–12120.
4. Amann, R.I.; Ludwig, W.; Schleifer, K.H. Phylogenetic identification and *in situ* detection of individual microbial cells without cultivation. *Microbiol. Rev.* **1995**, *59*, 143–169.
5. Venter, J.C.; Remington, K.; Heidelberg, J.F.; Halpern, A.L.; Rusch, D.; Eisen, J.A.; Wu, D.; Paulsen, I.; Nelson, K.E.; Nelson, W.; *et al.* Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **2004**, *304*, 66–74.
6. Rusch, D.B.; Halpern, A.L.; Sutton, G.; Heidelberg, K.B.; Williamson, S.; Yooseph, S.; Wu, D.; Eisen, J.A.; Hoffman, J.M.; Remington, K.; *et al.* The Sorcerer II Global Ocean Sampling expedition: Northwest Atlantic through eastern tropical Pacific. *PLoS Biol.* **2007**, *5*, e77.
7. Dinsdale, E.A.; Edwards, R.A.; Hall, D.; Angly, F.; Breitbart, M.; Brulc, J.M.; Furlan, M.; Desnues, C.; Haynes, M.; Li, L.; *et al.* Functional metagenomic profiling of nine biomes. *Nature* **2008**, *452*, 629–632.

8. Brodie, E.L.; Desantis, T.Z.; Joyner, D.C.; Baek, S.M.; Larsen, J.T.; Andersen, G.L.; Hazen, T.C.; Richardson, P.M.; Herman, D.J.; Tokunaga, T.K.; *et al*. Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation. *Appl. Environ. Microbiol.* **2006**, *72*, 6288–6298.

9. Palmer, C.; Bik, E.M.; Eisen, M.B.; Eckburg, P.B.; Sana, T.R.; Wolber, P.K.; Relman, D.A.; Brown, P.O. Rapid quantitative profiling of complex microbial populations. *Nucleic Acids Res.* **2006**, *34*, e5.

10. Japan Meteorological Agency. Ocean Current. Available online: www.data.kishou.go.jp/ db/kaikyo/jun/current_NK.html (accessed on 20 July 2012).

11. Japan Meteorological Agency. Monthly Mean Sea Surface Temperature. Available online: www.data.kishou.go.jp/db/kaikyo/monthly/sst.html (accessed on 20 July 2012).

12. Wale, N.; Karypis, G. Target fishing for chemical compounds using target-ligand activity data and ranking based methods. *J. Chem. Inf. Model* **2009**, *49*, 2190–2201.

13. Monfort, P.; Baleux, B. Distribution and survival of motile *Aeromonas* spp. in brackish water receiving sewage treatment effluent. *Appl. Environ. Microbiol.* **1991**, *57*, 2459–2467.

14. Jolivet-Gougeon, A.; Tamanai-Shacoori, T.; Sauvager, F.; Cormier, M. Production of *Escherichia coli* group I-like heat-labile enterotoxin by Enterobacteriaceae isolated from environmental water. *Microbios* **1997**, *90*, 209–218.

15. Alexopoulos, A.; Plessas, S.; Voidarou, C.; Noussias, H.; Stavropoulou, E.; Mantzourani, I.; Tzora, A.; Skoufos, I.; Bezirtzoglou, E. Microbial ecology of fish species ongrowing in Greek sea farms and their watery environment. *Anaerobe* **2011**, *17*, 264–266.

16. Lee, E.S.; Lee, M.Y.; Han, S.H.; Ka, J.O. Occurrence and molecular differentiation of environmental mycobacteria in surface waters. *J. Microbiol. Biotechnol.* **2008**, *18*, 1207–1215.

17. Jernigan, J.A.; Farr, B.M. Incubation period and sources of exposure for cutaneous Mycobacterium marinum infection: Case report and review of the literature. *Clin. Infect. Dis.* **2000**, *31*, 439–443.

18. Yu, Y.; Li, H.; Zeng, Y.; Chen, B. Extracellular enzymes of cold-adapted bacteria from Arctic sea ice, Canada Basin. *Polar Biol.* **2009**, *32*, 1539–1547.

19. Takami, H.; Inoue, A.; Fuji, F.; Horikoshi, K. Microbial flora in the deepest sea mud of the Mariana Trench. *FEMS Microbiol. Lett.* **1997**, *152*, 279–285.

20. Gihring, T.M.; Banfield, J.F. Arsenite oxidation and arsenate respiration by a new Thermus isolate. *FEMS Microbiol. Lett.* **2001**, *204*, 335–340.

21. Takai, K.; Hirayama, H.; Sakihama, Y.; Inagaki, F.; Yamato, Y.; Horikoshi, K. Isolation and metabolic characteristics of previously uncultured members of the order aquificales in a subsurface gold mine. *Appl. Environ. Microbiol.* **2002**, *68*, 3046–3054.

22. Hirayama, H.; Takai, K.; Inagaki, F.; Yamato, Y.; Suzuki, M.; Nealson, K.H.; Horikoshi, K. Bacterial community shift along a subsurface geothermal water stream in a Japanese gold mine. *Extremophiles* **2005**, *9*, 169–184.

23. Nakagawa, T.; Ishibashi, J.; Maruyama, A.; Yamanaka, T.; Morimoto, Y.; Kimura, H.; Urabe, T.; Fukui, M. Analysis of dissimilatory sulfite reductase and 16S rRNA gene fragments from deep-sea hydrothermal sites of the Suiyo Seamount, Izu-Bonin Arc, Western Pacific. *Appl. Environ. Microbiol.* **2004**, *70*, 393–403.

24. Kojima, H.; Teske, A.; Fukui, M. Morphological and phylogenetic characterizations of freshwater *Thioploca* species from Lake Biwa, Japan, and Lake Constance, Germany. *Appl. Environ. Microbiol.* **2003**, *69*, 390–398.

25. Haramoto, E.; Kitajima, M.; Katayama, H.; Asami, M.; Akiba, M.; Kunikane, S. Application of real-time PCR assays to genotyping of F-specific phages in river water and sediments in Japan. *Water Res.* **2009**, *43*, 3759–3764.

26. Colwell, R.R. Infectious disease and environment: Cholera as a paradigm for waterborne disease. *Int. Microbiol.* **2004**, *7*, 285–289.

27. Hrudey, S.E.; Hrudey, E.J. Published case studies of waterborne disease outbreaks—Evidence of a recurrent threat. *Water Environ. Res.* **2007**, *79*, 233–245.

28. Berger, M.; Shiau, R.; Weintraub, J.M. Review of syndromic surveillance: Implications for waterborne disease detection. *J. Epidemiol. Community Health* **2006**, *60*, 543–550.

29. Edwards, D.D.; Mcfeters, G.A.; Venkatesan, M.I. Distribution of *Clostridium perfringens* and fecal sterols in a benthic coastal marine environment influenced by the sewage outfall from McMurdo Station, Antarctica. *Appl. Environ. Microbiol.* **1998**, *64*, 2596–2600.

30. Abdelzaher, A.M.; Wright, M.E.; Ortega, C.; Solo-Gabriele, H.M.; Miller, G.; Elmir, S.; Newman, X.; Shih, P.; Bonilla, J.A.; Bonilla, T.D.; *et al*. Presence of pathogens and indicator microbes at a non-point source subtropical recreational marine beach. *Appl. Environ. Microbiol.* **2010**, *76*, 724–732.

31. Robben, P.M.; Sibley, L.D. Food- and waterborne pathogens: You are (infected by) what you eat! *Microb. Infect.* **2004**, *6*, 406–413.

32. Bhan, M.K.; Bahl, R.; Bhatnagar, S. Typhoid and paratyphoid fever. *Lancet* **2005**, *366*, 749–762.

33. Rich, V.I.; Pham, V.D.; Eppley, J.; Shi, Y.; Delong, E.F. Time-series analyses of Monterey Bay coastal microbial picoplankton using a "genome proxy" microarray. *Environ. Microbiol.* **2011**, *13*, 116–134.

34. National Institute of Genetics. DNA Data Bank of Japan (DDBJ). Available online: ftp://ftp.ddbj.nig.ac.jp/ddbj_database/16S (accessed on 30 March 2009).

35. Tan, J.C.; Patel, J.J.; Tan, A.; Blain, J.C.; Albert, T.J.; Lobo, N.F.; Ferdig, M.T. Optimizing comparative genomic hybridization probes for genotyping and SNP detection in *Plasmodium falciparum. Genomics* **2009**, *93*, 543–550.

36. Uda, A.; Tanabayashi, K.; Fujita, O.; Hotta, A.; Yamamoto, Y.; Yamada, A. Comparison of whole genome amplification methods for detecting pathogenic bacterial genomic DNA using microarray. *Jpn. J. Infect. Dis.* **2007**, *60*, 355–361.

37. Suzuki, K.; Takigawa, W.; Tanigawa, K.; Nakamura, K.; Ishido, Y.; Kawashima, A.; Wu, H.; Akama, T.; Sue, M.; Yoshihara, A.; *et al*. Detection of *Mycobacterium leprae* DNA from archaeological skeletal remains in Japan using whole genome amplification and polymerase chain reaction. *PLoS One* **2010**, *5*, e12422.

38. Suzuki, M.T.; Giovannoni, S.J. Bias caused by template annealing in the amplification of mixtures of 16S rRNA genes by PCR. *Appl. Environ. Microbiol.* **1996**, *62*, 625–630.

39. Kuroki, H.; Miyamoto, H.; Fukuda, K.; Iihara, H.; Kawamura, Y.; Ogawa, M.; Wang, Y.; Ezaki, T.; Taniguchi, H. *Legionella impletisoli* sp. nov. and *Legionella yabuuchiae* sp. nov., isolated from soils contaminated with industrial wastes in Japan. *Syst. Appl. Microbiol.* **2007**, *30*, 273–279.

40. Nubel, U.; Engelen, B.; Felske, A.; Snaidr, J.; Wieshuber, A.; Amann, R.I.; Ludwig, W.; Backhaus, H. Sequence heterogeneities of genes encoding 16S rRNAs in *Paenibacillus polymyxa* detected by temperature gradient gel electrophoresis. *J. Bacteriol.* **1996**, *178*, 5636–5643.

41. Akama, T.; Suzuki, K.; Tanigawa, K.; Kawashima, A.; Wu, H.; Nakata, N.; Osana, Y.; Sakakibara, Y.; Ishii, N. Whole-genome tiling array analysis of *Mycobacterium leprae* RNA reveals high expression of pseudogenes and noncoding regions. *J. Bacteriol.* **2009**, *191*, 3321–3327.

42. Xie, Y.; Wang, X.; Story, M. Statistical methods of background correction for Illumina BeadArray data. *Bioinformatics* **2009**, *25*, 751–757.

43. Affymetrix, Inc. Statistical Algorithms Description Document. Technical Report; Affymetrix: Santa Clara, CA, USA, 2002.

44. Eisen, M.B.; Spellman, P.T.; Brown, P.O.; Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 14863–14868.

45. Saldanha, A.J. Java Treeview—Extensible visualization of microarray data. *Bioinformatics* **2004**, *20*, 3246–3248.

46. Hollibaugh, J.T.; Bano, N.; Ducklow, H.W. Widespread distribution in polar oceans of a 16S rRNA gene sequence with affinity to Nitrosospira-like ammonia-oxidizing bacteria. *Appl. Environ. Microbiol.* **2002**, *68*, 1478–1484.