Journal of Microbiological Methods 75 (2008) 523–530

Contents lists available at ScienceDirect

Journal of Microbiological Methods

journal homepage: www.elsevier.com/locate/jmicmeth

Statistical superiority of genome-probing microarrays as genomic DNA–DNA hybridization in revealing the bacterial phylogenetic relationship compared to conventional methods ☆

Ho-Won Chang [a], Young-Do Nam [a], Min Young Jung [a], Kyoung-Ho Kim [a], Seong Woon Roh [a], Min-Soo Kim [a], Che Ok Jeon [b], Jung-Hoon Yoon [a], Jin-Woo Bae [a,c,*]

[a] Biological Resource Center, KRIBB, Daejeon 305-806, Republic of Korea
[b] Department of Life Science, Chung-Ang University, Seoul 156-756, Republic of Korea
[c] Environmental Biotechnology National Core Research Center, Gyeongsang National University, Jinju 660-701, Republic of Korea

ARTICLE INFO

Article history:
Received 19 May 2008
Received in revised form 7 August 2008
Accepted 8 August 2008
Available online 17 August 2008

Keywords:
DNA–DNA hybridization
Genome-probing microarray
Multi-locus sequence typing (MLST)

ABSTRACT

The genomic DNA–DNA hybridization (DDH) method has been widely used as a practical method for the determination of phylogenetic relationships between closely related biological strains. Traditional DDH methods have serious limitations including low reproducibility, a high background and a time-consuming procedure. The DDH method using a genome-probing microarray (GPM) has been recently developed to complement conventional methods and could be used to overcome the limitations that are typically encountered. It is necessary to compare the GPM-based DDH method to the conventional methods before using the GPM for the estimation of genomic similarities since all of the previous scientific data have been entirely dependent on conventional DDH methods. In order to address this issue we compared the DDH values obtained using the GPM, microplate and nylon membrane methods to multi-locus sequence typing (MLST) data for 9 *Salmonella* genomes and an *Escherichia coli* type strain. The results showed that the genome similarity values and the degrees of standard deviation obtained using the GPM method were lower than those obtained with the microplate and nylon membrane methods. The dendrogram from the cluster analysis of GPM DDH values was consistent with the phylogenetic tree obtained from the multi-locus sequence typing (MLST) data but was not similar to those obtained using the microplate and nylon membrane methods. Although the signal intensity had to be maximal when the targets were hybridized to their own probe, the methods using membranes and microplates frequently produced higher signals in the heterologous hybridizations than those obtained in the homologous hybridizations. Only the GPM method produced the highest signal intensity in homologous hybridizations. These results show that the GPM method can be used to obtain results that are more accurate than those generated by the other methods tested.

1. Introduction

Genomic DNA–DNA hybridization (DDH) has been widely used as a practical method for the estimation of the genomic similarity of multiple strains in viruses (Bond et al., 1978), bacteria (Johnson, 1973), yeasts (Bicknell and Douglas, 1970), fungi (Dutta et al., 1976), plants (Bendich and Bolton, 1967), insects (Sohn et al., 1975) and even primates (Sibley and Ahlquist, 1984). Indeed, analysis of prokaryotic genomic DNA provides a much higher resolution than SSU rDNA sequencing (Forney et al., 2004) and is considered to be critical for determining bacterial species boundaries (Wayne et al., 1987). Conventional DDH methods using micro-well plates (Christensen et al., 2000) or membrane filters (De Ley and De Smedt, 1975) have several unavoidable limitations: (i) Results obtained in different laboratories or even in replicated experiments in the same laboratory appear discordant (Selander et al., 1985). (ii) Reciprocal experiments often yield nonisomorphic values. (iii) Conventional DDH experiments can only provide information about the genome similarity of an isolated strain to one or a few reference strains rather than to a complete matrix of all strains of the same group because of the labor intensive process (Sneath, 1983). (iv) In contrast to microarray analysis, non-statistical analysis (calculation of the signal without the determination of the background intensity and standard deviation) and arbitrary exposure during image acquisition that is common using the conventional DDH methods has generated long-term scientific debates (Templeton, 1985).

We have recently developed a new format of microarray using bacterial genomes as probes for the detection and identification of bacteria in natural environments; we called this the genome-probing microarray

☆ The sequences reported in this study have been deposited in the GenBank database (accession numbers EU014630-EU014689).

* Corresponding author. Biological Resource Center, KRIBB, Daejeon 305-806, Republic of Korea. Tel.: +82 42 860 4628; fax: +82 42 860 4677.
E-mail address: baejw@kribb.re.kr (J.-W. Bae).

0167-7012/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.mimet.2008.08.003

(GPM) (Bae et al., 2005). We utilized the powerful characteristics of the microarray-style of the DDH method to ensure that its accuracy and precision would allow genomic DDH data to be highly reproducible, sensitive, and specific (Dorris et al., 2002). In order to achieve this, the genomic DNA target for the hybridization reaction was labeled with Cy-5 and the entire reaction was performed on a glass slide. The fluorescence intensity was detected and analyzed using a microarray scanner and automatic image analysis software. A low background and reliable resolution were achieved using a robot to automatically print the slides. Moreover, this method could be applied to simultaneously compare the genome similarity of large numbers of strains (Bae et al., 2005).

It was necessary to compare the microarray-based DDH method to the conventional DDH methods before using the GPM for the estimation of genomic similarities since all of the previous scientific data and the rule determining bacterial species boundaries (DDH values that are higher than 70% means that two strains are the same species) entirely depends on conventional DDH methods. In the current study we have compared three DDH methods (GPM, microplate, and nylon membrane) using nine *Salmonella* type strains and an *Escherichia coli* type strain which share extraordinarily close inter-genomic relationships. In order to determine which method provided the most accurate results, a similarity dendrogram of the DDH values obtained for each method was compared to multi-locus sequence typing (MLST) data, which is currently and widely used as the microbial typing method (Maiden, 2006).

2. Materials and methods

2.1. Bacterial strains and preparation of genomic DNA

In order to compare DDH methods, nine strains of *Salmonella* type species and one strain of *E. coli* were used in this study (Table 1). These strains were obtained from the Korean Collection for Type Cultures (KCTC), the American Type Culture Collection (ATCC), and the German Collection of Microorganisms and Cell Cultures (DSMZ). All of the strains were maintained and grown under the conditions suggested by the collections from which they were sourced. Cells were harvested at the exponential growth phase and frozen at −80 °C for the extraction of the genomic DNA. Genomic DNA was isolated using the bead-beating method as described previously (Yeates et al., 1998). DNA was treated with RNase A (Sigma, St. Louis, MO) and purified using the ethanol precipitation method. The DNA concentration was determined in triplicate using a spectrophotometer (Nanodrop Technologies, Rockland, DE).

2.2. DNA–DNA hybridization by GPM

DNA–DNA hybridization by GPM was performed by the method of Bae et al. (2005). In order to construct the GPM, genomic DNA was diluted to a final concentration of 400 ng/μl in 0.1×Tris-EDTA buffer. Five microliters of ten probe genomes were transferred to a 384-well microplate and mixed with 5 μl of 2×microarray spotting solution (ArrayIt™, Telechem International, Inc., Sunnyvale, CA) for printing. A GenePix® 4000B microarray scanner set (Axon instruments, Union City, CA) was used to scan the microarrays at a resolution of 10 μm. The local background signal was also automatically subtracted from the hybridization signal of each individual spot. The signal-to-noise ratio (SNR) for each spot was calculated based on following formula (Zong et al., 2003): SNR=(signal intensity−background)/SD of background, in which the "background" measurement refers to the local spot background intensity and the "SD of background" was calculated across all of the pixels, as measured by the GenePix® software.

Table 1
List of strains used in this study

Number	Culture collection number	Species
1	DSM 14848	*Salmonella choleraesuis* subsp. *indica*
2	DSM 14846	*Salmonella choleraesuis* subsp. *choleraesuis*
3	DSM 13772	*Salmonella bongori*
4	DSM 9386	*Salmonella choleraesuis* subsp. *arizonae*
5	DSM 9221	*Salmonella choleraesuis* subsp. *hountenae*
6	DSM 9220	*Salmonella choleraesuis* subsp. *salamae*
7	ATCC 13311	*Salmonella typhimurium*
8	ATCC 13076	*Salmonella enteritidis*
9	DSM 14847	*Salmonella choleraesuis* subsp. *diarizonae*
10	KCTC 2441	*Escherichia coli*

2.3. DNA–DNA hybridization using a microplate

The DDH experiment using a microplate was performed by the method of Ezaki et al. (1989). A Fluoroskan ascent plate reader (Thermo) was used for 0, 15, 30, 60, 90, and 120 min at a wavelength of 360 nm for excitation and 450 nm for emission in order to obtain a measurement of the fluorescence intensity; DDH values were calculated using the fluorescence values determined at 90 min. All reactions were performed in quintuplicate experiments. The highest and lowest values of fluorescence intensity were excluded and the average values of the remaining three intensities were used as the DDH value. The fluorescence intensity value for the homologous reaction well was considered to be the 100% relationship.

2.4. DNA–DNA hybridization using a nylon membrane

DNA–DNA slot-blot hybridization reactions using a nylon membrane were performed as described previously (Kafatos et al., 1979) with some modifications. Following hybridization and washing steps, the membrane was subsequently deposited on a development cassette together with the chemiluminescent substrate for alkaline phosphatase and exposed to Hyperfilm (Amersham Pharmacia Biotech, Piscataway, NJ) using the DIG High Prime DNA Labeling and Detection Starter Kit for 30 min at room temperature. Overdeveloped film was washed in flowing water and dried. The intensities of each blot were calculated using the program TINA 2.0 (Raytest, Straubenhardt, Germany). The DDH value was determined to be the average intensity of triplicate experiments and the density value of the homologous reaction blot was considered to be the 100% relationship.

2.5. Multilocus sequence typing (MLST) analysis

Six housekeeping genes (*manB* (phosphomannomutase), *putP* (proline permease), *gapA* (glyceraldehyde-3-phosphate dehydrogenase), *gyrB* (DNA gyrase subunit B), *mdh* (malate dehydrogenase), and *16S rRNA* (16S ribosome RNA)) were used for the MLST analysis. All of the PCR amplifications were carried out in a final volume of 20 μl, containing 1 μl (10 ng/μl) of template, 10 mM Tris HCl (pH 9.0), 40 mM KCl, 250 μM deoxynucleoside triphosphates (dNTPs), 1 unit (U) of *Taq* polymerase, 1.5 mM $MgCl_2$, and 10 pmole of each primer. Primer sets for the amplification of the six genes used in this study are summarized in Table 2. Sequences of all of the genes were combined following the gene order and aligned using the multiple sequence alignment program CLUSTAL_X (1·8) (Thompson et al., 1997). Phylogenetic relationships between the ten species were determined using MEGA version 2·1 software. Distance matrices were determined following the assumptions that were described previously (Kimura, 1980). These matrices were used to elaborate dendrograms using the neighbor-joining method (Saitou and Nei, 1987).

2.6. Comparison using percent similarity values

The construction of dendrogram based on percent similarity of the DNA–DNA relationships was performed using BioNumerics software (BioSystematica, UK). Each DDH value was converted to the visual

Table 2
List of primers for MLST analysis and amplified gene products

Gene	Forward	Reverse	Amplificon size	Reference
manB	5-CCG GCA CCG AAG AGA-3	5-CGC CGC CAT CCG GTC-3	893 bp	(Kotetishvili et al., 2002)
putP	5-ACC CCC CAT GGT GGT GGT TCC CAT-3	5-TGA CGG CGG AGC GGA ATG ATA ATG-3	1500 bp	(Nelson and Selander, 1992)
gapA	5-TAT GAC TAT CAA AGT AGG TAT-3	5-GTT GGA GTA ACC GGT TTC GT-3	924 bp	(Nelson et al., 1991)
gyrB	5-GAA GTC ATC ATG ACC GTT CTG CA-3	5-AGC AGG GTA CGG ATG TGC GAG CC-3	1200 bp	(Yamamoto and Harayama, 1995)
mdh	5-GAT GAA AGT CGC AGT CCT CG-3	5-TAT CCA GCA TAG CGT CCA GC-3	849 bp	(Boyd et al., 1994)
16S rRNA	5-AGA GTT TGA TCM TGG CTC AG-3	5-TAC GGY TAC CTT GTT ACG ACT T-3	1400 bp	(Klappenbach et al., 2000)

band with the representative color tone (white bands indicate lower values of percent similarity and black bands indicate higher values) using ArrayColor.exe (http://microarray.brc.re.kr/). One hundred bands converted were used for the phylogenetic analysis based on the band intensities. Ten bands of each lane were selected as positions for normalization and the lanes were normalized to compensate for differences in migration distance. The dendrograms were calculated on the basis of the Pearson product moment correlation coefficient by using the unweighted pair group method with the arithmetic averages clustering algorithm (UPGMA).

3. Results

3.1. Phylogenetic analysis of Salmonella species and E. coli with 16S rRNA gene sequences

Currently, the genus *Salmonella* includes five systematically valid species: *S. choleraesuis* (subsp. *choleraesuis*, *salamae*, *arizonae*, *diarizonae*, *hountenae*, and *indica*), *S. enteritidis*, *S. typhimurium*, *S. typhi*, and *S. bongori*. Although *S. enteritidis*, *S. typhimurium*, and *S. typhi* are almost genetically identical to *S. choleraesuis* subsp. *choleraesuis*, these species continue to be listed in the Bacterial Approved Lists and retain their unique nomenclatural because they are now very important human pathogens. In this study, the type strains of *S. enteritidis*, *S. typhimurium*, *S. bongori* and 6 subspecies (serovars) of *S. choleraesuis* and *E. coli* were selected to compare three DDH methods to each other. The 16S rDNA gene-based phylogeny revealed that the Salmonella type strains used in this study are very closely related within all species and subspecies, with the exception of S. bongori and *E. coli*, and share more than 98% 16S rDNA sequence similarity. S. bongori exhibited 97.2–98.0% sequence identity with the other *Salmonella* type strains while *E. coli*, which was used as an experimental control and an out-group in the phylogenetic tree, exhibited 96.7–97.9% similarity with *Salmonella*. The 16S rDNA distances between the Salmonella strains used in this study are plotted in Fig. 1.

3.2. Comparison of GPM to conventional DDH methods

We used the ten genomically similar type strains to compare the GPM-based DDH to the conventional methods that utilize either microplates or nylon membranes. When the genomes from nine *Salomonlla* and *Escherichia* type strains were hybridized to the genome probes on the GPM slide, the signal intensity values for cross-hybridization (genomic similarity values) using the GPM were lower than those obtained with the microplate and nylon membrane methods (Fig. 2). Additionally, when the genomes of the phylogenetically distant microbes *S. indica*, *S. bongori* and *E. coli* were hybridized to the probes, only the GPM method could distinguish them into different species boundaries (prokaryotic species are considered to be a group of strains that exhibit over 70% similarity in their DDH ratio) from the others (Fig. 2A). Given that only the type strains of *S. indica*, *S. bongori* and *E. coli* are different species from the other 7 strains, only DDH using the GPM method produced reliable results; when the genome of the *E. coli* strain was used as a target the DDH values were 1–14% similar to the other 9 *Salmonlla* probes. When the genomes of the *Salmonella* strains were used as GPM targets the DDH values from the heterologous probes ranged from 17.3% to 28.4% (*S. indica*), 4.1% to 75.9% (*S. choleraesuis*), 22.5% to 48.7% (*S. bongori*), 18.1% to 72% (*S. arizonae*), 26.5% to 90.6% (*S. hountenae*), 19.9% to 62.3% (*S. salamae*), 9.4% to 51.2% (*S. typhimurium*), 16.6% to 93.1% (*S. enteritidis*) and 18.6% to 76.7% (*S. diarizonae*). In the experiment using the microplate method the targets from the *Salmonella* genomes produced DDH values ranging from 50.8–115.1% and the *E. coli* targets produced 52.2–71.1% in heterologous hybridizations (Fig. 2B). In the DDH experiment using nylon membranes, the targets from nine strains

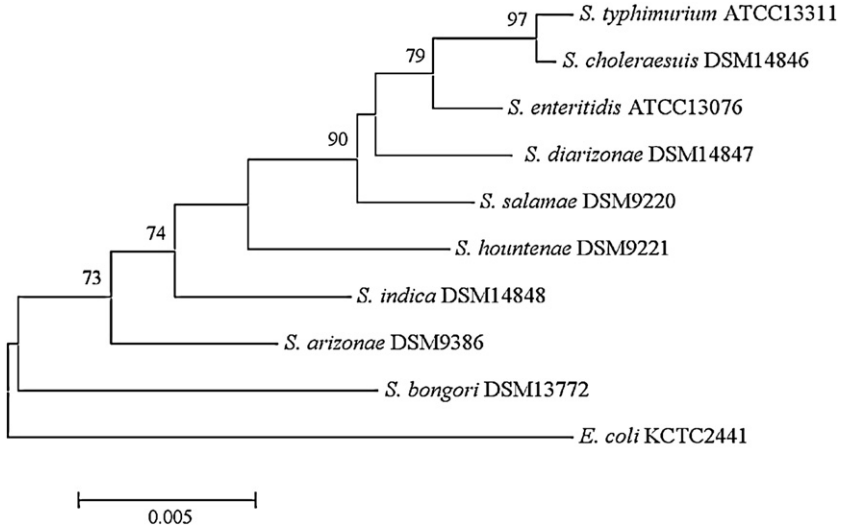

Fig. 1. A phylogenetic tree based on 16S rRNA gene sequences showing the position of each type strain in the genus *Salmonella*. The tree was generated by the neighbor-joining method. The numbers at the nodes indicate bootstrap values (1000 replications). Bar, 0.005 accumulated changes per nucleotide.

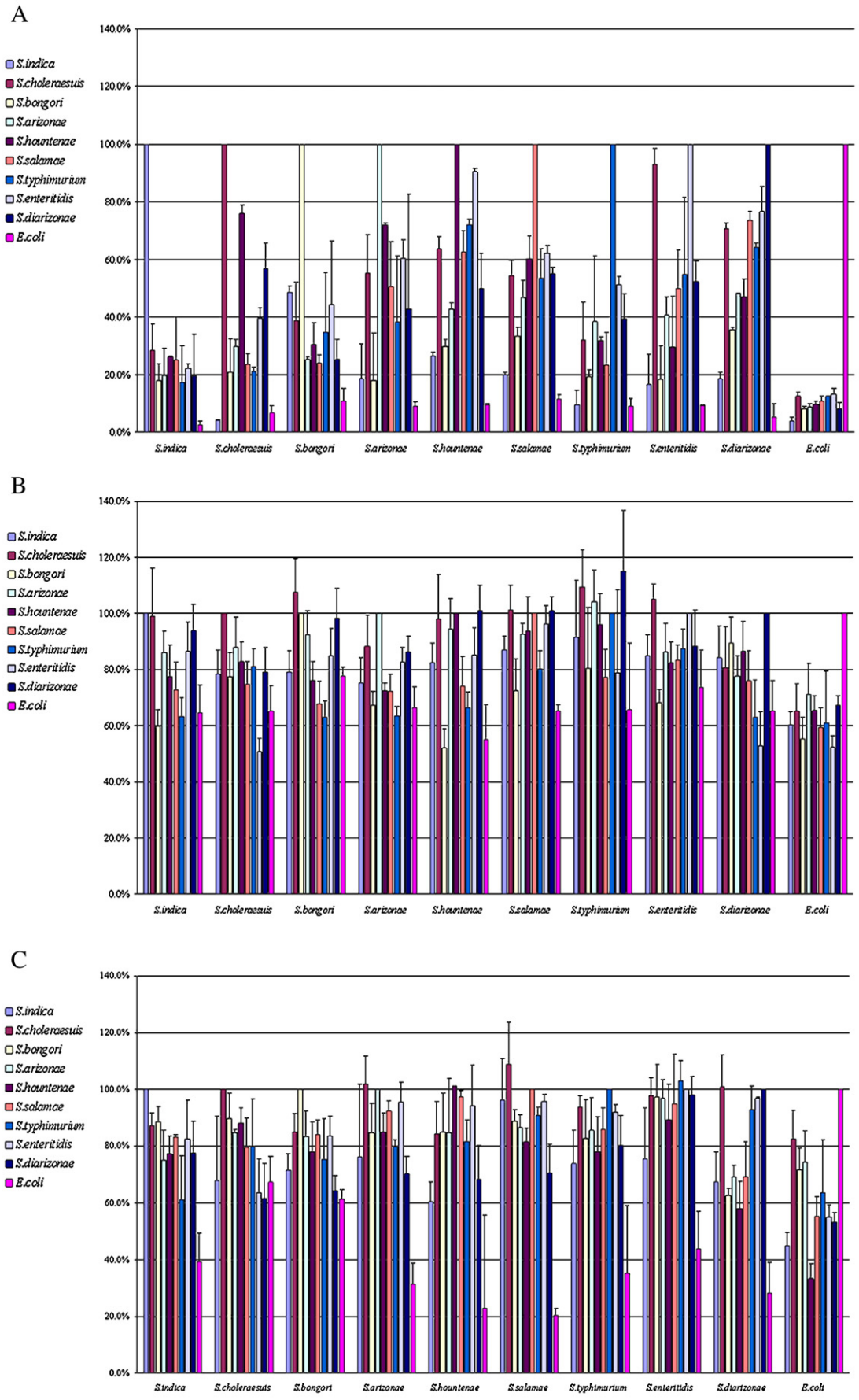

Fig. 2. Histogram plots showing DNA–DNA relationships using the DDH method with (A) GPM, with (B) microplate, and (C) membrane. The *y*-axis represents DDH values (%) and the *x*-axis represents the strains used.

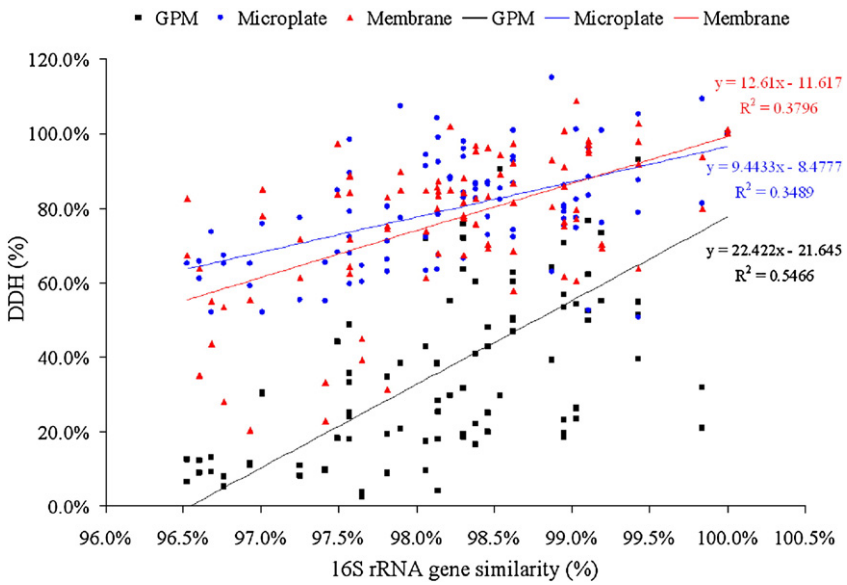

Fig. 3. Relationship between the similarity of the 16S rRNA gene and DDH values. The *y*-axis represents the values (%) obtained for each DDH method and the *x*-axis represents the 16S rRNA gene homology (%). The GPM data indicates a generally high R^2 value, low DDH values and a steeper slope than the other methods.

of the genus *Salmonella* produced values ranging from 57.9–108.9% and *E. coli* produced values of 33.3–82.7% in heterologous hybridizations (Fig. 2C). Although the signal intensity had to be maximal when the targets were hybridized to their own probe, the methods using membranes and microplates frequently produced higher signals in the heterologous hybridizations than those obtained in the homologous hybridizations (Fig. 2 B and C). Only the GPM method produced the highest signal intensity in homologous hybridizations (Fig. 2 A). The average values obtained from all of the heterologous DDH experiments using GPM, microplates and membranes were 31%, 71% and 69%, respectively.

The results obtained using the GPM method indicated that DDH values between the strains of different genus (*E. coli* vs 9 *Salmonella*) were very low (<14% similarity) and DDH values between the strains of different species (*S. bongori* vs 8 other *Salmonella*) exhibited less than 50% similarity. These values were an average of 48% lower than the values obtained using the microplate method and 46% lower than the ones obtained using the membrane method. The most prominent difference between the methods was their reproducibility in multiple experiments. The averages of the standard deviations of DDH using the GPM, microplate and nylon membrane methods obtained from quadruplicate experiments were 6%, 8.4% and 8.2%, respectively. The lower standard deviation obtained with the GPM method proved that the DDH GPM assay using glass slides was the most reproducible of the methods examined.

In Fig. 3, the DDH value from the GPM method was plotted versus the 16S rRNA gene similarity; this result produced a sharper angle than the others, suggesting that the GPM method had more resolving power to discriminate between genomic relationships. The angle of the lines describing 16S rRNA gene similarity versus DDH values obtained with the GPM, microplate and nylon membrane methods were 22.42, 9.44 and 12.61, respectively, and the R^2 values were 0.546, 0.348 and 0.379, respectively. The discriminating capability of the GPM method is also statistically supported by the average standard deviation.

3.3. Comparison of the phylogenetic difference by MLST

The MLST approach is known to provide accurate, reliable data and is a highly discriminating typing method that can be appropriate for genomically coherent biological organisms (Maiden, 2006). In order to determine which DDH method is the most accurate, MLST of the *Salmonella* strains was performed and compared to the DDH results. Six housekeeping genes (phosphomannomutase, proline permease, glyceraldehyde-3-phosphate dehydrogenase, DNA gyrase subunit B, malate dehydrogenase and 16S ribosomal RNA gene) were used for the MLST analysis. The similarity values of five genes that are conserved between *E. coli* and the other *Salmonella* strains were in an approximately 10% wider range than the 16S rRNA gene sequence similarity (data not shown). Comparison of the dendrogams based on each sequence with the results of the DDH values from each method was too complex; therefore, we merged 6 sequences from each organism, obtained imaginary 6811 bp sequences and determined their phylogenetic relationships. A phylogenetic tree based on six housekeeping genes in the ten strains is shown in Fig. 4. Given that the size of the genomic DNA for the *Salmonella* species is approximately 4 Mb, 6811 base pairs is more than 0.1% of the whole genome. When the size of imaginary sequences covers more than 0.1% of the whole genome, the MLST analysis could be the appropriate method for analysis of the whole genome (Maiden, 2006). The relationships that were based on the MLST analysis of the imaginary sequences between

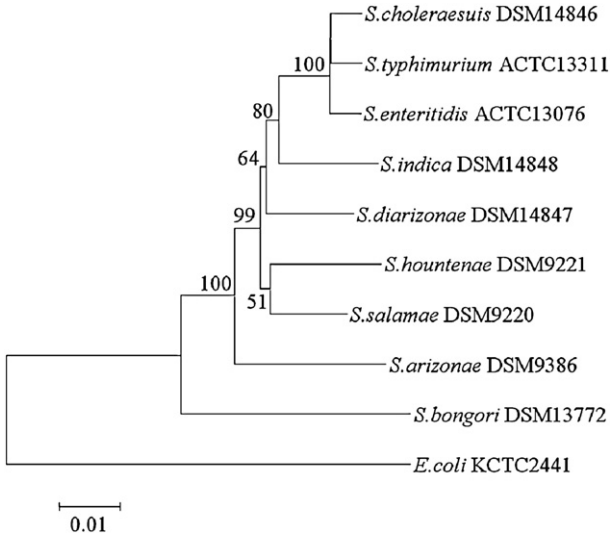

Fig. 4. Phylogenetic analysis based on a concatenated alignment of the six gene sequences (*manB*, *putP*, *gapA*, *gyrB*, *mdh*, *16S rRNA*) showing the relationship between strains used in this study. The phylogenetic tree was generated by the neighbor-joining method. The numbers at the nodes indicate bootstrap values (1000 replications). Bar, 0.01 accumulated changes per nucleotide.

the almost genetically identical *S. enteritidis*, *S. typhimurium* and *S. choleraesuis* were in the 98.9–99.0% range. The identities of the imaginary sequences from *S. arizonae*, *S. hountenae*, *S. salamae*, *S. diarizonae*, and *S. indica* were all in the range of 96.2–97.2%. The *S. bongori* sequence produced a 92.8–93.9% identity with those of the other eight *Salmonella* strains. *E. coli* produced 88–88.7% similarity with those from the other 9 *Salmonella* strains. The phylogenetic tree of the imaginary sequences was similar to that of the 16S rRNA gene.

In order to compare the DDH methods more accurately, a dendrogram based on cluster analysis was obtained from the genomic similarity (cross-hybridization) values obtained for each DDH method (Fig. 5). The DDH similarity values were converted numerically based on the Pearson product moment correlation coefficients and indicated as related values (r). When r is 100, the DDH value of the genome of a strain is identical to the compared one. In the dendrogram constructed from the GPM results, *E. coli* was separated from the other species by 5% of the r value and the group containing *S. bongori* and *S. indica* was separated from the other *Salmonella* by 30% of the r value. The other *Salmonella* strains were subsequently grouped by more than 60% of their r value. The dendrogram obtained from cluster analysis of the GPM DDH values was almost consistent with the phylogenetic tree obtained from the MLST data (Fig. 5A), with the exception of *S. arizonae* and *S. indica*. In contrast, the dendrograms generated by cluster analysis of DDH values obtained from the microplate and nylon membrane methods were not consistent with the phylogenetic tree obtained from MLST data and 16S rRNA gene sequences (Fig. 5B and C). These results show that in addition to an increased resolving power, the GPM method is considered to be more accurate than the other methods.

4. Discussion

Over the past three decades the 16S rRNA gene has been widely used to determine bacterial phylogenetic relationships because of the ability to rapidly estimate the identification of bacteria at the genus-level (Gevers et al., 2005). In practice, strains that share less than 97% of their 16S rRNA gene sequences are delineated into different species (Stackebrandt and Goebel, 1994). The classification of closely related species of bacteria is difficult to acquire through 16S rRNA gene sequence analysis (Christensen et al., 1998). Thus, DDH currently provides higher resolution than SSU rDNA sequencing (Forney et al., 2004) and is considered to be the cornerstone for determining species boundaries (Wayne et al., 1987). The strains that showed similarity

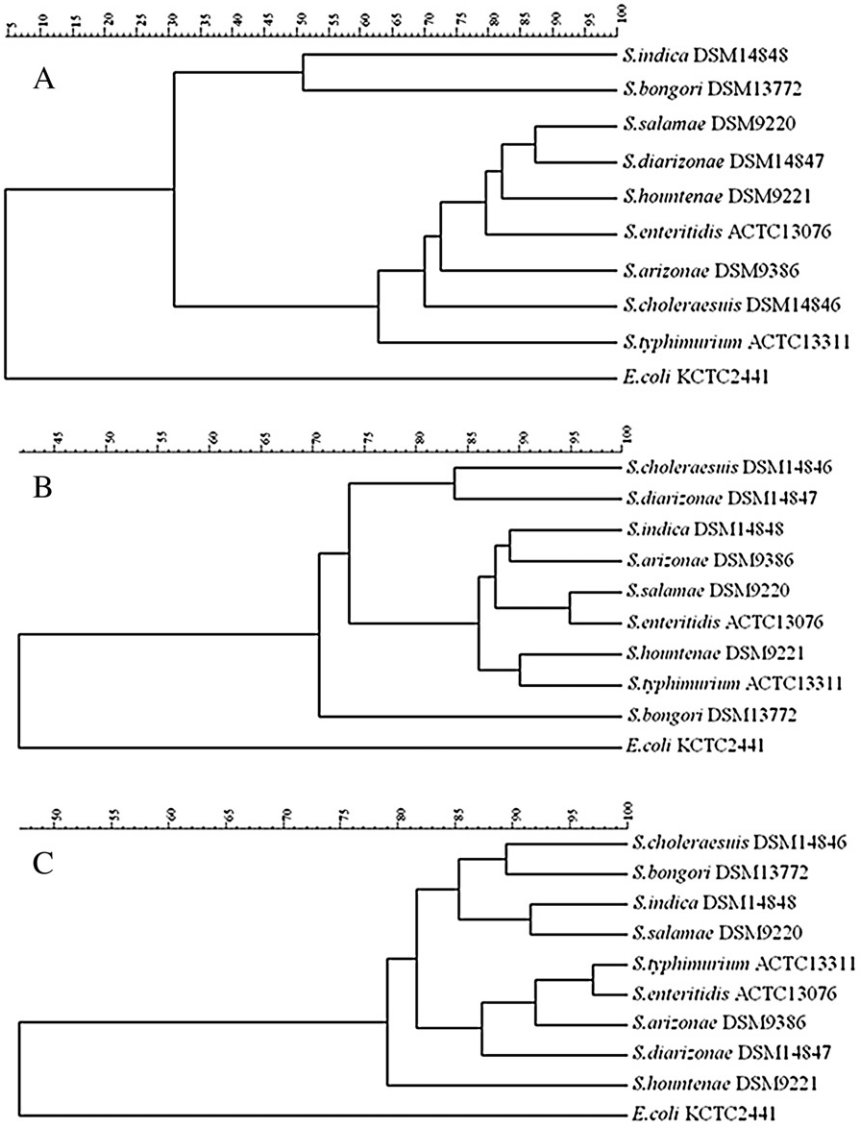

Fig. 5. Cluster analysis of the profiles obtained from DDH values with (A) GPM, (B) microplate, and (C) membranes. Dendrograms were calculated on the basis of the Pearson product moment correlation coefficient of similarity using the unweighted pair group method with the arithmetic averages clustering algorithm (UPGMA).

Table 3
Overview of the protocol of DDH methods

	GPM	Microplate	Membrane
Template	Non-porous superamine-coated glass slides	Non-treated polystyrene microplate	Positively charged nylon membrane, porous
DNA printing	Computerized robot	Manual	Manual
DNA fixing	Ion bond and UV cross-link	Ion bond	Ion bond and UV cross-link
Target labeling	Cy-5 using random prime amplification	Photobiotin with genomic DNA	Digoxigenin using random prime amplification
Target size (dominant)	400 to 700 bp	variable size	100 to 300 bp
Hybridization Tm.	37 °C	0.41×(GC mol% of target DNA)+24.3 °C	0.41×(GC mol% of probe DNA)+24.3 °C
Washing	Mixed SSC and SDS buffer	2×SSC buffer	Mixed SSC and SDS buffer
Substrate	None	4-Methylumbelliferyl-β-d-galactopyranoside	Anti-digoxigenin-alkaline phosphatase
Sensitive emulsion	None	None	Chemiluminescence
Development	None	None	X-ray film
Detection device	Microarray scanner	Fluorescent reader	Densitometry software
Excitation and emission	649 nm and 670 nm	360 nm and 465 nm	None
Signal calculation	SNR=(signal intensity−background)/ standard deviation of background	Fluorescent value	Density value

values greater than 70% by DDH are predominantly strains of the same species whereas strains that showed similarity values less than 70% are generally strains of different species (Johnson, 1973; Wayne et al., 1987).

The DDH value of 70% is critical to demarcate the boundary between bacterial species [DDH values were able to differentiate between strains sharing 67.8% of their DDH values with different species (Stan-Lotter et al., 2002)]; however, to date the conventional DDH methods have not been statistically verified. Research using the conventional microplate method of hybridization has been utilized less than the microarray glass slide method. In addition, it is virtually impossible to eliminate background from the porous surface of nitrocellulose membranes, and image acquisition via the arbitrary exposure of the conventional DDH methods can cause the resulting DDH values to be vague (Templeton, 1985). We found that the methods using membranes and microplates frequently produced higher signals in the heterologous hybridizations than the homologous hybridizations (Fig. 2 B and C), although the GPM method absolutely produced the highest signal intensity in homologous hybridizations (Fig. 2 A). In the results obtained using the conventional methods (Fig. 2 B and C), our results were not in agreement with the current bacterial classification criteria; these results indicated that the species *S. bongori* and the genus *E. coli* frequently belonged to the same *Salmonella* species.

The GPM method described here has several other merits over the conventional DDH methods (The protocols for each DDH method are compared in Table 3. The most prominent differences between the GPM method and the conventional methods were the quality of template, the manner of DNA probe printing, detection device and signal calculation method): (i) GPM can be made easily and avoids labor-intensive work when a large number of hybridizations are needed due to the accurate and precise printing robot. (ii) It is not necessary to use to fluorescent substrates because of the use of a fluorescently labeled DNA target. (iii) Fluorescent detection using a microarray scanner without a fluorescent substrate binding step can save time. (iv) The GPM could be simultaneously hybridized to the large number of probes on a glass slide. (v) The high reproducibility and self-consistent calculation method of the hybridization values using microarray scanner equipment and software could be consistent with results obtained from different laboratories. Automatic resolution is commonly considered to be accurate and allows higher throughput analysis and a walk-away processor (Jain et al., 2002). The other sides the MLST based sequence analysis has major drawback that was only compared to conserved housekeeping genes among closely related microorganisms and the use of the slowly evolving housekeeping genes might induce the ambiguous outcome (Maiden, 2006).

In order to solve the problems encountered using the conventional DDH methods, it is necessary to determine whether the background signal is global background signal obtained from random cross-hybridization affected by negative control or local background signal obtained at the neighborhood of the spot. It is also necessary to calculate the standard deviation of the background signal. One solution to obtain the quantitative evaluation of hybridization signals is to utilize an internal standard in which phylogenetically irrelative genomes are prepared, printed together and mixed with probes. In particular, the DDH method using microplates needs certain verification steps to confirm that the probes are sufficient and are equally printed on the microplate. Recently, Goris et al. (2007) compared DDH values using microplate with various parameters derived from whole-genome sequences. These researchers expressed that linear model was best delineated connection between DDH values and ANI. However, in this study, we reasoned that relationship between DDH values using GPM and parameters derived from whole-genome sequence might be well described by the new exponential model.

Comparison of whole genome sequences is an ideal method for the estimation of genomic relationships among biological organisms. After the success of the Human Genome Project, cost effective sequencing technologies have been rapidly developed (Shendure et al., 2004) and this has made it possible to compare bacteria with an extended number of whole genome sequences (Coenye et al., 2005). Whole genome sequencing of microorganisms is still expensive, given that more than 600 novel species of bacteria are now being added to the Approved list per year, but less than 100 bacterial genomes have been sequenced. DDH and 16S rRNA gene sequencing are therefore practical methods to define prokaryotic species until genomic sequencing becomes more widely available.

Acknowledgements

This work was supported by BDM0200726, the KRIBB Research Initiative Program, the Environmental Biotechnology National Core Research Center (KOSEF: R15-2003-012-02002-0), the Eco-technopia 21 project and the 21C Frontier Microbial Genomics and Application Center Program. The first author was supported by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD, Basic Research Promotion Fund) (KRF-2007-512-C00016).

References

Bae, J.W., Rhee, S.K., Park, J.R., Chung, W.H., Nam, Y.D., Lee, I., Kim, H., Park, Y.H., 2005. Development and evaluation of genome-probing microarrays for monitoring lactic acid bacteria. Appl. Environ. Microbiol. 71, 8825–8835.
Bendich, A.J., Bolton, E.T., 1967. Relatedness among plants as measured by the DNA-agar technique. Plant Physiol. 42, 959–967.
Bicknell, J.N., Douglas, H.C., 1970. Nucleic acid homologies among species of *Saccharomyces*. J. Bacteriol. 101, 505–512.
Bond, S.B., Howley, P.M., Takemoto, K.K., 1978. Characterization of K virus and its comparison with polyoma virus. J. Virol. 28, 337–343.
Boyd, E.F., Nelson, K., Wang, F.S., Whittam, T.S., Selander, R.K., 1994. Molecular genetic basis of allelic polymorphism in malate dehydrogenase (*mdh*) in natural populations of *Escherichia coli* and *Salmonella enterica*. Proc. Natl. Acad. Sci. 91, 1280–1284.
Christensen, H., Angen, O., Mutters, R., Olsen, J.E., Bisgaard, M., 2000. DNA–DNA hybridization determined in micro-wells using covalent attachment of DNA. Int. J. Syst. Evol. Microbiol. 50, 1095–1102.

Christensen, H., Nordentoft, S., Olsen, J.E., 1998. Phylogenetic relationships of Salmonella based on rRNA sequences. Int. J. Syst. Bacteriol. 48, 605–610.
Coenye, T., Gevers, D., Van de Peer, Y., Vandamme, P., Swings, J., 2005. Towards a prokaryotic genomic taxonomy. FEMS Microbiol. Rev. 29, 147–167.
De Ley, J., De Smedt, J., 1975. Improvements of the membrane filter method for DNA: rRNA hybridization. Antonie Van Leeuwenhoek. 41, 287–307.
Dorris, D.R., Ramakrishnan, R., Trakas, D., Dudzik, F., Belval, R., Zhao, C., Nguyen, A., Domanus, M., Mazumder, A., 2002. A highly reproducible, linear, and automated sample preparation method for DNA microarrays. Genome Res. 12, 976–984.
Dutta, S.K., Sheikh, I., Choppala, J., Aulakh, G.S., Nelson, W.H., 1976. DNA homologies among homothallic, pseudo-homothallic and heterothallic species of Neurospora. Mol. Gen. Genet. 147, 325–330.
Ezaki, T., Hashimoto, Y., Yabuuchi, E., 1989. Fluorometric deoxyribonucleic acid deoxyribonucleic acid hybridization in microdilution wells as an alternative to membrane-filter hybridization in which radioisotopes are used to determine genetic relatedness among bacterial strains. Int. J. Syst. Bacteriol. 39, 224–229.
Forney, L.J., Zhou, X., Brown, C.J., 2004. Molecular microbial ecology: land of the one-eyed king. Curr. Opin. Microbiol. 7, 210–220.
Gevers, D., Cohan, F.M., Lawrence, J.G., Spratt, B.G., Coenye, T., Feil, E.J., Stackebrandt, E., Van de Peer, Y., Vandamme, P., Thompson, F.L., Swings, J., 2005. Opinion: re-evaluating prokaryotic species. Nat. Rev. Microbiol. 3, 733–739.
Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., Tiedje, J.M., 2007. DNA–DNA hybridization values and their relationship to whole-genome sequence similarities. Int. J. Syst. Evol. Microbiol. 57, 81–91.
Jain, A.N., Tokuyasu, T.A., Snijders, A.M., Segraves, R., Albertson, D.G., Pinkel, D., 2002. Fully automatic quantification of microarray image data. Genome Res. 12, 325–332.
Johnson, J., 1973. Use of nucleic-acid homologies in the taxonomy of anaerobic bacteria. Int. J. Syst. Bacteriol. 23, 308–315.
Kafatos, F.C., Jones, C.W., Efstratiadis, A., 1979. Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. Nucleic Acids Res. 7, 1541–1552.
Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.
Klappenbach, J.A., Dunbar, J.M., Schmidt, T.M., 2000. rRNA operon copy number reflects ecological strategies of bacteria. Appl. Environ. Microbiol. 66, 1328–1333.
Kotetishvili, M., Stine, O.C., Kreger, A., Morris Jr., J.G., Sulakvelidze, A., 2002. Multilocus sequence typing for characterization of clinical and environmental salmonella strains. J. Clin. Microbiol. 40, 1626–1635.
Maiden, M.C., 2006. Multilocus sequence typing of bacteria. Annu. Rev. Microbiol. 60, 561–588.
Nelson, K., Selander, R.K., 1992. Evolutionary genetics of the proline permease gene (*putP*) and the control region of the proline utilization operon in populations of *Salmonella* and *Escherichia coli*. J. Bacteriol. 174, 6886–6895.
Nelson, K., Whittam, T.S., Selander, R.K., 1991. Nucleotide polymorphism and evolution in the glyceraldehyde-3-phosphate dehydrogenase gene (*gapA*) in natural populations of *Salmonella* and *Escherichia coli*. Proc. Natl. Acad. Sci. 88, 6667–6671.
Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.
Selander, R.K., McKinney, R.M., Whittam, T.S., Bibb, W.F., Brenner, D.J., Nolte, F.S., Pattison, P.E., 1985. Genetic structure of populations of *Legionella pneumophila*. J. Bacteriol. 163, 1021–1037.
Shendure, J., Mitra, R.D., Varma, C., Church, G.M., 2004. Advanced sequencing technologies: methods and goals. Nat. Rev., Genet. 5, 335–344.
Sibley, C.G., Ahlquist, J.E., 1984. The phylogeny of the hominoid primates, as indicated by DNA–DNA hybridization. J. Mol. Evol. 20, 2–15.
Sneath, P.H.A., 1983. Distortions of taxonomic structure from incomplete data on a restricted set of reference strains. J. Gen. Microbiol. 129, 1045–1073.
Sohn, U., Rothfels, K.H., Straus, N.A., 1975. DNA:DNA hybridization studies in black flies. J. Mol. Evol. 5, 75–85.
Stackebrandt, E., Goebel, B., 1994. Taxonomic note: a place for DNA:DNA reassociation and 16S rRNA sequence analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. 44, 846–849.
Stan-Lotter, H., Pfaffenhuemer, M., Legat, A., Busse, H.J., Radax, C., Gruber, C., 2002. *Halococcus dombrowskii* sp. nov., an archaeal isolate from a Permian alpine salt deposit. Int. J. Syst. Evol. Microbiol. 52, 1807–1814.
Templeton, A.R., 1985. The phylogeny of the hominoid primates: a statistical analysis of the DNA–DNA hybridization data. Mol. Biol. Evol. 2, 420–433.
Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.
Wayne, L., Brenner, D., Colwell, R., Grimont, P., Kandler, O., 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Bacteriol. 37, 463–464.
Yamamoto, S., Harayama, S., 1995. PCR amplification and direct sequencing of *gyrB* genes with universal primers and their application to the detection and taxonomic analysis of *Pseudomonas putida* strains. Appl. Environ. Microbiol. 61, 1104–1109.
Yeates, C., Gillings, M.R., Davison, A.D., Altavilla, N., Veal, D.A., 1998. Methods for microbial DNA extraction from soil for PCR amplification. Biol. Proced. Online 1, 40–47.
Zong, Y., Wang, Y., Zhang, S., Shi, Y., 2003. How to evaluate a microarray scanner. In: Hardiman, G. (Ed.), Microarrays Methods and Applications: Nuts & Bolts. DNA press, LLC, pp. 97–114.