Generation of subspecies level-specific microbial diagnostic microarrays using genes amplified from subtractive suppression hybridization as microarray probes

Jin-Woo Bae, Sung-Keun Rhee, Young-Do Nam and Yong-Ha Park

1Biological Resources Center, Korea Research Institute of Bioscience and Biotechnology (KIRIBB), Eundong 52, Yusong, Daejeon, Korea, 2Department of Microbiology, Chungbuk National University, 12 Gaeshin-dong Heungduk-gu, Cheongju, Korea and 3National Research Laboratory of Molecular Ecosystematics, Institute of Probionics, Probionic Corporation, Korea Research Institute of Bioscience and Biotechnology (KIRIBB), Eundong 52, Yusong, Daejeon, Korea

Received May 15, 2005; Revised June 17, 2005; Accepted June 30, 2005

ABSTRACT
The generation of microarray probes with specificity below the species level is an ongoing challenge, not least because the high-throughput detection of microorganisms would be an efficient means of identifying environmentally relevant microbes. Here, we describe how suppression subtractive hybridization (SSH) can be applied to the production of microarray probes that are useful for microbial differentiation at the subspecies level. SSH was used to initially isolate unique genomic sequences of nine Salmonella strains, and these were validated in quadruplicate by microarray analysis. The results obtained indicate that a large group of genes subtracted by SSH could serve together, as one probe, for detecting a microbial subspecies. Similarly, the whole microbial genome (not subjected to SSH) can be used as a species-specific probe. The detailed methods described herein could be used and adapted for the estimation of any cultivable bacteria from different environments.

INTRODUCTION
Microarrays are powerful tools for the parallel, high-throughput detection and quantification of nucleic acids (1). Currently, DNA microarrays, in particular, are broadly applied across most sectors of the life sciences, including environmental microbiology and microbial ecology and human, veterinary, food and plant diagnostics (2,3). While there is abundant opportunity to use DNA microarray technology to identify microorganisms, several practical limitations have slowed the implementation of this methodology (4). These include low sensitivity and poor resolution at the species and subspecies level. For the potential of DNA microarrays to be realized, these long-term methodological obstacles must be overcome. Because species-level resolution does not necessarily provide sufficient information for diagnosis, there is a need for subspecies-level genetic resolution markers in both clinical and environmental microbiology (1). In this study, sets of genes produced by suppression subtractive hybridization (SSH) were tested as subspecies-specific microarray probes for discriminating several strains that are closely related phylogenetically.

SSH is a widely used method for separating DNA sequences that distinguish two closely related genomic DNA (gDNA) libraries (5–8). A key feature of the method is simultaneous normalization and subtraction steps. The normalization step equalizes the abundance of DNA fragments within the target population, and the subtraction step excludes sequences that are common to the two populations being compared (7). Specific amplification of genes with SSH has allowed identification of minute genomic differences between closely related microbial strains (9–13) and enabled the profiling of genetic diversity in an environmental metagenome (14). Although there have been several studies applying tester-specific genes as microarray probes (15–17), they all used a single clone as an individual probe because their microarray was constructed for analysis of genome-wide differential gene expression. Recently, using nylon membrane-based macroarray hybridization, Li et al. (18) demonstrated that gene fragments produced by SSH could be used as a species-specific probe for the diagnosis of five species of the genus Dendrobrium. While that study provided valuable insights...
into gDNA subtraction between different species, we suggest that their approach may actually be subspecies-specific, given that the gDNA probes were absolutely species-specific in the reverse sample genome probing (RSGP) hybridization (19). In addition, while those authors reported that only 12.16% of SSH clones were species-specific, it is generally recognized that all fragments from SSH are absolutely tester-specific (14). In this study, therefore, SSH-amplified genes that are specific to a microbial strain were not separated into individual fragments by cloning. Rather, the sum of amplified genes was applied together as one microarray probe, which was used to specifically detect the individual strain. The diagnostic specificity of these microarray probes was evaluated by typing nine *Salmonella* type strains that are phylogenetically closely related and by comparing them to gDNA probes. For more accurate and precise analysis, miniaturized microarrays with guaranteed high reproducibility (20) were constructed by depositing DNA probes onto non-porous substrates with printing robots. The underlying rationale of SSH-microarray hybridization is discussed from a phylogenomic perspective.

**MATERIALS AND METHODS**

**Bacterial strains and preparation of gDNA**

Nine *Salmonella* type strains plus an *Escherichia coli* type strain were used in this study, as detailed in Table 1. They were obtained from the Korean Collection for Type Cultures (KCTC), the German Collection of Microorganisms and Cell Cultures (DSMZ) and the American Type Culture Collection (ATCC). All strains were maintained and grown under conditions suggested by the collection from which they were sourced. Cells at the exponential growth phase were quickly harvested and frozen at −80°C for the extraction of DNA. The gDNAs were isolated using a bead-beating method, as previously described (21). All DNA samples were treated with RNase A (Sigma, St Louis, MO) and were analyzed on agarose gels stained with ethidium bromide prior to SSH, microarray fabrication and hybridization. Concentrations of the obtained DNAs were determined in triplicate using a spectrophotometer (Nanodrop Technologies, Rockland, DE).

### Table 1. Microorganisms whose genomes were used as microarray probes

<table>
<thead>
<tr>
<th>Location in microarraya</th>
<th>Target microorganism used as a SSH tester</th>
<th>Genus and species name</th>
<th>Subspecies name</th>
<th>Culture collection number</th>
<th>SSH driver</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td><em>S</em>. <em>choleraeuis</em></td>
<td><em>indica</em></td>
<td></td>
<td>DSM 14848</td>
<td><em>S</em>. <em>choleraeuis</em> subsp. <em>choleraeuis</em></td>
</tr>
<tr>
<td>B</td>
<td><em>S</em>. <em>choleraeuis</em></td>
<td><em>choleraeuis</em></td>
<td></td>
<td>DSM 14846</td>
<td><em>S</em>. <em>typhimurium</em></td>
</tr>
<tr>
<td>C</td>
<td><em>S</em>. <em>bongori</em></td>
<td></td>
<td></td>
<td>DSM 13772</td>
<td><em>S</em>. <em>choleraeuis</em> subsp. <em>choleraeuis</em></td>
</tr>
<tr>
<td>D</td>
<td><em>S</em>. <em>choleraeuis</em></td>
<td><em>arizonae</em></td>
<td></td>
<td>DSM 9386</td>
<td><em>S</em>. <em>choleraeuis</em> subsp. <em>choleraeuis</em></td>
</tr>
<tr>
<td>E</td>
<td><em>S</em>. <em>choleraeuis</em></td>
<td><em>hountenae</em></td>
<td></td>
<td>DSM 9221</td>
<td><em>S</em>. <em>choleraeuis</em> subsp. <em>choleraeuis</em></td>
</tr>
<tr>
<td>F</td>
<td><em>S</em>. <em>choleraeuis</em></td>
<td><em>salamae</em></td>
<td></td>
<td>DSM 9220</td>
<td><em>S</em>. <em>choleraeuis</em> subsp. <em>choleraeuis</em></td>
</tr>
<tr>
<td>G</td>
<td><em>S</em>. <em>typhimurium</em></td>
<td></td>
<td></td>
<td>ATCC 13311</td>
<td><em>S</em>. <em>choleraeuis</em> subsp. <em>choleraeuis</em></td>
</tr>
<tr>
<td>H</td>
<td><em>S</em>. <em>enteritidis</em></td>
<td></td>
<td></td>
<td>ATCC 13076</td>
<td><em>S</em>. <em>choleraeuis</em> subsp. <em>choleraeuis</em></td>
</tr>
<tr>
<td>I</td>
<td><em>S</em>. <em>choleraeuis</em></td>
<td><em>diarizonae</em></td>
<td></td>
<td>DSM 14847</td>
<td><em>S</em>. <em>choleraeuis</em> subsp. <em>choleraeuis</em></td>
</tr>
<tr>
<td>b</td>
<td><em>E</em>. <em>coli</em></td>
<td></td>
<td></td>
<td>KCTC 2441</td>
<td><em>S</em>. <em>choleraeuis</em> subsp. <em>choleraeuis</em></td>
</tr>
</tbody>
</table>

*a* All gene probes were printed in quadruplicate. Microbial gDNAs were printed in the first row, and SSH probes in the second row (Figure 3).

bThe gDNA and SSH probes for *E*. *coli* were not printed together in the same section as those for *Salmonella*, in the interests of clarity.

### Suppressing subtractive hybridization

SSH was used to isolate DNA fragments present in the target microbial organisms but absent from the reference strains. The procedure was performed using the PCR-Select Bacterial Genome Subtraction Kit (Clontech), with minor modifications. *Salmonella choleraeuis* subsp. *choleraeuis* (Type species) was assigned as the SSH driver while the other eight strains were assigned as SSH testers. For SSH of *S*. *choleraeuis* subsp. *choleraeuis*, *Salmonella typhimurium*, which has the most similar 16S rRNA sequence, was assigned as a driver. Tester and driver samples were digested separately with RsAl. Adapter ligations using the Clontech kit were performed as described in the user manual. Ligation efficiency was analyzed also as described in the manual with the bacterial 1088r primer (22). PCR amplification of tester-specific fragments was performed using primers directed to tester ligated adaptor sequences. The AccuPower® PCR PreMix (Bioneer, Korea) and a Peltier Thermal Cycler (DNA Engine DYAD™, MJ Research) were used for both primary and secondary PCR amplifications. Both PCR amplifications were performed using the primers and recommended primer and template concentrations from the Clontech kit, adjusted to a total volume of 50 μl. The primary PCR cycling conditions were as follows: one incubation of 94°C for 5 min followed by 30 cycles of 94°C for 10 s, 66°C for 30 s and 72°C for 1.5 min. Secondary PCR cycling involved a similar program, but used an annealing temperature of 68°C. Five replicates of each secondary PCR product (250 μl) were purified using the AccuPrep® PCR Purification Kit (Bioneer, Korea).

**Microarray construction, labeling and hybridization**

Microarray construction, post-treatment, labeling and hybridization were performed as described previously (23) with minor modifications. The gDNAs and genes specifically amplified by SSH were diluted to a final concentration of 400 ng/μl in 0.1× TE. Five microliters aliquots of each probe 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. Each probe set was printed in quadruplicate. The exact location of each gDNA on the glass slide is listed in Table 1. After post-treatments
including cross-linking, blocking, heat-denaturing (immersing in D_{2}O at 95°C for 2 min), rinsing in 95% ethanol and air-drying, microarrays were stored dry in a clean slide box at room temperature. DNA was labeled with the BioPrime DNA Labeling System and 2.5 mM Cy5 dCTP (Amersham Pharmacia Biotech, Piscataway, NJ) by incubating at 37°C for 3 h. The labeled target DNA was purified using a QIAquick PCR purification column (Qiagen, Valencia, CA), concentrated in a Speedvac for 1 h, and resuspended in 4.35 µl of D_{2}O for hybridization. All microarray hybridizations were performed in triplicate (a total of 12 replicates per DNA probe), unless otherwise noted, to facilitate statistical analyses. After hybridization, each microarray slide was taken out and its cover slip was immediately removed under wash solution 1 (1x SSC and 0.2% SDS). Slides were washed using wash solution 1, wash solution 2 (0.1x SSC and 0.2% SDS), and wash solution 3 (0.1x SSC) for 5 min each, at ambient temperature prior to drying. The slide was dried by centrifugation, as described above.

Microarray scanning and data analysis

A GenePix® 4000B microarray scanner set (Axon instruments, Union City, CA) was used for scanning the microarrays at a resolution of 10 µm. For consistent scanning of all hybridized slides, the laser power and photomultiplier tube (PMT) gain were adjusted to 100%. Scanned image displays were analyzed by quantifying the pixel density (intensity) of each hybridization spot using the GenePix® software program (version 4.0; Axon instruments). Hybridization images presented here are representative images, which were automatically contrast-adjusted by the software. A grid of individual circles defining the location of each DNA spot on the array was superimposed onto the image to indicate each quantified fluorescent spot. Mean signal intensity was automatically determined for each spot. 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 (24):

\[ SNR = \frac{\text{signal intensity}}{\text{background}} \]

where the ‘background’ measurement refers to the local spot background intensity and the ‘SD of background’ was calculated across all pixels, as measured by the GenePix® software. Statistical analysis was performed using Excel 2003 (Microsoft) and Sigmaplot 8.0 (Jandel Scientific, San Rafael, CA). The SNRs from 12 replicate data sets were then averaged to obtain the SNR value of each individual probe. Relative SNRs were obtained by dividing the average SNR value by the mean value of the hybridized template SNR in the same microarray slide.

Phylogenetic analysis of the genus Salmonella

16S rDNA isolated from each Salmonella strain was amplified by PCR using two universal primers as described by Yoon et al. (25) and sequenced as described by Yoon et al. (26). Multiple alignments were performed using the Clustal X program (27). The evolutionary distances were calculated using the Jukes & Cantor method. The phylogenetic tree was constructed using a neighbor-joining method (28) in the MEGA 2 program (29). To compare similarities, a 16S rRNA homology matrix was constructed using PairProWin.exe (http://microarray.kaist.ac.kr) software, which is based on the Clustal W program.

RESULTS

Phylogenetic analysis on the genus Salmonella

Currently, the valid names of species and subspecies in the genus Salmonella are as follows: S.choleraesuis (subsp. choleraesuis, salamae, arizonae, diarisonae, hountaeae, and indica), S.enteritidis, S.typhimurium, Salmonella typhi and Salmonella bongori (30). Although Salmonella enteritidis, S.typhimurium and S.typhi are almost genetically identical to S.choleraesuis subsp. choleraesuis (31–33), these species continue to appear in the approved lists and retain their unique nomenclature because they are now very important human pathogens. In this study, the type strains of S.enteritidis, S.typhimurium, S.bongori and six subspecies (serovars) of S.choleraesuis were selected to determine whether SSH-generated microarray probes are specific at a subspecies level. To confirm the identity of these strains and to preliminarily investigate phylogenetic relationships within the genus Salmonella, 16S rDNA was sequenced for each strain, subjected to BLAST searching (34) and inter-strain comparisons were conducted. This 16S rDNA gene-based phylogeny revealed that the Salmonella type strains used in this study are very closely related with all species and subspecies, except S.bongori, sharing >98% 16S rDNA sequence similarity. S.bongori had 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, had 96.7–97.9% similarity with Salmonella. The 16S rDNA distances between the Salmonella strains used in this study are plotted in Figure 1.

SSH of gDNA from Salmonella species and subspecies

SSH was carried out between the type strains of the nine Salmonella species and subspecies to determine the usefulness of this technique for isolating DNA unique to a Salmonella subspecies. S.choleraesuis subsp. choleraesuis (type species of the genus) was assigned as the SSH driver while the other eight strains were assigned as SSH testers. Based on 16S rDNA homology comparison, the nearest phylogenetic neighbor, S.typhimurium, was assigned as the driver for S.choleraesuis subsp. choleraesuis (Figure 1). Most tester and driver samples were between 100 and 1500 bp in size when digested with Rsal (Figure 2A). Analysis of ligation efficiency, which was conducted by PCR amplification with the bacterial 1088r primer, produced a dim band of ~280 bp in size. This corresponded to the fragment size predicted on the basis of the priming and Rsal-cutting sites in the Salmonella 16S rRNA sequence (Figure 2B). After twice hybridization between testers and drivers, primary PCR and secondary nested PCR of tester-specific fragments were executed. Although the apparent intensity and size of ethidium bromide-stained primary products were similar for all tester species/subspecies, the size of secondary nested PCR products varied significantly between testers (Figure 2C and D). Concentrations of the purified secondary PCR products (50 µl) were in the range 62–119 ng/µl.
These products were further concentrated and used as probes for SSH microarrays.

Specificity of SSH microarrays

To obtain proof of principle, the specificity of SSH microarrays was evaluated by hybridizing them with the gDNA of each Salmonella type strain. As predicted, strong signals were only obtained for the probes corresponding to the labeled target (lower lines in Figure 3). Little or no cross-hybridization (0–4%) of SNRs was observed for non-target species and subspecies, indicating that, at least under the conditions applied here, strain-specific hybridization is achieved with SSH microarrays. In order to compare the specificity of SSH probes to intact genome probes, gDNA extracted from each genus Salmonella strain was also spotted on the same slide (upper lines in Figure 3). In the Salmonella gDNA hybridization, cross-hybridizations ranging from 1 to 92% of SNRs were observed between the target subspecies and its phylogenetic neighbor subspecies. Different species such as S.bongori and E.coli were clearly distinguished by genomic hybridization whereas almost identical species, such as S.typhimurium and S.enteritidis, could not be differentiated from S.choleraesuis subsp. choleraesuis (31). This indicated that strain or subspecies differentiation could not be achieved in gDNA-based hybridization, in contrast to the SSH probes, all of which were perfectly specific for their corresponding targets. Regarding cross-hybridization of gDNA probes, there was a clear trend for higher cross-hybridization SNR values among subspecies with close 16S rDNA homology (Figures 1 and 4). This strongly suggests that genomic hybridization on a microarray slide is very useful for estimating bacterial similarity, below the species level, for closely related strains.

Since real environments are generally composed of a variety of different microorganisms, evaluation of SSH probes with mixed genomes may enhance the applicability of SSH microarrays. To determine whether other non-target DNA interferes with the specificity of SSH microarray-based hybridization, three different gDNAs (500 ng per species) from S.choleraesuis subsp. indica, hountenae and diarizonae were mixed, labeled, and hybridized to the SSH microarray (J in Figure 3). Specific hybridization signals were observed for the corresponding probes, with these exhibiting similar SNRs of 286.3, 292.8 and 309.4, respectively. On the contrary, the non-target probes had SNRs below the background level. These results suggest that SSH probes could also be specific in the presence of other genomes and strongly validate the hypothesis that the subtracted genome only contains tester-specific genes.

DISCUSSION

Salmonella is one of the major bacterial agents that cause foodborne infections in humans worldwide (35). Members of the genus Salmonella have been divided into >2300 serovars (32) and their classification has provoked many arguments between microbiologists (30). The extraordinarily close inter-genomic relationships that exist between Salmonella species present a major problem for prohibiting their phylogeny from being unshakable (36). Since the major epidemiological markers of Salmonella strains are subspecies and serovars, which are determined by chromosomal genes and are not affected by extrachromosomal elements (37), high-throughput detection of Salmonella below the species level is crucial for studying human infections in detail. Existing methods for the molecular typing of Salmonella include plasmid typing, pulsed field gel electrophoresis, ribotyping and random-amplified polymorphic DNA analyses (38). However, the ever-increasing availability of genome sequences for these microorganisms implies that genotypic methods that are amenable to automation will be developed in time. In this...
regard it is plausible that DNA microarrays will be developed with specific probes for thousands of different microbial species or strains, enabling fast and reliable identification of microorganisms (39). Most previous Salmonella microarrays have been manufactured for the purpose of studying genetic relationships between Salmonella and/or other close relatives (40). Although a few 16S rRNA gene-based microarrays have been developed for high-throughput diagnosis of bacterial
pathogens including Salmonella (41,42), so far these have exhibited poor resolution at the species level as well as problems in detection sensitivity (43,44).

In this study, we differentiated seven subspecies of Salmonella using SSH-generated microarray probes. Their ability to distinguish S.enteritidis and S.typhimurium from the almost genetically identical organism, S.choleraesuis subsp. choleraesuis (and vice versa), indicated that SSH probes are strain-specific. Importantly, this specificity was also maintained in a mixed genome hybridization test. These results indicate that the generation and application of microarray probes with SSH may be useful for monitoring various microorganisms in real environments. Although Li et al. (18) reported that only 12.16% of SSH clones were species-specific, all the SSH fragments in this study were found to be absolutely tester-specific since no cross-hybridization was observed, even between almost identical strains. The high degree of specificity of SSH probes was not surprising, given that gDNA probes were also absolutely species-specific—an observation consistent with previous findings for RSGP hybridization (19).

Subtracting the tester genome with a driver genome is not an ideal approach for the construction of strain-specific probes, since it is possible that the remaining tester strain gDNA could still contain the genes similar to those of a third unknown organism. The potential paucity of SSH microarrays could be addressed by selecting the most closely related strain as a SSH driver for each test strain and using a nested probe approach such that probes of different taxonomic levels, such as gDNA, are combined to increase the confidence of the assay.

The use of microbial diagnostic microarrays is expected to accelerate our understanding of previously unrecognized complex microbial processes (2). However, high resolution microarray analysis techniques must also be developed on a high-throughput basis. This will provide a significant advance over current techniques since for many organisms, even

**Figure 3.** Fluorescence images showing hybridization specificity of gDNA and SSH probes. All gene probes were printed in quadruplicate (divided with white dotted lines). Microbial gDNAs were printed in the first row (gDNA), and SSH probes in the second row (SSH). Location of probes in the microarray is shown in Table 1. Each microarray was hybridized with 500 ng of Salmonella gDNA (see yellow indications on each image and Table 1), except the 3 mix. The microarray indicated as J (yellow indication) was hybridized with 500 ng of equally mixed gDNAs of S.choleraesuis subsp. indica (A), S.choleraesuis subsp. hountenae (E) and S.choleraesuis subsp. diarizonae (I).
species-level differentiation is difficult to achieve (1). The specificity of SSH probes observed in this study could clearly solve the existing methodological problems associated with microarrays. The ability to conduct such high resolution analyses has important implications across diverse areas including the monitoring of inoculated strains in bioreactors or the lactic acid bacteria used as a probiotics in gastrointestinal microflora, as well as the diagnosis of closely related microorganisms such as *Salmonella*.

**ACKNOWLEDGEMENTS**

This work was supported by grant BDM0200413, grant NNM0100411 and the NRL research programme (grant M1014000294-01J000012800) from the Korean Ministry of Science and Technology (MOST). Funding to pay the Open Access publication charges for this article was provided by KRIBB Research Initiative Program.

**Conflict of interest statement.** None declared.

**REFERENCES**


**Figure 4.** Histogram plots showing relative SNRs of the hybridization specificity images (Figure 3) of gDNA (A) and SSH probes (B). The probe sources of gDNA and SSH tester were indicated in the left of the figure and hybridized templates were shown below the grouped bars.


