J. Microbiol. Biotechnol. (2007), **17**(6), 945–951

JOURNAL OF MICROBIOLOGY AND BIOTECHNOLOGY

Relationship Between Genome Similarity and DNA-DNA Hybridization Among Closely Related Bacteria

KANG, CHEOL-HEE[1], YOUNG-DO NAM[2], WON-HYONG CHUNG[3], ZHE-XUE QUAN[2,4], YONG-HA PARK[2], SOO-JE PARK[1], RACHEAL DESMONE[5], XIU-FENG WAN[5], AND SUNG-KEUN RHEE[1]*

[1]*Department of Microbiology and Biotechnology Research Institute, Chungbuk National University, Cheongju 361-763, Korea*
[2]*Biological Resources Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon 305-333, Korea*
[3]*National Genome Information Center, Korea Research Institute of Bioscience and Biotechnology (KRIBB), Daejeon 305-333, Korea*
[4]*Department of Microbiology and Microbial Engineering, School of Life Sciences, Fudan University, Shanghai, China*
[5]*Department of Microbiology, Miami University, Oxford, OH 45056, U.S.A.*

Received: December 15, 2006
Accepted: January 25, 2007

Abstract DNA-DNA hybridization has been established as an important technology in bacterial species taxonomy and phylogenetic analysis. In this study, we analyzed how the efficiency with which the genomic DNA from one species hybridizes to the genomic DNA of another species (DNA-DNA hybridization) in microarray analysis relates to the similarity between two genomes. We found that the predicted DNA-DNA hybridization based on genome sequence similarity correlated well with the experimentally determined microarray hybridization. Between closely related strains, significant numbers of highly divergent genes (<55% identity) and/or the accumulation of mismatches between conserved genes lowered the DNA-DNA hybridization signal, and this reduced the hybridization signals to below 70% for even bacterial strains with over 97% 16S rRNA gene identity. In addition, our results also suggest that a DNA-DNA hybridization signal intensity of over 40% indicates that two genomes at least shared 30% conserved genes (>60% gene identity). This study may expand our knowledge of DNA-DNA hybridization based on genomic sequence similarity comparison and further provide insights for bacterial phylogeny analyses.

Keywords: Genome similarity, predicted DNA-DNA hybridization signal, DNA microarray

The phylogenetic definition of a species relies on DNA-DNA hybridization data along with several other features, including chemotaxonomic and physiological data. DNA-DNA hybridization is regarded as a reliable reference method to decide whether a strain represents a new species in combination with other parameters [28], since it employs the entire genome for comparison. DNA-DNA hybridization has also been criticized, however, for its lack of robustness, as it has been found that the DNA-DNA hybridization data obtained in different laboratories may not correlate well, partly because of the use of different hybridization methods [8] and conditions [16, 20–22, 24, 26, 27]. Thus, a more rapid and reliable DNA-DNA hybridization method is in demand for genomic similarity measurement.

DNA microarray analysis is a recently emerged technology, frequently used to explore genome-wide transcriptional profiles. Its applications have recently extended into the fields of environmental microbiology and microbial ecology [34]. For example, in recent studies, genome-probing microarrays consisting of an individual microbial genome immobilized to a particular spot were used to specifically identify the genomes present in environmental samples [1, 30]. The genome-probing microarray technology was shown to be a reliable DNA-DNA hybridization approach to determine species similarity [4, 5].

Currently, bacterial species were determined based on a 70% hybridization signal ratio [14, 22], which is rather arbitrarily selected, although the phylogenetic definition of a species relies on several other features, including various morphological, physiological, and chemotaxonomic data. How to define taxa using DNA-DNA hybridization technology is a nontrivial problem. A fundamental challenge is to select the cutoff value of genome similarity for a new taxon. In order to solve this problem, in this study, we researched the potential factors affecting genome similarity and DNA-DNA hybridization using microarray. Our study may shed some light on the principles underlying DNA-

*Corresponding author
Phone: 82-43-261-2300; Fax: 82-43-264-9600;
E-mail: rhees@chungbuk.ac.kr

DNA hybridization in genomic scale, which underlines the validity and utility of this method to define bacterial species.

MATERIALS AND METHODS

Calculation of Pairwise Gene Identity Between Closely Related Strains

The genomic sequences in the microbial genome database (http://cmr.tigr.org/) were analyzed to select appropriate genome pairs for the DNA-DNA hybridization experiments. All sequences have been downloaded from "Batch Download Query Page" for those organisms: *Bacillus cereus* 10987, *Bacillus cereus* ATCC 14579, *Bacillus subtilis* 168, *Escherichia coli* K12-MG1655, *Shigella flexneri* 2a 2457T, and *Salmonella typhimurium* LT2 SGSC1412. To compare genome similarity, the pairwise gene identities on a genomic scale were calculated between each pair of genomes. To calculate pairwise gene identity, a specific gene from one strain (the query gene) was used to BLAST-search all of the genes in the other targeted strain. The highest scored hit based on e-value then served together with the query gene to constitute the gene-pair sharing the highest identity. Pairwise gene identity was calculated using the following fomula:

$$\text{Pairwise gene identity} = 100 \times (L_{align} - N_{mis} - N_{gap}) / L_{ShortSequence}$$

Pairwise gene identity, expressed as a percentage, is the pairwise identity between genes. L_{align} is the aligned length of the sequence of the query and the best-hit gene, N_{mis} is the number of nucleotides in the alignment that are mismatched, N_{gap} is the gap size, and $L_{ShortSequence}$ is the gene size of the smaller gene between the query gene and subjected gene.

DNA-DNA Hybridization Based on Microarray

All strains used in this hybridization study were previously analyzed for pairwise genomic sequence comparison (see above). These strains were obtained from the American Type Culture Collection (ATCC) and the German Collection of Microorganisms and Cell Cultures (DSMZ): *Bacillus cereus* (ATCC 10987), *Bacillus cereus* (ATCC 14579), *Bacillus subtilis* 168 (DSM 401), *Escherichia coli* K12 (DSM 5911), *Shigella flexneri* 2a (DSM 4782), and *Salmonella typhimurium* LT2 (ATCC 19585). Each strain was grown under the conditions recommended by the culture collections centers. Bacterial cells at the exponential phase were quickly harvested and frozen at −80°C prior to DNA extraction [33]. The microarray fabrication and hybridization protocols used followed those described by Bae *et al.* [1] Kim *et al.* [11], and Oh *et al.* [15]. The genomic DNAs were diluted to a final concentration of 400 ng/μl in 0.1× Tris-EDTA buffer and mixed with an equal volume of 2× microarray spotting solution (ArrayIt; Telechem International, Inc., Sunnyvale, CA, U.S.A.) for printing. The probes were arrayed onto 25- by 75-mm Superamine glass slides (Telechem) with one pin using a PixSys 5500 printer (Cartesian Technologies, Inc., Irvine, CA, U.S.A.) as triplicates at 55 to 58% relative humidity. The slides were cross-linked by exposure to 120 mJ of UV irradiation (UV Stratalinker 1800; Stratagene, La Jolla, CA, U.S.A.). The probe DNA was denatured by immersion of the slides in deionized water at 95°C for 2 min and then rinsed briefly in 95% ethanol, air dried at room temperature, and stored dry in a clean slide box at room temperature.

In order to label the genomic DNA, the BioPrime DNA Labeling System was modified as follows [9, 10]: 15 μl of various concentrations of DNA was mixed with 20 μl of 2.5× Random Primers solution in the kit and was then denatured by boiling for 2 min and immediately chilled on ice. The denatured genomic DNA solution was then mixed with 15 μl of a labeling reaction solution containing 5 mM dATP, 5 mM dTTP, 5 mM dGTP, 2.5 mM dCTP (New England Biolabs, Beverly, MA, U.S.A.), 2.5 mM Cy5 dUTP (Amersham Pharmacia Biotech, Piscataway, NJ, U.S.A.), and 40 U of Klenow fragment (Invitrogen, Carlsbad, CA, U.S.A.). The reaction mixture was incubated at 37°C for 3 h. The labeled target DNA was purified using a QIAQuick PCR purification column (QIAGEN, Valencia, CA, U.S.A.), concentrated in a Speedvac for 1 h, and resuspended in 4.35 μl of deionized water for hybridization.

The microarray hybridization solution contained 4.35 μl of labeled DNA, 8.75 μl of formamide (50%, vol/vol), 3× SSC (1× SSC is 150 mM NaCl and 15 mM trisodium citrate), 1.25 μg of unlabeled herring sperm DNA (Promega, Madison, WI, U.S.A.), and 0.3% sodium dodecyl sulfate (SDS) in a total volume of 17.5 μl. A reduced volume (7.5 μl) of the hybridization mixture was deposited directly onto the slides and covered with a coverslip (10 by 15 mm; Sigma). Fifteen μl of 3× SSC was dispensed into the hydration wells on either side of the hybridization chambers (Corning, Inc., Corning, NY, U.S.A.). The microarray slide was placed into a hybridization chamber, boiled for 5 min to denature the hybridization solution, and immediately plunged into the water bath at 37°C for 12 h. After hybridization, each microarray slide was taken out, and the coverslip was immediately removed in wash solution 1 (1× SSC and 0.2% SDS) at room temperature. Slides were washed using wash solution 1, wash solution 2 (0.1× SSC and 0.2% SDS), and wash solution 3 (0.1× SSC) for 5 min each at room temperature prior to drying. The slides were dried by centrifugation.

After hybridization, a GenePix 4000A microarray scanner set (Axon instruments, Union City, CA, U.S.A.) was used to scan the slides at a resolution of 10 μm, and the pixel density (intensity) of each spot was quantified by using the GenePix version 4.0 software (Axon Instruments, Union City, CA, U.S.A.). The signal-to-noise ratio (SNR) of the

probe for each spot was calculated by the following formula [13]:

$$SNR=[I_p-(I_N-I_{NLB})]/I_{PLB}$$

where I_p is the mean pixel intensity of specific probe spots, I_N is the mean pixel intensity of the nonsense probe spots, I_{NLB} is the mean pixel intensity of the local background area around the nonsense probe spots, and I_{PLB} is the mean pixel intensity of the local background area around the specific probe spots, as measured by the GenePix software. Genomic DNA of *Lactobacillus brevis* ATCC 14869 served as the negative control probe. The signal obtained from this negative probe was regarded as the nonsense probe signal. In each experiment, the signal obtained from hybridization using the genome used for a target spot as probe served as the positive control and was used for normalization as 100%. The relative signal value between these negative and positive signal values were calculated as the DNA-DNA hybridization signal. The SNRs from 12 replicate data sets were then averaged to yield the mean SNR of a particular probe.

Calculation of Predicted DNA-DNA Hybridization Signal (PHS)

To assess the correlation between the genome similarity and the DNA-DNA hybridization signals, we calculated the PHS, which represents the predicted hybridization signal intensity (%) of all gene pairs as follows:

$$PHS=\frac{\sum_{i=1}^{n} Phs_i}{N}$$

PHS is the predicted genome hybridization signal between genomes. Phs_i is the predicted gene hybridization signal for the i^{th} gene pair, and N is the total number of gene pairs. The predicted hybridization signal intensity of each gene pair, Phs_i, was calculated on the basis of both its pairwise gene identity and the expected cDNA cross-hybridization values for given pairwise identities; the latter values were determined by using one of two different models of the relationship between the hybridization signal and gene sequence identity. These models are those described by Wu *et al.* [29], Girke *et al.* [6], and Xu *et al.* [32]. Intergenic regions that comprised about 13.4±4.2% of the genomes were excluded from this calculation.

To compare the relationship between PHS with the experimental DNA-DNA hybridization signal (EHS), the average of relative error (ARE) of the four genome pairs was calculated.

$$ARE=\frac{\sum_{i=1}^{n} \frac{|PHSi-EHSi|}{PHSi} \times 100}{N}$$

PHS_i is the predicted hybridization signal for the i^{th} genome pair, EHS_i is the experimental DNA-DNA hybridization signal for the i^{th} genome pair, and N is the total number of genome pairs.

RESULTS AND DISCUSSION

Screening of Closely Related Strains with High Genome Similarities

Closely related strains with high genome similarities were applied to study the correlation between DNA-DNA hybridization and genome similarity. The genomic sequences in the microbial genome database were analyzed to select appropriate genome pairs for the DNA-DNA hybridization experiments and genome similarity calculations. After calculating the pairwise gene identities between each pair of genomes, we prepared a histogram of the distribution of all the pairwise gene identities using 5% intervals. This

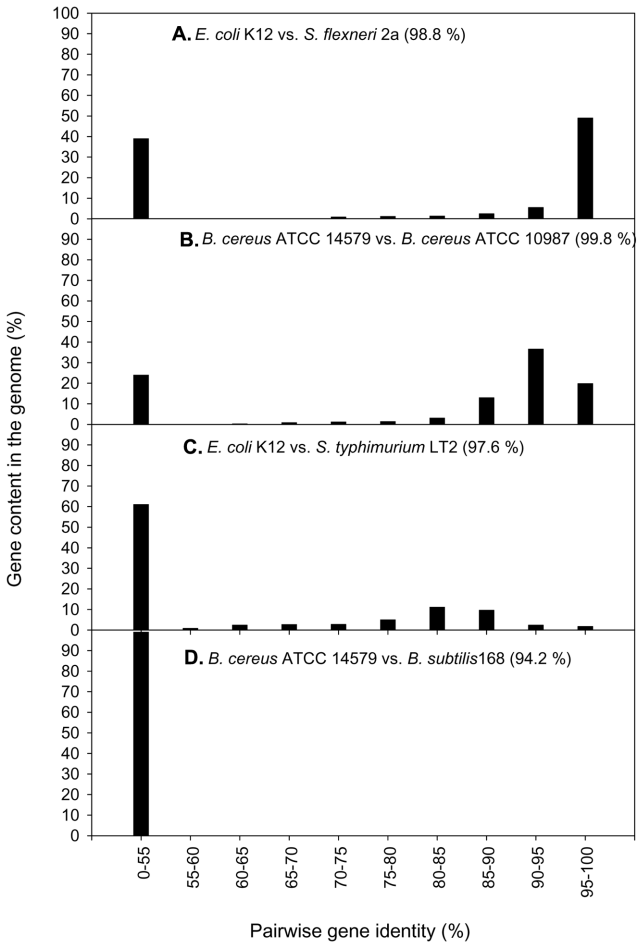

Fig. 1. Distribution patterns of the pairwise gene identities between different genome pairs.

The pairwise gene identity in each genome pair was calculated and the percentage of gene pairs that share the indicated degree of identity are shown. The numbers in parentheses in each panel indicates the 16S rRNA gene identity of the compared genomes.

yielded three distinct patterns (Fig. 1). Gene pairs with less than 55% identity were collectively described, since those contained few genes of the cluster of orthologous group (COG). In pattern 1, most of the genes had either more than 95% or less than 55% pairwise gene identity (Fig. 1A). In pattern 2, significant numbers of the genes had between 95% and 55% pairwise gene identity (Figs. 1B, 1C). In pattern 3, most of the gene pairs shared less than 55% pairwise gene identity (Fig. 1D). Thus, these three patterns can be ranked, in terms of the degree of genome similarity they represent, as pattern 1>pattern 2>pattern 3.

Through such overall analyses of these genomes, a few groups of closely related genome pairs in human pathogenic microorganisms, including *Bacillus cereus*-related and *Escherichia coli*-related organisms, were identified, with all of these three distinct patterns indicated in Fig. 1. The genome pairs from the *Bacillus cereus*-related and *Escherichia coli*-related groups were selected for microarray hybridization. Specifically, the *Bacillus cereus*-related group contained Gram-positive *Bacillus cereus* ATCC 10987, *Bacillus cereus* ATCC 14579, and *Bacillus subtilis* 168; the *Escherichia coli*-related group included Gram-negative *Escherichia coli* K12, *Shigella flexneri* 2a, and *Salmonella typhimurium* LT2.

Microarray-Based DNA-DNA Hybridization Analysis

We were able to detect genomic DNA in the range between 2.5 and 1,000 ng by using our hybridization conditions. In this range, the hybridization signal was dynamic and quantifiable. Thus, 500 ng of each target genomic DNA described above was subjected to Cy5 labeling for the following experiments. As shown in Table 1 and Fig. 2, strong signals were observed only when the target genomic DNA was hybridized to itself or to its closely related strain. Notably, the hybridization signals obtained when *S. flexneri* 2a was hybridized to *E. coli* K12 or when *B. cereus* ATCC 14579 was hybridized to ATCC 10987 were only about 70% (where 100% corresponds to the hybridization observed when the probe genome is hybridized to itself) (Table 1). The strain pairs *S. flexneri* 2a and *E. coli* K12 and the strain pairs *B. cereus* ATCC 14579 and ATCC 10987 share 98.8% and 99.8% 16S rRNA gene identity, respectively. These genome pairs contain a significant amount of distantly related genes in addition to highly conserved genes (see Fig. 1A).

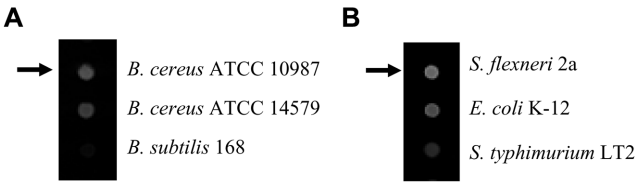

Fig. 2. Fluorescence images showing microarray-based DNA-DNA hybridization.
The probe genomic DNAs of *B. cereus* ATCC 10987 (Panel A) and *S. flexneri* 2a (Panel B) were labeled with Cy5 and then permitted to hybridize with the indicated genomic DNAs. The arrow indicates where the probe DNA was hybridized to its own genomic DNA.

On the other hand, our results showed that if the 16S rRNA gene identity is less than 95%, no significant signals were detected, which was shown when we hybridized the *B. cereus* ATCC 10987 strain with *B. subtilis* 168 (Fig. 2 and Table 1). Brenner *et al.* [2] and Crosa *et al.* [3] have obtained similar DNA-DNA hybridization signals for the *S. typhimurium* LT2 and *E. coli* K12 pair and for the *S. flexneri* 2a and *E. coli* K12 pair by using a traditional hybridization method (see Table 1).

Comparison Between Experimental DNA-DNA Hybridization Signal (EHS) and the Predicted DNA-DNA Hybridization Signal (PHS)

To assess the correlation between the genome similarity and the DNA-DNA hybridization signals that were obtained upon microarray analysis, we compared the EHS and the

Table 1. Reciprocal DNA-DNA hybridization values.

Strain	DNA-DNA hybridization (%) with					
	B. cereus ATCC 14579	*B. cereus* ATCC 10987	*B. subtilis* 168	*E. coli* K12	*S. flexneri* 2a	*S. typhimurium* LT2
1. *B. cereus* ATCC 14579	100	70.4±10.6				
2. *B. cereus* ATCC 10987	73.5±7.6	100				
3. *B. subtilis* 168			100			
4. *E. coli* K12				100	67.3±16.0 (84)	45.3±2.6 (45)[b]
5. *S. flexneri* 2a				71.9±8.0	100	45.5±4.5
6. *S. typhimurium* LT2				41.6±4.6 (46)[a]	42.2±6.0 (39)[a]	100

Hybridization signals of less than 5% are not indicated.
Numbers in parentheses are the DNA-DNA hybridization values of the indicated pairs that have been determined previously by DNA reassociation experiments using hydroxyapatite: [a]Crosa *et al* [3]; [b]Brenner *et al* [2]. DNA-DNA hybridization values of the strains of *B. cereus* used in this study have not been determined previously.
The 16S rRNA gene identity between the indicated strain pairs is as follows: 99.8%, *B. cereus* ATCC 14579 and *B. cereus* ATCC 10987; 97.6%, *E. coli* K12 and *S. typhimurium* LT2; 98.8%, *E. coli* K12 and *S. flexneri* 2a; 96.9%, *S. flexneri* 2a and *S. typhimurium* LT2.

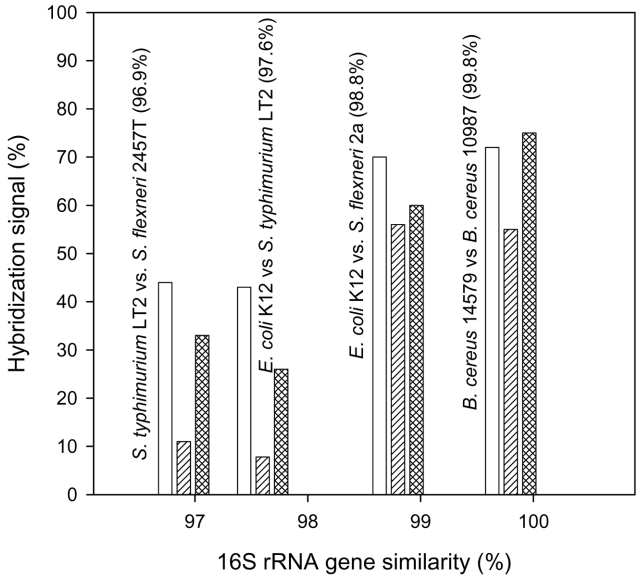

Fig. 3. Relationship between 16S rRNA gene identity, DNA-DNA hybridization, and the PHS.
For the genome pairs indicated in the vicinity of each set of bars, the DNA-DNA hybridization (shown by an open bar), the PHS predicted by the model of Xu *et al.* [32], and Girke *et al.* [6] (shown by the diagonally striped bar), and the PHS predicted by the model of Wu *et al.* [29] (shown by the crosshatched bar) were plotted against the 16S rRNA gene identity. Shown are the averages of reciprocal experiments or calculations.

PHS. Fig. 3 shows comparison data of PHS and EHS of four genome pairs. Of the two models used to calculate Phs_i, the model from Wu *et al.* [29] obtained a lower ARE (20%) than the model from Girke *et al.* [6], and Xu *et al.* [32], which had an ARE of 50%. However, even when the model of Wu *et al.* [29] is used, the relative error between the PHS and the EHS increases as genome similarity decreases. For example, the relative error is only 4% when the *B. cereus* strains are compared, whereas the relative error rises to 25% when *S. typhimurium* LT2 and *S. flexneri* 2a are compared. This indicates that cross-hybridization occurs at a significantly greater level when genomes with lower similarity are hybridized. This may be due to a short continuous stretch of perfectly matched bases (as few as 15 bp), which can induce a strong microarray cross-hybridization signal in gene pairs with low overall nucleotide identity [6, 29]. Additionally, noncoding regions of genome such as intergenic sequences might affect the genomic DNA hybridization signal. As expected, we observed the highest ARE when the PHS was obtained by using another model that had been formulated on the basis of cross-hybridization experiments with 50-mer oligonucleotide probes [18, 25] (data not shown).

Genomic Factors Influencing DNA-DNA Hybridization

Our observations suggest that the accumulation of mutations and lateral gene transfer (LGT) are two major factors that affect the measurement of genome similarity by DNA-DNA hybridization. As discussed in more detail below, LGT was evident in the gene pair identity pattern in Fig. 1A (pattern 1), whereas the accumulation of mutations was significant in the gene pair identity pattern in Figs. 1B and 1C (pattern 2).

Accumulation of Mutations in Conserved Genes. As revealed by the pairwise gene identity distribution profiles shown in Fig. 1, excluding the *B. subtilis* 168/*B. cereus* ATCC 14579 pair, the highest gene number (HGN) was at 80–85% (Fig. 1C), 90–95% (Fig. 1B), or 95–100% (Fig. 1A) pairwise gene identity. This observation led us to screen pattern 1- (Figs. 1B and 1C) and pattern 2-like (Fig. 1A) gene identity distributions in the microbial genome database (http://cmr.tigr.org/) to determine the degree of gene identity of the HGNs (Fig. 4). Some examples were shown that closely related strains could have differential genome similarities (Fig. 4). Interestingly, we did not find any genome pairs having HGNs below 80% gene pair

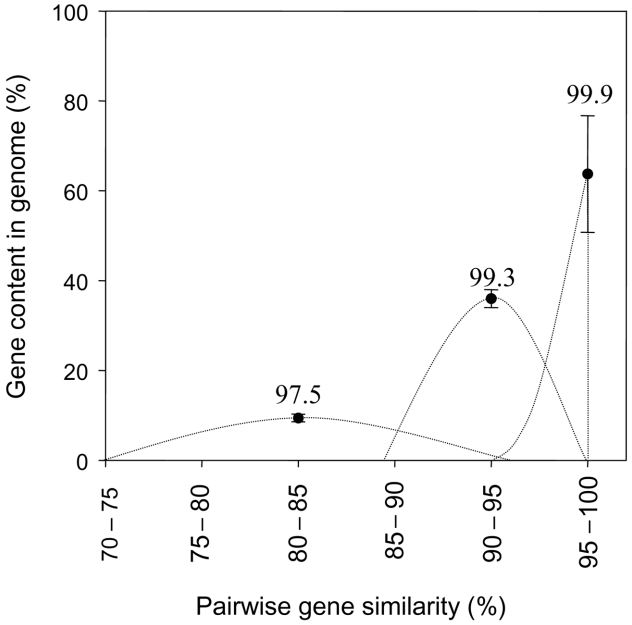

Fig. 4. Pairwise gene identity distribution profiles of closely related genomes.
See Fig. 1 for details. Dotted lines indicate the distribution frequencies of gene identities, which were drawn arbitrarily for illustrative purposes. The HGNs are at the peak of each distribution profile. The number at each peak indicates the average identity of the 16S rRNA gene pairs. The pairs of genomes showing HGN at 80–85% gene identity were *E. coli* K12/*S. typhimurium* LT2, and *S. typhimurium* LT2/*S. flexneri* 2a. The pairs of genomes showing HGN at 90–95% gene identity were *S. mitis*/*S. pneumoniae* R6, *S. miti*/*S. pneumoniae* TIGR4, and *B. cereus* ATCC 10987/*B. cereus* ATCC 14579. The pairs of genomes showing HGN at 95-100% gene identity were *S. pneumoniae* R6/*S. pneumoniae* TIGR4, *S. pyogens* MGAS8232/*S. pyogenes* MGAS315, *S. pyogenes* MGAS8232/*S. pyogenes* SF370, *S. pyogens* MGAS8232/*S. pyogenes* SSI-1, *B. anthraces Ames*/*B. cereus* ATCC 10987, *E. coli* K12/*E. coli* O157 EDL933, *E. coli* K12/*E. coli* O157 VT2, and *E. coli* K12/*S. flexneri* 2a.

identity. Our analyses also showed that, as genome similarity decreased, not only were the HGNs at a lower gene pair identity, but the associated HGN value also decreased. This would help decrease the hybridization signal. Under our experimental conditions, hybridization signal intensities greater than 40% indicate HGNs at above 80% in a pairwise gene identity distribution profile with more than 30% of genes at over 60% pairwise gene identity (*e.g.*, see the *E. coli* K12 and *S. typhimurium* LT2 pair of Table 1 and Fig. 1C).

Acquisition of New Genes. The second factor influencing genome similarity and hybridization signal is the presence of highly divergent genes (with <55% gene identity) between closely related strains. For instance, when the highly related *E. coli* pairs *E. coli* O157 EDL933/*E. coli* K12 and *E. coli* O157 VT2/*E. coli* K12 are compared, about 38% of the genes are divergent. This divergence usually results from LGT [7]. In these cases, most of the genes (>90%) can be divided into two categories, which thus lead to the pairwise gene identity distribution pattern (pattern 1) of the *E. coli* K12/*S. flexneri* 2a pair in Fig. 1A. Most genome pairs with gene identity distribution pattern 1 showed about 70% EHS (*i.e.*, genome similarity), as indicated in Fig. 4. Divergent genes (with <55% gene identity) of pattern 1 make it difficult to obtain a hybridization signal (and thus a genome similarity) greater than 70%, even among closely related strains showing very high 16S rRNA gene identity (>97%). Bacteria have obtained a significant proportion of their genes through the acquisition of sequences from distantly related organisms, unlike eukaryotes, which evolve principally through the modification of existing genetic information. As a result, LGT produces extremely dynamic genomes in which significant amounts of genes are introduced into and deleted from the genome. These LGTs have effectively changed the ecological and pathogenic character of bacterial strains. Different ecotypes or pathovars may be generated by such LGT-mediated acquisition of novel genes during the environmental adaptation of a species to invade a new environmental niche [19].

At present, several methods based on genomic sequences have been used to determine bacterial phylogeny; these methods replace those that are based on only one or just a few genes [12, 17, 31]. Most of the genomic sequence-based methods compare COG genes across distantly related taxa and can be useful for constructing phylogenetic trees above the species level. However, our data show that the frequency of genes obtained by LGT (non-COG genes) could be more important in determining genome similarity than pairwise COG gene identity below the species level.

According to the current criteria generally used for discriminating between species, a DNA-DNA hybridization signal that is lower than 70% is considered not to be important in current taxonomy. However, our observations suggest that DNA-DNA hybridization signals of between 70% and 40% could still be indicative of a significant level of genome similarity, since the presence of pattern 1 in a pairwise gene identity distribution profile could be evidence that the strains in question are in the process of speciation (diversification). Thus, although a 70% hybridization signal is the cutoff for distinguishing between species, we suggest that hybridization signals down to 40% might be indicative of significant genome-relatedness.

In summary, by determining the correlation between genome sequence similarity and the DNA-DNA hybridization signal, we have expanded our knowledge of the genomic principles underlying DNA-DNA hybridization.

Acknowledgment

This work was supported by grant MG05-0104-3-0 (the 21C Frontier Microbial Genomics and Application Center Program) from the Ministry of Science and Technology (MOST), Republic of Korea.

REFERENCES

1. Bae, J. W., S. K. Rhee, J. R. Park, W. H. Chung, Y. D. Nam, I. Lee, H. Kim, and Y. H. Park. 2005. Development and evaluation of genome-probing microarrays for monitoring lactic acid bacteria. *Appl. Environ. Microbiol.* **71:** 8825–8835.
2. Brenner, D. J., G. R. Fanning, F. J. Skerman, and S. Falkow. 1972. Polynucleotide sequence divergence among strains of *Escherichia coli* and closely related organisms. *J. Bacteriol.* **3:** 953–965.
3. Crosa, J. H., D. J. Brenner, W. H. Ewing, and S. Falkow. 1973. Molecular relationships among the *Salmonelleae*. *J. Bacteriol.* **115:** 307–315.
4. Dorris, D. R., R. Ramakrishnan, D. Trakas, F. Dudzik, R. Belval, C. Zhao, A. Nguyen, M. Domanus, and A. Mazumder. 2002. A highly reproducible, linear, and automated sample preparation method for DNA microarrays. *Genome Res.* **12:** 976–984.
5. Fukushima, M., K. Kakinuma, H. Hayashi, H. Nagai, K. Ito, and R. Kawaguchi. 2003. Detection and identification of *Mycobacterium* species isolates by DNA microarray. *J. Clin. Microbiol.* **41:** 2605–2615.
6. Girke, T., J. Todd, S. Ruuska, J. White, C. Benning, and J. Ohlrogge. 2000. Microarray analysis of developing *Arabidopsis* seeds. *Plant Physiol.* **124:** 1570–1581.
7. Gogarten, J. P. and J. P. Townsend. 2005. Horizontal gene transfer, genome innovation and evolution. *Nat. Rev. Microbiol.* **3:** 679–687.
8. Grimont, P. A. D., M. Y. Popoff, F. Grimont, C. Coynault, and M. Lemelin. 1980. Reproducibility and correlation study of three deoxyribonucleic acid hybridization procedures. *Curr. Microbiol.* **4:** 325–330.

9. Hwang, K. O. and J. C. Cho. 2006. Evaluation of DNA microarray approach for identifying strain-specific genes. *J. Microbiol. Biotechnol.* **16:**1773–1777.
10. Kim, B. S., S. J. Kang, S. B. Lee, W. Hwang, and K. S. Kim. 2005. Simple method to correct gene-specific dye bias from partial dye swap information of a DNA microarray experiment. *J. Microbiol. Biotechnol.* **15:** 1377–1383.
11. Kim, I. S., H. S. Yun, H. Shimisu, E. Kitagawa, H. Iwahashi, and I. Jin. 2005. Elucidation of copper and asparagine transport systems in *Saccharomyces cerevisiae* KNU5377 through genome-wide transcriptional analysis. *J. Microbiol. Biotechnol.* **15:** 1240–1249.
12. Konstantinidis, K. T. and J. M. Tiedje. 2005. Towards a genome-based taxonomy for prokaryotes. *J. Bacteriol.* **187:** 6258–6264.
13. Loy, A., A. Lehner, N. Lee, J. Adamczyk, H. Meier, J. Ernst, K. H. Schleifer, and M. Wagner. 2002. Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. *Appl. Environ. Microbiol.* **68:** 5064–5081.
14. Maynard, S. J. 1995. *Do Bacteria Have Population Genetics?* Cambridge University Press, Cambridge.
15. Oh, M. K., M. J. Cha, S. G. Lee, L. Rohlin, and J. C. Liao. 2006. Dynamic gene expression profiling of *Escherichia coli* in carbon source transition from glucose to acetate. *J. Microbiol. Biotechnol.* **14:** 543–549.
16. Paster, B. J. and F. E. Dewhirst. 1988. Phylogeny of campylobacters, wolinellas, *Bacteroides gracilis*, and *Bacteroides ureolyticus* by 16S ribosomal ribonucleic acid sequencing. *Int. J. Syst. Bacteriol.* **38:** 56–62.
17. Qi, J., B. Wang, and B. I. Hao. 2004. Whole proteome prokaryote phylogeny without sequence alignment: A K-string composition approach. *J. Mol. Evol.* **58:** 1–11.
18. Rhee, S. K., X. Liu, L. Wu, S. C. Chong, X. Wan, and J. Zhou. 2004. Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. *Appl. Environ. Microbiol.* **70:** 4303–4317.
19. Rocap, G., F. W. Larimer, J. Lamerdin, S. Malfatti, P. Chain, N. A. Ahlgren, A. Arellano, M. Coleman, L. Hauser, W. R. Hess, Z. I. Johnson, M. Land, D. Lindell, A. F. Post, W. Regala, M. Shah, S. L. Shaw, C. Steglich, M. B. Sullivan, C. S. Ting, A. Tolonen, E. A. Webb, E. R. Zinser, and S. W. Chisholm. 2003. Genome divergence in two *Prochlorococcus* ecotypes reflects oceanic niche differentiation. *Nature* **424:** 1042–1047.
20. Roop, R. M., R. M. Smibert, J. L. Johnson, and N. R. Krieg. 1984. Differential characteristics of catalase-positive campylobacters correlated with DNA homology groups. *Can. J. Microbiol.* **30:** 938–951.
21. Roop, R. M., R. M. Smibert, J. L. Johnson, and N. R. Krieg. 1994. DNA homology studies of the catalase-negative campylobacters and *Campylobacter fecalis*, an emended description of *Campylobacter sputorum*, and proposal of the neotype strain of *Campylobacter sputorum*. *Can. J. Microbiol.* **31:** 846–849.
22. Rossello-Mora, R. and R. Amann. 2001. The species concept for prokaryotes. *FEMS Microbiol. Rev.* **25:** 39–67.
23. Somerville, H. J. and M. L. Jones. 1972. DNA competition studies within the *Bacillus cereus* group of bacilli. *J. Gen. Microbiol.* **73:** 257–265.
24. Stackebrandt, E. and B. M. Goebel. 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.
25. Taroncher-Oldenburg, G., E. M. Griner, C. A. Francis, and B. B. Ward. 2003. Oligonucleotide microarray for the study of functional gene diversity in the nitrogen cycle in the environment. *Appl. Environ. Microbiol.* **69:** 1159–1171.
26. Ursing, J., M. Walder, and K. Sandstedt. 1983. Base composition and sequence homology of deoxyribonucleic acid of thermotolerant *Campylobacter* from human and animal sources. *Curr. Microbiol.* **8:** 307–310.
27. Vandamme, P., E. Falsen, R. Rossau, B. Hoste, P. Segers, R. Tytgat, and J. D. Ley. 1991. Revision of *Campylobacter*, *Helicobacter*, and *Wolinella* taxonomy: Emendation of generic descriptions and proposal of *Arcobacter* gen. nov. *Int. J. Syst. Bacteriol.* **41:** 88–103.
28. Wayne, L.G., D. J. Brenner, R. R. Colwell, P. A. D. Grimont, O. Kandler, M. I. Krichevsky, L. H. Moore, W. E. C. Moore, R. G. E. Murray, E. Stackebrandt, M. P. Starr, and H. G. Trüper. 1987. Report of the *ad hoc* committee on reconciliation of approaches to bacterial systematics. *Int. J. Syst. Bacteriol.* **37:** 463–464.
29. Wu, L., D. K. Thompson, G. Li, R. A. Hurt, J. M. Tiedje, and J. Zhou. 2001. Development and evaluation of functional gene arrays for detection of selected genes in the environment. *Appl. Environ. Microbiol.* **67:** 5780–5790.
30. Wu, L., D. K. Thompson, X. Liu, M. W. Fields, C. E. Bagwell, J. M. Tiedje, and J. Zhou. 2004. Development and evaluation of microarray-based whole-genome hybridization for detection of microorganisms within the context of environmental applications. *Environ. Sci. Technol.* **38:** 6775–6782.
31. Wu, X., X. F. Wan, G. Wu, D. Xu, and G. Lin. 2006. Whole genome phylogeny via complete composition vectors. *Int. J. Bioinform. Res. Applic.* **2:** 219–248.
32. Xu, W., S. Bak, A. Decker, S. M. Paquette, R. Feyereisen, and D. W. Galbraith. 2001. Microarray-based analysis of gene expression in very large gene families: The cytochrome P450 gene superfamily of *Arabidopsis thaliana*. *Gene* **272:** 61–74.
33. Zhang, L., U. Srinivasan, C. F. Marrs, D. Ghosh, J. R. Gilsdorf, and B. Foxman. 2004. Library on a slide for bacterial comparative genomics. *BMC Microbiol.* **4:** 12.
34. Zhou, J. and D. K. Thompson. 2002. Challenges in applying microarrays to environmental studies. *Curr. Opin. Biotechnol.* **13:** 204–207.