

ORIGINAL ARTICLE

Microarray-based whole-genome hybridization as a tool for determining procaryotic species relatedness

Liyou Wu^{1,2,3}, Xueduan Liu^{2,4}, Matthew W Fields⁵, Dorothea K Thompson⁶, Christopher E Bagwell⁷, James M Tiedje⁸, Terry C Hazen⁹ and Jizhong Zhou^{1,2}

¹Department of Botany and Microbiology, Institute for Environmental Genomics, Stephenson Research and Technology Center, University of Oklahoma, Norman, OK, USA; ²Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA; ³College of Bio-Safety Science and Technology, Hunan Agricultural University, Changsha, Hunan, PR China; ⁴School of Minerals Processing & Bioengineering, Central South University, Changsha, Hunan, PR China; ⁵Department of Microbiology, Montana State University, Bozeman, MT, USA; ⁶Department of Biological Sciences, Purdue University, West Lafayette, IN, USA; ⁷Environmental Biotechnology Section, Westinghouse Savannah River Company, Aiken, SC, USA; ⁸Center for Microbial Ecology, Michigan State University, East Lansing, MI, USA and ⁹Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA

The definition and delineation of microbial species are of great importance and challenge due to the extent of evolution and diversity. Whole-genome DNA–DNA hybridization is the cornerstone for defining procaryotic species relatedness, but obtaining pairwise DNA–DNA reassociation values for a comprehensive phylogenetic analysis of procaryotes is tedious and time consuming. A previously described microarray format containing whole-genomic DNA (the community genome array or CGA) was rigorously evaluated as a high-throughput alternative to the traditional DNA–DNA reassociation approach for delineating procaryotic species relationships. DNA similarities for multiple bacterial strains obtained with the CGA-based hybridization were comparable to those obtained with various traditional whole-genome hybridization methods ($r=0.87$, $P<0.01$). Significant linear relationships were also observed between the CGA-based genome similarities and those derived from small subunit (SSU) rRNA gene sequences ($r=0.79$, $P<0.0001$), *gyrB* sequences ($r=0.95$, $P<0.0001$) or REP- and BOX-PCR fingerprinting profiles ($r=0.82$, $P<0.0001$). The CGA hybridization-revealed species relationships in several representative genera, including *Pseudomonas*, *Azoarcus* and *Shewanella*, were largely congruent with previous classifications based on various conventional whole-genome DNA–DNA reassociation, SSU rRNA and/or *gyrB* analyses. These results suggest that CGA-based DNA–DNA hybridization could serve as a powerful, high-throughput format for determining species relatedness among microorganisms.

The ISME Journal (2008) 2, 642–655; doi:10.1038/ismej.2008.23; published online 28 February 2008

Subject Category: integrated genomics and post-genomics approaches in microbial ecology

Keywords: DNA–DNA reassociation; community genome array; *gyrB*; procaryotic species; REP-PCR and BOX-PCR; SSU rRNA

Introduction

Microorganisms are the most abundant and diverse life forms on the Earth and play integral, often unique, roles in the biogeochemical cycling of elements that are crucial to ecosystem function

and sustainability (Whitman *et al.*, 1998). However, the definition of procaryotic species is arbitrary in that it is more of a function of the methodology than of any actual relationship (Harayama and Kasai, 2006; Stackebrandt and Goebel, 1994; Whitman *et al.*, 1998). By contrast, the definition of species in multicellular organisms is well defined: two organisms that reproduce sexually to produce fertile offspring. The definition of species in bacteria is problematic, because reproduction is generally asexual, and the exchange of genetic material can be horizontal. As such, there have been a number of attempts to proffer a working definition of what

Correspondence: J Zhou, Institute for Environmental Genomics and the Department of Botany/Microbiology, Stephenson Research and Technology Center, University of Oklahoma, 101 David L Boren Boulevard, Norman, OK 73019, USA.
E-mail: jzhou@ou.edu

Received 18 December 2007; revised 4 February 2008; accepted 4 February 2008; published online 28 February 2008

constitutes a bacterial species (Wayne *et al.*, 1987; Cohan, 2002; Harayama and Kasai, 2006). One well-accepted definition is based on whole-genome DNA–DNA hybridization, by which a procaryotic species is defined as an entity comprised of strains sharing an approximately 70% or greater reassociation value (Wayne *et al.*, 1987). Although this definition is pragmatic and universally applicable within the procaryotic world (Rossello-Mora and Amann, 2001; Stackebrandt *et al.*, 2002; Konstantinidis and Tiedje, 2004), there is no theoretical foundation for setting the value of $\geq 70\%$ reassociation value as a criterion for species designation. Nonetheless, it is still considered as the ‘gold standard’ for delineating procaryotic species (Harayama and Kasai, 2006; Stackebrandt, 2006). Almost 5000 bacterial species have been described (Garrity *et al.*, 2004), many of them classified predominantly on the basis of whole-genome DNA–DNA reassociation experiments (Rossello-Mora, 2006). The practice of DNA–DNA hybridization has allowed the establishment of a relatively stable and operative classification system for procaryotes (Stackebrandt *et al.*, 2002).

Over the last 50 years, various hybridization-based techniques have been developed for microbial taxonomy (Rossello-Mora, 2006). All of these methods measure the extent and/or stability of complementary DNAs to renature under stringent conditions. Two main strategies have been used for performing re-association experiments: one is to carry out the hybridization reaction in free solution (Brenner *et al.*, 1969; Crosa *et al.*, 1973; Popoff and Coynault, 1980; Ziemke *et al.*, 1998) and the other is to immobilize the test DNA onto a solid surface, such as membrane filters and microtiter plates (De Lay and Tijtjat, 1970; Ezaki *et al.*, 1989; Amann *et al.*, 1992; Adnan *et al.*, 1993; Kaznowski, 1995; Cardinali *et al.*, 2000; Gade *et al.*, 2004; Mehlen *et al.*, 2004). Relative binding ratio, the proportion of double-stranded hybrid DNA for a given pair of genomes relative to that of the reference DNA under identical renature conditions, and/or the increment of melting temperature (ΔT_m) can be used to measure genome similarities. However, DNA–DNA hybridization is extremely labor intensive, tedious and time consuming because only pairwise comparisons can be made per experiment, thus prohibiting implementation of this method for analysis of large culture collections. More effective high-throughput approaches are needed to allow large-scale parallel analysis of many microbial genomes.

Compared to the conventional nucleic acid hybridization with porous membranes, real-time PCR and other molecular approaches, microarray-based hybridization offers advantages of high-throughput and parallel detection. DNA microarrays have been widely used to analyze gene expression in pure-culture studies (Schena *et al.*, 1995, 1996; Lockhart *et al.*, 1996; DeRisi *et al.*, 1997; Ye *et al.*, 2000; Liu *et al.*, 2003; Gao *et al.*, 2004), environmental

microbial community analysis (Wu *et al.*, 2001; Loy *et al.*, 2002; Bodrossy *et al.*, 2003; Rhee *et al.*, 2004; Tiquia *et al.*, 2004; Schadt *et al.*, 2005; Zhang *et al.*, 2007) and the comprehensive comparison of genomes among closely related species (Behr *et al.*, 1999; Salama *et al.*, 2000; Dong *et al.*, 2001; Kato-Maeda *et al.*, 2001; Murray *et al.*, 2001). Two types of microarrays have been used for determining species relatedness. One is to use the whole-genome open-reading frame array-based hybridization approach to reveal genome diversity and relatedness among closely related organisms (Murray *et al.*, 2001). The other is to use DNA microarrays containing random genomic fragments to delineate species relationships where the genome sequence does not exist (Cho and Tiedje, 2001). These types of microarrays are useful in revealing the genomic diversity and relatedness of closely related organisms with higher resolution compared to traditional DNA–DNA hybridization methods (Zhou, 2003). These approaches, however, are more time consuming and costly to develop and such arrays would have more limited use because many of the arrayed probes would be used for each reference microorganism (Zhou, 2003).

We have previously developed a novel type of microarray, termed the community genome array (CGA), that contained whole-genomic DNA of different microorganisms (Wu *et al.*, 2004). This prototype array was tested for specificity, sensitivity, quantitation and environmental applications with respect to microbial community analysis (Wu *et al.*, 2004). Our results suggested that CGA-based hybridization has potential as a specific, sensitive and quantitative tool for the detection and identification of microorganisms in environmental samples. Additionally, CGA-based hybridization is potentially useful for high-throughput parallel determination of species relatedness of many microorganisms. To test this possibility, we evaluated the CGA-based hybridization as an alternative method to measure DNA relatedness, and compared predicted relationships with traditional whole-genome DNA–DNA hybridization methods, small subunit (SSU) rRNA and *gyrB* gene analysis, and genomic fingerprinting. Our results indicated that the CGA-based hybridization is comparable to traditional DNA–DNA hybridization methods in terms of species delineation and could serve as a powerful, high-throughput format for determining species relatedness of microorganisms.

Materials and methods

Bacterial strains and genomic DNA isolation

Both closely and distantly related representative bacterial strains ($n = 66$) were selected for this study on the basis of phylogenetic relationships, GC content, availability of DNA–DNA reassociation data and other molecular studies, and/or the

accessibility of the strains (Supplementary Table_SM1). Fifty-five of the 66 representative bacterial strains were from the laboratory culture collections of JZ and JT, and included 16 *Shewanella* strains, 30 *Pseudomonas* strains and 9 *Azoarcus* strains. The genotypic and phenotypic descriptions and taxonomic classification of some of these bacteria have been reported elsewhere (Rossello *et al.*, 1991; Zhou *et al.*, 1995; Song *et al.*, 1999; Venkateswaran *et al.*, 1999; Cladera *et al.*, 2004). The strain collection also included one *Halomonas* strain, one α -*proteobacterium* strain, five *Marinobacter* strains and two *Bacillus* strains (Braker *et al.*, 2000). *Escherichia coli* S17-1/pir and a *Saccharomyces cerevisiae* strain were also included as negative controls.

The genomic DNAs were isolated from pure cultures using previously described protocols (Zhou *et al.*, 1995). All genomic DNA samples were treated with RNase A (Sigma, St Louis, MO, USA) and analyzed on agarose gels stained with ethidium bromide prior to microarray fabrication. DNA concentration was determined in the presence of ethidium bromide by fluorometric measurement of the excitation at 360 nm and emission at 595 nm using a HTS700 BioAssay Reader (Perkin-Elmer, Norwalk, CT, USA).

DNA sequencing and genomic fingerprinting

Small subunit rRNA gene sequences for all of the strains used in this study were retrieved from the Ribosomal Database Project (<http://rdp.cme.msu.edu/html/>). DNA topoisomerase subunit B (*gyrB*) gene sequences of some representative strains were PCR amplified as described elsewhere (Yamamoto and Harayama, 1995), and the products cloned directly into the pCR2.1 vector according to the manufacturer's instructions (Invitrogen, Carlsbad, CA, USA). The *gyrB* sequences were determined using the BigDye Terminator kit (Applied Biosystem, Foster City, CA, USA) with a 3700 DNA analyser (Perkin-Elmer, Wellesley, MA, USA) according to the manufacturer's instructions. Sequences were assembled and edited manually using Sequencher 4.7 (Gene Codes Corp., Ann Arbor, MI, USA) and aligned in ClustalW (Thompson *et al.*, 1994). Bootstrap values (500 re-samplings) supported neighbor joining phylogenies that were constructed in MEGA 3.1 (Kumar *et al.*, 2003) from Poisson correction distances to account for multiple substitution events/site using pairwise deletion of gaps and missing data.

The genomic fingerprinting methods of BOX-PCR and REP-PCR were performed according to a previously described protocol (Versalovic *et al.*, 1994). The REP-PCR and BOX-PCR genomic fingerprints for each isolate were linearly combined using Molecular Analyst 1.6 (BioRad, Hercules, CA, USA) with a resolution of 400 as described previously (Rademaker *et al.*, 2000). The similarity between

pairs of combined fingerprints was calculated, and cluster analysis of the similarity values was determined using a UPGMA (unweighted pair group method using averages) algorithm.

Microarray construction, probe labeling and hybridization

The genomic DNA probe concentration used for printing was 200 ng μl^{-1} , and each DNA sample was suspended in 50% DMSO. Five *S. cerevisiae* genes were also included on the arrays as negative controls. All 71 probes (including negative controls) were arranged as a matrix of 15 rows \times 5 columns (denoted columns a–e). Genomic DNA samples were prepared for deposition and printed as described previously (Wu *et al.*, 2001). Each glass slide contained three replicates of genomic DNA from individual strains. Following printing, glass slides were post-processed and evaluated for spot quality as described previously (Wu *et al.*, 2001).

Representative strains from different bacterial groups, including three *Shewanella* strains, two *Azoarcus* strains, nine *Pseudomonas* strains, one *Halomonas* strain, one *Bacillus* strain, one *Marinobacter* strain and one α -*proteobacterium*, were labeled and hybridized against the CGAs. For convenience, these labeled strains are referred to as reference strains. The genomic DNAs (1 μg) from the reference strains were fluorescently labeled using the random priming method and purified as described previously (Wu *et al.*, 2001). Microarray experiments were carried out in triplicate (a total of nine replicates per genomic DNA probe) to enable statistical analyses. Microarray hybridization was performed as described previously (Wu *et al.*, 2004).

Microarray scanning and signal intensity quantitation

A ScanArray 5000 Microarray Analysis System (Perkin-Elmer) was used for scanning microarrays. A quick scan at a resolution of 50 μm was performed prior to the real scanning at a resolution of 5 μm and laser power and photomultiplier tube gain were adjusted to avoid saturation of spots. Scanned image displays were saved as 16-bit TIFF files and analyzed by quantifying the pixel density (intensity) of each spot using ImaGene version 4.0 (Bio-discovery Inc., Los Angeles, CA, USA). Mean signal intensity was determined for each spot and the local background signals were subtracted automatically from the hybridization signal of each spot. Fluorescence intensity values for all replicates of the negative control genes (five yeast genes) were averaged and then subtracted from the background-corrected intensity values for each hybridization signal. The signal-to-noise ratio (SNR) was also calculated based on the formula derived from Verdnik *et al.* (2002), where $\text{SNR} = (\text{Signal Intensity} - \text{Background}) / \text{Standard Deviation of Background}$. Spots with SNRs lower than 2 were defined as

empty spots. Empty spots and poor spots defined by the program based on the morphology and signal saturation of the spots were removed prior to data analysis. The outliers, represented by the data points that were not consistently reproducible and had a disproportionately large effect on the statistical result, were detected and removed at $P < 0.01$. When the absolute value of a data point minus the mean was larger than 2.90σ , this data point was determined as an outlier and removed.

DNA similarity calculation and cluster analysis

Routine statistical analysis was performed using SigmaPlot 5.0 (Jandel Scientific, San Rafael, CA, USA). DNA similarity of the strain of interest relative to a reference strain (S_i) was calculated using the following formula: $S_i = f_i/f_r \times 100$, where f_i is the average (nine replicated spots) signal intensity of the spots corresponding to the strain of interest, and f_r is the average fluorescence signal intensity of the spots of the reference strain. Relationships among different strains from the microarray hybridizations were determined using hierarchical cluster analysis (CLUSTER) and visualized with TREEVIEW (Eisen *et al.*, 1998).

Results

Species relatedness revealed by the CGA-based hybridization

To determine whether the CGA-based hybridization can be used to determine the relationships of closely related species, the genomic DNAs from the selected representative strains were fluorescently labeled and hybridized to the CGA. The DNA similarity data based on hybridization signal intensities were then used to visualize the relationships among the organisms spotted on the arrays. As expected, all strains examined were well separated at different taxonomic levels, which were more or less consistent with previous classifications, especially at the genus/species levels (Figure 1). For instance, the Gram-positive *Bacillus* species (Figure 1 branch D) were separated from most *Proteobacterial* species (Figure 1). Different subdivisions of the *Proteobacteria* were also fairly well separated from each other (Figure 1). '*Pseudomonas*' sp. G179 (an α -*proteobacterium*), for example, did not cluster with true *Pseudomonas* species (γ -*Proteobacteria*). Species in the *Azoarcus* genus (Figure 1, branch B) of the β -*Proteobacteria* were well separated from the γ -*Proteobacteria*, represented by *Shewanella* (Figure 1, branch A) and *Pseudomonas* species (Figure 1, branch C). In addition, different species or strains within a genus formed distinctive cohesive clusters, such as *Azoarcus*, *Pseudomonas* and *Shewanella* clusters, and were clearly separated from each other. Finally, species relatedness was also revealed based on the labeled strains, which formed distinct clusters as expected (Figure 1, top panel).

Genome differences among different species within different genera were also clearly revealed by CGA-based hybridization. For instance, all *Azoarcus* strains showed very strong hybridization to the labeled *Azoarcus* reference strains (Td-1, Td-21) but weaker hybridization to other labeled reference strains (Figure 1). Within the *Azoarcus* genus, three subclusters were formed (Figure 1, branch A). *A. toluityticus* strains Td17 and Td21 grouped together as the first subcluster (Figure 1, b1); Td1, Td2, BL-11, Td15 formed the second subcluster (Figure 1, b2), whereas Td3 and Td19 grouped together as the third subcluster (Figure 1, b3). Although some strains of *Azoarcus toluityticus* could not be unambiguously resolved under hybridization conditions of 55 °C and 50% formamide, they could be differentiated at higher temperatures (65 and 75 °C with 50% formamide) (data not shown) (Wu *et al.*, 2004). These results are consistent with those based on conventional DNA–DNA hybridization and/or SSU rRNA gene analysis (see below).

Similarly, all *Pseudomonas* strains showed very strong hybridization to the labeled *P. stutzeri* strains (B2-2, E4-2 and ATCC 17587), but not to other labeled reference strains (Figure 1). All *Pseudomonas stutzeri* strains clustered together (Figure 1, branch C). Interestingly, the strain *P. balearica* DSM 6038 (Figure 1, branch C) formed a distinct cluster with all *P. stutzeri* strains, and this strain was originally classified as a *Pseudomonas stutzeri* strain (Rossello *et al.*, 1991). Four subclusters were observed within *P. stutzeri* (Figure 1, c1–c4). ATCC strains 17595, 17592 and 17587 as well as marine isolates B2-2, E4-2 and F9-2 grouped within the first subcluster (Figure 1, c1), while ATCC strains 17594, 11256, 27591 and DNSP21 formed the second subcluster (Figure 1, c2). *P. stutzeri* ATCC 50238, *P. balearica* DSM 6083 and marine isolate C5-1 formed the third subcluster (Figure 1, c3). *P. stutzeri* ATCC 50238 showed relatively higher hybridization to the reference strain ATCC 17587, while C5-1 as well as *P. balearica* DSM 6038 showed weaker hybridization to this reference strain (Figure 1), indicating that the marine isolate C5-1 is less related to *P. stutzeri* than it is to *P. balearica* DSM 6038. The two isolates D7-6 and D8-12 from the same marine environment formed the fourth subcluster (Figure 1, c4). Both strains showed very strong hybridization with the reference strain ATCC 17587 (Figure 1). Interestingly, while strain D8-12 showed weaker hybridization to the labeled reference strains (B2-2, E4-2), strain D7-6 had stronger hybridization signals to these reference strains (Figure 1), indicating the existence of some genetic differences among these strains. The above results are consistent with available DNA–DNA hybridization data using conventional approaches (see below).

All *Shewanella* strains hybridized relatively well to the labeled *Shewanella* reference strains (BrY, MR-1 and MR-4) (Figure 1). *S. algae* strains BrY and OK-1, *S. oneidensis* MR-1, and *Shewanella* sp. MR-4 showed very strong hybridization with

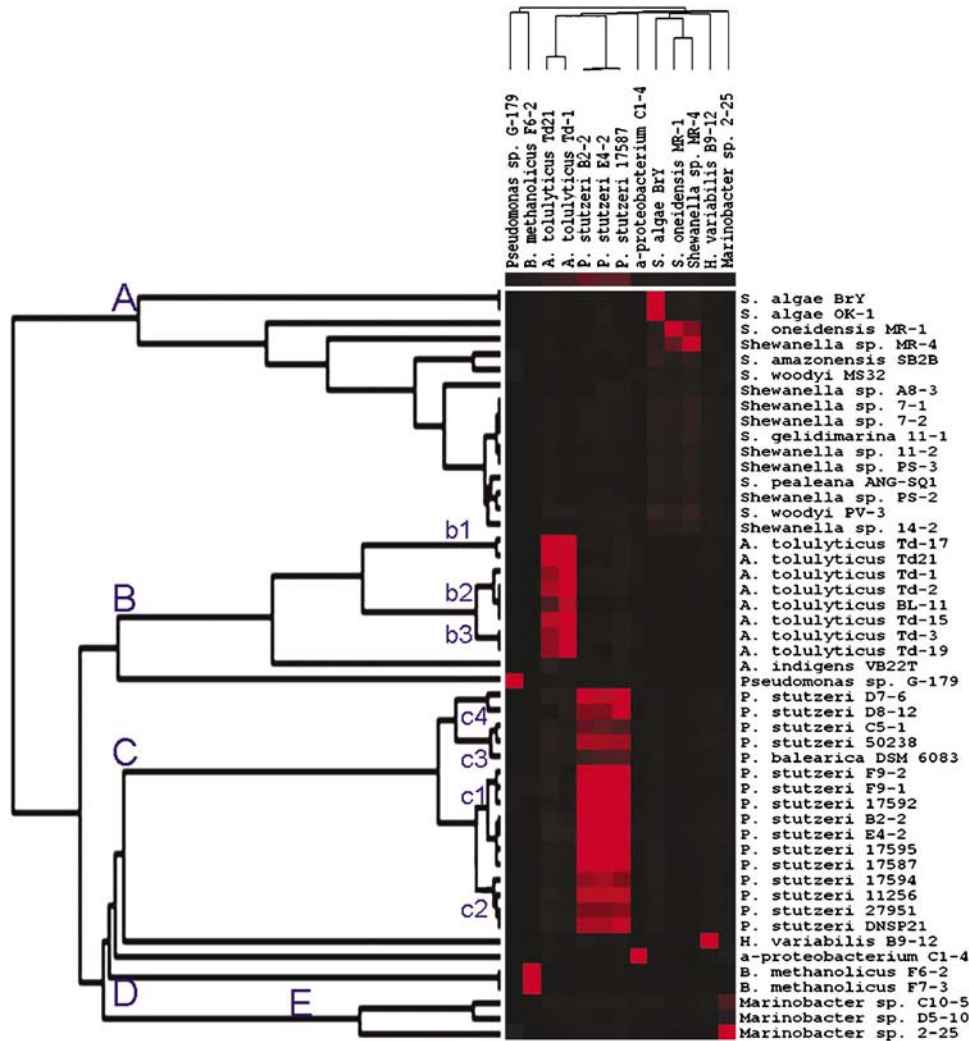


Figure 1 Hierarchical cluster analysis of species relationships based on DNA similarities obtained from the CGA-based hybridization. This figure was generated using hierarchical cluster analysis (CLUSTER) and visualized with TREEVIEW. (A) Microarray hybridization patterns at 55 °C plus 50% formamide with the labeled genomic DNAs from the selected species indicated in each column. Each row represents the hybridization signal observed for each species (that is, the test strain) when the genomic DNA from the species (that is, the reference strain) indicated in the column was fluorescently labeled and used for hybridization. Black represents no detectable hybridization above background levels, while red represents positive hybridization signals. The color darkness indicates differences in hybridization signal intensity. The columns correspond to the hybridization patterns obtained with Cy5-labeled genomic DNA from the following species: *Pseudomonas* sp. G179, *Bacillus methanolicus* F6-2, *Azoarcus toluyliticus* Td-21, *A. toluyliticus* Td-1, *P. stutzeri* B2-2, *P. stutzeri* E4-2, *Pseudomonas stutzeri* ATCC 17587, an unknown α -proteobacterium C1-4, *Shewanella algae* BrY, *S. oneidensis* MR-1, *Shewanella oneidensis* MR-4, *Halomonas variabilis* B9-12, and *Marinobacter* sp. 2-25.

their corresponding labeled reference strains, while all other *Shewanella* strains examined showed weak hybridization (Figure 1), but their hybridization signals were stronger than those with non-*Shewanella* strains. These results indicated the relatively high genome diversity of the *Shewanella* strains examined. Within the *Shewanella* genus, MR-1 is more closely related to MR-4 as expected, while *S. algae* strain BrY is more closely related to strain OK-1. *S. woodyi* strain MS32 and *S. amazonensis* SB2B formed a separate subcluster. Most of the isolates from marine sediments (Stapleton *et al.*, 2005) formed a distinct subcluster except for *Shewanella* sp. A8-3. In addition, the three *Marinobacter* sp. strains, 2-25 (Braker *et al.*, 2000), C10-5

and D5-10, formed a distinct cluster (Figure 1), which is consistent with SSU rRNA gene-based analysis.

In summary, all of the above results indicated that the CGA-based hybridization could be used to reliably visualize the relationships among different microorganisms at the genus and species levels. However, it could be difficult to use CGA-based hybridization to provide reliable classification for distantly related species at higher taxonomic levels. For instance, as illustrated in Figure 1, the *Bacillus* strains more closely grouped with *Pseudomonas* strains than with *Shewanella* strains. This is consistent with the conventional whole-genome DNA-DNA hybridization approach.

Comparison of the CGA-based hybridization to conventional DNA–DNA hybridizations

To determine whether the species relatedness revealed by the CGA-based whole-genome hybridization is similar to that by traditional DNA–DNA reassociation approaches, DNA similarities determined by the CGA-based hybridization were compared to previously reported DNA–DNA hybridization data (Rossello *et al.*, 1991; Song *et al.*, 1999; Venkateswaran *et al.*, 1999; Sepulveda-Torres *et al.*, 2001; Cladera *et al.*, 2004). Detailed DNA similarity data among *Pseudomonas* strains with various reference strains, *P. stutzeri* KC, *P. chlororaphis* ATCC17811, *P. putida* ATCC12633, *P. fluorescens* 13525 and *P. auruginosa* 15692, were also listed (Supplementary Table_SM2). Significant overall linear relationships were observed between DNA–DNA reassociation values and similarities based on CGA-based hybridization data ($n=43$, $r=0.82$, $P<0.0001$; Figure 2a) among all strains examined, thus suggesting that the resolving power of CGA-based hybridization could be comparable to that of traditional whole-genome DNA–DNA hybridization. The 70% DNA–DNA reassociation value corresponds to about 62% of similarity based on CGA hybridization data under the conditions examined (Figure 2a). However, the majority of the DNA–DNA hybridization data points have less than 40% similarity. A reliable linear relationship did not extend above a similarity of 40% based on the traditional DNA–DNA hybridization method, which corresponds to about 50% similarity obtained using the CGA-based hybridization method under the conditions examined.

When the comparisons were made among individual groups, even higher significant linear relationships ($r=0.9–0.99$) were obtained. DNA similarities from the CGA-based whole-genome DNA–DNA hybridization for various *Pseudomonas* strains were consistent with those determined using hydroxyapatite and/or the dot filter method (Supplementary Table_SM2). For example, differences in ΔT_m derived from DNA–DNA hybridization with strain 17591 using the hydroxyapatite and/or dot filter method (Rossello *et al.*, 1991) correlated well with microarray-based similarities ($n=7$, $r=-0.92$, $P<0.01$; Figure 2b). The CGA-based hybridization revealed that the strains 17592 and 17595 were 73.5% and 78.2% similar, respectively, to reference strain 17587 (Supplementary Table_SM2). This is consistent with previously reported DNA differences (0.5 and 1 °C ΔT_m , respectively) of these strains compared to strain 17591, which is identical to the reference strain 17587 (Rossello *et al.*, 1991) used in this study. Generally, the 70% DNA–DNA reassociation value corresponds to 5 °C ΔT_m . As shown in Figure 2b, the data suggested that the 5 °C ΔT_m is equivalent to 57% similarity from CGA-based hybridization. In addition, significant correlations of DNA similarities ($r=0.90$, $P<0.01$; Figure 2c) were observed between the CGA-based approach

and the traditional membrane filter method for the strains *P. fluorescens* 13525, *P. stutzeri* KC, *P. auruginosa* 15692 and *P. putida* 12633. However, the DNA similarities derived from microarray hybridization were also somewhat lower than those from the membrane methods (Supplementary Table_SM2). Similarly, the 70% DNA–DNA reassociation values for these strains corresponds to about 57.5% of similarity based on CGA hybridization data under the conditions examined (Figure 2c).

Phenotypic and genotypic heterogeneity among *P. stutzeri* strains has been documented in previous studies (Palleron *et al.*, 1970). Actually, Rossello *et al.* (1991) divided *P. stutzeri* into seven clear-cut genomic groups with significant DNA differences. The DNA similarities derived from CGA-based hybridization agreed with Rossello's classification of the *P. stutzeri* strains. As shown in Figure 1, the genetic group II (17595, 17592 and 17587) strains in *P. stutzeri* are distinctly separated from group I strains (ATCC strains 17594, 11256). In addition, ATCC strain 27591 was not included when various *P. stutzeri* strains were classified into different genetic groups (Rossello *et al.*, 1991). However, the phylogenetic analysis based on *gyrB* sequence similarities showed a close relationship with genetic group I (Cladera *et al.*, 2004). According to the DNA similarities from the CGA-based hybridizations, ATCC strain 27591 also clustered with genetic group I strains (ATCC strains 17594 and 11256) (Figure 1, c2). Finally, CGA-based hybridization supports the previous results that strain KC should belong to a new species other than *P. stutzeri* (Sepulveda-Torres *et al.*, 2001), because its CGA hybridization-based DNA similarities to all other *P. stutzeri* strains were relatively low, ranging from 12.8% to 47.8% (Supplementary Table_SM2 and 3).

The CGA hybridization-derived DNA similarities matched well with the DNA reassociation values using the S1 DNA nuclease method for different *A. toluolyticus* strains ($r=0.99$, $P<0.0001$; Figure 2d and Table 1). CGA-based hybridization revealed 89.7% similarity of Td-17 to the reference strain Td-21, which is very similar to that (89%) determined by the S1 method. The DNA similarities from CGA-based hybridization of Td-2 and BL-11 with the reference strain Td-1 were 97.4% and 55%, respectively, which are consistent with the reassociation values derived from S1 method (99% and 56%, respectively). The microarray-based DNA similarities between the other *Azoarcus* strains and the reference strains (Td-1 and Td-21) ranged from 15% to 35%, which also agreed well with the corresponding DNA–DNA reassociation values (Table 1). The 70% DNA–DNA reassociation values for the *Azoarcus* strains correspond to about 69% similarity based on CGA hybridization data under the conditions examined (Figure 2d).

The DNA similarities obtained from CGA-based whole-genome DNA–DNA hybridization supported a different classification of the *A. toluolyticus*

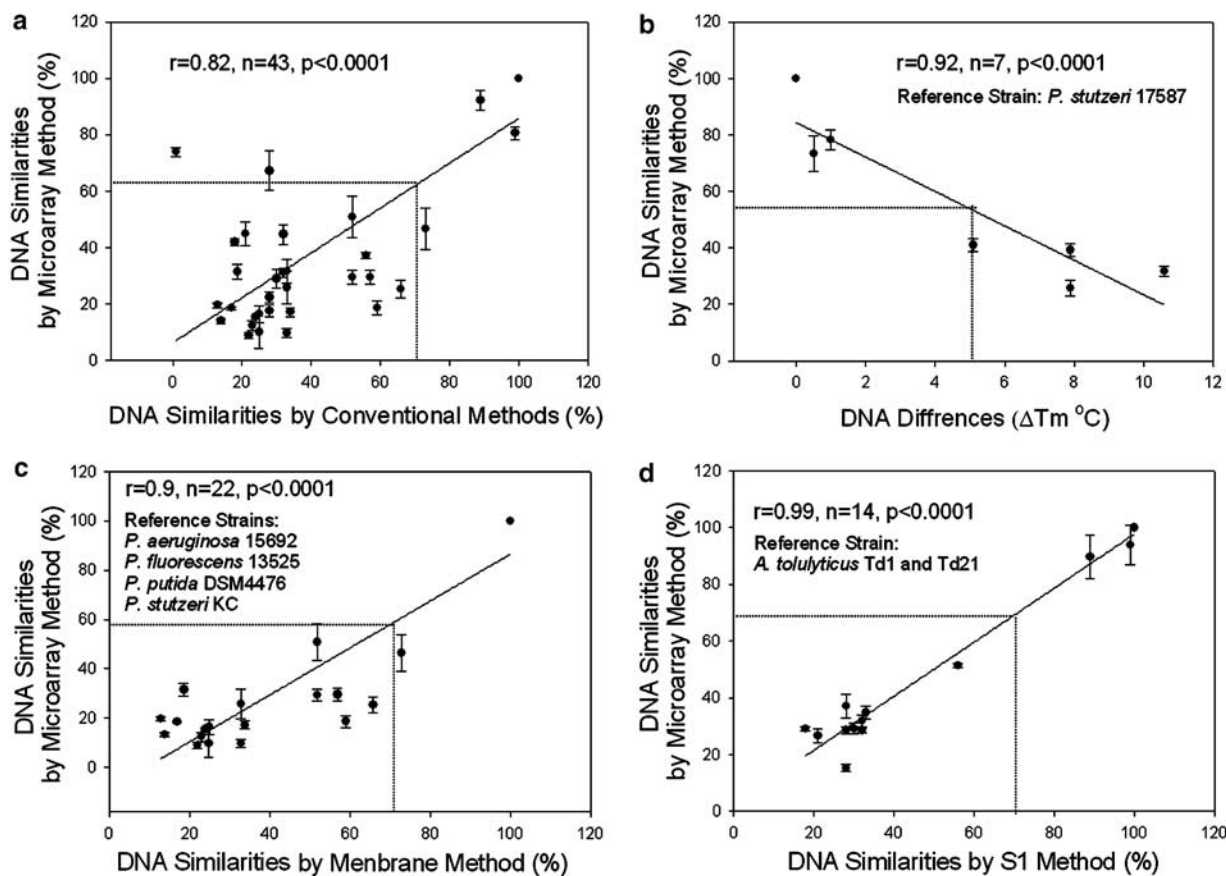


Figure 2 Relationship between DNA similarities (%) obtained from the CGA-based hybridization and DNA similarities (%) reported previously using different methods. (a) Overall comparison was made for DNA similarities by the microarray method to all available DNA similarities by conventional methods (DNA differences were transformed to similarities.). Then, DNA similarities by the microarray method was compared to DNA differences (b) by hydroxyapatite and dot filter methods [45] for reference strain *P. stutzeri* 17587 (identical to 17591), and to DNA similarities by (c) membrane filter methods, for reference strains *P. aeruginosa* 15692, *P. fluorescens* 13525, *P. putida* DSM 4467 and *P. stutzeri* KC, and (d) S1 nuclease method for reference strain *A. tolylyticus* Td-1 and Td-21.

Table 1 Comparison of DNA similarities from microarray hybridization and traditional methods for *Azoarcus* strains

	<i>A. tolylyticus</i> Td-1		<i>A. tolylyticus</i> Td-21	
	Microarray method (%)	S1 nuclease methods (%) ^a	Microarray method (%)	S1 nuclease methods (%) ^a
<i>A. tolylyticus</i> Td-1	100	100	28.5	28
<i>A. tolylyticus</i> Td-2	93.9	99	34.9	33
<i>A. tolylyticus</i> BL-11	51.4	56	15.2	28
<i>A. tolylyticus</i> Td-17	29.0	18	89.7	89
<i>A. tolylyticus</i> Td-21	37.1	28	100	100
<i>A. tolylyticus</i> Td-3	26.6	21	29.1	30
<i>A. tolylyticus</i> Td-19	31.8	32	28.6	32
Correlation to microarray method	—	0.99	—	0.99

^aDNA-DNA reassociation value obtained using S1 nuclease method by Song *et al.* (1999).

strains capable of aromatic compound utilization under denitrifying conditions into different genomic groups or new species (Zhou *et al.*, 1995; Song *et al.*, 1999). The strains Td-17 and Td-21 were considered to be well separated from genomic group 1 of *A. tolylyticus* and were reclassified as a new species (that is, *A. toluvorans*) based

on the DNA similarities measured by DNA-DNA hybridization using S1 method and SSU rRNA gene similarities (Song *et al.*, 1999). The CGA-based hybridization is able to delineate these differences as well. As shown in Figure 1, strain Td-17 and Td-21 were distinctly separated from Td-1.

Comparison of the CGA hybridization-derived similarities to SSU rRNA gene sequence similarities

The SSU rRNA genes are commonly used as markers for establishing phylogenetic relationships among the *Bacteria* and *Archaea*. To further understand whether the CGA-based hybridization could be used to determine species relatedness, the relationships between the CGA hybridization-derived similarities and the nucleotide sequence similarities of SSU rRNA genes were examined. Overall, a significant linear relationship was obtained between the DNA similarities derived from the CGA-based hybridization and the similarities based on SSU rRNA gene sequences ($n=54$, $r=0.79$, $P<0.0001$; Figure 3a). Such linear relationships became lower for those microorganisms exhibiting SSU rRNA gene sequence similarities of $>92\%$ ($n=47$, $r=0.77$, $P<0.0001$, Supplementary Table SM3). This is because the SSU rRNA gene sequences of the *A. tolytolicus* strains with distant DNA relatedness are almost identical (Song *et al.*, 1999). However, the relationships between DNA similarities derived from the CGA-based hybridization and those from SSU rRNA genes vary considerably among different individual reference strains (Supplementary Table SM_3). Also, the tree topology based on the CGA-hybridization data is more or less congruent with that based on SSU rRNA gene sequences (Supplementary Figure SM1 and Figure 1), although noticeable differences were observed at high taxonomic levels and fine-scale strain levels. These results also indicated that the percent similarity value based on the CGA hybridization could be a good predictor of fine-scale phylogenetic distances among different microorganisms.

Previous studies showed that the 70% DNA-DNA reassociation value corresponds roughly to 97% homology based on SSU rRNA genes (Stackebrandt and Goebel, 1994). This value is generally used to define a new prokaryotic species based on SSU rRNA gene sequences (Venter *et al.*, 2004; Harayama and Kasai, 2006). From the data available in this study, the 97% homology of SSU rRNA genes corresponds to about 47% similarity based on CGA hybridization data under the conditions examined (Figure 3a). This result also indicated that the 97% homology of SSU rRNA genes as a cutoff for defining new species is a very conservative estimation.

Comparison of DNA similarities from CGA-based hybridizations to gyrB gene sequence similarities

Although SSU rRNA gene sequences are widely used as a phylogenetic marker for bacterial systematics and ecology, it is difficult to use them to obtain fine-scale resolution at species and strain levels (Yamamoto and Harayama, 1995), because the rate of evolution of this molecule is low. An alternative phylogenetic marker, *gyrB*, which encodes the β -subunit of DNA gyrase, has been widely used for differentiating closely related species/strains

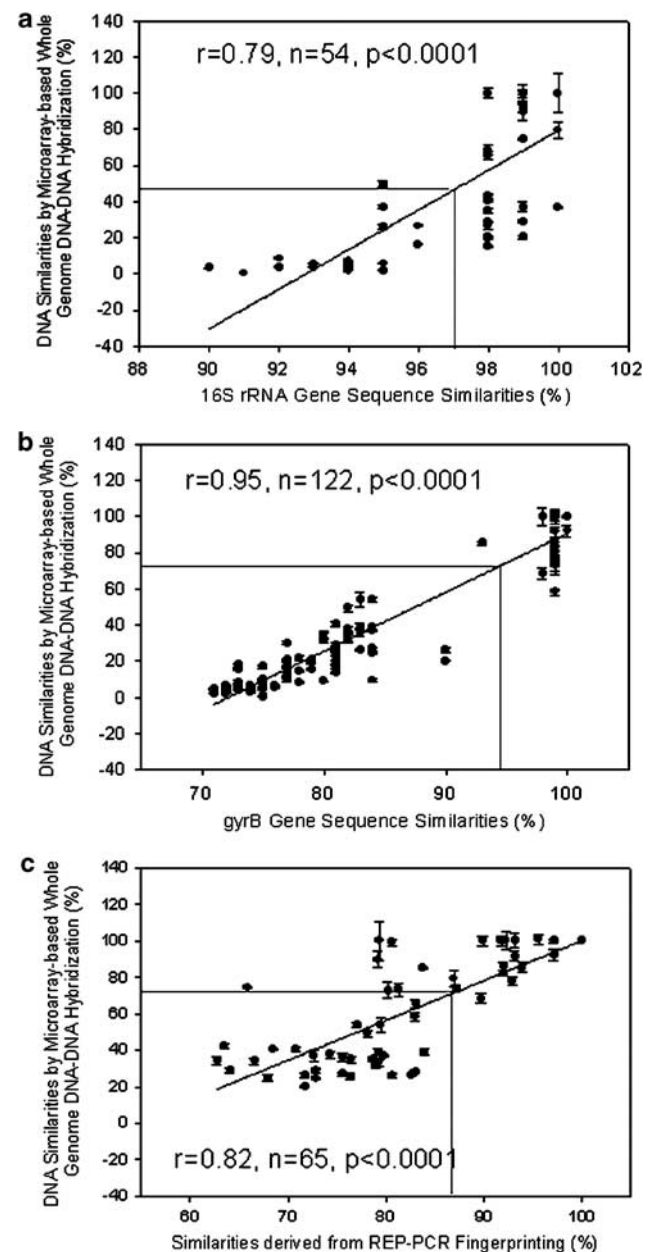


Figure 3 Relationship between DNA similarities by the CGA-based hybridization and percent similarity in nucleotide sequences for SSU rRNA and *gyrB* genes and fingerprinting patterns. DNA similarities was compared to (a) the SSU rRNA gene sequence similarities in the range of $>90\%$; (b) the *gyrB* gene sequence similarities in the range of $>70\%$; and (c) the REP-PCR similarities (combined REP-PCR and BOX-PCR), for the closely related strains.

(Yamamoto and Harayama, 1995, 1996, 1998; Yamamoto *et al.*, 1999; Rodrigues *et al.*, 2003; Yan *et al.*, 2003; Holmes *et al.*, 2004; Maeda *et al.*, 2006). A phylogenetic tree based on *gyrB* sequences results in a 10-fold greater resolution than a tree based on SSU rRNA genes. To further understand whether the CGA-based hybridization could be used to determine species relatedness, the relationships between the CGA hybridization-based similarities

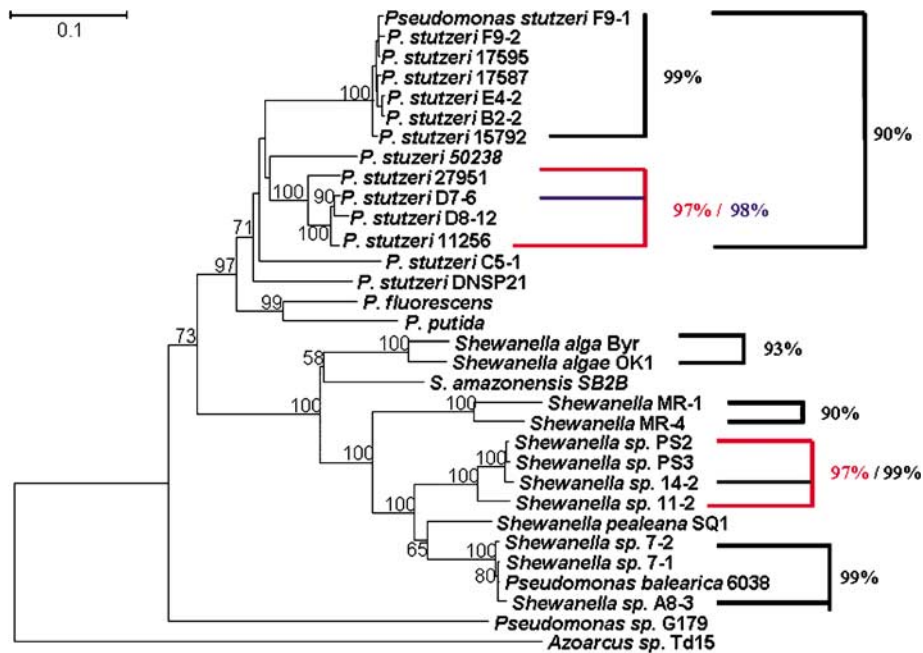


Figure 4 Phylogenetic relationship of *gyrB* gene nucleotide sequences of *Shewanella* and *Pseudomonas* strains (partial, 780 bp). The dendrogram was generated by phylogenetic distance analysis with a neighbor-joining algorithm with *Azoarcus sp. Td-15* as the outgroup. Values indicate the percentage of 100 replicate trees supporting the branching order. Bootstrap values below 50 were omitted. Scale bar, 10 mutations per 100 sequence positions. The percentage and color indicate the nucleotide sequence similarities among the strains in groups.

and the nucleotide sequence similarities of *gyrB* genes were also compared at the species/strain levels. Significant overall linear relationships ($n = 122$, $r = 0.95$, $P < 0.0001$) between the CGA hybridization-based similarities and the *gyrB* gene sequence similarities were observed for strains that exhibited $> 70\%$ similarities of *gyrB* gene sequences (Figure 3b). The comparisons were also made for the hybridizations with individual reference strains. All correlations ($n = 13-17$, $r = 0.88-0.96$, $P < 0.01$; Supplementary Table_SM3) were significant, although such correlations vary among different reference strains. Based on the data available in this study, the 70% DNA-DNA reassociation values for various *Shewanella* and *Pseudomonas* strains corresponds to about 94% homology based on *gyrB* genes, which is also equivalent to 71% similarity based on CGA hybridization data under the conditions examined (Figure 3b).

The phylogenetic trees based on *gyrB* gene sequences for *Shewanella* and *Pseudomonas* strains were constructed (Figure 4). The tree topology based on *gyrB* sequences is comparable to that derived from the CGA-based hybridizations. For instance, in the *gyrB* gene tree, *Pseudomonas sp.* G179 is distant from other *Pseudomonas* strains, and a similar result was obtained with the CGA-based DNA similarity trees (Figure 1). *Pseudomonas stutzeri* strains 17595, 17587 and 17592 grouped in the same cluster, which is also found in the DNA similarity tree. *Pseudomonas stutzeri* strain 27951 grouped in the same branch with strain 11256, as revealed by

both the *gyrB* gene and DNA similarity tree. In both *gyrB* gene and DNA similarity tree, *S. algae* strains BrY and OK-1 grouped together. Also both trees revealed the close relationship between *S. oneidensis* MR-1 and *Shewanella sp.* MR-4 (Figures 1 and 4). Most of the *Shewanella* marine isolates clustered in a similar way in both the *gyrB* gene and DNA similarity trees, although there were fine differences (Figures 1 and 4).

Comparison of CGA-based similarities to those derived from genomic fingerprinting (REP-PCR and BOX-PCR)
REP-PCR and BOX-PCR (repetitive DNA sequence of the BOXA subunit of the BOX element of *Streptococcus pneumoniae*) are commonly used methods for differentiating bacterial strains (Versalovic *et al.*, 1994; Rademaker *et al.*, 2000). To further understand whether different bacterial strains could be differentiated using CGA-based hybridization, the relationship between the CGA hybridization-derived similarities and the similarities based on combined REP-PCR and BOX-PCR fingerprint patterns were examined for closely related *Pseudomonas*, *Shewanella* and *Azoarcus* strains. Similarly, a linear relationship was obtained for these two methods when the REP-PCR and BOX-PCR values were above 60% ($n = 65$, $r = 0.82$, $P < 0.0001$) (Figure 4c). The linear relationship was stronger for the strains within species ($n = 65$, $r = 0.82$, $P < 0.0001$, Figure 3c) than for relationships within a genus ($n = 76$, $r = 0.73$, $P < 0.01$, data not shown).

The results suggested that the CGA-based hybridization could distinguish closely related bacterial strains of *Pseudomonas*, *Shewanella* and *Azoarcus* species. Based on the data available in this study, the 70% DNA–DNA reassociation values for various *Shewanella*, *Pseudomonas* and *Azoarcus* strains corresponded to about 87% similarity based on genome fingerprinting, which is also equivalent to 71% similarity based on CGA hybridization data under the conditions examined (Figure 3c).

Discussion

Determination of procaryotic species and the degree of their relationship is a great challenge for microbiologists. In the last 50 years, many different molecular methods, including whole-genome DNA–DNA hybridization, SSU rRNA sequencing, multiple locus sequencing of protein encoding genes (for example, *gyrB*, *recA*) and average nucleotide identity, have been proposed for delineating bacterial species. Although the SSU rRNA gene-based method is a valuable, convenient and rapid tool for the determination of the phylogenetic relationships among different microorganisms, it provides poor resolution at the species and subspecies levels (Yamamoto and Harayama, 1998). Also, many procaryotic species have virtually identical SSU rRNA gene sequences but only have 25% DNA similarity (Stackebrandt and Goebel, 1994). Thus, the SSU rRNA analysis method is not a valid approach for determining species/strain relationships. Protein-coding genes could provide high resolution for species/strain determination, but the difficulty in using sets of protein coding genes for phylogenetic evaluation lies in selecting appropriate gene targets and designing amplification primers useful for large sets of microorganisms (Harayama and Kasai, 2006). In addition, phylogenetic analyses based on complete microbial genome sequences are possible, but the likelihood that all the sequenced genomes needed for comparison will be available is not feasible (Brutlag, 1998). Finally, whole-genome sequence analysis is a powerful approach for resolving the major problems of evolution, phylogeny and systematics of living organisms, but its use in general taxonomic studies is not currently practical (Tourova, 2000). Such analyses will require larger genomic data sets and more carefully designed sampling of natural populations (Konstantinidis and Tiedje, 2007). The average nucleotide identity of the shared genes between two strains was proposed to be a robust approach to determine genetic relatedness among different strains (Konstantinidis and Tiedje, 2004). The average nucleotide identity value of 94% corresponded to the traditional 70% DNA–DNA reassociation standard of the current species definition. Although the average nucleotide identity approach is simple, it still relies on the availability

of whole-genome sequences and hence it will have a limited use. Nevertheless, whole-genome DNA–DNA hybridization is still considered to be the cornerstone for bacterial species determination and will have to be used to circumscribe procaryotic species (Rossello-Mora, 2006).

The development and application of microarray-based genomic technology for microbial detection and community analysis have received a great deal of attention. Because of its high-density and high-throughput capacity, it is expected that microarray-based genomic technologies will revolutionize the detection, identification and characterization of microorganisms. Therefore, in this study, we have developed the CGA-based hybridization approach for determining species relationships. Experimental comparisons of the CGA hybridization-based results with available traditional DNA–DNA hybridization data, SSU rRNA and *gyrB* gene sequences and genome fingerprinting methods indicate that the CGA-based hybridization could be a useful alternative to the traditional whole-genome DNA–DNA hybridization approaches for determining procaryotic species relationships.

Overall, DNA similarities from the CGA-based hybridizations were comparable to those from various conventional whole-genome DNA–DNA hybridization approaches. When the actual values for genome relatedness were compared between different methods, the results from the CGA-based hybridization were more consistent to those from the S1 method than the membrane filter method (Goris *et al.*, 1998), as indicated by smaller average differences (<15%) between the similarity values derived from CGA-based hybridization and S1 methods for *A. tolulyticus* strains. DNA similarities from the CGA-based hybridizations for reference strain *P. stutzeri* ATCC 17587 also matched well with the ΔT_m values of the *P. stutzeri* strains by hydroxyapatite and/or dot filter method with the reference strain *P. stutzeri* ATCC 17591 (identical to 17587). However, DNA similarities from CGA-based hybridizations with multiple *Pseudomonas* strains were significantly lower than those from membrane filter methods, although strong linear relationships were observed. One possible explanation might be related to differences in hybridization stringency. Higher similarity values are expected if the hybridization is carried out at relatively low stringent conditions. The hybridization conditions may need to be optimized for different target microbial groups by considering GC content and genome size. It is also important to point out that DNA similarities determined by the traditional hybridization methods vary significantly among different methods (Goris *et al.*, 1998).

Although significant relationships between the CGA hybridization-based similarities and those from SSU rRNA genes, *gyrB* genes or genomic fingerprinting were observed, the degree of correlations is considerably different. The

correlations of the CGA hybridization-derived similarities to *gyrB* sequences are stronger than those to SSU rRNA sequences and genomic fingerprinting. It appears that the taxonomic resolution of the CGA-based hybridization is similar to or slightly higher than *gyrB* sequence analysis. While SSU rRNA sequence analysis can provide reliable information about species relationships at higher taxonomic levels (for example, genus or above), whole-genome DNA–DNA hybridizations are useful in providing insight into phylogenetic relationships at the species/strain levels. For example, the SSU rRNA gene sequence similarities among the *A. toluityticus* strains tested in this study are all over 98%–100%; however, the DNA similarities determined by both the CGA-based method and S1 method varied across a wide range (15.2%–93.9% for the CGA method, 18%–99% for S1 method) (Song *et al.*, 1999). Owing to differences in resolving phylogenetic relationships, integrating CGA-based hybridization with SSU rRNA and *gyrB* gene sequence analysis could provide a reliable, rapid approach for delineating procaryotic species relationships.

Genome fingerprinting analysis is suitable for the elucidation of strain-level relationships (Versalovic *et al.*, 1994), and was shown to be highly correlated to DNA–DNA reassociation values for xanthomonads (Rademaker *et al.*, 2000). In this study, strong correlations between CGA hybridization-based similarities and the similarities derived from fingerprinting approaches were observed among closely related strains of *P. stutzeri* and *A. toluityticus*, but not among distantly related species from *Pseudomonas*, *Azoarcus* or *Shewanella* genera. The genomic fingerprinting similarity values above 60% correlated well with CGA hybridization values for the tested *P. stutzeri* and *A. toluityticus* strains. These results indicated that CGAs could provide meaningful insight into relationships between closely related strains. But the power of phylogeny to resolve relationships at the strain level will be lower using CGA-based hybridization than genome fingerprinting approaches. For instance, based on the CGA hybridization results, *A. toluityticus* Td-3 was not separated from Td-19, and *P. stutzeri* DNSP21 (genetic group IV) could not be separated from *P. stutzeri* ATCC 11256 (genetic group I) (Figure 1), but they were well separated based on genome fingerprinting methods. Lower resolution was also observed for some *Shewanella* species (data not shown).

Compared to the traditional DNA–DNA reassociation approach, CGAs have several advantages for the determination of species relatedness, including high-throughput capacity, parallel analyses and quantitation. CGA hybridization differs from membrane filter-based hybridization approaches in that the non-porous surface has advantages of miniaturization, hybridization kinetics, sample volume, reagent absorption, signal detection approaches

and reproducibility (Schna and Davis, 2000). The capability of accurate and precise miniaturization with robots on non-porous substrates with the use of fluorescence-based detection offers significant advantages. In addition, multiple pairwise comparisons can be done with smaller amounts of genomic DNA (that is, 1 µg). This is important for determining the relationships between procaryotic species that are difficult to cultivate. CGAs could provide a high-throughput means for rapid identification of microbial species/strains. Because of its high capacity, one can construct a CGA containing bacterial type strains plus appropriately related strains. By hybridizing genomic DNA from unknown strains with this type of microarray, one should be able to quickly and reliably identify unknown strains, provided a suitably related probe is on the array. Generally, SNRs for hybridizations with perfect match DNAs are significantly higher than those with mismatch DNAs from other strains of the same species (Wu *et al.*, 2004). Thus, species identification can be achieved based on the differences in hybridization intensity. However, as a low level of cross-hybridization could occur among different strains, establishing appropriate SNR thresholds to differentiate self-hybridization, cross-hybridization and background hybridization among different strains should be useful. In addition, when using CGAs for species identification, lower stringent hybridization conditions (for example, 42 °C and 50% formamide) should be used first to ensure that good hybridization signals can be obtained for distantly related target species.

Microbial diversity is extremely high and the majority of microorganisms are as-yet uncultivable. This could be a limitation in using CGA-based hybridization to determine species relatedness of uncultured microorganisms. However, the CGA-based hybridization itself does not require culturing. With the recent advances in environmental genomics, high-molecular-weight DNA from uncultivated microorganisms could be accessed through bacterial artificial chromosomes or fosmid cloning. High-molecular-weight bacterial artificial chromosomes/fosmid clones could also be used to fabricate CGAs, thus allowing the determination of relationships of target strain/clones to the uncultivated components of a complex microbial community. Because the size of bacterial artificial chromosomes/fosmid clones is generally 50- to 200-fold less than that for an entire genome, it is expected that microarrays fabricated with high-molecular-weight bacterial artificial chromosomes/fosmid clones should have similar performance characteristics as CGAs.

Acknowledgements

This research was supported by The United States Department of Energy under Genomics: GTL program through the Virtual Institute of Microbial Stress and

Survival (VIMSS; <http://vimss.lbl.gov>), and the Environmental Remediation Science Program (ERSP), and was supported by Oklahoma Center for the Advancement of Science and Technology under Oklahoma Applied Research Support Program. A part of XL's efforts was supported by Chinese National Science Foundation (NSFC) Joint Research Fund for Overseas Chinese Young Scholars (30428014), NSFC Fund for Creative Research (code 50321402) and Chinese National Basic Research (2004CB619201).

References

- Adnan S, Li N, Miura H, Hashimoto Y, Yamamoto H, Ezaki T. (1993). Covalently immobilized DNA plate for luminometric DNA-DNA hybridization to identify viridans streptococci in under 2 h. *FEMS Microbiol Lett* **106**: 139–142.
- Amann RI, Lin CH, Key R, Montgomery L, Stahl DA. (1992). Diversity among fibrobacter isolates—towards a phylogenetic classification. *Syst Appl Microbiol* **15**: 23–31.
- Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S *et al.* (1999). Comparative genomics of BCG vaccines by whole-genome DNA microarray. *Science* **284**: 1520–1523.
- Bodrossy L, Stralis-Pavese N, Murrell JC, Radajewski S, Weilharter A, Sessitsch A. (2003). Development and validation of a diagnostic microbial microarray for methanotrophs. *Environ Microbiol* **5**: 566–582.
- Braker G, Zhou JZ, Wu LY, Devol AH, Tiedje JM. (2000). Nitrite reductase genes (*nirK* and *nirS*) as functional markers to investigate diversity of denitrifying bacteria in Pacific northwest marine sediment communities. *Appl Environ Microbiol* **66**: 2096–2104.
- Brenner DJ, Fanning GR, Rake AV, Johnson KE. (1969). Batch procedure for thermal elution of DNA from hydroxyapatite. *Anal Biochem* **28**: 447–459.
- Brutlag DL. (1998). Genomics and computational molecular biology. *Curr Opin Microbiol* **1**: 340–345.
- Cardinali G, Liti G, Martini A. (2000). Non-radioactive dot-blot DNA reassociation for unequivocal yeast identification. *Int J Syst Evol Microbiol* **50**: 931–936.
- Cho JC, Tiedje JM. (2001). Bacterial species determination from DNA-DNA hybridization by using genome fragments and DNA microarrays. *Appl Environ Microbiol* **67**: 3677–3682.
- Cladera AM, Bennasar A, Barcelo M, Lalucat J, Garcia-Valdes E. (2004). Comparative genetic diversity of *Pseudomonas stutzeri* genomovars, clonal structure, and phylogeny of the species. *J Bacteriol* **186**: 5239–5248.
- Cohan FM. (2002). What are bacterial species? *Annu Rev Microbiol* **56**: 457–487.
- Crosa JH, Brenner DJ, Falkow S. (1973). Use of a single-strand specific nuclease for analysis of bacterial and plasmid deoxyribonucleic acid homoduplexes and heteroduplexes. *J Bacteriol* **115**: 904–911.
- De Lay J, Tijtgat R. (1970). Evaluation of membrane filter methods for DNA-DNA hybridization. *Antonie Van Leeuwenhoek J Microbiol* **36**: 461–474.
- DeRisi JL, Iyer VR, Brown PO. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**: 680–686.
- Dong YM, Glasner JD, Blattner FR, Triplett EW. (2001). Genomic interspecies microarray hybridization: rapid discovery of three thousand genes in the maize endophyte, *Klebsiella pneumoniae* 342, by microarray hybridization with *Escherichia coli* K-12 open reading frames. *Appl Environ Microbiol* **67**: 1911–1921.
- Eisen MB, Spellman PT, Brown PO, Botstein D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* **95**: 14863–14868.
- 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.
- Gade D, Schlesner H, Glockner FO, Amann R, Pfeiffer S, Thomm A. (2004). Identification of planctomycetes with order-, genus-, and strain-specific 16S rRNA-targeted probes. *Microb Ecol* **47**: 243–251.
- Gao HC, Wang Y, Liu XD, Yan TF, Wu LY, Alm E *et al.* (2004). Global transcriptome analysis of the heat shock response of *Shewanella oneidensis*. *J Bacteriol* **186**: 7796–7803.
- Garrity GM, Bell JA, Lilburn TG. (2004). Taxonomic outline of the Prokaryotes. In: Boone DR, Castenholz RW, Garrity GM (eds). *Bergey's Manual of Systematic Bacteriology*. Springer: Berlin, Heidelberg, New York, pp 163–164.
- Goris J, Suzuki K, De Vos P, Nakase T, Kersters K. (1998). Evaluation of a microplate DNA-DNA hybridization method compared with the initial renaturation method. *Can J Microbiol* **44**: 1148–1153.
- Harayama S, Kasai H. (2006). Bacterial phylogeny reconstruction from molecular sequences. In: Stackebrandt E (ed). *Molecular Identification, Systematics, and Population Structure of Prokaryotes*. Springer: New York, pp 105–140.
- Holmes DE, Nevin KP, Lovley DR. (2004). Comparison of 16S rRNA, *nifD*, *recA*, *gyrB*, *rpoB* and *fusA* genes within the family Geobacteraceae fam. nov. *Int J Syst Evol Microbiol* **54**: 1591–1599.
- Kato-Maeda M, Rhee JT, Gingeras TR, Salamon H, Drenkow J, Smittipat N *et al.* (2001). Comparing genomes within the species *Mycobacterium tuberculosis*. *Genome Res* **11**: 547–554.
- Kaznowski A. (1995). A method of colorimetric DNA-DNA hybridization in microplates with covalently immobilized DNA for identification of *Aeromonas* spp. *Med Microbiol Lett* **4**: 362–369.
- Konstantinidis KT, Tiedje JM. (2004). Trends between gene content and genome size in prokaryotic species with larger genomes. *Proc Natl Acad Sci USA* **101**: 3160–3165.
- Konstantinidis KT, Tiedje JM. (2007). Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead. *Curr Opin Microbiol* **10**: 504–509.
- Kumar S, Tamura K, Nei M. (2003). MEGA3: an integrated software for molecular evolutionary genetic analysis and sequence alignment. *Integr Comp Biol* **43**: 947.
- Liu YQ, Zhou JZ, Omelchenko MV, Beliaev AS, Venkateswaran A, Stair J *et al.* (2003). Transcriptome dynamics of *Deinococcus radiodurans* recovering from ionizing radiation. *Proc Natl Acad Sci USA* **100**: 4191–4196.
- Lockhart DJ, Dong HL, Byrne MC, Follettie MT, Gallo MV, Chee MS *et al.* (1996). Expression monitoring by

- hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* **14**: 1675–1680.
- Loy A, Lehner A, Lee N, Adamczyk J, Meier H, Ernst J *et al.* (2002). Oligonucleotide microarray for 16S rRNA gene-based detection of all recognized lineages of sulfate-reducing prokaryotes in the environment. *Appl Environ Microbiol* **68**: 5064–5081.
- Maeda Y, Shinohara H, Kiba A, Ohnishi K, Furuya N, Kawamura Y *et al.* (2006). Phylogenetic study and multiplex PCR-based detection of *Burkholderia plantarii*, *Burkholderia glumae* and *Burkholderia gladioli* using *gyrB* and *rpoD* sequences. *Int J Syst Evol Microbiol* **56**: 1031–1038.
- Mehlen A, Goeldner M, Ried S, Stindl S, Ludwig W, Schleifer KH. (2004). Development of a fast DNA–DNA hybridization method based on melting profiles in microplates. *Syst Appl Microbiol* **27**: 689–695.
- Murray AE, Lies D, Li G, Neelson K, Zhou J, Tiedje JM. (2001). DNA/DNA hybridization to microarrays reveals gene-specific differences between closely related microbial genomes. *Proc Natl Acad Sci USA* **98**: 9853–9858.
- Palleron NJ, Doudorof M, Stanier RY, Solanes RE, Mandel M. (1970). Taxonomy of aerobic pseudomonads—properties of *Pseudomonas-stutzeri* group. *J Gen Microbiol* **60**: 215–231.
- Popoff M, Coynault C. (1980). Use of deae-cellulose filters in the s1 nuclease method for bacterial deoxyribonucleic-acid hybridization. *Ann Microbiol A* **131**: 151–155.
- Rademaker JLW, Hoste B, Louws FJ, Kersters K, Swings J, Vauterin L *et al.* (2000). Comparison of AFLP and rep-PCR genomic fingerprinting with DNA–DNA homology studies: *Xanthomonas* as a model system. *Int J Syst Evol Microbiol* **50**: 665–677.
- Rhee SK, Liu XD, Wu LY, Chong SC, Wan XF, Zhou JZ. (2004). Detection of genes involved in biodegradation and biotransformation in microbial communities by using 50-mer oligonucleotide microarrays. *Appl Environ Microbiol* **70**: 4303–4317.
- Rodrigues JLM, Silva-Stenico ME, Gomes JE, Lopes JRS, Tsai SM. (2003). Detection and diversity assessment of *Xylella fastidiosa* in field-collected plant and insect samples by using 16S rRNA and *gyrB* sequences. *Appl Environ Microbiol* **69**: 4249–4255.
- Rossello R, Garcavaldes E, Lalucat J, Ursing J. (1991). Genotypic and phenotypic diversity of *Pseudomonas-stutzeri*. *Syst Appl Microbiol* **14**: 150–157.
- Rossello-Mora R. (2006). DNA–DNA reassociation methods applied to microbial taxonomy and their critical evaluation. In: Stackebrandt E (ed). *Molecular Identification, Systematics, and Population Structure of Prokaryotes*. Springer Verlag: Berlin, Heidelberg, New York, pp 23–50.
- Rossello-Mora R, Amann R. (2001). The species concept for prokaryotes. *Fems Microbiol Rev* **25**: 39–67.
- Salama N, Guillemin K, McDaniel TK, Sherlock G, Tompkins L, Falkow S. (2000). A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proc Natl Acad Sci USA* **97**: 14668–14673.
- Schadt CW, Liebich J, Chong SC, Gentry TJ, He ZL, Pan HB *et al.* (2005). Design and use of functional gene microarrays (FGAs) for the characterization of microbial communities. In: Savidge T, Pothulakis C (eds). *Microbial Imaging*. Elsevier Academic Press: Amsterdam, pp 331–368.
- Schena M, Davis RW. (2000). Technology standards for microarray research. In: Schena M (ed). *Microarray Biochip Technology*. Eaton Publishing: Natick, pp 301–306.
- Schena M, Shalon D, Davis RW, Brown PO. (1995). Quantitative monitoring of gene-expression patterns with a complementary-DNA Microarray. *Science* **270**: 467–470.
- Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW. (1996). Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci USA* **93**: 10614–10619.
- Sepulveda-Torres LD, Zhou JZ, Guasp C, Lalucat J, Knaebel D, Plank JL *et al.* (2001). *Pseudomonas* sp strain KC represents a new genomovar within *Pseudomonas stutzeri*. *Int J Syst Evol Microbiol* **51**: 2013–2019.
- Song B, Haggblom MM, Zhou JZ, Tiedje JM, Palleroni NJ. (1999). Taxonomic characterization of denitrifying bacteria that degrade aromatic compounds and description of *Azoarcus toluvorans* sp nov and *Azoarcus toluclasticus* sp nov. *Int J Syst Bacteriol* **49**: 1129–1140.
- Stackebrandt E. (2006). Exciting times: the challenge to be a bacterial systematist. In: Stackebrandt E (ed). *Molecular Identification, Systematics, and Population Structure of Prokaryotes*. Springer: New York, pp 1–16.
- Stackebrandt E, Frederiksen W, Garrity GM, Grimont PAD, Kampf P, Maiden MCJ *et al.* (2002). Report of the *ad hoc* committee for the re-evaluation of the species definition in bacteriology. *Int J Syst Evol Microbiol* **52**: 1043–1047.
- Stackebrandt E, Goebel BM. (1994). A Place for DNA–DNA reassociation and 16 s ribosomal-RNA sequence-analysis in the present species definition in bacteriology. *Int J Syst Bacteriol* **44**: 846–849.
- Stapleton RD, Sabree ZL, Palumbo AV, Moyer CL, Devol AH, Roh Y *et al.* (2005). Metal reduction at cold temperatures by *Shewanella* isolates from various marine environments. *Aquat Microb Ecol* **38**: 81–91.
- Thompson JD, Higgins DG, Gibson TJ. (1994). Clustal-W—improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**: 4673–4680.
- Tiquia SM, Wu LY, Chong SC, Passovets S, Xu D, Xu Y *et al.* (2004). Evaluation of 50-mer oligonucleotide arrays for detecting microbial populations in environmental samples. *Biotechniques* **36**: 664–670.
- Tourova TP. (2000). The role of DNA–DNA hybridization and 16S rRNA gene sequencing in solving taxonomic problems by the example of the order Haloanaerobiales. *Microbiology* **69**: 623–634.
- Venkateswaran K, Moser DP, Dollhopf ME, Lies DP, Saffarini DA, MacGregor BJ *et al.* (1999). Polyphasic taxonomy of the genus *Shewanella* and description of *Shewanella oneidensis* sp. nov. *Int J Syst Bacteriol* **49**: 705–724.
- Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA *et al.* (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304**: 66–74.
- Verdnik D, Handran S, Pickett S. (2002). Key considerations for accurate microarray scanning and image

- analysis. In: Shah GKas (ed). *DNA Array Image Analysis Nuts and Bolts*. DNA Press: Eagleville, PA, pp 83–98.
- Versalovic J, Schneider M, de Bruijn FJ, Lupski JR. (1994). Genomic fingerprinting of bacteria using repetitive sequence-based polymerase chain reaction. *Methods Mol Cell Biol* **5**: 25–40.
- Wayne LG, Brenner DJ, Colwell RR, Grimont PAD, Kandler O, Krichevsky MI *et al*. (1987). Report of the *ad hoc* committee on reconciliation of approaches to bacterial systematics. *Int J Syst Bacteriol* **37**: 463–464.
- Whitman WB, Coleman DC, Wiebe WJ. (1998). Prokaryotes: the unseen majority. *Proc Natl Acad Sci USA* **95**: 6578–6583.
- Wu LY, Thompson DK, Li GS, Hurt RA, Tiedje JM, Zhou JZ. (2001). Development and evaluation of functional gene arrays for detection of selected genes in the environment. *Appl Environ Microbiol* **67**: 5780–5790.
- Wu LY, Thompson DK, Liu XD, Fields MW, Bagwell CE, Tiedje JM *et al*. (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.
- Yamamoto S, Bouvet PJM, Harayama S. (1999). Phylogenetic structures of the genus *Acinetobacter* based on *gyrB* sequences: comparison with the grouping by DNA–DNA hybridization. *Int J Syst Bacteriol* **49**: 87–95.
- 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.
- Yamamoto S, Harayama S. (1996). Phylogenetic analysis of *Acinetobacter* strains based on the nucleotide sequences of *gyrB* genes and on the amino acid sequences of their products. *Int J Syst Bacteriol* **46**: 506–511.
- Yamamoto S, Harayama S. (1998). Phylogenetic relationships of *Pseudomonas putida* strains deduced from the nucleotide sequences of *gyrB*, *ropD* and 16S rRNA genes. *Int J Syst Bacteriol* **48**: 813–819.
- Yan TF, Fields MW, Wu LY, Zu YG, Tiedje JM, Zhou JZ. (2003). Molecular diversity and characterization of nitrite reductase gene fragments (*nirK* and *nirS*) from nitrate- and uranium-contaminated groundwater. *Environ Microbiol* **5**: 13–24.
- Ye RW, Tao W, Bedzyk L, Young T, Chen M, Li L. (2000). Global gene expression profiles of *Bacillus subtilis* grown under anaerobic conditions. *J Bacteriol* **182**: 4458–4465.
- Zhang L, Hurek T, Reinhold-Hurek B. (2007). A *nifH*-based oligonucleotide microarray for functional diagnostics of nitrogen-fixing microorganisms. *Microb Ecol* **53**: 456–470.
- Zhou JH. (2003). Microarrays for bacterial detection and microbial community analysis. *Curr Opin Microbiol* **6**: 288–294.
- Zhou JZ, Fries MR, Cheesanford JC, Tiedje JM. (1995). Phylogenetic analyses of a new group of denitrifiers capable of anaerobic growth on toluene and description of *Azoarcus toluolyticus* sp. nov.. *Int J Syst Bacteriol* **45**: 500–506.
- Ziemke F, Hofle MG, Lalucat J, Rossello-Mora R. (1998). Reclassification of *Shewanella putrefaciens* Owen's genomic group II as *Shewanella baltica* sp. nov. *Int J Syst Bacteriol* **48**: 179–186.

Supplementary Information accompanies the paper on The ISME Journal website (<http://www.nature.com/ismej>)