Title: Analysis of microbial community composition using oligonucleotide fingerprinting of ribosomal RNA genes

Running Title: Oligonucleotide fingerprinting of ribosomal RNA genes

Lea Valinsky[1§], Gianluca Della Vedova[2§], Alexandra J Scupham[1], Sam Alvey[3], Andres Figueroa[4], Bei Yin[1], R. Jack Hartin[1], Marek Chrobak[4], David E. Crowley[3], Tao Jiang[4], and James Borneman[1*]

[1]Department of Plant Pathology, University of California, Riverside, CA 92521, USA. [2]Dipartimento di Statistica, Università degli Studi di Milano-Bicocca, via Bicocca degli Arcimboldi, 8, I-20126, Milano, Italy. [3]Department of Environmental Sciences, University of California, Riverside, CA 92521, USA. [4]Department of Computer Science and Engineering, University of California, Riverside, CA 92521, USA

[*]Corresponding author. Mailing address: Department of Plant Pathology, University of California, Riverside, 92506. Phone: 909-787-3584. Fax: 909-787-4294. E-mail: borneman@ucrac1.ucr.edu.

[§]These authors contributed equally to this work

ABSTRACT

One of the first steps in characterizing an ecosystem is to describe the organisms inhabiting it. For microbial studies, experimental limitations have hindered the ability to depict these diverse communities. Here we describe oligonucleotide fingerprinting of ribosomal RNA genes (OFRG), a method that permits the identification of arrayed ribosomal RNA genes (rDNA) through a series of hybridization experiments using small DNA probes. To demonstrate this strategy, we examined the bacteria inhabiting two different soils. Analysis of 1536 rDNA clones revealed 766 clusters grouped into 5 major taxa: *Bacillus*, *Actinobacteria*, *Proteobacteria* and two undefined assemblages. Soil-specific taxa were identified and then independently confirmed through cluster-specific PCR of the original soil DNA. Near species level resolution was obtained by this analysis as clones with average sequence identities of 97% were grouped into the same cluster. A comparison of these OFRG results with those obtained from a denaturing gradient gel electrophoresis (DGGE) analysis of the same two soils demonstrated the significance of this methodological advance. OFRG provides a cost effective means to extensively analyze microbial communities and should have application in medicine, biotechnology and ecosystem studies.

INTRODUCTION

How diverse are microbial communities? Does microbial diversity lead to ecosystem stability? What are the relationships between microbial community composition and ecosystem function? These questions as well as many others remain unresolved because of limitations with current experimental capabilities, including the difficulty of simply describing the organisms inhabiting an environment. Traditionally, microorganisms have been classified by characterizing their morphological and physiological traits in pure culture. However, such traits do not provide a meaningful framework for evolutionary classifications. Moreover, this approach is limited by its reliance on pure culture techniques, which detect only a fraction of extant microorganisms (1). In the 1970s, Woese and colleagues described the use of comparative ribosomal RNA (rRNA) analysis for phylogenetic studies (13, 24, 31, 32). This work not only provided an evolutionary basis for prokaryotic taxonomy, but it also led to the three-domain organization of the living world – Archaea, Bacteria, and Eucarya (31, 33). The subsequent development of strategies to analyze rRNA molecules and genes (rDNA) obtained from the environment provided a culture-independent means to examine the immense diversity of microorganisms inhabiting the natural world (3, 14, 21, 29).

Numerous rDNA-based strategies have been developed for microbial community analysis. Currently, most investigators analyze rDNA with methods such as denaturing gradient gel electrophoresis (DGGE) (20), terminal random fragment length polymorphisms (T-RFLP) (19), and ribosomal intergenic spacer analysis (RISA) (5). Although these methods permit rapid analysis of numerous samples, they generate only superficial descriptions of microbial community composition. Thorough depictions of community composition can be obtained by extensive sequence analysis of rDNA clone libraries. Yet, this approach is not commonly practiced because of the high costs associated with examining such diverse communities. The recent development of array based methods, which permit thousands of hybridization events to be examined in parallel, have brought great promise to the field of microbial ecology. In this approach, labeled rRNA or rDNA from environmental samples are analyzed by their hybridization to oligonucleotide probes attached to a substrate. While some successes have been reported (4, 15, 22, 23, 28), none of the methods described in these studies demonstrate the potential to facilitate thorough depictions of microbial community composition. The most significant unresolved obstacle at this juncture appears to be probe design. For this approach to work, each oligonucleotide probe must hybridize to a specific rDNA sequence or group of sequences. However, the development of such probes remains a significant challenge because of the highly conserved nature of rDNA and the extensive diversity of microbial life. An additional problem will be designing probes for the multitude of microorganisms that have yet to be described.

In this report, we describe oligonucleotide fingerprinting of ribosomal RNA genes (OFRG), an alternative array based approach in which the rDNA, not the oligonucleotide probes, are attached to a solid substrate. OFRG is an adaptation of a method used for gene expression profiling (9, 10, 18). Briefly, clone libraries are constructed from rDNA molecules that have been PCR amplified from environmental DNA (Fig. 1). The rDNA clones are then arrayed on nylon membranes and subjected to a series of hybridization experiments, each using a single DNA probe 10 nucleotides long. For every probe, the signal intensities are transformed into three discrete values 0, 1, and N, where 0 and 1 respectively specify negative and positive hybridization events and N designates an uncertain assignment. This process creates a hybridization fingerprint for each clone, which is a vector of values resulting from its hybridizations with all probes. The clones are identified by clustering their hybridization fingerprints with those of known sequences as well as by nucleotide sequence analyses of representative clones within a cluster.

MATERIALS AND METHODS

Soil collection and DNA extraction. Soil (top 10 cm) was collected from two adjacent agriculture fields at the Agriculture Research Station at the University of California, Riverside. Five replicate samples were collected from each site. The samples were dried by an overnight incubation at 30^{o}C and then stored at -70^{o}C. DNA was extracted from each soil sample (0.5 g) using the FastDNA Spin Kit for Soil (Bio101, Vista, CA) (7). DNA from the five replicate samples were pooled for analysis of bacterial composition as described below.

rDNA library construction. Bacterial small subunit rDNA were PCR amplified from soil DNA in 10 µl glass capillary tubes using a 1002 RapidCycler (Idaho Technologies, Idaho Falls, Idaho). The 100 µl reactions contained 50 mM Tris (pH 8.3), 500 µg/ml BSA, 2.5 mM $MgCl_2$, 250 µM of each dNTP, 400 nM of each bacterial SSU rDNA primer (27F, GAGCTCAGAGTTTGATCMTGGCTCAG and 1492R, CACGYTACCTTGTTACGACTT) (17), 5 U *Taq* DNA polymerase and 10 µl soil DNA, composed of equal volumes of DNA from each of the replicate soil samples. The cycling parameters were: 94°C for 2 min; 35 cycles of

94°C for 5 s, 48°C for 20 s, and 72°C for 40 s; followed by 72°C for 2 min. PCR products were gel isolated and purified with QIAquick PCR Purification Kit (Qiagen, Chatsworth, CA), ligated into pGEM-T (Promega), transformed into competent *Escherichia coli* JM109 (Promega), and plated on LB agar plates containing 100 µg/ml ampicillin that were surface spread with X-GAL and IPTG. For each soil treatment, 768 white colonies were randomly picked into 384-well culture plates, each well containing 30 µl LB (100 µg/ml ampicillin) except for the perimeter ones which were filled with 60 µl to prevent drying. For array construction (see below), these plates were shaken (300 rpm) overnight at 37°C in a plastic bag to avoid drying. For long-term storage, the bacteria were grown overnight at 37°C without shaking in a plastic bag, each well containing 30 µl LB broth (100 µg/ml ampicillin); the next day, the plates were stored at –70°C after adding 30 µl LB + 30% glycerol.

Array construction. The arrays were constructed by spotting PCR amplified rDNA onto nylon membranes. The 35 µl PCR reactions contained 50 mM Tris (pH 8.3), 500 µg/ml BSA, 2.5 mM $MgCl_2$, 250 µM of each dNTP, 400 nM of each primer (T725, GGCCCGACGTCGCATGCTC and SP650, TGGTCGACCTGCAGGCGGC, which anneal to regions of the multiple cloning site within the vector) and 1.75 U *Taq* DNA polymerase. (Note that rDNA-specific primers were not used here because they would have also amplified *E. coli* rDNA from the host cell.) The reagents (35 µl per well) were added to 384-well PCR plates (Marsh Bio Products, Rochester, NY). Freshly grown, overnight cultures (0.1 µl of each) of the rDNA clones (described in the previous paragraph) were added to the PCR reagents using a 384-pin solid pin replicator (V & P Scientific, Inc. San Diego, CA). The plates were sealed with Thermo-Seal foil (Marsh Bio Products, Rochester, NY) using a preheated Combi Thermo-sealer (Advanced Biotechnologies, Ltd, UK) for 4 s. PCR was then performed by alternately submerging the PCR plates in two water baths. The cycling parameters were: 94°C for 10 min; 35 cycles of 94°C for 1 min, 72°C for 2 min; followed by 72°C for 5 min. The PCR products were spotted onto dry 11 x 8 cm Hybond N+ membranes (Amersham Pharmacia Biotech) with a surfactant coated 0.5 µl slot pin replicator and a Multi-Print replication registration device (V&P Scientific, San Diego, CA). One µl of each PCR product was delivered to the membrane by 2 sequential spotting applications. The Multi-Print device allows the contents of four different 384-well plates to be printed onto a single 11 x 8 cm membrane, resulting in an array of 1536 clones.

Array hybridization. The membranes were fixed by UV crosslinking (70 mJ). Immediately before hybridization, the membranes were denatured with 0.5 N NaOH/1.5 M NaCl (2 x 5 min on absorbent paper), neutralized with 50 mM Na phosphate pH 7.2 (3 x 3 min on absorbent paper), covered with boiling 0.1% SDS, allowed to cool for 15 min, and then dried for 30 min. DNA oligonucleotides were end labeled with T4 Polynucleotide Kinase (T4 PNK) (New England Biolabs); the 4 µl reactions contained 1 µM oligonucleotide, 6 µCi gamma^{33}P–ATP, 0.4 µl 10X T4 PNK buffer, 2.6 U T4 PNK and were incubated at 37°C for 30 min. Hybridization solution (1.8 ml 5% sarcosyl/0.2 M Na phosphate pH 7.2 containing 1 nM ^{33}P end-labeled DNA oligonucleotide probe) was applied to the membranes, covered with plastic sheeting (102 µm), and incubated overnight at 12°C (11). Arrays were washed twice in SSC (0.1-4X) for 5-30 min at 12°C (12); the washing conditions were determined empirically for each probe. After washing, membranes were briefly placed on absorbent paper to remove any excess fluids and then enclosed with plastic wrap to prevent drying. The membranes were exposed to an Imaging Screen (BioRad) for 16 hours and then scanned with a Personal Molecular Imager FX (BioRad).

The signal intensities with background correction were obtained using ImaGene 4.0 software (Biodiscovery). Membranes were reused up to five times. To remove the probe between experiments, the membranes were covered with boiling Stripping Buffer (1X SSC, 0.1% SDS, 200 mM Tris pH 7.5), allowed to cool for 15 min, and then dried for 30 min.

Oligonucleotide probes. The discriminating oligonucleotide probes were: 1, GTTGGGTTAA; 2, GTAACCTGCC; 3, GAAAGCCTGA; 4, AATTCGATGC; 5, TTCGGATTGT; 6, CGAAAGCGTG; 7, CGGCCCAGAC; 8, TTGATCCTGG; 9, CACATGCAAG; 10, GGTAATGGCC; 11, GGGCGCAAGC; 12, TGAAATGCGT; 13, ATTCGATGCA; 14, GCAAGCCTGA; 15, TCAGTTCGGA; 16, GAGGATGGCC; 17, GGGTAAAGGC; 18, CACCATGGGA; 19, AGCTAACGCG; 20, GTTGGTGAGG; 21, GTGAAAGCCC; 22, GTAAACGATG; 23, ATGGCCCTTA; 24, GAACGGGTGA; 25, ACACCATGGG; 26, GAAGCTAACG; 27, AAGTGGGGGA. The reference probe (no. 28) was GCTGCTGGCA. These probes were designed using a previously described simulated annealing algorithm (6). Simulated annealing is a popular heuristic method for efficiently solving difficult optimization problems (16). Our goal was to design a probe set that could discriminate 1158 bacterial small subunit rDNA obtained from GenBank. However, since some of the “theoretical” probes did not hybridize in a consistent and predictable manner in the actual experiments, the probes used in this study were a collection of oligonucleotides originating from several different sets, but which produced strong signal intensities and hybridized to the control clones in the expected manner. Even though this probe set was generated through suboptimal means, it was still able to produce near species level resolution (see Results and Discussion section). Future refinements of the probe selection algorithm, which allow replacement of ineffective probes, should increase the resolution of this approach.

Data analysis. The signal intensities were normalized by dividing the values obtained from the discriminating probes by those from the reference probe (probe 28, which is expected to hybridize to all rDNA clones and is derived from the universal rDNA primer 530F (GTGCCAGCMGCCGCGG) (17). These normalized values were then classified as 0, 1, or N, using the intensity values of control clones. For this experiment, 1536 clones were arrayed, 26 of which had defined nucleotide sequences and served as positive and negative controls for each hybridization experiment. Clones with intensity values less than or equal to those of control clones expected to not hybridize with the probe were given a 0 classification. Clones with intensity values greater than or equal to those of control clones expected to hybridize to the probe were given a 1 classification. All other clones were given an N classification. The process created a hybridization fingerprint for each clone, which is a vector of values resulting from its hybridizations with all probes. The fingerprints were clustered using UPGMA (Unweighted Pair Group Method with Arithmetic Mean), default parameters, PAUP 4.0 beta 8. Each cluster was defined as a group of clones with the same fingerprint (with Ns consistently resolved).

Sequence analysis. Nucleotide sequences of the rDNA clones were obtained using the ABI PRISM® BigDye^{TM} Terminators v3.0 Cycle Sequencing Kit and were performed at the DNA lab of Arizona State University. Plasmid DNA was extracted using QIAprep Spin Miniprep Kit (Qiagen, Chatsworth, CA). Sequencing primers used were T725 and SP650 (see above), 907R, CCGTCAATTCMTTTRAGTTT and 1392R, ACGGGCGGTGTGTRC) (17). rDNA sequences were assembled using ContigExpress (Vector NTI). Sequence identities were determined using BLAST (NCBI) and Align X (Vector NTI).

Cluster specific PCR. Specific PCR primers were designed using conserved sequences within selected clusters (see Fig. 3 for cluster designations): 1, F = TGTTAGGGAAGAACAAGTACC, R = TTGCAGCCCTTTGTACCA; 2, F = ATGGTGACAGTTGTTAAAGC, R = TTTCACAACTGACTTGCG; 3, F = AATCTGCCCTTCACTCT, R = CCATCTCTGATGCTTTC; 4, F = GCAAGTCGAACGAGGTGCTT, R = CACGTAGTTAGCCGAGA; 5, F = GAACGGTAACAGGAAGCA, R = GCACATCCGATGGCAA; 6, F = GGAACGTGTCCTCTTGT, R = GCGTTACTAAGGGATTTAACT; 7, F = TCTTTTACCCGGGATGATA, R = TTACAAAGCCGCCTACG; 8, F = AGCTAACGCATTAAACATTC, R = CTGAGATGGCTTTTGGA. Ten µl PCRs were performed on 9C and 9E soil DNA (0.2 µl) using the protocol described above in the rDNA Library Construction section. Annealing temperatures were determined empirically for each primer set (1, 68^{o}C; 2, 60^{o}C; 3, 62^{o}C; 4, 65^{o}C; 5, 68^{o}C; 6, 60^{o}C; 7, 62^{o}C; 8, 62^{o}C; 9, 60^{o}C). The PCR products were resolved on 1% agarose gels, stained with ethidium bromide and photographed under UV illumination.

Richness and diversity estimations. UPGMA trees were constructed for each soil as described above. Clusters were used to represent species. The number of clones within a cluster represented abundance. From these trees, estimates of species richness and diversity were determined by Chao1 and Shannon analyses (R. K. Colwell, 1997. EstimateS: Statistical estimation of species richness and shared species from samples. Version 5. User's Guide and application published at: http://viceroy.eeb.uconn.edu/estimates). Diversity was also estimated by summing the branch lengths from these trees.

DGGE. DGGE was performed as previously described (34).

Nucleotide sequence data. The sequences reported in this paper have been deposited in the GenBank database. These data are in the format, cluster: clone, accession. Cluster designations refer to Fig. 3A. Cluster 1: 335-1, AF423249; 432-1, AF423263; 911-1, AF423299; 1104-1, AF423209; 1200-1, AF423214. Cluster 2: 375-2, AF423254; 456-2, AF423267. Cluster 3: 41-1, AF423260; 228-1, AF423240. Cluster 4: 572-2, AF423278; 573-2, AF423305; 666-2, AF423285; 746-2, AF423290; 838-2, AF423295; Cluster 5: 336-1, AF423250; 431-1, AF423262; 624-1, AF423283; 1389-1, AF423223. Cluster 6: 646-2, AF423284; 1326-2, AF423222; 1506-2. Cluster 7: 367-2, AF423253; 565-2, AF423277; 739-2, AF423289; 749-2, AF423291; 845-2, AF423296. Cluster 8: 60-2, AF423281; 353-2, AF423251; 1306, AF423219; 1315-2, AF423220. Additional clones from the *Bacillus* clade (Fig. 3B) include: 2-1, AF423237; 125-1, AF423216; 14-1, AF423226; 154-2, AF423232; 408-1, AF423258; 1005-1, AF423203; 309-1, AF423246; 4-1, AF423259; 10-1, AF423204; 38-1, AF423256; 21-1, AF423238; 26-1, AF423244; 102-1, AF423205; 720-1, AF423288; 511-1, AF423270; 1488-1, AF423230; 9-1, AF423298; 513-1, AF423272; 16-1, AF423233; 48-1, AF423304; 1421-2, AF423227.

RESULTS AND DISCUSSION

To demonstrate OFRG, we examined 1536 rDNA clones from two agricultural soils with an analysis employing 27 oligonucleotide probes. The principal factor influencing the discriminating power of this analysis is probe design. For this project, the probe set was developed using a simulated annealing algorithm (6). The discriminating nature of the probes is visualized by the differential patterns produced when a common set of clones was hybridized with each of the 27 probes (Fig. 2A). The hybridization pattern obtained from a single probe applied to all 1536 soil rDNA clones is shown in Fig. 2B. The signal intensities obtained from these experiments reflect the number of occurrences of the probe sequence in each clone as well as variations in DNA concentration resulting from the array construction process. To adjust for this variation, these data were normalized using signal intensities from a reference probe hybridized to the same array of clones (Fig. 2A, probe 28). Analysis with all 27 probes produced a hybridization fingerprint for every clone. A UPGMA (Unweighted Pair Group Method with Arithmetic Mean) analysis of the fingerprints was then used to generate a taxonomic depiction of the bacteria inhabiting the two soils (Fig. 3). Clones were identified from this analysis by their association with known rDNA sequences within the tree (Fig. 3). This was accomplished by including fingerprints from known rDNA sequences in the UPGMA analysis and by nucleotide sequence analysis of representative clones distributed throughout the tree. This process generated 766 clusters, which were grouped into 5 major taxa: *Bacillus*, *Actinobacteria*, *Proteobacteria* and two assemblages containing clones with relatively low identities to a variety of known cultured bacteria. In this report, clusters were defined as groups of clones with the same fingerprint.

The OFRG analysis identified several potential differences in bacterial community composition between the two soils. These soils were chosen because they possess different functional properties, which may be associated with their microflora. The 9E soil suppresses the plant parasitic nematode *Heterodera schachtii*, whereas the 9C soil does not suppress the pathogen but comes from an adjacent field and has similar chemical and physical properties (30). Among the numerous compositional differences, members of the *Bacillus* and *Enterobacteria* taxa were predominantly found in the 9C soil while α-*Proteobacteria* were almost exclusively identified in the 9E soil (Table 1 and Fig. 3). Even though compositional differences between the two soils were identified, correlations between specific bacterial populations and pathogen suppressiveness cannot be made from this experiment because of the limited number of samples analyzed and because these adjacent soils share neither common cropping histories nor weed and pest management practices. Nonetheless, these results demonstrate the potential of this approach for facilitating examinations into community structure-function relationships, a topic of interest in many areas of microbiology.

The resolution of this OFRG analysis was evaluated by examining the nucleotide sequences of the clones within six clusters distributed throughout the UPGMA tree (Fig. 3A, clusters 1, and 4-8). Clusters 2 and 3 (Fig. 3A) were not used for this analysis because only two clones were fully sequenced from these groups. For each cluster, pairwise sequence analysis showed that clones with an average sequence identity of 97% were grouped into the same cluster. The range of identities was 92.7-100%, 79% of which were between 96 and 98%. Thus, this OFRG analysis approximated species level resolution because DNA-DNA reassociation experiments have been used to show that bacterial rDNA with sequence identities of 97% or higher are likely to come from the same species (25). In general, the resolution of OFRG will depend on probe set properties such as the number of probes, their discriminatory power, and the suitability of the set for a specific environment. Further refinements of probe selection algorithms (6), coupled with the ever-expanding collections of rDNA sequences, should facilitate the design of probe sets for all types of microorganisms and environments. Additionally, we anticipate these future developments will also allow the design of probe sets capable of differentiating microorganisms at varying degrees of resolution (ie. identity levels of 90%, 95%, 97% …).

Several strategies have been used to estimate microbial species richness and diversity, most of which are based on considerable extrapolation. The difficulty of such endeavors stems from the considerable diversity of organisms found in most environments, which is exemplified by estimates suggesting that 4,000-40,000 different bacterial species can inhabit a single gram of soil (26, 27). Utilization of OFRG may lead to more confident estimates of richness and diversity because it offers the potential of identifying most, if not all, of the rDNA sequences in an environment. In this work, we used the distribution of clones within UPGMA trees to estimate these parameters. Two trees were constructed, one from each of the soils. Clusters were designated to represent species because they contained clones with an average sequence identity of 97%; since this number was less than 97%, the value that typically delineates bacterial species (25), our analyses will likely lead to underestimations of bacterial richness and diversity. With this approach, a Chao1 analysis was used to estimate bacterial richness to be 751 (9C) and 1602 (9E). Shannon’s index was used to generate diversity values of 5.0 and 6.1 for the 9C and 9E soils respectively. In addition, we also assessed diversity by summing the branch lengths from the trees (9C = 16.7, 9E = 33.4). In all of these analyses, the suppressive 9E soil exhibited greater bacterial species richness and diversity than the non-suppressive 9C soil. This result suggests that these parameters may contribute to the *H. schachtii* suppressiveness, a general phenomenon that has also been described in other plant pathogen systems (2, 8).

Several experiments were performed to corroborate the results obtained by the OFRG analysis. In the first approach, we examined OFRG's ability to accurately identify rDNA clones. An examination of 102 clones distributed throughout the UPGMA tree showed that the identities obtained from the UPGMA analysis of the hybridization fingerprints were consistent with those obtained by a BLAST (NCBI) analysis of the nucleotide sequences (Fig. 3, some of the data not shown because this figure depicts a partial tree). In the second approach, we attempted to corroborate the soil-specific compositional differences identified from the OFRG analysis by determining if these same variations were present in the original soil DNA. This was accomplished through the development of specific PCR primers for eight clusters distributed throughout the tree (Fig. 3A). PCR amplifications performed on the 9C and 9E soil DNA using these primers confirmed the compositional differences, as the relative intensity of PCR products correlated with the soil-specific distribution of the clones within each of the clusters (Fig. 4). That is, primers derived from clusters containing only 9C clones generated either no PCR product from the 9E soil DNA or a relatively weak band. These experiments support the bacterial community composition depicted by OFRG, confirming the value of this approach.

To demonstrate the impact that OFRG may have on microbial community studies, we examined the same two agricultural soils with DGGE (20), currently the most commonly used method for rDNA analysis. Here, no definitive DNA bands were obtained from the pathogen-suppressive 9E soil, suggesting a high level of bacterial diversity but providing little other community composition data (Fig. 5). For the non-suppressive 9C soil, thirteen DNA bands depicted the entire bacterial community. The difference between these DGGE results and those obtained by ORFG is striking, underscoring the potential of this new method for improving our understanding of microbial communities. This approach should also be useful for investigations of other microorganisms such as fungi and have application in other disciplines, including medicine, biotechnology, and ecosystem studies.

ACKNOWLEDGEMENTS

This research was funded by in part by grants from the UCR Vice Chancellor (seed grant to J.B. and D.C.), UC Biotechnology Research & Training Program (J.B. and D.C.), and NSF ITR program (G.D.V. and T.J.). L.V. was supported by a Vaddia-BARD Postdoctoral Award No. FI-306-00 from BARD, The United States – Israel Binational Agricultural Research and Development Fund. G.D.V. was supported by MURST grant “Bioinformatica e Ricerca Genomica". A.F. was supported by a Ph.D fellowship from CONACYT. M.C. was supported by NSF grant CCR-9988360.

We thank Woody Smith and Mark Springer for their assistance with data processing and UPGMA tree construction and Brian Lanoil for his critical review of the manuscript.

REFERENCES

1. Amann, R. I., W. Ludwig, and K. H. Schleifer. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. **59**:143-169.
2. Baker, R. 1990. Diversity in Biological Control. Crop Prot. **10**:85-94.
3. Barns, S. M., R. E. Fundyga, M. W. Jeffries, and N. R. Pace. 1994. Remarkable archaeal diversity detected in a Yellowstone National Park hot spring environment. Proc. Natl. Acad. Sci. USA **91**:1609-1613.
4. Behr, T., C. Koob, M. Schedl, A. Mehlen, H. Meier, D. Knopp, E. Frahm, U. Obst, K. H. Schleifer, R. Niessner, and W. Ludwig. 2000. A nested array of rRNA targeted probes for the detection and identification of enterococci by reverse hybridization. Syst. Appl. Microbiol. **23**:563-572.
5. Borneman, J., and E. W. Triplett. 1997. Molecular microbial diversity in soils from eastern Amazonia: Evidence for unusual microorganisms and microbial population shifts associated with deforestation. Appl. Environ. Microbiol. **63**:2647-2653.
6. Borneman, J., M. Chrobak, G. Della Vedova, A. Figueroa, and T. Jiang. 2001. Probe selection algorithms with applications in the analysis of microbial communities. Bioinformatics **17**(Suppl. 1):S39-48.
7. Borneman, J., P. W. Skroch, K. M. O'Sullivan, J. A. Palus, N. G. Rumjanek, J. L. Jansen, J. Nienhuis, and E. W. Triplett. 1996. Molecular microbial diversity of an agricultural soil in Wisconsin. Appl. Environ. Microbiol. **62**:1935-1943.
8. Cook, R. J., and K. F. Baker. 1983. *In* The nature and practice of biological control of plant pathogens, American Phytopathological Society, St. Paul, Minn.

9. Drmanac, R. 1999. cDNA screening by array hybridization. Methods Enzymol. **303**:165-178.

10. Drmanac, R., G. Lennon, S. Drmanac, I. Labat, R. Crkvenjakov, and H. Lehrach. 1991. p. 60-75. *In* C. Cantor, and H. Lim (ed.), Proceedings of the First International Conference on Electrophoresis, Supercomputing and the Human Genome. World Scientific, Singapore.

11. Drmanac, R., Z. Stezoska, I. Labat, S. Drmanac, and R. Crkvenjakov. 1990. Reliable hybridization of oligonucleotides as short as six nucleotides. DNA Cell Biol. **9**:527-534.

12. Drmanac, S., N. A. Stavropoulos, I. Labat, J. Vonau, B. Hauser, M. B. Soares, and R. Drmanac. 1996. Gene-representing cDNA clusters defined by hybridization of 57,419 clones from infant brain libraries with short oligonucleotide probes. Genomics **37**:29-40.

13. Fox, G. E., K. R. Pechman, and C. R. Woese. 1977. Comparative cataloging of 16S ribosomal ribonucleic acid: molecular approach to prokaryotic systematics. Int. J. Syst. Bacteriol. **27**:44-57.

14. Giovannoni, S. J., T. B. Britschgi, C. L. Moyer, and K. G. Field. 1990. Genetic diversity in Sargasso Sea bacterioplankton. Nature **345**:60-63.

15. Guschin, D. Y., B. K. Mobarry, D. Proudnikov, D. A. Stahl, B. E. Rittmann, and A. D. Mirzaberkov. 1997. Oligonucleotide microchips as genosensors for determinative and environmental studies in microbiology. Appl. Environ. Microbiol. **63**:2397-2402.

16. Kirkpatrick, S., C. D. Gelatt, and M. Vecchi. 1983. Optimization by simulated annealing. *Science* 4598:671-680.

17. Lane, D. J. 1991. 16S/23S rRNA Sequencing, p. 115-175. *In* E. Stackebrandt, and M. Goodfellow (ed.), Nucleic Acid Techniques in Bacterial Systematics, Wiley, New York.

18. Lennon, G. S., and H. Lehrach. 1991. Hybridization Analyses of Arrayed Complementary DNA Libraries. Trends Genet. **7**:314-317.

19. Liu, W. T., T. L. Marsh, H. Cheng, and L. J. Forney. 1997. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. Appl. Environ. Microbiol. **63**:4516-4522.

20. Muyzer, G., E. D. De Waal, A. G. Uitterlinden. 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction amplified genes coding for 16S rRNA. Appl. Environ. Microbiol. **59**:695-700.

21. Pace, N. R. 1997. A molecular view of microbial diversity and the biosphere. Science **276**:734-740.

22. Rudi, K., O. M. Skulberg, R. Skulberg, and K. S. Jakobsen. 2000. Application of sequence-specific labeled 16S rRNA gene oligonucleotide probes for genetic profiling of cyanobacterial abundance and diversity by array hybridization. Appl. Environ. Microbiol. **66**:4004-4011.

23. Small, Jack; Call, Douglas R.; Brockman, Fred J.; Straub, Timothy M. Chandler, Darrell P. Direct detection of 16S rRNA in soil extracts by using oligonucleotide microarrays. Applied and Environmental Microbiology. October, 2001. 67(10):4708-4716.

24. Sogin, S. J., M. L. Sogin, and C. R. Woese. 1972. Phylogenetic measurement in prokaryotes by primary structural characterization. J. Mol. Evol. **1**:173-184.

25. 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.

26. Tiedje, J. M. 1995. Approaches to the comprehensive evaluation of prokaryote diversity of a habitat, p. 73-87. *In* D. Allsopp, R. R. Colwell, and D. L. Hawksworth (ed.) Microbial Diversity and Ecosystem Function, Cab International, Wallingford.

27. Torsvik, V., J. Goksoyr, and F. L. Daae. 1990. High diversity in DNA of soil bacteria. Appl. Environ. Microbiol. **56**:782-787.

28. Urakawa, Hidetoshi; Noble, Peter A.; El Fantroussi, Said; Kelly, John J. Stahl, David A. Single-base-pair discrimination of terminal mismatches by using oligonucleotide microarrays and neural network analyses. Applied and Environmental Microbiology. January, 2002. 68(1):235-244.

29. Ward, D. M., M. M. Bateson, R. Weller, and A. L. Ruff-Roberts. 1992. Ribosomal RNA analysis of microorganisms as they occur in nature. Adv. Microb. Ecol. **12**:219-286.

30. Westphal, A., and J. O. Becker. 1999. Biological suppression and natural population decline of Heterodera schachtii in a California field. Phytopathology **89**:434-440.

31. Woese, C. R., and G. E. Fox. 1977. Phylogenetic structure of the prokaryotic domain: the primary kingdoms. Proc. Natl. Acad. Sci. USA **74**:5088-5090.

32. Woese, C. R., G. E. Fox, L. Zablen, T. Urchida, L. Bonen, K. Pechman, B. J. Lewis, and D. Stahl. 1975. Conservation of primary structure in 16S ribosomal RNA. Nature **254**:83-86.

33. Woese, C. R., O. Kandler, and M. L. Wheelis. 1990. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, Eucarya. Proc. Natl. Acad. Sci. USA **87**:4576-4579.

34. Yang, C. H., and D. E. Crowley. 2000. Rhizosphere microbial community structure in relation to root location and plant iron nutritional status. Appl. Environ. Microbiol. **66**:345-351.

FIGURE LEGENDS.

FIG. 1. Schematic for oligonucleotide nucleotide fingerprinting of ribosomal RNA genes (OFRG).

FIG. 2. Arrayed bacterial rDNA clones hybridized with ^{33}P-labeled DNA oligonucleotide probes: (A) discriminating probes no. 1-27 and reference probe no. 28 hybridized to a common set of clones, (B) discriminating probe no. 4 hybridized to 1536 clones.

FIG. 3. Taxonomic depiction of soil bacteria produced by oligonucleotide fingerprinting of ribosomal RNA genes (OFRG). The UPGMA tree was constructed from rDNA clones hybridization fingerprints derived from two soils. (A) Complete UPGMA tree. Numbers indicate cluster designations. (B) Detailed depiction of the *Bacillus* clade. This tree is composed of three segments; * and ● indicate connection sites. The full-length tree can be obtained through the corresponding author. rDNA clones are designated by a number followed by a space and then either 1 or 2, which designate their 9C or 9E soil origin respectively. Clones whose nucleotide sequences were determined are designated by the suffix S.

FIG. 4. Amplification of soil DNA using cluster-specific PCR primers. Numbers indicate cluster designations (see Fig. 3). C and E designate soil type. Clusters 1, 3, 5 contain 9C clones only. Clusters 2, 4, 6, 7, 8 contain 9E clones only. M = DNA ladder. Arrows designate the DNA fragments amplified by each PCR primer pair.

FIG. 5. Denaturing gradient gel electrophoresis (DGGE) analysis of soil bacterial rDNA. 9C and 9E designate soil type.

TABLE

TABLE 1. Taxonomic distribution of bacterial rDNA clones obtained from two agricultural soils

Taxon	9C*	9E*
Bacillus	405	35
Cytophaga/Flexibacter/Bacteroides	5	25
Actinobacteria	130	185
Proteobacteria	299	237
α-subdivision	10	142
β- and γ-subdivsions	162	87
Enterobacteria	127	8

*number of clones in each soil; these data were obtained by summing the clones within each taxonomic group from UPGMA tree (Figure 3).