APPLIED AND ENVIRONMENTAL MICROBIOLOGY, June 1996, p. 1935–1943 Vol. 62, No. 6
0099-2240/96/$04.00+0

Molecular Microbial Diversity of an Agricultural Soil in Wisconsin

JAMES BORNEMAN,[1] PAUL W. SKROCH,[2] KATHERINE M. O'SULLIVAN,[1] JAMES A. PALUS,[1] NORMA G. RUMJANEK,[1] JENNIFER L. JANSEN,[1] JAMES NIENHUIS,[2] AND ERIC W. TRIPLETT[1]*

Departments of Agronomy[1] and Horticulture,[2] University of Wisconsin—Madison, Madison, Wisconsin 53706

Received 21 December 1995/Accepted 22 March 1996

A culture-independent survey of the soil microbial diversity in a clover-grass pasture in southern Wisconsin was conducted by sequence analysis of a universal clone library of genes coding for small-subunit rRNA (rDNA). A rapid and efficient method for extraction of DNA from soils which resulted in highly purified DNA with minimal shearing was developed. Universal small-subunit-rRNA primers were used to amplify DNA extracted from the pasture soil. The PCR products were cloned into pGEM-T, and either hypervariable or conserved regions were sequenced. The relationships of 124 sequences to those of cultured organisms of known phylogeny were determined. Of the 124 clones sequenced, 98.4% were from the domain *Bacteria*. Two of the rDNA sequences were derived from eukaryotic organelles. Two of the 124 sequences were of nuclear origin, one being fungal and the other a plant sequence. No sequences of the domain *Archaea* were found. Within the domain *Bacteria*, three kingdoms were highly represented: the *Proteobacteria* (16.1%), the *Cytophaga-Flexibacter-Bacteroides* group (21.8%), and the low-G+C-content gram-positive group (21.8%). Some kingdoms, such as the *Thermotogales*, the green nonsulfur group, the *Fusobacteria*, and the *Spirochaetes*, were absent. A large number of the sequences (39.4%) were distributed among several clades that are not among the major taxa described by Olsen et al. (G. J. Olsen, C. R. Woese, and R. Overbeek, J. Bacteriol., 176:1–6, 1994). From the alignments of the sequence data, distance matrices were calculated to display the enormous microbial diversity found in this soil in two ways, as phylogenetic trees and as multidimensional-scaling plots.

An enormous amount of effort is being made worldwide by microbial ecologists to identify microorganisms in environmental samples. The keen interest in this topic is based on the observations of several laboratories that most bacteria from natural environments cannot be cultured with current techniques (for reviews, see references 2, 36, 45, and 50). In soil, estimates are that 80 to 99% of the microorganisms remain unidentified (1). The use of the PCR to amplify a gene common to all organisms now allows the identification of these previously unknown organisms. The gene commonly amplified for this purpose codes for the RNA sequence of the small subunit (SSU) of the ribosome (24, 54).

A recent paper by Boivin-Jahns et al. (6) illustrates the value of rRNA sequence analysis in the identification of bacteria compared with an analysis of phenotypic traits. Boivin-Jahns et al. (6) isolated 74 bacterial strains from a mine gallery and found that misidentification of bacteria was far less common with the gene coding for SSU rRNA (SSU rDNA) sequence than with more-traditional methods of microbial identification such as morphology, Gram stain, enzyme activities, and the utilization of several substrates as sole carbon and energy sources. When misidentification occurs with the SSU rDNA sequences, the cause is usually the lack of rDNA sequences from close relatives in the databases.

One possible use of rDNA sequence analysis is identification of soil microorganisms. The diversity of microorganisms in soil is critical to the maintenance of good soil health, because microorganisms are involved in many important functions such as soil formation, toxin removal, and elemental cycles of carbon, nitrogen, phosphorus, and others (7, 12, 23). Environmental stresses, however, can alter microbial populations and therefore endanger soil health. Recent studies have shown that pesticides and herbicides can decrease microbial respiration, biomass (14, 53), and diversity (3, 4). Genetic diversity is essential to life, since it permits adaptation through the creation of new organisms by genetic transfer and mutations (51).

A glimpse at the extensive microbial diversity in soil has been previously provided by renaturation experiments. These studies estimated that there are approximately 4×10^3 (47) to 10^4 (17) species per g of soil. Thus far, attempts to determine the taxonomic identification of this enormous diversity have been accomplished only by using traditional culturing methods. Since microscopic analyses have demonstrated that only 1 to 10% of soil organisms can be isolated by these techniques (1), other methods must be employed to accurately determine diversity.

To date, the microbial diversities of two soil samples have been analyzed by using the culture-independent approach of isolating DNA from soil, amplifying a portion of the SSU rDNA with conserved primers, sequencing selected clones from libraries of amplified SSU rDNA fragments, and comparing the resulting sequences with the more than 6,500 sequences present in the databases (24, 44, 48). Thirty such sequences from a subtropical soil from Queensland, Australia, were analyzed and found to be primarily from α-proteobacteria, with some planctomycetes and distant relatives of the planctomycetes (24, 44). This result was particularly surprising since one of the two primers was designed to specifically amplify the SSU rDNA fragment of streptomycetes (24, 44).

Sequences of 17 SSU rDNA clones from a soil sample collected from a soybean field in Japan were found to include a diverse group of proteobacteria, green sulfur bacteria, an archeaon, a high-G+C-content gram-positive strain, and several bacteria of novel phylotypes within the domain *Bacteria* (48). In that work, Ueda et al. (48) sequenced a fragment of 220 nucleotides of the SSU rDNA molecule amplified by using

* Corresponding author. Phone: (608) 262-9824. Fax: (608) 262-5217. Electronic mail address: ewtriple@facstaff.wisc.edu.

universal primers. Ueda et al. (49) have also taken a similar approach to analyze the diversity of nitrogen-fixing organisms in the rhizosphere of rice roots, using universal *nifD* primers.

In this paper, we provide a strategy to estimate soil microbial diversity which includes an efficient soil DNA extraction method and sequence analyses of 124 clones from a universal SSU rDNA clone library. We use this strategy to depict the diversity of life in a pasture soil from southern Wisconsin in two ways, showing the enormous diversity of the dominant microorganisms found in this soil and demonstrating that this diversity is not randomly distributed among major taxa.

MATERIALS AND METHODS

Site chosen for examination: WICST. The Wisconsin Integrated Cropping Systems Trial (WICST) at the Arlington Agricultural Research Station of the University of Wisconsin was begun in 1989 at two agricultural research stations for the purpose of testing agroecological questions by using treatments with increasing rotation complexity. The center of this large, 125-ha site at Arlington, Wis., is located at 43°18′23″N, 89°19′83″W. Of the six cropping systems established at the WICST, we chose the clover-grass mixture pasture for our analysis of diversity. The pasture site was chosen because it contains the highest level of diversity of plant species among the six treatments of this experiment and thus might have the greatest microbial diversity as well.

Soil sampling, analyses, and site climate. The soil at the Arlington site is a Plano silt-loam. The 20-cm-deep A horizon is a silt-loam and contains 4.4% organic matter. The loess mantel is >1.25 m deep. Four 2.5-cm-diameter soil cores were taken from the top 10 cm of a clover-grass pasture at the Arlington Agricultural Research Station. The soil samples were immediately placed on dry ice, mixed, and then stored at −70°C prior to DNA extraction. Soil analysis was done by the Soil Testing Laboratory of the University of Wisconsin—Madison as described by Schulte et al. (40). The soil sample contained 13% sand, 70% silt, 17% clay, 4.4% organic matter, 0.3% total N, 400 ppm of K^+, and 98 ppm of P. The soil pH was 6.5. The site is well drained, with groundwater more than 25 m below the surface. Two-thirds of the 79-cm annual rainfall occurs from April to October. The site has an average of 165 frost-free days.

DNA extraction from soil. DNA extraction utilized a new bead beater, the FastPrep System, developed by Bio 101 (Vista, Calif.) and Savant (Farmingdale, N.Y.). A 978-ml volume of 200 mM sodium phosphate buffer (pH 8.0), 122 ml of MT buffer (Bio 101, catalog no. 6010-450), and 500 mg of soil were added to a FastDNA tube containing a matrix designed to lyse most cell types (Bio 101, catalog no. 6530-401). The mixture was shaken in the FastPrep instrument for 30 s at 5.5 m/s and then centrifuged at 16,000 × *g* for 30 s. One milliliter of supernatant was removed and mixed with 250 μl of protein precipitating solution (Bio 101, catalog no. 6550-203). This mixture was centrifuged at 16,000 × *g* for 5 min at room temperature. The supernatant was then collected and stored at −20°C.

DNA purification. A volume of 250 μl of the soil supernatant was added to a Spinfilter (Bio 101, catalog no. 6540-407) with 500 μl of binding matrix (Bio 101, catalog no. 6540-403). This tube was gently inverted five times, incubated for 5 min at room temperature, and then centrifuged for 30 s at 16,000 × *g*. For this step and all other purification steps, the eluate in the catch tube was discarded after centrifugation. The pellet in the Spinfilter was washed twice. Each wash was done by adding 500 μl of salt-ethanol wash solution (Bio 101, catalog no. 6540-404) and then centrifuging for 30 s at 16,000 × *g*. The Spinfilter was then centrifuged for 1 min at 16,000 × *g* to dry the pellet. The DNA was eluted by transferring the Spinfilter to a new catch tube (Bio 101, catalog no. 2080-401), adding 50 μl of DNA elution solution (Bio 101, catalog no. 6540-406), gently flicking the tube five times, and then centrifuging for 1 min at 16,000 × *g*. To minimize DNA shearing, vortex mixing was avoided.

DNA quantitation. Soil DNA and a *Hin*dIII digest of λ DNA were resolved on a 0.7% agarose gel, stained with ethidium bromide, and photographed. Densitometric measurements of the λ DNA were used to create a standard curve, and soil DNA concentrations were determined by interpolation (27).

DNA extraction efficiency. By the method described above, DNA was extracted from both *Rhizobium leguminosarum* bv. viceae 128C1 cells and sterile soil amended with *R. leguminosarum* bv. viceae 128C1. The DNA was extracted immediately after the 128C1 cells were mixed with the soils. Extraction efficiency was determined by comparing the DNA yields from the two extractions. No DNA was obtained when only the unamended sterile soil was extracted.

Universal SSU rDNA library construction. Soil DNA was amplified by PCR using an Air Thermo-Cycler (Idaho Technologies). The 60-μl reaction mixture contained the following final concentrations or total amounts: 6 μl of DNA (65 ng/μl), 50 mM Tris (pH 8.3), 250 μg of bovine serum albumin (BSA) per ml, 2.5 mM $MgCl_2$, 200 nM deoxynucleoside triphosphates, 200 nM each universal SSU rRNA primer, and 3 U of *Taq* polymerase. All reagents were mixed and then heated to 92°C for 1 min. Forty cycles of PCR were then run at 92°C for 0 s, 50°C for 30 s, and 72°C for 60 s followed by 72°C for 3 min. When an Air Thermo-Cycler is used, the reaction mixtures are placed in capillary tubes that permit rapid heat exchange. Thus, a very short denaturation time is sufficient.

The primers were designed as follows to amplify most SSU ribosomal genes: 530F, 5′-TGACTGACTGAGTGCCAGCMGCCGCGG-3′; and 1494R, 5′-TGACTGACTGAGGYTACCTTGTTACGACTT-3′ (20). Each primer also contained an additional leader sequence (underlined) that has the stop codon, TGA, in all three reading frames. The SSU rDNA library was produced by gel isolating the amplified genes, cloning them into the pGEM-T vector (Promega), and then transforming the ligated plasmids into competent *Escherichia coli* DH5α (38). The clones were screened for α-complementation by using X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) and IPTG (isopropyl-β-D-thiogalactopyranoside) (38). The stop codons were added to prevent transcription through the insert and *lacZ* gene resulting in false blue colonies. The T-tails of the SSU rDNA inserts allowed direct cloning into pGEM-T without restriction digestion. This removes bias associated with certain restriction sites found in some, but not all, clones.

Template production and sequencing. Plasmid DNA was isolated from randomly picked bacterial colonies by the alkaline lysis miniprep method (38). PCR was then used to produce the sequencing templates. The 15-μl reaction volumes contained the following final concentrations or total amounts: 1 μl of DNA (diluted 1:100 in H_2O), 50 mM Tris (pH 8.3), 250 μg of BSA per ml, 2.5 mM $MgCl_2$, 200 nM deoxynucleoside triphosphates, 200 nM each T7 and SP6 primer, and 0.75 U of *Taq* polymerase. All reagents were mixed and then heated to 92°C for 1 min. Thirty-five cycles of PCR were then run at 92°C for 5 s, 50°C for 30 s, and 72°C for 60 s, followed by 72°C for 3 min. These PCR products were purified by using Wizard PCR Preps (Promega). The sequencing reactions were done by using a PRISM Ready Reaction DyeDeoxy Terminator Cycle Sequencing kit (Perkin-Elmer). The 15-μl reaction volumes contained 7.2 μl of reaction mix, 0.8 μl of BSA (5 mg/ml), 2 μl of T7 primer (10 ng/ml), and 150 ng of template DNA. Twenty-five cycles of PCR were run at 94°C for 0 s, 50°C for 10 s, and 60°C for 3 min. Excess dye terminators were removed from the reaction mixture with Centri-Sep spin columns (Princeton Separations). The reactions were run on an ABI 377 sequencer (Perkin-Elmer).

Sequence analyses. Since the SSU rDNA PCR fragments were not directionally cloned into the pGEM-T vector (above), clones 100 to 147 (group A) were sequenced through a conserved region and clones 1 to 76 (group B) were sequenced through the hypervariable region V9 and a portion of V8 (28). The conserved and variable regions were sequenced at approximately nucleotides 531 to 785 and nucleotides 1253 to 1492, respectively, according to *E. coli* numbering. One strand of each clone was sequenced with an error rate of 3% determined by the error rate for 1,138 bases of pGEM-T vector subjected to the same template production and sequencing protocols (described above). We decided that it was better to sequence one strand of many samples and tolerate a small error rate than to sequence both strands with redundancy of each clone to ensure 100% accuracy of each sequence. The use of more clones sequenced with a low error rate gives a better impression of the diversity in the soil than does 100% accurate sequencing of a few clones.

Preparation of phylogenetic trees. The rDNA clone sequences were separated into two groups, A and B, and aligned with representative rRNA sequences from GenBank by using PILEUP (Genetics Computer Group [GCG]). Distance matrices and phylogenetic trees were constructed by using the Jukes-Cantor algorithm (15) and the neighbor-joining method (37), respectively, with the MEGA package (19). Both trees were rooted to an archeaon, *Methanococcus thermolithotrophicus.*

Phylogenetic assignments. Identification of the rDNA clones was done by two methods. First, we constructed two phylogenetic trees (see Fig. 2) from 124 partial SSU rDNA sequences. Partial SSU rDNA fragments have previously been shown to yield accurate phylogenetic assignments (21, 39). Second, taxonomic assignments were done by comparing the soil clone sequences with the nonredundant nucleotide database at GenBank by using BLAST (NCBI). Only clones whose identities were corroborated by both methods were given a taxonomic assignment; all others were classified as unknown soil organisms. Unknown sequences are represented in Fig. 2 by a clone number, followed by their closest known relative along with the percent homology to that relative as determined by BLAST from NCBI.

MDS. Relationships among rDNA sequences were also displayed in two dimensions on the basis of a multidimensional scaling (MDS) analysis of the distance matrix used to compute the phylogenetic tree. The MDS plot complements the information in the phylogenetic tree, because it allows visualization of the distribution of phylotypes within and among groups. The MDS plot allows the identification of relationships that are often difficult to see in the dendrogram. No additional biological assumptions are needed for MDS analysis. For a matrix describing the relationships among objects, the MDS procedure finds the coordinates for each object, in a reduced number of dimensions that results in the closest fit between the distance matrices corresponding to the original and reduced dimensions (for a review, see reference 55). Relationships among rDNA sequences were estimated by plotting the sequences in two dimensions by the first two MDS coordinates resulting from an MDS analysis of the same distance matrices used for phylogenetic tree construction. MDS analysis was performed by the procedure in the SYSTAT statistical package (52). Five sequences were removed from the 5′ and 3′ distance matrices before MDS analysis was performed. All three nuclear sequences and sequences 129 and 133 were removed. This was done because these sequences were very different from all other se-

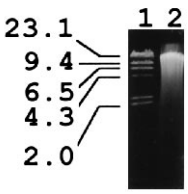


FIG. 1. DNA extracted from a Wisconsin pasture soil. Lanes: 1, 100 ng of lambda DNA digested with *Hin*dIII; 2, 1.5 μg of pasture soil DNA. The DNA was resolved on a 0.8% agarose gel.

quences and their inclusion reduced our ability to display the relationships among the remaining sequences.

Nucleotide sequence accession numbers. The sequences of the soil rDNA clones were deposited in GenBank. Clones 1 to 76 and 100 to 147 were assigned accession numbers U42850 to U42925 and U42927 to U42974, respectively.

RESULTS

DNA extraction method. The DNA extraction method described above is at least five times faster than previous methods (22, 27) and results in very pure DNA suitable for restriction digestion, cloning, and PCR amplification. This method also provides far less shearing of DNA (predominantly 9 to 23 kb) (Fig. 1) than previous bead beater methods used to extract DNA from sediments (predominantly 0.5 to 9 kb) (22, 29). The extraction efficiency of this new soil procedure was examined by two methods. First, an extraction efficiency of 98% was determined by comparing the DNA yields obtained from *R. leguminosarum* 128C1 cells and from sterile soil inoculated with 128C1 (data not shown). The second measure of efficiency was the diverse group of phylotypes identified by the rDNA sequence analysis (below). The DNA was also colorless and sufficiently purified for restriction digestion and PCR amplification (data not shown).

Amplification of hypervariable versus less variable regions of rDNA for diversity estimates. In the cloning strategy for the amplified SSU rDNA molecules, the inserts were not directionally cloned into pGEM-T. As a result, sequences from either end of the ~1,000-bp rDNA insert were obtained. The sequences representing the 3′ end of the rDNA molecule (hypervariable) are depicted in Fig. 2A and 3A, while the 5′-end sequences (more conserved) are shown in Fig. 2B and 3B. This allowed us to determine which region was more useful in a diversity analysis and whether directional cloning of the insert DNA would be necessary for future analyses of this kind. Because the ~250-bp regions sequenced from each end of the molecule do not overlap, sequences from both regions cannot be depicted on one phylogenetic tree or MDS plot. As a result, two phylogenetic trees and two MDS plots were necessary to illustrate the data.

In Table 1, we show the distribution of rDNA sequences among major taxonomic groups of bacteria for each of the two ends of the rDNA molecules that were amplified and sequenced. These distributions of phylotypes within the taxa of *Bacteria* are very similar but not identical between the two regions sequenced. Although either dendrogram shows that the same three dominant groups are present in this soil (the *Proteobacteria*, the *Cytophaga-Flexibacter-Bacteroides* group, and the low-G+C gram-positive group), the proportion of clones that were unrelated to known major taxa within the domain *Bacteria* was higher with the conserved region than with the hypervariable region. Thus, a sequence analysis of a conserved region may reduce the assignment of certain clones to specific known taxa.

Molecular microbial diversity of the pasture soil. Table 1 shows the distribution of the soil rDNA clones among all major taxa of the domain *Bacteria* as defined by Olsen et al. (30). Of the 124 clones sequenced, 98.4% were from the domain *Bacteria*. Two of the rDNA sequences listed in Table 1 are derived from a plant chloroplast and an insect mitochondrion and are most closely related to the cyanobacteria and α-proteobacteria, respectively. Two of the 124 sequences were of nuclear origin, one being fungal and the other a plant sequence. No *Archaea* sequences were found.

The relationships of these sequences to cultured organisms of known phylogeny are illustrated with phylogenetic trees (Fig. 1) and MDS plots (Fig. 3). Among the *Bacteria*, three major taxa were highly represented: the class *Proteobacteria* (16.1%), the *Cytophaga-Flexibacter-Bacteroides* group (21.8%), and the low-G+C gram-positive group (21.8%). Some groups, such as the *Thermotogales*, the green nonsulfur group, the Kingdom *Fusobacteria*, and the Kingdom *Spirochaetes*, were absent. A large number of the sequences (39.4%) were distributed among several clades that are not among the major taxa described by Olsen et al. (30).

Phylogenetic analyses. Algorithms which can be done quickly on a personal computer, rather than a large mainframe computer, were used. This conserved time and resources. The phylogenetic trees generated by the algorithms used here were validated by the addition of known sequences and by BLAST analysis of each sequence. These algorithms have been used by others to construct dendrograms containing a similar breadth of phylotypes (10).

MDS plots. As the same distance matrices were used to generate the phylogenetic trees and MDS plots, it is not surprising that the two illustrations of diversity give similar results. In each case, large clusters of soil phylotypes are found within the *Cytophaga-Flexibacter-Bacteroides* group and the low-G+C gram-positive group (Fig. 2 and 3). However, the relationships among different clusters of clones are almost impossible to discern in the dendrograms because of multiple branches in the tree. For example, the α-, β-, γ-, and δ-*Proteobacteria* cluster in different groups in the two trees, and the arrangements of these clusters are not consistent for conserved and hypervariable regions. However, as shown by the ellipses drawn in the MDS plots (Fig. 3), the *Proteobacteria* and other major taxa consistently form contiguous clusters that exclude nearly all other clones. The description of the relationships among and within clusters in the two-dimensional MDS plot is also more biologically realistic, since variability of organisms within groups is more easily visualized.

Chimeras. We are aware that the amplification of SSU rRNA genes from DNA isolated from an environmental sample can result in the formation of chimeras (5, 8, 18, 25). The frequency of chimeras within rDNA libraries has varied from 4% to as high as 20% (5, 8, 18). Among the 124 rDNA sequences from this soil analyzed to date, we have found only 2 that appear to be chimeras, clones 54 and 129. These sequences can be split into two fragments which show a reasonably high degree of homology to two very different bacteria. None of the other sequences showed this phenomenon. All of this was done manually by using the BLAST program from NCBI or the FASTA program from GCG. The CHECK_CHIMERA program of the Ribosomal Database Project (http://rdp.life.uiuc.edu/RDP/commands/chimera.html) at the University of Illinois was not used for this analysis because it cannot be used for sequences shorter than 400 bp. In addition, the CHECK_CHIMERA program is by no means foolproof for detecting chimeric molecules (34).

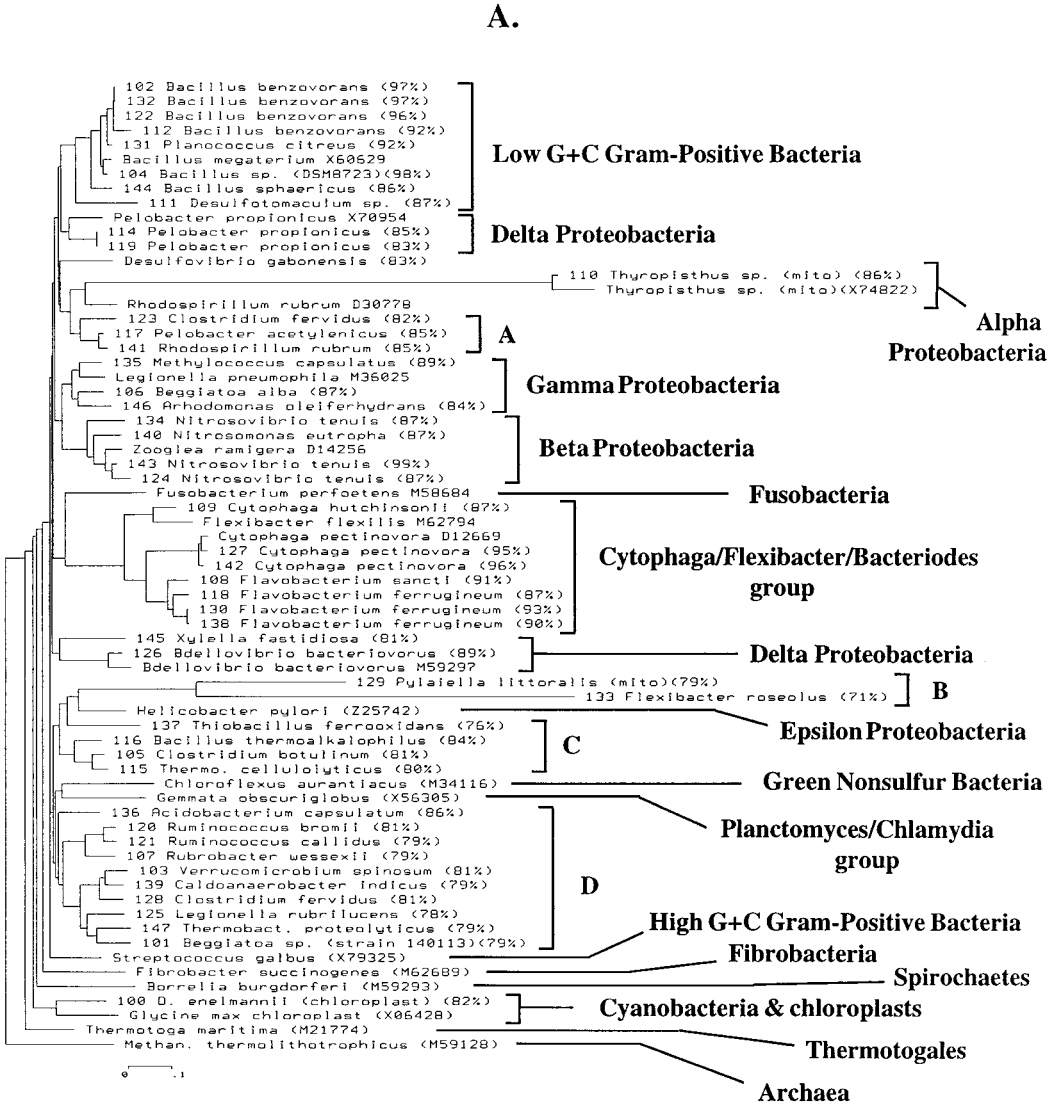


FIG. 2. Phylogenetic relationships of partial SSU rDNA sequences from 124 unknown soil clones and 41 sequences from identified bacteria in the databases. The sequences were aligned by PILEUP (GCG). The distance matrices and phylogenetic trees were calculated by the Jukes-Cantor (15) and neighbor-joining (37) algorithms, respectively. Unknown sequences are represented by a clone number followed by their closest known relative along with the percent homology to that relative as determined by BLAST from NCBI. Database sequence accession numbers for known organisms are indicated. Names of the major taxa are shown in boldface. Clades A to F are groups of sequences that cannot be assigned to any known group of organisms within the domain *Bacteria*. Sequences from nucleotides 531 to 785 (A) and 1253 to 1492 (B), according to *E. coli* 16S rRNA numbering, are represented. The sequences in panel B are from a more hypervariable region than those in panel A. The number of changes per sequence position is shown on the distance scale, assuming a median rate of change.

Sequence analysis of partial versus full-length rDNA molecules. For our purpose of surveying the dominant microorganisms found in this soil, our sequence analyses compared approximately 250 bp of each rDNA clone. Two previous studies which compared the phylogenetic trees created from partial and full-length rRNA sequences support this approach (21, 39). These reports showed that the phylogenetic assignments obtained from the partial and the full-length sequences were very similar (21, 39). Although some of the more deeply branching orders of the groups differed, the groups that were established were the same (21, 39). Since we were interested only in determining the dominant groups of soil microorganisms, a partial sequence analysis was justified. Sequencing smaller regions also has several advantages over full-length sequencing. The partial sequencing is faster and less expensive, reduces computer time during analysis, and may reduce chimeras in the analysis.

Little redundancy among clones sequenced. The degree of evenness of the microbial diversity in this soil can be noted by the fact that only 4% of sequences may be duplicates. This was determined by considering the 3% sequencing error rate and the distance matrices. Thus, only two sequences from the conserved region and three from the hypervariable region were considered potential duplicates, since their sequences were ≥97% homologous.

DISCUSSION

Soil DNA extraction. The rapid isolation of DNA from soil saves time, which can be subsequently used to analyze repli-

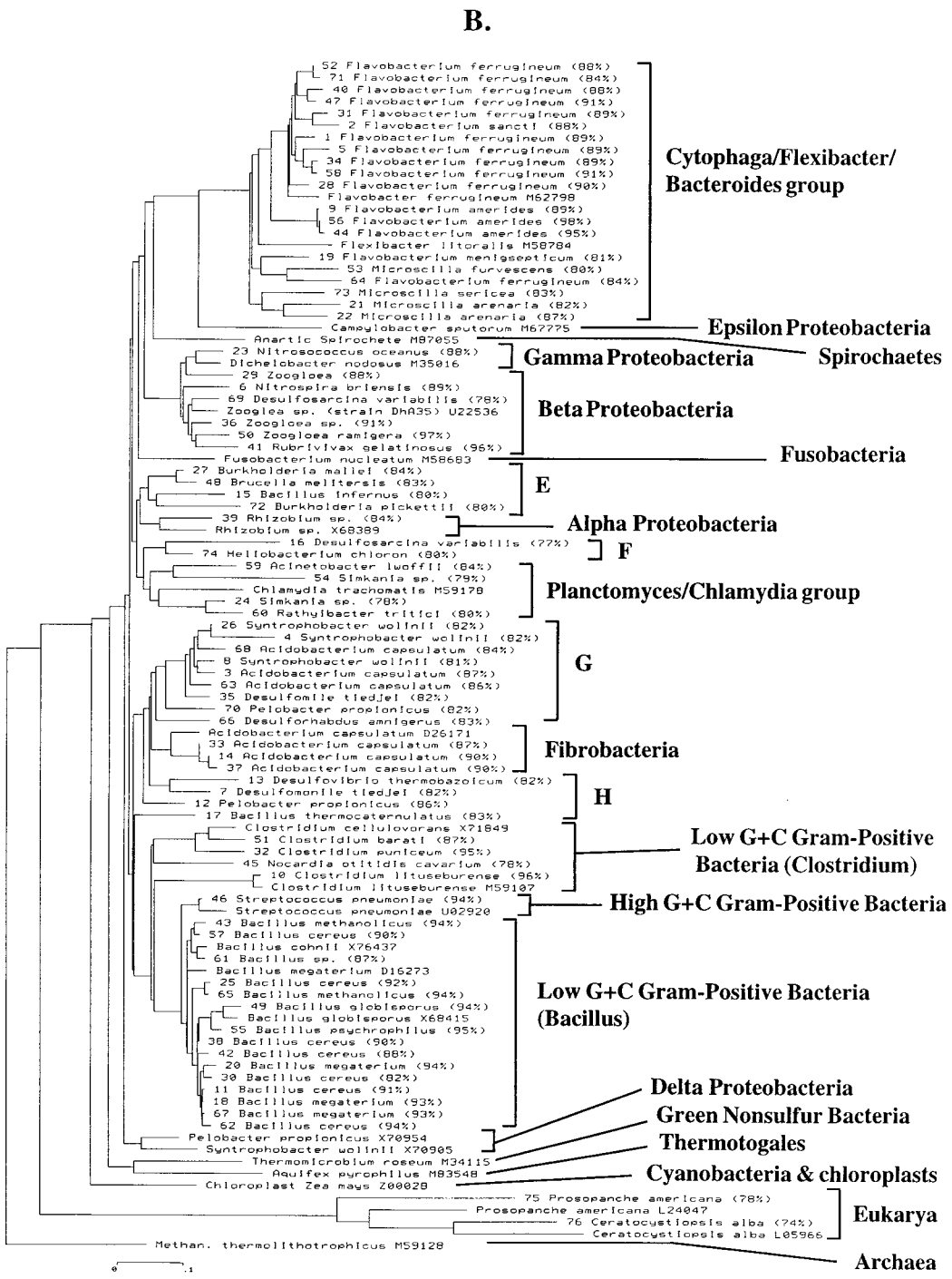
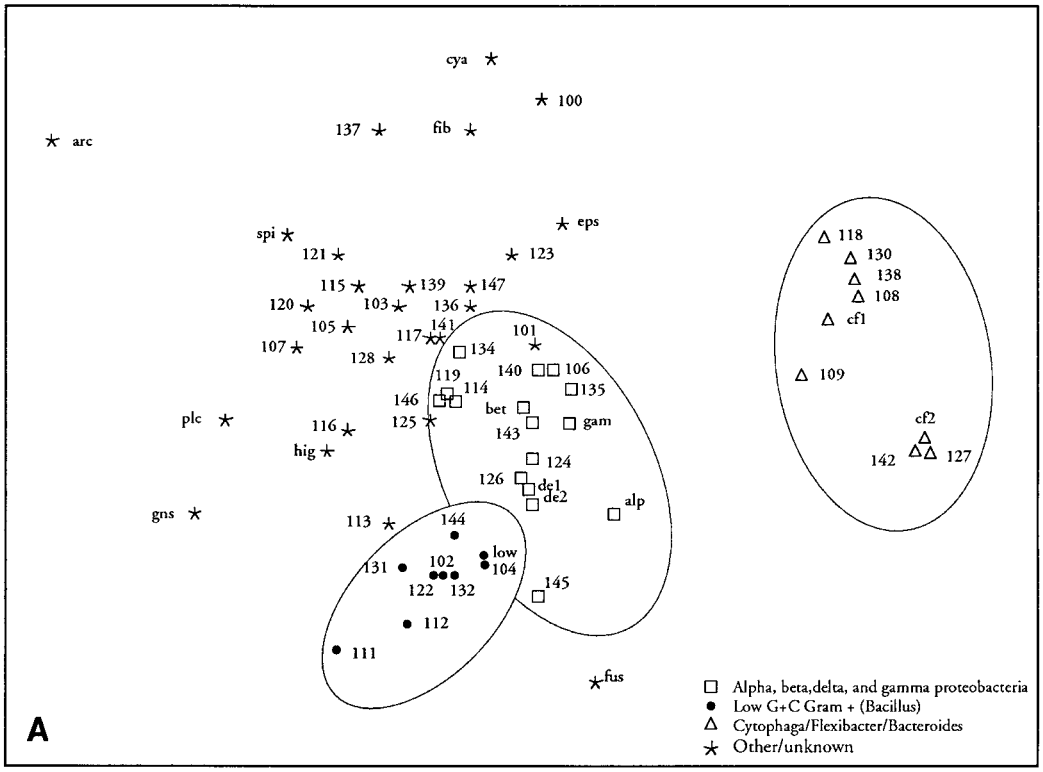


FIG. 2—*Continued.*

cates or perform hypothesis testing. Previous methods for rapid soil DNA isolation required several hours per sample (22, 27). We have reduced this time to 25 min.

The resulting DNA is of sufficient purity for restriction digestion, PCR amplification, and cloning. We are also confident that the degree of efficiency is very high, given the large diversity of phylotypes obtained in our analysis of the SSU rDNA clones from the pasture plots. The sequence analysis of the rDNA library shows that this method allows the isolation of DNA from organisms that are difficult to rupture, such as *Bacillus* spp. and insects.

Sequences identified from the domain *Eucarya*. Of the 124 clones sequenced, 4 are most similar to known eukaryotic sequences. One of these was most similar to a chloroplast rDNA sequence that is present within the *Cyanobacteria*-chloroplast clade. Another was most similar to a mitochondrial rDNA sequence from an insect. Only 2 of the 124 rDNA sequences were of eukaryotic nuclear origin, one fungal and one



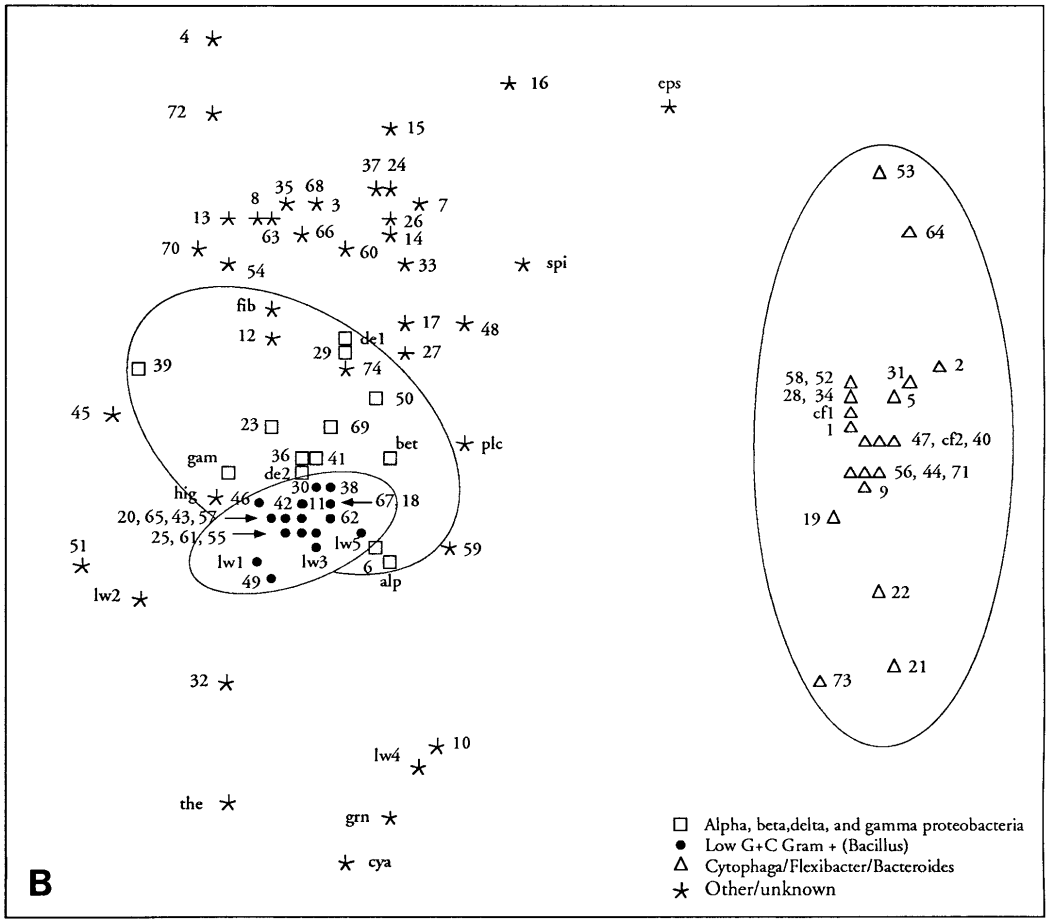


FIG. 3. (A and B) MDS plots of the SSU rDNA sequence data from Fig. 2A and B, respectively. Sequences from the database are shown as letters, while the unknown sequences amplified from soil DNA are given the same clone numbers as in Fig. 2. Note that the *Cytophaga-Flexibacter-Bacteroides* group clusters on the right-hand side of the plot and that the low-G+C gram-positive group is tightly clustered in the left center of the plot. The abbreviations for the sequences from cultured reference strains are as follows. (A) arc, M59128; alp, X68389; bet, U22536; cf1, M62798; cf2, M58784; cya, Z00028; de1, X70905; d32, X70954; eps, M67775; fib, D26171; fus, M58683; gam, M35016; grn, M34115; hig, U02920; lw1, X68415; lw2, X71849; lw3, X76437; lw4, M59107; lw5, D16273; plc, M59178; spi, M87055; and the, M83548. (B) alp, D30778; arc, M59128; bet, D14256; Cf1, M62794; cf2, D12669; cya, X06428; de1, X70954; de2, M59297; eps, Z25742; fib, M62689; fus, M58684; gam, M36025; gns, M34116; hig, X79325; low, X60629; plc, X56305, spi, M59293; and the, M21774.

plant. We were not surprised that so few sequences from eukaryotic microorganisms were identified. This is in agreement with the observation that the number of protozoa in soil can be 4 to 5 orders of magnitude below the number of bacteria, consistent with their relative positions in the food chain (1).

Few fungal sequences in sequence analysis. As fungi often constitute a large amount of soil biomass, we were surprised that so few of the sequences in our survey were from fungi. One sequence with a low level of homology to Zygomycetes was identified. As pesticides are not used at the pasture plots from which the soil was collected, the absence of fungi in our survey cannot be explained by fungicide use. An explanation of this dichotomy is that there is very little DNA present in fungal hyphae (13). Although soils can contain as much as 205 m of fungal hyphae per g of soil, the amount of fungal DNA extractable from soil is at least 10-fold smaller than that obtained from bacteria (13). Therefore, fungal hyphae may be viewed as long tubes with very little cytoplasm or few nuclei.

No members of the domain *Archaea* found in sequence analysis. None of the sequences were of the domain *Archaea*. *Archaea* may be present in this soil because *Archaea*-specific primers (5, 9) do amplify SSU rDNA sequences with soil DNA used as a template. However, we cannot be certain of the presence of *Archaea* at this site until these amplification products are cloned and sequenced. Nevertheless, since none of the 124 sequences we analyzed were similar to those of the *Archaea*, *Archaea* are probably not a dominant group in this soil. This is in contrast to similar analyses of marine environments in which 4 to 34% of the SSU rDNA sequences were from the domain *Archaea* (9).

TABLE 1. Distribution of organisms identified in the rDNA library made from Wisconsin soil DNA within the major taxa[a]

Major taxon and group	No. (%) of organisms identified		
	Hypervariable region	Less variable region	Total
Bacteria[b]	74 (97.4)	48 (100)	122 (98.4)
Thermotogales	0	0	0
Green nonsulfur	0	0	0
Cyanobacteria-chloroplasts	0	1 (2.1)[c]	1 (0.8)
Low-G+C gram positive	19 (25.0)	8 (16.7)	27 (21.8)
Fusobacteria	0	0	0
High-G+C gram positive	1 (1.3)	0	1 (0.8)
Cytophaga-Flexibacter-Bacteroides	20 (26.3)	7 (14.6)	27 (21.8)
Fibrobacteria	3 (3.9)	0	3 (2.4)
Spirochaetes	0	0	0
Planctomyces-Chlamydia	4 (5.3)	0	4 (3.2)
Proteobacteria	8 (10.5)	12 (25.0)	20 (16.1)
α	1 (1.3)	1 (2.1)[d]	2 (1.6)
β	6 (7.9)	4 (8.3)	10 (8.1)
γ	1 (1.3)	3 (6.3)	4 (3.2)
δ	0	4 (8.3)	4 (3.2)
Unknown	19 (25.0)	20 (41.6)	39 (31.4)
Archaea	0	0	0
Eucarya	2 (2.6)	0	2 (1.6)
Fungi	1 (1.3)	0	1 (0.8)
Plants	1 (1.3)	0	1 (0.8)

[a] Distributions are listed for each of the two regions sequenced and for the total of the two regions sequenced.
[b] All major taxa of the domain *Bacteria* are listed as defined by Olsen et al. (30).
[c] This sequence is a chloroplast rDNA sequence from a plant.
[d] This sequence is a mitochondral rDNA sequence from an insect.

Comparison with culture techniques. Most of the phylotypes (59.7%) identified in this work are members of the *Cytophaga-Flexibacter-Bacteroides* group (21.8%), the low-G+C-content gram-positive group (21.8%), or the class *Proteobacteria* (16.1%). These taxonomic groups have been commonly found in other studies of soil microbial diversity that were based on the identification of cultured organisms (for a review, see reference 1). However, these studies have also shown that the high-G+C gram-positive group can comprise as much as 92% of the cultured isolates from soil. With respect to specific genera, dominant microorganisms found in soil by culture methods were of the genera *Arthrobacter* (5 to 60%), *Bacillus* (7 to 67%), *Pseudomonas* (3 to 15%), *Agrobacterium* (1 to 20%), *Alcaligenes* (1 to 20%), and *Flavobacterium* (1 to 20%) and the order *Actinomycetales* (5 to 20%) (3). Of these groups, only *Bacillus* (19%) and *Actinomycetales* (0.8%) were found in our analysis. Comparisons of our results with those obtained by culturing techniques, however, are difficult since we have made no attempt to identify bacteria in this pasture soil by culture techniques. Nevertheless, differences between traditional culturing studies and rDNA sequence analyses are expected, since most bacteria (90 to 99%) in environmental samples are not culturable (1, 2, 36, 45). In addition, over one-third of our sequences were in clades outside the known major taxa of *Bacteria*. As most bacteria in soil are not culturable, it is not surprising that some hitherto-unknown taxa would be found among the doninant groups of bcteria in soil.

Despite the fact that this soil shows enormous diversity among the *Bacteria*, this diversity is not random. A majority of the sequences fall into just three major taxa. In addition, since the unidentified sequences generated in this study (Fig. 2, groups A to H) have low-level sequence homology (71 to 87%) to identified sequences in the GenBank, it is apparent that the diversity of microorganisms in soil is extensive and that the phylogenies of many dominant soil bacteria remain uncharacterized.

PCR amplification bias. Biases can also occur during the PCR amplification of SSU rDNA sequences from a mixture of DNA from several organisms (25, 32, 43). Farrelly et al. (11) showed that genome size and copy number of the SSU rDNA affect the amount of product obtained in PCR amplification of the SSU rDNA. Since the genome size and copy number of the SSU rDNA are not known for any of the uncultured organisms, Farrelly et al. (11) argue that the number of species in SSU rDNA libraries made from environmental samples can never be determined. Although a determination of the exact number of species is not feasible, the direct extraction of DNA from

soil and the amplification of rDNA genes provide a method of identifying many previously unknown microorganisms.

Comparison with the molecular microbial diversities in a subtropical Australian soil and a temperate soil from Japan. Our results are very different from the results of a similar analysis of soils from Australia and Japan. To date, these soils are the only other soil samples that have been similarly analyzed (24, 44, 48). With the Australian soil, the majority of sequences obtained were from the α-*Proteobacteria*. In our clover-grass pasture library, which should contain many cells of *R. leguminosarum* (a member of the α-*Proteobacteria*), only a small percentage (1.6%) of the rDNA clones were identified as α-*Proteobacteria*. There are many reasonable explanations for the differences observed between the Queensland and Wisconsin soils. First, these differences may be real. These two sites are very far apart geographically and very different in soil type, and a subtropical soil may be less diverse than a mixed pasture. Second, different primers were used by each laboratory for their amplification of the rDNAs. The SSU rDNA from the Australian soil was amplified by using one universal primer and a second primer designed to amplify streptomycetes (44), while we used primers that are regarded as universal. This allowed us to detect eukaryotes as well as prokaryotes. Third, we collected our sample from the top 10 cm of soil, whereas the Stackebrandt group took soil not from the top 10 cm but from a zone 5 to 10 cm below the surface (24, 44).

Traditional mathematical analyses of biodiversity. Over several decades, ecologists have developed superb tools with which to measure biodiversity and have described two components of diversity: species richness and species evenness (for a review, see reference 16). Species richness is the number of species compared with the total number of individual organisms in a given area. Species evenness is a measure of the distribution of individual organisms among species. Estimates of species evenness and richness are then used to calculate a diversity index. Several diversity indices are described in the literature (26, 31, 41, 42). In summary, diversity increases as richness and evenness increase. Richness and evenness are not mutually exclusive measures. For example, in most cases, if richness is high, evenness will also be high, because as the number of species in a given area increases, there tend to be fewer individuals of each species.

We have given very careful thought to how best to describe diversity among a group of organisms (in this case, microorganisms) in an environment in which most of the organisms are unknown and often unculturable. Unknown organisms, whose identity is known only through a portion of SSU rDNA sequence, usually have little homology to known organisms. In our experience, the average SSU rDNA clone from an SSU rDNA library made from soil DNA has about 85% homology with the closest relative in the databases. This is not sufficient information to define these organisms at the genus level, let alone the species level. More information on the identity of these phylotypes would require culturing of the organism and analysis of physiological and morphological characteristics. As this is either not possible or very difficult with most bacteria in bulk soil, species or genus classification of most soil bacteria is not possible. As a result of the inability to identify uncultured bacteria to the species or genus level, it is not possible to use the diversity indices above, which are for known organisms.

MDS analysis. In addition to the traditional use of dendrograms or phylogenetic trees, we have adapted MDS analysis of distance matrices to describe relationships among clones. MDS was first widely applied in psychology by Richardson (33), was refined substantially by Torgerson (46), and is now widely used in the social sciences. MDS analysis can be used with any data set for which a distance matrix is available. No additional biological assumptions are needed for MDS analysis. Alignments of homologous sequences are ideally suited for MDS analysis because distance matrices are easily and routinely calculated with such data. Principal-component analysis, which has been used in the past to describe microbial communities in soil by using physiological and biochemical analyses of cultured organisms (35), could also be used to describe the relationships among rDNA sequences. However, MDS results in a better correlation of distances, in reduced dimensions, with the original distance matrix (data not shown). The MDS plot allows identification of relationships that are often difficult to see in the dendrogram and allows visualization of the distribution of organisms found in the site, and this diversity is particularly noticeable among the organisms that cannot be assigned to any major taxon.

Summary. Our objective in this work was to obtain an understanding of the microbial diversity of this Wisconsin soil, using a culture-independent molecular approach. We did not intend to describe new phylogenetic groups. In so doing, we decided that it was better to sequence short fragments of many clones than to sequence 1,000 nucleotides of several clones. Four aspects of this survey of the microbial diversity in a Wisconsin soil reduced the bias in our analysis. First, culturing bias was removed by analyzing rDNA sequences amplified from DNA isolated from soil. Second, a rapid method for extraction of DNA from soil was developed to minimize DNA shearing and to ensure that microorganisms from many taxa would be lysed efficiently. Third, primer sequences that are highly conserved among the domains *Bacteria*, *Archaea*, and *Eucarya* were chosen for the amplification of the rDNA. Fourth, the rDNA PCR products were cloned directly into the vector without the need for any restriction digestion of the products. This removes any bias that might result from the presence of a particular restriction site in any soil microorganism.

The phylogenetic trees and MDS plots depict the enormous diversity of life located in the clover-grass pasture. Diverse representatives of both the *Bacteria* and the *Eucarya* were found in the rDNA library. Only 4% of the 124 rDNA clones were considered possible duplicates, and several clades of microorganisms had no close relatives in the database.

ACKNOWLEDGMENTS

This work was supported by funds from USDA NRI grant 94-37305-0767, USDA Risk Assessment Program grant 94-33120-0433, USEPA Risk Assessment cooperative agreement CR-822882-01-0, and Hatch project 5201 from the College of Agricultural and Life Sciences, University of Wisconsin—Madison. The WICST is maintained with support from USDA NRI grant 94-37108-0919 and from Hatch project no. 3300 from the College of Agricultural and Life Sciences, University of Wisconsin—Madison.

We thank Josh Posner and other colleagues at the College of Agricultural and Life Sciences, University of Wisconsin—Madison, for initiating and maintaining the WICST and allowing us access to one of the sites of that trial for this work.

REFERENCES

1. **Alexander, M.** 1977. Introduction to soil microbiology, p. 472. John Wiley & Sons, New York.
2. **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.
3. **Atlas, R. M.** 1984. Diversity of microbial communities. Adv. Microb. Ecol. **7:**1–47.
4. **Atlas, R. M., A. Horowitz, M. Krichevsky, and A. K. Bej.** 1991. Response of microbial populations to environmental disturbance. Microb. Ecol. **22:**249–256.
5. **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.
6. **Biovin-Jahns, V., A. Bianchi, R. Ruimy, J. Garcin, S. Daumas, and R. Christen.** 1995. Comparison of phenotypical and molecular methods for the identification of bacterial strains isolated from a deep subsurface environment. Appl. Environ. Microbiol. **61:**3400–3406.
7. **Brock, T. D., D. W. Smith, and M. T. Madigan.** 1984. Biology of microorganisms, 4th ed. Prentice-Hall, Inc., Englewood Cliffs, N.J.
8. **Choi, B. K., B. J. Paster, F. E. Dewhirst, and U. B. Gobel.** 1994. Diversity of cultivable and uncultivable oral spirochetes from a patient with severe destructive periodontis. Infect. Immun. **62:**1889–1895.
9. **DeLong, E. F., K. Y. Wu, B. B. Prezelin, and R. V. Jovine.** 1994. High abundance of *Archaea* in Antarctic marine picoplankton. Nature (London) **371:**695–697.
10. **Embley, T. M., R. P. Hirt, and D. M. Williams.** 1994. Biodiversity at the molecular level: the domains, kingdoms, and phyla of life. Phil. Trans. R. Soc. Lond. Ser. B **345:**21–33.
11. **Farrelly, V., F. A. Rainey, and E. Stackebrandt.** 1995. Effect of genome size and *rrn* gene copy number on PCR amplification of 16S rRNA genes from a mixture of bacterial species. Appl. Environ. Microbiol. **61:**2798–2801.
12. **Fredrickson, J. K., and C. Hagedorn.** 1992. Overview: identifying ecological effects from the release of genetically engineered microorganisms and microbial pest control agents, p. 559–578. *In* M. A. Levin, R. J. Seidler, and M. Rogul (ed.), Microbial ecology: principles, methods and applications. McGraw-Hill, New York.
13. **Harris, D.** 1994. Analyses of DNA extracted from microbial communities, p. 111–118. *In* K. Ritz, J. Dighton, and K. E. Giller (ed.), Beyond the biomass. John Wiley & Sons, Chichester, United Kingdom.
14. **Hassink, J., G. Lebbink, and J. A. van Veen.** 1991. Microbial biomass and activity of a reclaimed polder soil under conventional or reduced-input farming system. Soil Biol. Biochem. **23:**507–513.
15. **Jukes, T. H., and C. R. Cantor.** 1969. Evolution of protein molecules, p. 21–132. *In* H. N. Munro (ed.), Mammalian protein metabolism. Academic Press, New York.
16. **Kennedy, A. C., and K. L. Smith.** 1995. Soil microbial diversity and the sustainability of agricultural soils. Plant Soil **170:**75–86.
17. **Klug, M. J., and J. M. Tiedje.** 1994. Response of microbial communities to changing environmental conditions: chemical and physiological approaches, p. 371–378. *In* R. Guerrero and C. Pedros-Alio, Trends in microbial ecology. Spanish Society for Microbiology, Barcelona, Spain.
18. **Kopczynski, E. D., M. M. Bateson, and D. M. Ward.** 1994. Recognition of chimeric small-subunit ribosomal DNAs composed of genes from uncultivated microorganisms. Appl. Environ. Microbiol. **60:**746–748.
19. **Kumar, S., K. Tomura, and M. Nei.** 1993. MEGA: Molecular Evolutionary Genetics Analysis, version 1.0. The Pennsylvania State University, University Park.
20. **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.
21. **Lane, D. J., B. Pace, G. J. Olsen, D. A. Stahl, M. L. Sogin, and N. R. Pace.** 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. Proc. Natl. Acad. Sci. USA **82:**6955–6959.
22. **Leff, G. L., J. R. Dana, J. V. McArthur, and L. J. Shimkets.** 1995. Comparison of methods of DNA extraction from stream sediments. Appl. Environ. Microbiol. **61:**1141–1143.
23. **Leung, K., L. S. England, M. B. Cassidy, J. T. Trevors, and S. Weir.** 1994. Microbial diversity in soil: effect of releasing genetically engineered microorganisms. Mol. Ecol. **3:**413–422.
24. **Liesack, W., and E. Stackebrandt.** 1992. Occurrence of novel groups of the domain *Bacteria* as revealed by analysis of genetic material isolated from an Australian terrestrial environment. J. Bacteriol. **174:**5072–5078.
25. **Liesack, W., H. Weyland, and E. Stackebrandt.** 1991. Potential risks of gene amplification by PCR as determined by 16S rDNA analysis of a mixed-culture of strict barophilic bacteria. Microb. Ecol. **21:**191–198.
26. **MacArthur, R.** 1960. On the relative abundance of species. Am. Nat. **94:**25–35.
27. **More, M. I., J. B. Herrick, M. C. Silva, W. C. Ghiorse, and E. L. Madsen.** 1994. Quantitative cell lysis of indigenous microorganisms and rapid extraction of microbial DNA from sediment. Appl. Environ. Microbiol. **60:**1572–1580.
28. **Neefs, J.-M., Y. Van de Peer, L. Hendricks, and R. de Wachter.** 1990. Compilation of small ribosomal subunit RNA sequences. Nucleic Acids Res. **18:**2237–2317.
29. **Ogram, A. V., G. S. Sayler, and T. J. Barkay.** 1987. The extraction and purification of microbial DNA from sediments. J. Microbiol. Methods **7:**57–66.
30. **Olsen, G. J., C. R. Woese, and R. Overbeek.** 1994. The winds of (evolutionary) change: breathing new life into microbiology. J. Bacteriol. **176:**1–6.
31. **Peet, R. K.** 1984. The measurement of species diversity. Annu. Rev. Ecol. Sys. **5:**285–307.
32. **Reysenbach, A.-L., L. J. Giver, G. S. Wickham, and N. R. Pace.** 1992. Differential amplification of rRNA genes by polymerase chain reaction. Appl. Environ. Microbiol. **58:**3417–3418.
33. **Richardson, M. W.** 1938. Multidimensional psychophysics. Psychol. Bull. **35:**659–660.
34. **Robison-Cox, J. F., M. M. Bateson, and D. M. Ward.** 1995. Evaluation of nearest-neighbor methods for detection of chimeric small-subunit rRNA sequences. Appl. Environ. Microbiol. **61:**1240–1245.
35. **Rosswall, T., and E. Kvillner.** 1978. Principal components and factor analysis for the description of microbial populations. Adv. Microb. Ecol. **2:**1–48.
36. **Roszak, D. B., and R. R. Colwell.** 1987. Survival strategies of bacteria in the natural environment. Microbiol. Rev. **51:**365–379.
37. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for constructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
38. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
39. **Schmidt, T. M., E. F. DeLong, and N. R. Pace.** 1991. Analysis of a marine phytoplankton community by 16S rRNA gene cloning and sequencing. J. Bacteriol. **173:**4371–4378.
40. **Schulte, E. E., J. B. Peters, and P. R. Hodgson.** 1987. Wisconsin procedures for soil testing, plant analysis, and feed and forage analysis. Department of Soil Science, University of Wisconsin—Madison.
41. **Shannon, C. E., and W. Weaver.** 1949. The mathematical theory of communication, p. 117. University of Illinois Press, Urbana.
42. **Simpson, E. H.** 1949. Measurement of diversity. Nature (London) **163:**688.
43. **Stackebrandt, E., and B. M. Goebel.** 1994. Taxonomic note: a place for DNA-DNA reassociation and 16S rRNA analysis in the present species definition in bacteriology. Int. J. Syst. Bacteriol. **44:**846–849.
44. **Stackebrandt, E., W. Liesack, and B. M. Goebel.** 1993. Bacterial diversity in a soil sample from a subtropical Australian environment as determined by 16S rDNA analysis. FASEB J. **7:**232–236.
45. **Staley, J. T., and A. Konopka.** 1985. Measurement of *in situ* activities of nonphotosynthetic microorganisms in aquatic and terrestrial habitats. Annu. Rev. Microbiol. **39:**321–346.
46. **Torgerson, W. S.** 1952. Multidimensional scaling. I. Theory and method. Psychometrica **17:**401–419.
47. **Torsvik, V., J. Goksoyr, and F. L. Daae.** 1990. High diversity in DNA of soil bacteria. Appl. Environ. Microbiol. **56:**782–787.
48. **Ueda, T., Y. Suga, and T. Matsuguchi.** 1995. Molecular phylogenetic analysis of a soil microbial community in a soybean field. Eur. J. Soil Sci. **46:**415–421.
49. **Ueda, T., Y. Suga, N. Yahiro, and T. Matsuguchi.** 1995. Genetic diversity of N_2-fixing bacteria associated with rice roots by molecular evolutionary analysis of a *nifD* library. Can. J. Microbiol. **41:**235–240.
50. **Ward, D. M., R. Weller, and M. M. Bateson.** 1990. 16S rRNA sequences reveal numerous uncultured microorganisms in a natural community. Nature (London) **345:**63–65.
51. **Weinberg, S. R., and G. Stotzky.** 1972. Conjugation and genetic recombination of *Escherichia coli* in soil. Soil Biol. Biochem. **4:**171–180.
52. **Wilkinson, L.** 1992. SYSTAT: the system for statistics. Version 5.2. SYSTAT, Evanston, Ill.
53. **Winter, K., J. Filser, and F. Beese.** 1991. Mikrobielle Biomasse und Aktivität in Hopfengarten unterschiedlicher Bewirtschaftung. Verh. Ges. Okol. **XX/1:**49–54.
54. **Woese, C. R.** 1987. Bacterial evolution. Microbiol. Rev. **51:**221–271.
55. **Young, F. W.** 1987. Multidimensional scaling: history, theory, and applications, p. 317. Lawrence Erlbaum Publishers, Hillsdale, N.J.