Characterization of Microbial Diversity by Determining Terminal Restriction Fragment Length Polymorphisms of Genes Encoding 16S rRNA

WEN-TSO LIU,[1] TERENCE L. MARSH,[1] HANS CHENG,[2] AND LARRY J. FORNEY[1]*

Department of Microbiology and Center for Microbial Ecology, Michigan State University,[1] and USDA-ARS Avian Disease & Oncology Laboratory,[2] East Lansing, Michigan 48824

A quantitative molecular technique was developed for rapid analysis of microbial community diversity in various environments. The technique employed PCR in which one of the two primers used was fluorescently labeled at the 5′ end and was used to amplify a selected region of bacterial genes encoding 16S rRNA from total community DNA. The PCR product was digested with restriction enzymes, and the fluorescently labeled terminal restriction fragment was precisely measured by using an automated DNA sequencer. Computer-simulated analysis of terminal restriction fragment length polymorphisms (T-RFLP) for 1,002 eubacterial sequences showed that with proper selection of PCR primers and restriction enzymes, 686 sequences could be PCR amplified and classified into 233 unique terminal restriction fragment lengths or "ribotypes." Using T-RFLP, we were able to distinguish all bacterial strains in a model bacterial community, and the pattern was consistent with the predicted outcome. Analysis of complex bacterial communities with T-RFLP revealed high species diversity in activated sludge, bioreactor sludge, aquifer sand, and termite guts; as many as 72 unique ribotypes were found in these communities, with 36 ribotypes observed in the termite guts. The community T-RFLP patterns were numerically analyzed and hierarchically clustered. The pattern derived from termite guts was found to be distinctly different from the patterns derived from the other three communities. Overall, our results demonstrated that T-RFLP is a powerful tool for assessing the diversity of complex bacterial communities and for rapidly comparing the community structure and diversity of different ecosystems.

Species richness (the number of species within a community) and species evenness (the sizes of species populations within a community) are two essential parameters for defining community structure and diversity. Quantitation of these two parameters is severely limited by conventional culture-dependent methods by virtue of the fact that large fractions (>85 to 99.999%) of the organisms existing in nature appear to be refractory to cultivation (1, 3, 11, 30, 35). Another limitation of these approaches is that any departure from the original environmental parameters during cultivation can alter the community structure through the imposition of new selective conditions (8). In effect, a new community structure evolves, which may not accurately reflect the original structure.

Several molecular approaches now provide powerful adjuncts to the culture-dependent techniques. One approach in particular that couples PCR and rRNA-based phylogeny has been effective in the exploration of microbial environments and the identification of uncultured organisms. The stepwise strategy of this approach is to isolate total community DNA and use this DNA as a template for PCR amplification of 16S rRNA genes with universal or domain-specific primers. This is usually followed by construction of a clone library for genes encoding rRNA (rDNAs) and rapid screening of the library based on sequence differences (4) or by determining restriction fragment length polymorphisms (RFLPs) of the rDNAs (19). The richness and evenness of a community are qualitatively estimated based on the number of unique clones and the relative frequencies of the various "ribotypes" detected. While this approach obviates the requirement for cultivation, construction and screening of clone libraries are laborious. More recently, techniques dependent on DNA melting behavior (20, 21) or single-strand DNA conformation (14) have been developed as a means of circumventing library construction. These methods have been employed to assess community structure and to provide relatively crude estimates of species diversity, but are limited by the comparatively insensitive staining technologies employed and do not provide information concerning the specific phylogenetic groups that comprise a microbial community.

Here we report on a quantitative molecular method for rapid analysis of complex microbial communities in which an existing technology was extended. RFLP analysis of 16S rDNA, otherwise known as amplified rDNA restriction analysis (ARDRA), has been used for several years as a method for rapid comparison of rDNAs (13, 19). Briefly, rDNAs are obtained by PCR amplification by using universal primers, and the product is digested with restriction enzymes with 4-bp recognition sites. The typical analysis of restriction digests for isolates or clones is performed on relatively low-resolution agarose gels. For community analysis, the potentially large number of fragments can be resolved by using polyacrylamide gels to produce a community-specific pattern (17, 18). However, ARDRA is of limited use for demonstrating the presence of specific phylogenetic groups or for estimating species richness and evenness. In the technique reported on here, the initial steps of DNA isolation, PCR amplification, and restriction were similar to those used for ARDRA. However, one of the primers used was labeled with a fluorescent dye so that when the preparation was analyzed with an automated DNA sequencer, the sizes of only the terminal restriction fragment (T-RF) could be determined and the amount could be quan-

---

* Corresponding author. Mailing address: Department of Microbiology, 69 Giltner Hall, Michigan State University, East Lansing, MI 48824. Phone: (517) 432-1365. Fax: (517) 432-3770. E-mail: forney @pilot.msu.edu.

TABLE 1. Sequences of primers used in this study

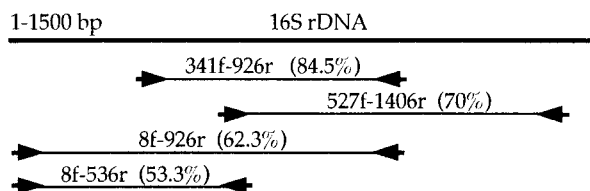| Primer	Specificity	Sequence (5′→3′)	Reference(s)
8f	Domain	AGAGTTTGATCCTGGCTCAG	1
341f	Domain	CCTACGGGAGGCAGCAG	21
536r	Universal	CAGCMGCCGCGGTAATWC	1, 11
527f	Universal	ACCGCGGCCKGCTGGC	This study
926r	Domain	CCGTCAATTCCTTTRAGTTT	20
1406r	Universal	ACGGGCGGTGTGTRC	1

FIG. 1. Primer pairing and target sites used in this study. The fraction of the database that could be matched by a particular primer pair (values in parentheses) was determined by using pattern-matching software (Pat Scan). The settings for mismatch/insertion/deletion in Pat Scan that determine primer specificity were as follows: 0/0/0 at the last four nucleotides from the 3′ end and 3/0/0 at any other nucleotide.

tified. Due to the terminal RFLPs (T-RFLPs) found in 16S rDNAs, the method can be used with DNA from complex microbial communities and can provide a sensitive and rapid means to assess community diversity and obtain a distinctive fingerprint of a microbial community.

MATERIALS AND METHODS

16S rRNA sequence database. The Ribosomal Database Project (RDP) database (release 5.0) contains almost 2,900 aligned 16S rRNA sequences (16). For the analysis of T-RFLPs, a subset of sequences containing at least 1,420 bp (*Escherichia coli* numbering) was selected from the RDP database; this subset included all of the sequences with fewer than two contiguous unassigned positions and with primer sites complementary to the primers selected. The final subset chosen included 1,007 eubacterial sequences and 95 archaebacterial sequences.

Computation of 16S rDNA primer pair specificity and terminal fragment lengths from the RDP database. A pattern-matching program, Pat Scan, developed by R. Overbeek (Argonne National Laboratory, Chicago, Ill.), was used for computing 16S rDNA primer specificity and for predicting the theoretical lengths of T-RFs. Primer specificity was estimated by considering the specificity of forward and reverse primer sequences, the individual primer annealing specificity, and the length of the sequence between the forward and reverse primers. The sequences and annealing sites of the forward and reverse primers considered in this study are shown in Fig. 1 and Table 1. The specificity of each pair of primers was expressed as the percentage of the total number of sequences in the database that were matched. The matched sequences were then used as a subset of the database for computing the lengths of T-RFs that would be obtained after digestion with a given restriction enzyme. The following 10 restriction enzymes with a 4-bp recognition site were used in this simulation: *Alu*I (AG′CT), *Bst*UI (CG′CG), *Dde*I (C′TNAG), *Hae*III (GG′CC), *Hha*I (GCG′C), *Hin*fI (G′ANTC), *Mbo*I (′GATC), *Msp*I (C′CGG), *Rsa*I (GT′AC), and *Taq*I (T′CGA). The frequency distribution of T-RF lengths was statistically analyzed by using SAS software (SAS Institute, Inc., Cary, N.C.).

Model and natural microbial communities. The six bacteria in a model microbial community were *Alcaligenes faecalis* ALC1, *Ralstonia eutropha*, *Arthrobacter globiformis* ATCC 8010, *Bacillus subtilis* BAC11, *Escherichia coli* HB101, and *Pseudomonas aeruginosa* ATCC 27853. All of these bacteria were obtained from the Department of Microbiology culture collection at Michigan State University. The samples of natural microbial communities studied included samples of (i) activated sludge from the aeration tank of a local municipal wastewater treatment plant in East Lansing, Mich. (ELAS sample); (ii) enriched sludge from a laboratory scale bioreactor obtained after 3 months of cultivation with substrate containing (per liter) 1 g of glucose, 1 g of peptone, and 1 g of KH_2PO_4 (carbon loading, 0.67 kg/m^3 · day) (MITI sample); (iii) aquifer sand from the groundwater interface of a trichloroethylene- and jet fuel-contaminated site; and (iv) the gut DNA of a termite, *Reticulitermes flavipes*, graciously supplied by T. Lilburn and J. Breznak. These samples were lyophilized or stored at −20°C prior to DNA extraction.

DNA extraction. Genomic DNA of each bacterium and total DNAs from environmental samples were isolated by using a previously described protocol (27), with minor modifications. Approximately 10 mg (dry weight) of biomass or 5 g of aquifer sand was homogenized with a tissue grinder or a sterilized mortar and suspended in 9.0 ml of lysis buffer (100 mM Tris-HCl, 100 mM EDTA, 0.75 M sucrose). Lysozyme (1.0 mg/liter) and achromopeptidase (100 μg/liter) were added, and the suspension was incubated at 37°C for 30 min. After the addition of proteinase K (200 μg/liter) and sodium dodecyl sulfate (1%, wt/vol), the samples were incubated at 37°C for 2 h. During that period, each tube was gently inverted several times every 30 min. The mixtures were then subjected to three freeze-thaw cycles. Hexadecyltrimethyl ammonium bromide and sodium chloride were added to final concentrations of 1.0% (wt/vol) and 0.7 M, respectively, and the mixtures were incubated for 20 min at 65°C to precipitate polysaccharides and residual proteins. Each precipitate was removed by extraction with an equal volume of phenol-chloroform-isoamyl alcohol (25:24:1), followed by extraction with chloroform-isoamyl alcohol (24:1). DNA was recovered by the addition of 1 volume of isopropanol followed by centrifugation at 10,000 × g for 10 min. The pellet was washed with 70% ethanol, dried, and resuspended in 200 to 400 μl of distilled water. Additional purification, if necessary, was by electrophoresis in a low-melting-point agarose gel (26).

PCR conditions. Reaction mixtures for PCR contained 1× PCR buffer, each deoxynucleoside triphosphate at a concentration of 200 μM, 1.5 mM MgCl₂, each primer at a concentration of 0.1 μM, and 2.5 U of *Taq* DNA polymerase (Gibco BRL, Gaithersburg, Md.) in a final volume of 100 μl. The primers used, 8f-Hex and 926r-TET, were labeled at the 5′ end with the phosphoramidite dyes 5-hexachlorofluorescein and 5-tetrachlorofluorescein (Operon, Inc., Alameda, Calif.). DNA amplification was performed with a model 9600 thermal cycler (Perkin-Elmer, Norwalk, Conn.) by using the following program: a 3-min hot start at 94°C, followed by 35 cycles consisting of denaturation (30 s at 94°C), annealing (45 s at 56°C), and extension (2 min at 72°C) and a final extension at 72°C for 3 min. Amplified DNA was verified by electrophoresis of aliquots of PCR mixtures (2 μl) in 1.0% agarose in 1× TAE buffer.

16S rDNA RFLP. Fluorescently labeled PCR products (100 μl) were purified by using WIZARD PCR purification columns (Promega, Madison, Wis.) and were eluted in a final volume of 50 μl. Aliquots (10 μl) of amplified 16S rDNAs of six bacterial strains were separately digested with *Hha*I (Boehringer Mannheim Biochemicals [BMB] Indianapolis, Ind.) according to the manufacturer's instructions. Aliquots (5 μl) of each digested product or a mixture of the six digested products were mixed with 2 μl of 5× loading dye buffer and resolved by electrophoresis through an 8% nondenaturing acrylamide gel (acrylamide–*N*,*N*-methylenebisacrylamide, 29:1 [17]). The RFLPs were visualized by staining with SYBR green I dye according to the instructions of the manufacturer (FMC Bioproducts, Rockland, Maine) and then photographed. For natural communities, 6-μl aliquots of purified PCR products were digested with 20 U of *Hha*I, *Msp*I, *Rsa*I, or *Hha*I plus *Msp*I (BMB), and the T-RFLP fingerprint of each community was determined as described below.

16S rDNA T-RFLP. The precise lengths of the T-RFs from the amplified rDNA products were determined by electrophoresis with a model 373A automated sequencer (Applied Biosystems Instruments [ABI], Foster City, Calif.), as follows. For the model community, 2 μl of digested DNA from each strain or a composite of the digests was mixed with 2 μl of deionized formamide, 0.5 μl of loading buffer (ABI), and 0.5 μl of DNA fragment length standard (Rox 2500 or Tamara 2500; ABI). The sizes of the DNA fragments in these mixtures were 37, 94, 109, 116, 172, 186, 222, 233, 238, 269, 286, 361, 470, 490, 536, 827, 1,115, 1,181, 1,722, 2,008, 2,162, 2,465, 2,481, 2,860, 4,529, 4,771, 5,099, and 14,079 bp. This mixture was then denatured at 94°C for 5 min and immediately chilled on ice prior to electrophoresis with the automated DNA sequencer in the GeneScan mode. Aliquots (2.5 μl) of the mixtures were loaded onto a 36-cm 6% denaturing polyacrylamide gel. Electrophoresis was done for up to 20 h with limits of 2,500 V and 40 mA. After electrophoresis, the lengths of fluorescently labeled T-RFs were determined by comparison with internal standards by using GeneScan software (ABI).

Species richness, community signature, and community similarity. The similarity of communities could be estimated by visual comparison of the electropherograms or by numerically analyzing the pattern of T-RFLPs in gel images with GelCompare software (version 3.1; Applied Maths, Kortrijk, Belgium). Following the normalization (no background subtraction) and image analysis steps of the software, a stacked band pattern of three T-RFLPs from the same terminus was generated for each community. Two different band-matching coefficients were used to produce a similarity matrix for the pattern fragments in the samples. The Jaccard coefficient considered the presence or absence of bands and the number of T-RFs in common in communities, as well as the total number of T-RFLPs observed. The area-sensitive coefficient took the Jaccard coefficient into account, as well as the difference in the relative abundance (peak area) of each T-RF. The unweighted pair group method using average linkages was applied to cluster patterns to obtain a similarity dendrogram for each coefficient. For both coefficients a value of 0 means that the community fingerprints are completely different from one another, and a value of 1 indicates that they are identical.

TABLE 2. Potential number of 5′- and 3′-T-RF lengths or ribotypes generated for sequences that could be annealed from the 1,102 complete 16S rRNA sequences in the RDP database with four different primer pairs

| Restriction enzyme | No. of ribotypes with[a]: | | | | | | | |
| | Primers 8f and 536r | | Primers 8f and 926r | | Primers 341f and 926r | | Primers 527f and 1406r | |
	5′ end	3′ end	5′ end	3′ end	5′ end	3′ end	5′ end	3′ end
*Alu*I	129	92	155	108	129	84	86	56
*Bst*UI	159	—[b]	173	72	33	67	—	61
*Dde*I	132	61	144	69	114	59	89	87
*Hae*III	180	74	159	109	93	80	84	45
*Hha*I	136	92	194	127	103	99	87	62
*Hin*fI	124	73	158	105	75	82	111	20
*Mbo*I	—	126	—	129	100	93	138	76
*Msp*I	170	41	221	91	80	102	59	36
*Rsa*I	133	35	194	27	106	32	72	79
*Taq*I	18	140	36	113	110	83	73	66

[a] The numbers of sequences which could be annealed from the 1,102 complete 16S rRNA sequences with primer sets 8f-536r, 8f-926r, 341f-926r, and 527f-1406r were 587, 686, 931, and 772.
[b] —, a restriction enzyme site was found in the primer sequence.

RESULTS AND DISCUSSION

T-RFLP in the 16S rRNA database. The ability of T-RFLP analysis to distinguish phylogenetic groups of bacteria was theoretically evaluated through a computer simulation of the T-RF size distribution for 1,102 complete bacterial sequences deposited in the RDP database. Four different pairs of domain-specific or universal PCR primers were selected to anneal to as many 16S rDNA sequences in the RDP database as possible. After digestion with any of 10 different restriction enzymes, the sizes of the terminal fragments amplified with the PCR primers were determined. The combination of PCR primers and enzyme(s) that gave the largest number of terminal fragments with unique sizes from either the 5′ or 3′ termini of the rDNAs of the amplifiable sequences was identified. This analysis indicated that the 341f-926r primer pair could theoretically anneal to the largest fraction (84.5%) of the 16S rRNA sequences in the RDP database (Fig. 1); however, the resulting PCR product was relatively short, and there were too few unique terminal fragments (from either the 5′ terminus or the 3′ terminus) produced upon digestion of the product with any of the 10 enzymes considered (Table 2). Similar results were obtained with the 527f-1406r primer pair. However, primer pairs 8f-536r and 8f-926r annealed to 587 and 686 of the 1,102 16S rRNA sequences, respectively, when the annealing criteria described above were used. By mapping the *Msp*I sites in these database subsets we identified 170 and 221 distinct 5′-terminal fragments, respectively. The high extent of length variation among the predicted T-RFs was mainly due to the occurrence of multiple insertions and deletions and true restriction site polymorphism (6). When primer pair 8f-926r and a combination of two enzymes (either *Hha*I plus *Hae*III, *Hha*I plus *Msp*I, or *Hha*I plus *Rsa*I) were used, the number of unique 5′-terminal fragments increased to 233. In contrast, digestion of the PCR products with either *Mbo*I or *Taq*I resulted in a low number of fragments with unique sizes because of the high conservation of restriction site positions.

Based on the simulations described above, we concluded that amplification with primer pair 8f-926r followed by digestion with *Hha*I or *Msp*I was the simplest way to classify the largest number of gene sequences into the largest number of unique 5′ T-RFs. However, levels of microbial diversity based on T-RFLPs would have been underestimated by a factor of three to four based on the results from the database sampled (233 terminal fragments from 686 sequences). 16S rRNA

genes from phylogenetically related organisms can result in the production of T-RFs of identical size (2, 31). For example, amplification with primer pair 8f-926r followed by digestion with either *Hha*I or *Hha*I plus *Msp*I resulted in the production of 194 or 233 5′ T-RFs, respectively, from the 16S rDNAs amplified (Fig. 2). The number of these 16S rDNAs with *Hha*I restriction sites in common ranged from 0 (unique) to 23. In the most extreme case, 23 species belonging to four genera of the *Bacteroides* group and the genus *Campylobacter* and 23 species belonging to 12 different genera of the *Neisseria* group (4 genera) and the *Vibrio* group (8 genera) had T-RFs that were 102 and 374 bp long, respectively. As a consequence of the variable conservation of restriction site positions in 16S rDNA, the resolution of T-RFLP analysis is often reduced from the level of species (unique restriction site position in each species) to that of higher-order groups. This effectively
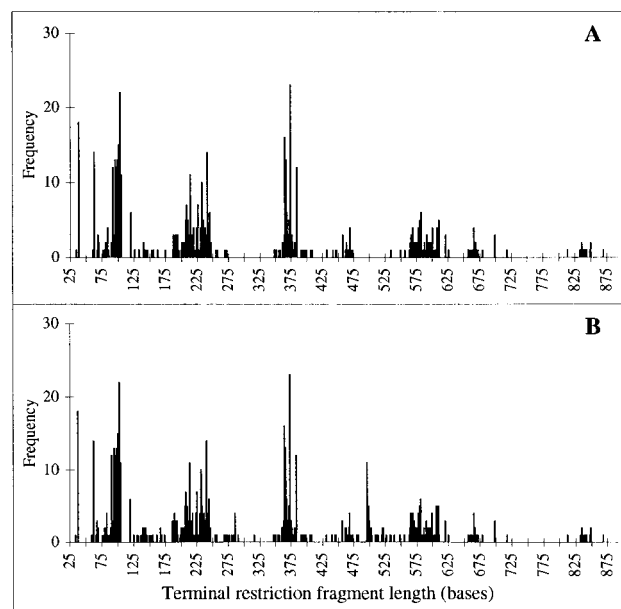

FIG. 2. Histograms of the predicted 5′ T-RF lengths of 686 rRNA sequences corresponding to restriction enzyme digestion with *Hha*I (A) and *Hha*I plus *Msp*I (B). The 686 sequences were a subset of sequences that could be hybridized from 1,102 complete 16S rRNA sequences by using the 8f-927r primer pair.
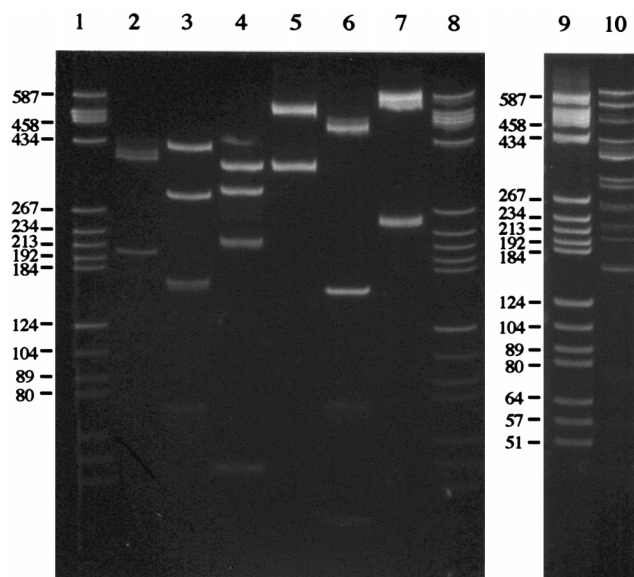
FIG. 3. RFLP analysis of 16S rDNA (ARDRA) of the six-member bacterial model community corresponding to *Hha*I digestion. Lanes 1, 8, and 9, marker V (Boehringer); lane 2, *Alcaligenes faecalis*; lane 3, *Ralstonia eutropha*; lane 4, *Arthrobacter globiformis*; lane 5, *B. subtilis*; lane 6, *E. coli*; lane 7, *Pseudomonas aeruginosa*; lane 10, mixture of the six individual digestion preparations.

FIG. 4. Electropherograms of the 5′ (A) and 3′ (B) T-RFLPs of *Hha*I-digested 16S rRNAs amplified from a six-member bacterial community corresponding to *Hha*I digestion.

reduces the complexity of the profile of T-RFs obtained from a community so that estimates of diversity within communities can readily be made.

Analysis of species diversity in a model bacterial community by T-RFLP. The computer-simulated analysis suggested that polymorphisms among T-RF lengths could be used to characterize microbial communities. To test this approach, a model community containing six known bacterial strains was constructed. PCR amplification in which primers 8f and 926r were labeled with different fluorescent dyes was used to amplify an internal region of the 16S rDNA templates of individual bacterial strains. The PCR products were digested with *Hha*I, and the products were separated with an automated sequencer, which yielded two T-RFs (5′ and 3′) for each bacterial strain. The terminal location of the fluorescent tag on the PCR products ensured that only terminal fragments were measured. The predicted outcome based on a computer-simulated restriction analysis of sequences from the RDP database indicated that the six strains used should have different RFLP (ARDRA) patterns and unique 5′ and 3′ T-RFs. The ARDRA patterns of the six strains were indeed different (Fig. 3, lanes 2 to 7). However, if the DNA fragments derived from the six strains were mixed together (lane 10), the pattern became too complex to estimate the species diversity, let alone derive phylogenetic information. In contrast, T-RFLP analysis readily resolved the six-member community by limiting the analysis to T-RFs (Fig. 4). Six unique T-RFs were observed for both the 5′
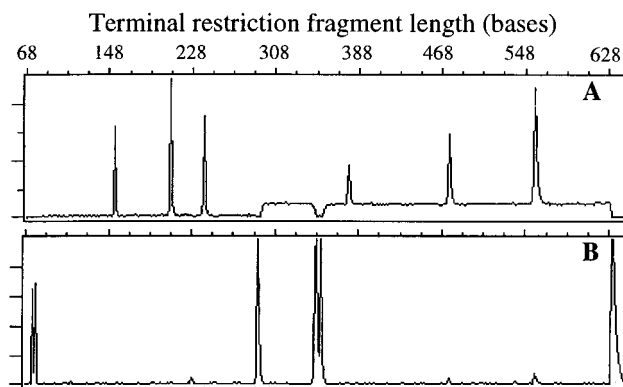
and 3′ termini of the rDNA following *Hha*I digestion, which led us to conclude there were at least six bacterial strains in the model community.

There were discrepancies between the results expected based on database analyses and the empirical data obtained. Two of the 12 T-RFs observed did not have the predicted lengths (Table 3). The observed length for the T-RF of *Alcaligenes faecalis* was found to correspond to the length from the terminus to the second *Hha*I site in the RDP sequence. The amount of *rrn* sequence variation among strains of the same species has been shown to be 0 to 5%, and thus differences in T-RFs, while not common, are also not unexpected (7, 28). This is problematic for T-RFLP analysis and several other molecular techniques (4, 14, 17–21) if species richness in a community is to be estimated correctly but should be partially overcome as more 16S rDNA sequences are included in the database. Similarly, the observed size of the 3′ T-RF of *B. subtilis* was different than the size predicted, but this was most likely due to difficulty in precisely determining the sizes of fragments larger than 600 bp with an automated sequencer. The sizing accuracy could be improved by increasing the size range of fragments used as standards. However, even in our analysis, T-RFs from 37 to 600 bp long were accurately sized to ±2 bp.

Analysis of microbial diversity of complex bacterial communities by T-RFLP. The species richness of natural communities was estimated by determining the number of unique T-RFs or ribotypes observed in digests of 16S rDNAs amplified by PCR from total community DNA. The T-RFLP pattern observed (referred to as the "community fingerprint") is a composite of the number of fragments with unique lengths and the relative abundance of each fragment as reflected by the size of each peak in the electropherogram. Samples of natural microbial communities, including ELAS and MITI samples and samples from a contaminated aquifer site and termite guts (TGs), were studied by T-RFLP analysis. Based on an analysis of electro-

TABLE 3. Observed and expected lengths of the 5′ and 3′ T-RFs for the six organisms in the model community

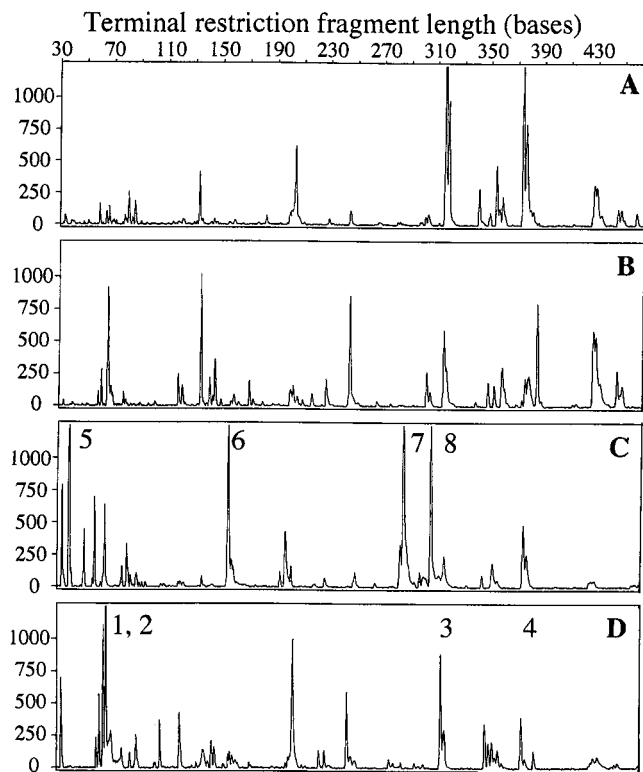| Terminal end	Observed (predicted) T-RF lengths (bp)					
	Ralstonia eutropha	*Alcaligenes faecalis*	*Arthrobacter globiformis*	*Bacillus subtilis*	*Escherichia coli*	*Pseudomonas aeruginosa*
5′	208 (207)	569 (67)	474 (472)	239 (240)	373 (373)	154 (155)
3′	291 (292)	72 (70)	76 (74)	632 (690)	350 (352)	346 (348)

Terminal restriction fragment length (bases)

FIG. 5. Electropherograms of the 5′ T-RFLPs of *Hha*I-digested 16S rDNAs amplified from four different complex communities. (A) ELAS sample. (B) Aquifer sand sample. (C) TG sample. (D) MITI sample.

pherograms from the sequencing gel images, the 5′ T-RFLP patterns revealed 72, 61, 45, and 61 unique T-RFs among these four communities after digestion with *Hha*I, *Msp*I, *Rsa*I, and *Hha*I plus *Msp*I, respectively (data not shown). These results suggested that there are high degrees of complexity in "ribotype diversity" (i.e., different T-RFs) among the four communities. A T-RFLP analysis in which only *Hha*I digestion was used showed that there were 21, 33, 33, and 36 different T-RFs in ELAS, MITI, aquifer sand, and TG samples, respectively. Based on the fluorescence intensity of each T-RF in a T-RFLP electropherogram, the ribotype present in the lowest amount accounted for about 0.4% of the total rDNA amplified.

In addition to estimating species richness or diversity, the T-RFLP pattern of a community could be viewed as a community fingerprint and used to assess the similarity of different communities. Figure 5 shows an electropherogram of the 5′ T-RFLP patterns for ELAS, MITI, aquifer sand, and TG samples following digestion of PCR products with *Msp*I plus *Hha*I. Each T-RFLP pattern constitutes a distinct phylogenetic fingerprint of a community based on the number of T-RFs and their intensities. Differences in the numerically dominant ribotypes between two T-RFLP patterns were clearly observed. For example, the dominant T-RFs of the ELAS and MITI samples were different (Fig. 5A and D). Fragments 1 to 4 in the MITI sample were either absent or present in low amounts in the ELAS sample. Since the MITI sample was originally seeded with activated sludge from different wastewater treatment processes and subsequently acclimated in a bioreactor under defined growth and substrate feed conditions for more than 3 months, the difference in numerically dominant T-RFs could have been due to differences in inocula or growth con-

ditions. This could have been anticipated since wastewater treatment and bioreactor operations are likely to select different functional and phylogenetic groups of organisms, such as nitrifying (32, 33), denitrifying (9), and uncultured phosphorus-removing bacteria (4, 34).

Certain major T-RFs (Fig. 5C, fragments 5 to 8) observed in the TG sample were not observed in the T-RFLP fingerprints of the other three communities. Among these was a 38-bp 5′ T-RF in the fingerprint of the TG sample produced following digestion with *Hha*I (Fig. 5C). Analysis of the 16S rRNA gene sequences in the RDP database showed that there were 14 sequences from spirochetes and related organisms that could theoretically be amplified with primers 8f and 926r and yield a 38-bp 5′ T-RFLP following digestion with *Hha*I. Moreover, the electropherograms of the TG sample following digestion with *Msp*I or *Rsa*I also had 5′ and 3′ T-RFs that were predicted based on the sequences of 16S rRNA genes of spirochetes (data not shown). Thus, the T-RFLP fingerprints of TGs suggest that spirochetes are a dominant ribotype in this community, and this suggestion is consistent with results of previous studies that have shown that spirochetes with distinctive morphological features, metabolic functions, and 16S rDNA sequences are among the most abundant groups of bacteria in TGs (5, 22). However, it should be noted that nine other 16S rRNA gene sequences in the database would theoretically also yield a 38-bp 5′ T-RF following digestion of their 16S rRNA genes with *Hha*I plus *Msp*I; these sequences include those of the 16S rRNA genes of *Spromusa paucivorans*, *Atopobium minutum*, *Polyangium cellulosum*, two *Eubacterium* spp., two *Chondromyces* spp., and two *Geotoga* spp. However, by determining the expected sizes of the 5′ T-RFs following simulated digestion of these sequences with *Msp*I or *Rsa*I, it was possible to conclude that the 38-bp T-RF in the electropherogram of the TG sample following digestion with *Hha*I was probably not derived from these organisms. These data illustrate the utility of T-RFLP analyses for presumptively identifying bacterial groups within a microbial community.

The degrees of similarity among the communities were quantified by numerically analyzing the band patterns of T-RFLPs. Each community was represented by a restriction fragment pattern constructed by stacking three individual 5′ T-RFLP patterns corresponding to three restriction enzyme digestions by using image analysis software (Fig. 6). Similarity dendrograms based on two different band-based coefficients both indicated that the composite T-RFLP pattern of the TG community was very different from the T-RFLP patterns of the three other communities, which suggested that there were significant differences in the microbial community structures (Fig. 6). The similarity dendrograms also suggested that the community structure of the MITI sample was more similar to the community structure of the aquifer sand sample than to the community structure of the ELAS sample. As indicated above, this could have been caused by selective enrichment after activated sludge was seeded into the bioreactor fed glucosepeptone. Although based on community similarity the MITI and aquifer sand samples were most closely related, the degree of similarity was much lower than that between two size references, and the intensities of the major T-RFs in the communities were apparently very different (Fig. 5).

The extent to which T-RFLP analysis can take advantage of rRNA-based phylogenies is dependent on the actual universalness or specificity of primer pairs. Universal primers, by definition, should be complementary to all known sequences. However, none of the known universal primers for 16S rDNA have been shown to hybridize to or amplify all sequences from the eukaryal, bacterial, and archaeal domains (6, 36). Further-
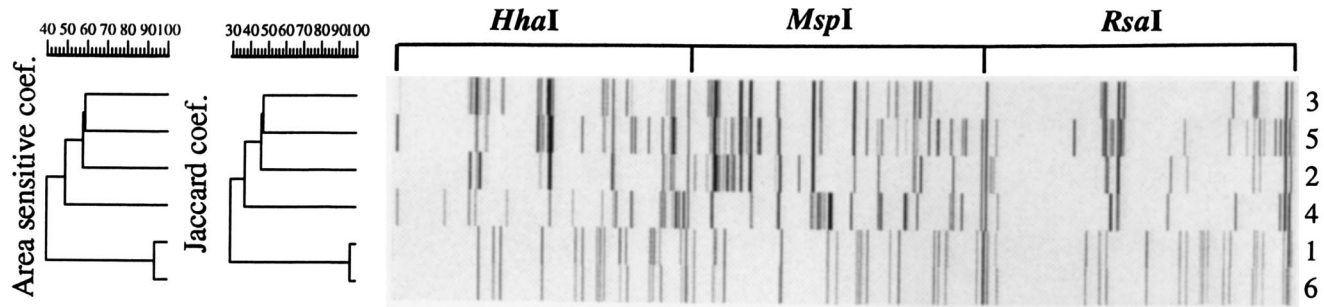
FIG. 6. Dendrograms of community relatedness of four different complex communities based on two different band-based coefficients, the Jaccard and area-sensitive coefficients. Each community was represented by a phylogenetic signature constructed by stacking three individual 5′ T-RFLP patterns corresponding to the *Hha*I, *Msp*I, and *Rsa*I digests by using image analysis software.

more, the universal primers that are currently used were designed by using the existing 16S rRNA database, which may represent only a portion of the total species diversity of the microbial world (1, 30). The primers that we used for community analysis were eubacterium-specific primers that were previously described, and our analysis was limited to this domain. A search of the subset of RDP sequences used for our database analysis clearly shows that primer pair 8f-926r, which we used for most of the T-RFLP community analyses, excludes the *Archaea*. This finding is based on a search that permitted no mismatches in the five residues at the 3′ terminus of the primer but permitted up to three mismatches in the remainder of the primer. This primer pair appears to exhibit a potential bias within the bacterial domain. A search of the bacterial domain by using the criteria described above indicated that chloroplasts, the α group, the *Aeromonas* subgroup, and the *Bifidobacterium* subgroup are recognized at levels of 23, 5.6, 21, and 1%, respectively, by the 8f-926r primer pair. However, the extent to which these phylogenetic groups are amplified under our PCR conditions is not known. The remainder of the database is recognized by this primer pair at levels of greater than 75% under the conditions specified above.

Data obtained by using T-RFLPs should be cautiously interpreted. For example, the number of populations represented in the fingerprint of any given community depends on the rank abundance of each population. Microbial populations that are not numerically dominant are not represented, because the template DNAs from these populations represent a small fraction of the total community DNA. Consequently, the species diversity of the microbial community is underestimated. Moreover, differences in gene copy number between species (10) and biases introduced during cell lysis, DNA extraction, and PCR amplification (12, 15, 23–25, 29) may yield a mixture of products that do not accurately reflect the rank abundance of the original community DNA template, thus skewing the apparent abundance of different populations.

In summary, the protocol developed here was shown to be useful for estimating the phylogenetic diversity and composition of communities in various environments. It provides a means to determine the number and abundance of numerically dominant ribotypes within a community of PCR-amplified 16S rDNAs. Using universal or domain-specific primers, one can draw only very general phylogenetic inferences from the results. The degree to which detailed phylogenetic information can be obtained with this approach reflects the level of phylogenetic specificity that can be imparted to the PCR primer. The method does produce distinct community signatures that can be used to assess the similarity of different communities. This approach should provide a facile means to assess micro-

bial diversity and to assess changes in microbial community structure that occur on temporal or spatial scales or that occur in response to environmental perturbations.

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. **Avaniss-Aghajani, E., K. Jones, A. Holtzman, T. Aronson, N. Glover, M. Boian, S. Froman, and C. F. Brunk.** 1996. Molecular technique for rapid identification of mycobacteria. J. Clin. Microbiol. **34:**98–102.
3. **Barns, S. M., R. E. Fundyga, M. W. Jefferies, 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. **Bond, P. L., P. Hugenholtz, J. Keller, and L. L. Blackall.** 1995. Bacterial community structures of phosphate-removing and non-phosphate-removing activated sludge from sequencing batch reactors. Appl. Environ. Microbiol. **61:**1910–1916.
5. **Breznak, J. A.** 1984. Hindgut spirochetes of termites and *Cryptocercus punctulatus*, p. 67–70. *In* N. R. Krieg and J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 1. The Williams & Wilkins Co., Baltimore, Md.
6. **Brunk, C. F., E. Avaniss-Aghajani, and C. A. Brunk.** 1996. A computer analysis of primer and probe hybridization potential with bacterial small-subunit rRNA sequences. Appl. Environ. Microbiol. **62:**872–879.
7. **Clayton, R. A., G. Sutton, P. S. Hinkle, C. Bult, and C. Fields.** 1995. Intraspecific variation in small-subunit rRNA sequences in GenBank: why single sequences may not adequately represent prokaryotic taxa. Int. J. Syst. Bacteriol. **45:**595–599.
8. **Dunbar, J., S. White, and L. J. Forney.** 1997. Genetic diversity through the looking glass: effect of enrichment bias. Appl. Environ. Microbiol. **63:**1326–1331.
9. **Evans, P. J., D. T. Mang, and L. Y. Young.** 1991. Degradation of toluene and *m*-xylene and transformation of *o*-xylene by denitrifying enrichment cultures. Appl. Environ. Microbiol. **57:**450–454.
10. **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.
11. **Giovannoni, S. J., T. B. Britschgi, C. L. Moyer, and K. G. Field.** 1990. Genetic diversity in Sargasso Sea bacterioplankton. Nature (London) **345:**60–63.
12. **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.

13. **Laguerre, G., M. Allard, F. Revoy, and N. Amarger.** 1994. Rapid identification of rhizobia by restriction fragment length polymorphism analysis of PCR-amplified 16S rRNA genes. Appl. Environ. Microbiol. **60:**56–63.

14. **Lee, D.-H., Y.-G. Zo, and S.-J. Kim.** 1996. Nonradioactive method to study genetic profiles of natural bacterial communities by PCR-single-strand-conformation polymorphism. Appl. Environ. Microbiol. **62:**3112–3120.

15. **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.

16. **Maidak, B. L., G. J. Olsen, N. Larsen, R. Overbeek, M. J. McCaughey, and C. R. Woese.** 1996. The Ribosomal Database Project (RDP). Nucleic Acids Res. **24:**82–85.

17. **Martinez-Murcia, A. J., S. G. Acinas, and F. Rodriguez-Valera.** 1995. Evaluation of prokaryotic diversity by restrictase digestion of 16S rDNA directly amplified from hypersaline environments. FEMS Microbiol. Ecol. **17:**247–256.

18. **Massol-Deya, A. A., D. A. Odelson, R. F. Hickey, and J. M. Tiedje.** 1995. Bacterial community finger printing of amplified 16S and 16-23S ribosomal DNA gene sequences and restriction endonuclease analysis (ARDRA). Mol. Microb. Ecol. Man. **3.3.2:**1–8.

19. **Moyer, C. L., F. C. Dobbs, and D. M. Karl.** 1994. Estimation of diversity and community structure through restriction fragment length polymorphism distribution analysis of bacterial 16S rRNA genes from a microbial mat at an active, hydrothermal vent system, Loihi seamount, Hawaii. Appl. Environ. Microbiol. **60:**871–879.

20. **Muyzer, G., A. Teske, and C. O. Wirsen.** 1995. Phylogenetic relationships of *Thiomicrospira* species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. Arch. Microbiol. **164:**165–172.

21. **Muyzer, G., E. C. de Waal, and 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.

22. **Paster, B. J., F. E. Dewhirst, S. M. Cooke, V. Fussing, L. K. Poulsen, and J. A. Breznak.** 1996. Phylogeny of not-yet-cultured spirochete from termite guts. Appl. Environ. Microbiol. **62:**347–352.

23. **Picard, C., C. Ponsonnet, E. Paget, X. Nesme, and P. Simonet.** 1992. Detection and enumeration of bacteria in soil by direct DNA extraction and polymerase chain reaction. Appl. Environ. Microbiol. **58:**2717–2722.

24. **Rainey, F. A., N. Ward, L. I. Sly, and E. Stackebrandt.** 1994. Dependence of the taxon composition of clone libraries for PCR amplified, naturally occurring 16S rDNA, on the primer pair and the cloning system used. Experientia **50:**796–797.

25. **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.

26. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed., vol. I. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

27. **Schmidt, T. M., E. F. DeLong, and N. R. Pace.** 1991. Analysis of a marine picoplankton community by 16S rRNA gene cloning and sequencing. J. Bacteriol. **173:**4371–4378.

28. **States, D. J.** 1992. Molecular sequence accuracy: analyzing imperfect data. Trends Genet. **8:**52–55.

29. **Tebbe, C. C., and W. Vahjen.** 1993. Interference of humic acids and DNA extracted directly from soil in detection and transformation of recombinant DNA from bacteria and a yeast. Appl. Environ. Microbiol. **59:**2657–2665.

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

31. **Totsch, M., E. Brommelkamp, A. Stücker, M. Fille, R. Gross, P. Wiesner, K. W. Schmid, W. Bocker, and B. Dockhorn-Dworniczak.** 1995. Identification of mycobacteria to the species level by automated restriction enzyme fragment length polymorphism analysis. Virchows Arch. **427:**85–89.

32. **van de Graaf, A. A., A. Mulder, P. de Bruijn, M. S. Jetten, L. A. Robertson, and J. G. Kuenen.** 1995. Anaerobic oxidation of ammonium is a biologically mediated process. Appl. Environ. Microbiol. **61:**1246–1251.

33. **Volsch, A., W. F. Nader, H. K. Geiss, G. Nebe, and C. Birr.** 1990. Detection and analysis of two serotypes of ammonia-oxidizing bacteria in sewage plants by flow cytometry. Appl. Environ. Microbiol. **56:**2430–2435.

34. **Wagner, M., R. Erhart, W. Manz, R. Amann, H. Lemmer, D. Wedi, and K. H. Schleifer.** 1994. Development of an rRNA-targeted oligonucleotide probe specific for the genus *Acinetobacter* and its application for in situ monitoring in activated sludge. Appl. Environ. Microbiol. **60:**792–800.

35. **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.

36. **Zheng, D., E. W. Alm, D. A. Stahl, and L. Raskin.** 1996. Characterization of universal small-subunit rRNA hybridization probes for quantitative molecular microbial ecology studies. Appl. Environ. Microbiol. **62:**4504–4513.