APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Feb. 2006, p. 1270–1278 Vol. 72, No. 2
0099-2240/06/$08.00+0 doi:10.1128/AEM.72.2.1270–1278.2006

New Threshold and Confidence Estimates for Terminal Restriction Fragment Length Polymorphism Analysis of Complex Bacterial Communities†

Catherine A. Osborne,[1] Gavin N. Rees,[2] Yaniv Bernstein,[3] and Peter H. Janssen[1]*

Department of Microbiology and Immunology, University of Melbourne, Parkville, Victoria 3010, Australia[1]; Murray-Darling Freshwater Research Centre, P.O. Box 991, Wodonga, Victoria 3689, Australia[2]; and School of Computer Science and Information Technology, RMIT University, Melbourne, Victoria 3001, Australia[3]

Received 24 October 2005/Accepted 6 December 2005

Terminal restriction fragment length polymorphism (T-RFLP) analysis has the potential to be useful for comparisons of complex bacterial communities, especially to detect changes in community structure in response to different variables. To do this successfully, systematic variations have to be detected above method-associated noise, by standardizing data sets and assigning confidence estimates to relationships detected. We investigated the use of different standardizing methods in T-RFLP analysis of PCR-amplified 16S rRNA genes to elucidate the similarities between the bacterial communities in 17 soil and sediment samples. We developed a robust method for standardizing data sets that appeared to allow detection of similarities between complex bacterial communities. We term this the variable percentage threshold method. We found that making conclusions about the similarities of complex bacterial communities from T-RFLP profiles generated by a single restriction enzyme (RE) may lead to erroneous conclusions. Instead, the use of multiple REs, each individually, to generate multiple data sets allowed us to determine a confidence estimate for groupings of apparently similar communities and at the same time minimized the effects of RE selection. In conjunction with the variable percentage threshold method, this allowed us to make confident conclusions about the similarities of the complex bacterial communities in the 17 different samples.

The 16S rRNA gene is the target of the majority of microbial ecological surveys because of its usefulness as a prokaryotic phylogenetic marker (17). Rapid community profiling techniques that allow an insight into the range of 16S rRNA genes present are being applied to a wide range of microbial habitats (21). One of these community-profiling techniques, terminal restriction fragment length polymorphism (T-RFLP), separates sequence variants in a population of genes based on differences in restriction endonuclease (RE) cut sites in different alleles (19, 22, 25). Differences in the positions of RE cut sites mean that restriction fragments of different lengths can be generated from different alleles. By end labeling the amplified products during PCR by using a fluorescently tagged primer, each different allele is reduced to one end-labeled terminal restriction fragment (T-RF), visualized as a peak on the resulting electrophoretically generated profile. When used in conjunction with gene sequence information that allows prediction of T-RF sizes and therefore assignment of identity to individual peaks in a profile, the technique can be an effective tool for analyzing microbial communities (3, 6, 16, 20, 23, 24, 44).

The use of T-RFLP has, however, been seen by some to lack the degree of resolution required for analyzing complex microbial communities, such as those found in soil (9–11, 30), because of the difficulty in assigning accurate identity to each T-RF in complex profiles of 16S rRNA genes. Individual soil samples contain a large diversity of microorganisms, with recent estimates suggesting that a gram of soil may contain many thousands of different bacterial species (7). Each peak in a profile generated from DNA extracted from a soil sample must therefore represent multiple T-RFs of the same size originating from different 16S rRNA genes. This limitation was demonstrated in a study of a manure-treated soil reported by Sessitsch et al. (37), in which some T-RFs could have been generated by members of at least three different bacterial phyla. Data presented by Engebretson and Moyer (11) suggested that a set of about 4,600 16S rRNA gene sequences would generate T-RFLP profiles with a mean of 9.1 to 18.5 different sequences contributing to each T-RF, depending on which of 18 different REs was selected. The inference that each unique T-RF can be defined as an operational taxonomic unit was tested, and it was found that "by choosing the appropriate number and type of restriction endonucleases" the profiles generated would "more accurately reflect the natural diversity of microbial populations within a sampled community" (11). In other investigations (33; C. A. Osborne and P. H. Janssen, unpublished data), the use of different REs to generate multiple T-RFLP profiles for each sample was found to yield more information that could then be used to determine if a sequence type was present or absent from complex communities.

While assignment of identities may be uncertain, it does not preclude the use of the technique to compare whole communities. Profiles generated from different soil samples could be compared to assess the similarity of soil bacterial commu-

* Corresponding author. Mailing address: Department of Microbiology and Immunology, University of Melbourne, Victoria 3010, Australia. Phone: 61 (3) 8344 5706. Fax: 61 (3) 9347 1540. E-mail: pjanssen@unimelb.edu.au.

† Supplemental material for this article may be found at http://aem.asm.org/.

nities, allowing spatial or temporal changes to be detected without the need to know the identity of every peak in every profile. This approach has been used to compare the effects of physical and biotic factors on soil microbial communities (4, 9, 18, 24, 42).

T-RFLP has been shown to be reproducible between PCR replicates and gel runs (23, 30, 35). However, variations in the amount of DNA loaded affect the profile, and this has been dealt with by applying normalization procedures (2, 10, 30, 35). There is, as yet, no agreed-upon method for normalizing samples with differing amounts of DNA, which would allow easy comparison of profiles with different total amounts of fluorescent label (2). There is a need to distinguish between major T-RFs that represent abundant members of the community and minor T-RFs that may be detected when large amounts of DNA are analyzed but may be below the detection limit when small amounts of DNA are analyzed. An appropriate method for calculating a threshold, baseline, or minimum fluorescence cutoff needs to be determined. The application of such a threshold is of utmost importance when the Sorensen-Dice pairwise similarity coefficient (8, 40), or any similar measure based on the presence or absence of peaks in each profile, is the basis of comparison. In such studies, the presence of small peaks that result as a consequence of noise or the amount of DNA analyzed may have an impact on the conclusions drawn.

In this study, we compare two published threshold determination protocols (10, 35), with a novel approach to threshold determination, and explore a means of assigning a confidence value to similarities detected between different bacterial communities, by creating more information for T-RFLP comparison from six separate REs. This study provides another perspective on the unresolved area of analyzing T-RFLP profiles (15) and describes a simple but rigorous approach that allows complex communities to be compared.

MATERIALS AND METHODS

Samples and DNA extraction. Fifteen soil and two sediment samples, collected between June 1997 and November 2003, were frozen at −20°C within 24 h of sample collection, after which total community DNA was extracted using the protocol of O'Farrell and Janssen (27). The samples used were: A, marine sediment from Williamstown Hatt Reserve seagrass bed; B, marine sediment from below the Williamstown Anglers Club pier; C, rhizosphere soil from a eucalypt in Wombat State Forest; D, rhizosphere soil from a pine tree on Moonee Ponds Creek; E, soil from a vineyard at Bunyip; F, rhizosphere soil from a camellia bush in Ascot Vale; G, soil from a heavily manured vegetable garden in Bunyip; H, soil from a pasture (0 to 10 cm below ground) at Ellinbank; HL, soil from a pasture (3 to 10 cm below ground) at Ellinbank; HR, roots and associated soil from pasture grasses at Ellinbank; I, soil from a petrol- and oil-contaminated grassed area in Ascot Vale; J, soil from a native grassland in Sunbury; KB, soil from a forage brassica plot at Ginninderra; KW, soil from a wheat plot at Ginninderra; L0, rhizosphere soil from a eucalypt on Moonee Ponds Creek; L2, soil under grass (2 m away from sample L0) on Moonee Ponds Creek; and L4, soil under grass (4 m away from sample L0) on Moonee Ponds Creek. All but the two Ginninderra sites, which were located in the Australian Capital Territory, were in Victoria, Australia.

Generation of T-RFLP profiles and data sets. T-RFLP profiles were generated using the primers FAM27f (5′-GAGTTTGATCMTGGCTCAG-3′), labeled at the 5′ terminus with 6-carboxyfluorescein, and 519r (5′-GWATTACCGCGG CKGCTG-3′) and otherwise followed the protocol of Sait et al. (35). Unincorporated primers and reaction components were removed using a Wizard SV Gel and PCR Clean-Up System column (Promega, Annandale, New South Wales, Australia) according to the manufacturer's instructions. Approximately 100 ng of purified PCR product was digested overnight, at the specified temperature, with 5 U of one of the following restriction endonucleases: BstUI, HaeIII, HhaI, HinfI, MspI, or Sau96I (New England Biolabs, Inc., Beverly, Mass.). Digested PCR products were then precipitated and analyzed on a model 377 DNA sequencer (Applied Biosystems, Foster City, Calif.) at the Australian Genome Research Facility, Parkville, Victoria, Australia (35) to generate profiles of fragments up to 530 nucleotides (nt) long.

A separate profile was generated for each sample with each of the six restriction endonucleases. Each profile consisted of T-RFs that each had a reported fragment length, in nucleotides, and a reported peak area of fluorescently labeled product, in fluorescence units (FU). The raw data set of fragment lengths and corresponding peak areas obtained from the Genotyper software (Applied Biosystems) was first compiled so that all T-RFs in each profile of the data set were aligned with the T-RFs of the same inferred length (rounded to the nearest nucleotide) in every other T-RFLP profile generated by that restriction endonuclease.

All of the profiles generated with any one RE constituted a data set. The fluorescence integrated under any one peak is referred to as the area of that peak, and the total area for any one profile is the sum of the areas of all of the peaks excluding those generated by fragments of less than 30 nt or greater than 500 nt. To standardize data sets, small peaks that may have been detected in a profile as a result of loading large amounts of DNA were removed by application of a threshold area. Peaks with an area smaller than this threshold area were removed from the data set, using three different standardization methods: the constant percentage threshold, the constant baseline threshold, and the variable percentage threshold.

Constant percentage threshold calculations. Sait et al. (35) determined the threshold area for a data set by applying increasing area thresholds, as percentages of the total area on each trace, until the minimum percentage that resulted in independence of the number of peaks remaining in each profile and the total area of each profile before threshold application was found. This standardization method is referred to as the constant percentage threshold calculation. All peaks that contributed less than this constant percentage threshold to any profile within a data set were removed before analyzing relationships between the profiles in the data set. A constant percentage threshold value was determined for each data set (referred to as constant percentage [different]) and also by combining all the profiles in all six data sets into a global data set to calculate a global constant percentage (referred to as constant percentage [global]).

Constant baseline threshold calculations. The second method for standardization was the constant baseline threshold calculation of Dunbar et al. (10), applied to peak areas. The total areas in all of the profiles in a data set were normalized to the same value as that of the profile having the smallest total area, and then all of the peaks in each trace were reduced proportionally by the factor required to yield that normalized total area. The first constant baseline threshold was set as the smallest peak area detected in the unmanipulated data sets (rounded up to 50 fluorescence units), and peaks with an area equal to or smaller than this were removed from profiles after normalization (referred to as constant baseline [50 FU]), before analyzing relationships between traces in the data set. A second analysis was performed on data sets where the constant baseline threshold was set at 100 fluorescence units (referred to as constant baseline [100 FU]).

Variable percentage threshold calculations. The third method for standardization was termed the variable percentage threshold method. The total area of each profile was divided by different values (divisors) to yield numbers that were used as percentage thresholds. For each divisor, all peaks that contributed less than the percentage threshold calculated for that profile were removed. Then, for each divisor, the remaining number of peaks was plotted against the total area, so that each profile contributed one point on that plot. We have found that, sometimes, if there is very little variation in the total area of the profiles, there is no detectable relationship between the number of peaks and the total area, and no threshold needs to be applied (M. Sait and P. H. Janssen, unpublished data). If standardization is required, a useful divisor to start with is 1,000 times the mean total area for the data set. Different divisors were then tested, and the divisor that resulted in the weakest relationship between the number of peaks remaining and the initial total area was considered to be the optimal divisor (Fig. 1). The unique percentage threshold value for each profile was calculated by dividing the total area of that profile by the optimal divisor. Peaks that contributed less than that percentage threshold were removed from that profile before analyzing relationships between traces in the data set.

Comparison of profiles. Matrices of Sorensen-Dice pairwise similarity coefficients (8, 40) were calculated for all possible comparisons of profiles within a data set using a program written in the C++ programming language. These similarities were converted to distances, where distance = 1 − similarity, to produce distance matrices. These matrices were represented graphically as dendrograms using the Fitch-Margoliash least-squares (14) and neighbor-joining (36) algorithms in PHYLIP (13). Consensus dendrograms were generated from

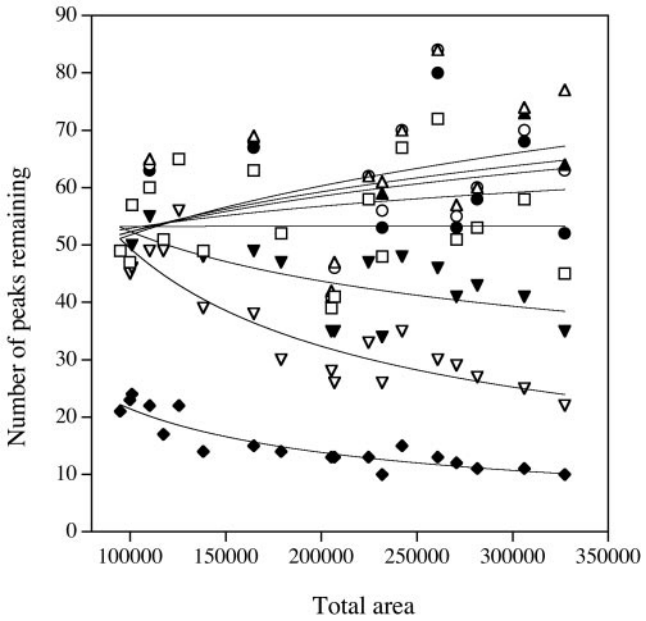

FIG. 1. Estimation of an appropriate divisor for the calculation of the variable percentage threshold for 20 profiles generated using HinfI. The optimum divisor was 2×10^8, which resulted in the weakest relationship between the total area on the original profiles and the number of peaks remaining after application of the threshold using that divisor. The curves were fitted as power functions. The divisors shown are as follows: △, no divisor, i.e., unmanipulated data; ◆, 1×10^7; ▽, 5×10^7; ▼, 1×10^8; □, 2×10^8; ●, 3×10^8; ○, 4×10^8; and ▲, 5×10^8.

the six data sets using the CONSENSE program in PHYLIP, and the consensus values reported are the numbers of REs that generated coherent clusters of samples. Shannon-Wiener diversity indices (41) were calculated using the peak areas, expressed as a proportion of the total area, as measures of abundance.

Multivariate statistical analysis. Matrices of Sorensen-Dice similarity coefficients were generated for each T-RFLP data set from each RE and for each threshold calculation procedure. A first-stage nonmetric dimensional scaling (MDS) plot was then derived from each of the similarity matrices. The second-stage MDS (2sMDS) analysis was then used to examine similarities between each of the first-stage MDS plots. For the second-stage analysis, a Spearman rank correlation was computed between the corresponding elements of each pair of first-stage similarity matrices. The resulting Spearman rank correlations then became the elements of a second similarity matrix, which was used to generate the 2sMDS plot. Each point on the 2sMDS represents a first-stage MDS plot and therefore essentially provides a visual interpretation of the similarity between all of the original MDS plots.

We also calculated a relative dispersion index for each data manipulation procedure (5, 39). The dispersion index compares the average rank similarity across a given group of samples. In our case, we defined each of the data manipulations or REs as the sample groupings. For interpretation, the lower the dispersion index, the tighter is the grouping of samples within the given treatment.

All multivariate statistical routines were carried out with the PRIMER 5 software package (Primer-E Ltd., Plymouth, United Kingdom).

RESULTS AND DISCUSSION

Experimental design. In this study, we used only one method of determining the size of peaks (by area according to Genotyper software outputs), one method of comparing T-RFLP profiles (Sorensen-Dice similarity coefficients), and two methods of representing the manipulated data sets (dendrograms and 2sMDS). We tested how different ways of calculating the threshold area and how the choice and number of REs used can affect the conclusions drawn about the similarity of samples. Peaks with areas smaller than the threshold area are removed from the data set to produce a manipulated data set. This is required to standardize profiles so that differences due to sample loading, resulting in differences in sensitivity for different samples, do not affect comparisons based on the presence or absence of peaks.

DNA was extracted from the 17 samples, and PCR products were generated from all DNA samples. The pooled products from two PCRs from each DNA template were separated into six aliquots and digested using six different REs. To create T-RFLP profiles that were similar to some of these, additional PCRs were performed using three of the DNA samples (D, F, and G), and these were also aliquoted and digested with the same six REs. In this way, a total of 120 T-RFLP profiles were generated. We chose REs that spanned the range of the number of T-RFs of 50 to 500 nt predicted for a set of about 4,600 16S rRNA gene sequences and therefore should produce profiles that vary in the number of sequences theoretically contributing to any one peak (11).

To ensure that T-RFLP profiles more accurately represented the bacterial community of a sample and to reduce method-associated noise, it has been suggested that the products from multiple PCRs from the same template be pooled to minimize the effects of amplification bias in individual PCRs (6) and that consensus data sets be generated from multiple electrophoretic separations of the digestion products from one sample and RE to minimize variations due to profile generation (10). However, for this study we relied on variations in PCRs and electropherograms to generate less-than-ideal data sets. We reasoned that such data sets would allow us to test the effects of thresholds and choices of REs, since we knew that the pairs of replicates (D and D′, F and F′, and G and G′) were derived from the same communities. Analyses that resulted in these replicates being deemed to be similar can therefore be considered more useful than analyses that failed to recognize replicates as being similar. This avoided a priori assumptions of sample similarity to test analysis methods, which may not be valid (15).

Similarities between raw T-RFLP profiles. The twenty profiles (17 samples and 3 extra replicates) generated with each RE were compiled into one data set per RE. The similarity of each pair of profiles within each data set was expressed as a Sorensen-Dice pairwise similarity coefficient. These values were then used to generate distances, and dendrograms were constructed from these distance values to depict the similarities between different profiles. Profiles that were similar to each other grouped closely in these dendrograms. The samples and replicates grouped differently in the analyses of different raw data sets, so that few groupings were consistently found in analyses of all six data sets, i.e., the use of different REs generally resulted in different pairs of profiles being deemed to be similar. For example, comparison of the data sets generated using BstUI and MspI reveals that few groupings are shared between the two (Fig. 2A and B). The two sediment samples (A and B), the two Ginninderra agricultural soil samples (KB and KW), and the two parkland soil samples separated by 2 m (L2 and L4) did group together in the analyses of both data sets. In fact, these three pairings were the only ones found consistently in the analyses with all six REs (Fig. 3A).

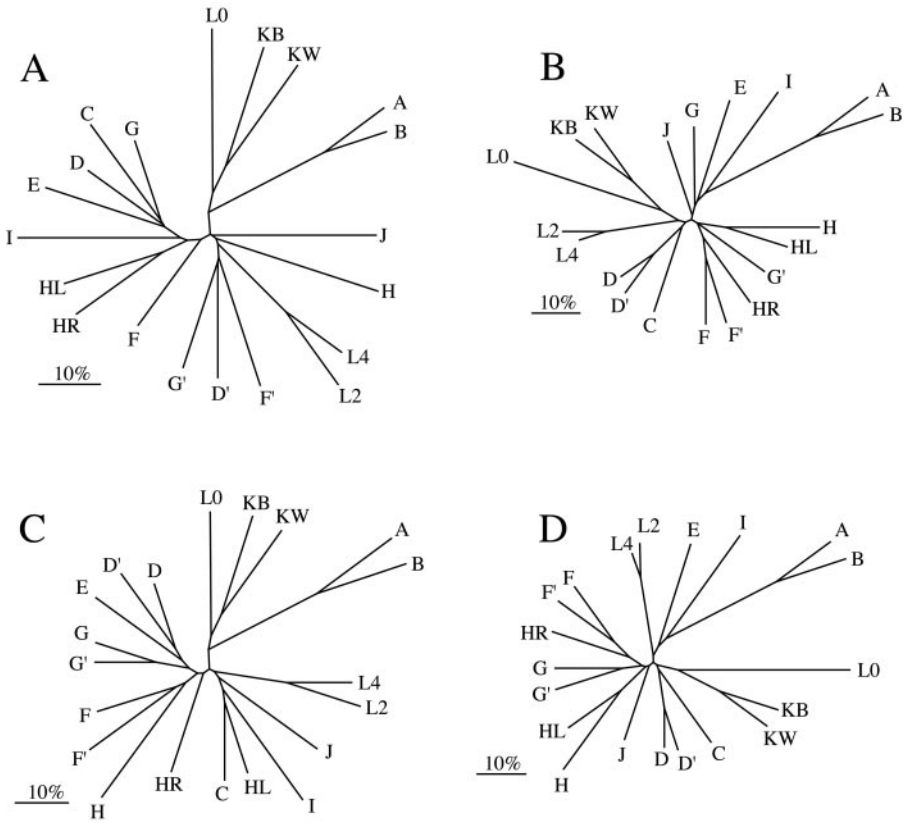

FIG. 2. Examples of relationships between soil bacterial communities elucidated using T-RFLP analysis with different REs and data manipulations. The panels display the differences calculated from Sorenson-Dice similarity coefficients. (A) Unmanipulated (raw) data generated using BstUI; (B) unmanipulated (raw) data generated using MspI; (C) data generated using BstUI after application of the variable percentage threshold; and (D) data generated using MspI after application of the variable percentage threshold. The samples are indicated by letter codes at the branch termini, and the replicate samples are indicated as D′, F′, and G′. Bars represent a 10% difference.

Two of the replicate PCR pairs, D and D′ and F and F′, were paired only by four of the six REs, while the third replicate pair, G and G′, paired in only one data set (Fig. 3A). The only RE to group all three replicates in analyses of the raw data sets was HaeIII. D and D′ were not grouped by BstUI and HhaI, but F and F′ were not grouped by BstUI and Sau96I. The different groupings of replicates meant that it was difficult to be sure whether some of the groupings formed by nonreplicate samples were meaningful, and, even with HaeIII, it could not be ruled out that some of the pairings of nonreplicates detected might be fortuitous rather than reflective of real similarities between the original bacterial communities.

Effects of different threshold calculations. We attempted to standardize the data sets and so produce greater similarity between the replicates. Three different methods were used to calculate a threshold area value to remove minor peaks that may have been detected purely as a result of the amount of DNA applied to the separation gels. A constant percentage (different) threshold (35) was applied to each data set from each individual RE. In addition, this method was also used to calculate a constant percentage (global) threshold area for all of the data sets combined into one data set, so that one threshold area was applied to all six data sets. Sait et al. (35) expressed the area under each peak as a percentage of the total area for all of the peaks in that profile and determined the minimum threshold as a percentage for the data set, so the profiles have comparable numbers of peaks while still retaining enough peaks to analyze the microbial community. Constant baseline thresholds, based on the method of Dunbar et al. (10), were applied to each data set from each RE, using the threshold areas of 50 and 100 FU. Dunbar et al. (10) actually compared peak heights, rather than area, but their method of proportionally reducing the heights of each peak in larger profiles and removing peaks that then fell below the threshold can also be used on peak areas. The third method combined elements of both of these published methods, and we term this the variable percentage threshold method. A unique threshold area, as a percentage of the total area, was calculated for each trace in a way that is dependent on the total peak area for each individual profile. A greater total peak area in any profile will increase the threshold area value of the baseline for that profile. This compensates for the increased sensitivity due to more labeled product being present in that profile. The constant percentage and variable percentage methods require enough samples to allow a reasonable estimate of the relationship between total peak area and the number of peaks in each profile to be determined (Fig. 1).

The variable percentage threshold method appeared to be the best for determining threshold values with these samples because it grouped all three pairs of PCR replicates (D and D′,

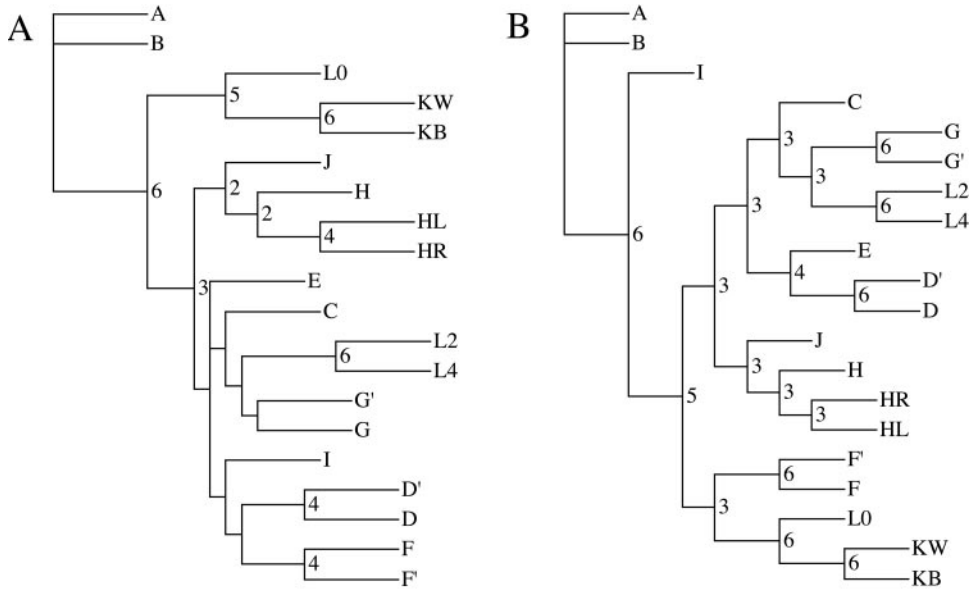

FIG. 3. Consensus dendrograms illustrating the number of times groupings of samples were recovered in analyses using six different REs, indicated as a number at the nodes defining the clusters. Nodes without a number were recovered only once. (A) Analysis of the unmanipulated (raw) data sets. (B) Analysis after the application of variable percentage thresholds to the data sets. The samples are indicated by letter codes at the branch termini, and the replicate samples are indicated as D′, F′, and G′.

F and F′, and G and G′) in all six data sets generated with different REs when the Sorensen-Dice similarity coefficient was represented by the Fitch-Margoliash least-squares treeing method (Fig. 3B and Table 1). The other methods of threshold determination resulted in the replicates grouping together less often. When the groupings were analyzed using the neighbor-joining method of depicting the outcomes as trees, the constant percentage (global) method resulted in the most consistent grouping of replicates, but some of the other methods also appeared satisfactory (Table 1). The variable percentage threshold method also resulted in the largest number of REs generating similar groupings between profiles generated from different samples of all the methods used (Table 1). Groupings uncovered in analyses generated using a number of different REs are more likely to consist of truly similar samples than groupings found using only one RE, indicating that this threshold method may be better for standardizing data sets and subsequently detecting true similarities between complex communities.

MDS is a robust approach to examining community data that has been used extensively to analyze data sets based on macrofaunal community structures (5) and, more recently, to analyze microbial community structures (32, 43). Here we used an extended MDS approach, termed 2sMDS, as an additional method to examine T-RFLP data for each RE and for each threshold calculation procedure. To date, 2sMDS analysis has been applied only to the study of marine benthic invertebrates (28, 29, 38) but has proven to be very useful in allowing metascale pattern analysis. Applying this technique to our data allowed us to place our results within a robust statistical framework. Each data point on second-stage ordination essentially represents an MDS ordination plot for one of the REs and one of the standardization procedures, and the position of each point is determined by the similarity of that MDS plot to all

TABLE 1. Effects of different threshold methods on number of T-RFs remaining for analysis, consensus of groupings in cluster analyses, and consistency of groupings in 2sMDS across the six data sets generated (one for each RE)

Threshold method	Mean no. of T-RFs/profile (SD)[c]	Mean consensus grouping of three replicates (out of six REs)		Mean consensus grouping of 17 samples (out of six REs)		Relative dispersion index in 2sMDS[b]
		Fitch-Margoliash[a]	Neighbor-joining[a]	Fitch-Margoliash	Neighbor-joining	
Original data	63 (13)	3.0	4.3	3.4	3.7	1.33
Constant percentage (different)	40 (17)	4.7	5.3	3.2	3.8	1.38
Constant percentage (global)	54 (10)	5.0	5.0	3.5	3.8	0.76
Constant baseline (50 FU)	61 (11)	3.3	4.3	3.1	3.4	1.10
Constant baseline (100 FU)	54 (10)	4.7	5.0	3.6	3.7	0.77
Variable percentage	55 (11)	6.0	5.0	4.1	3.8	0.66

[a] Fitch-Margoliash (14) and neighbor-joining (36) refer to the algorithms used to generate dendrograms from which the consensus values were calculated.
[b] From Fig. 4A; lower dispersion indices indicate greater agreement in the relationships uncovered by all six REs.
[c] $n = 120$.

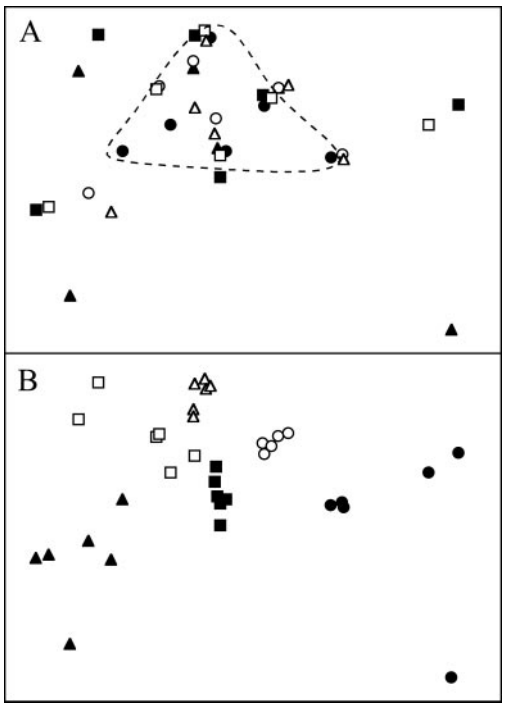

FIG. 4. 2sMDS plots showing the relationships between data sets generated with different REs after manipulation using different threshold methods. (A) Effects of different threshold methods. Symbols: ■, no threshold (raw, unmanipulated data); ▲, constant percentage (different) threshold; ○, constant percentage (global) threshold; □, constant baseline (50 FU) threshold; △, constant baseline (100 FU) threshold; and ●, variable percentage threshold. The data points indicating the positions of the variable percentage results are bounded by a dashed line. (B) Effects of different REs. Symbols: ○, HaeIII; ■, MspI; △, HinfI; □, Sau96I; ▲, HhaI; and ●, BstUI.

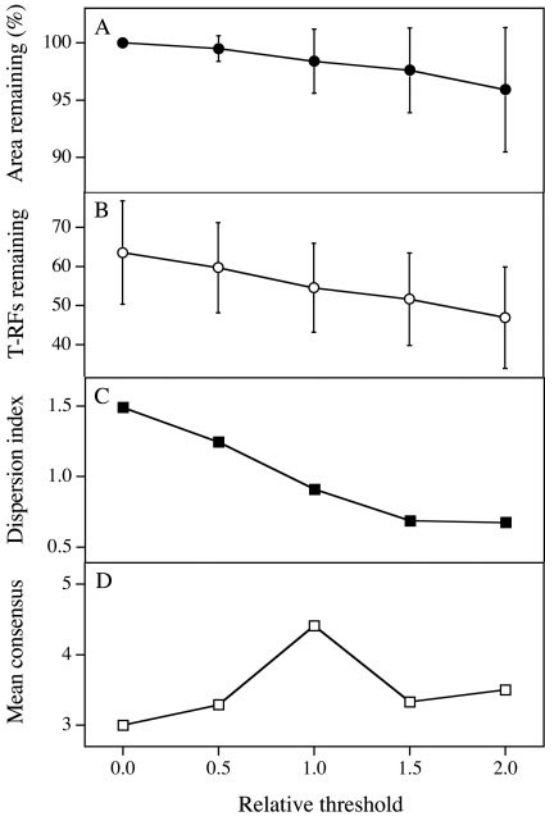

FIG. 5. Effects of suboptimal (<1 and >1) thresholds relative to the threshold calculated using the optimal divisor with the variable percentage threshold method on (A) the mean area remaining, (B) the number of T-RFs remaining after threshold application, (C) the dispersion index in 2sMDS, and (D) the mean consensus between the six REs on the grouping of the 20 samples in clustering dendrograms, after threshold application. The threshold calculated using the variable percentage threshold is shown as 1.0 for all six data sets to allow direct comparisons to be made; 0.0 represents the raw data sets without threshold application. The vertical bars indicate one standard deviation (SD) on either side of each mean. Each point represents 120 profiles.

other plots. The most consistent grouping is indicated in such plots by the smallest spread of points within a given treatment.

The variable percentage threshold method gave rise to the most consistent grouping of the samples (Fig. 4A) and therefore represents the preferred approach to analyzing data. That is, the relationship of the samples to each other as determined by the six REs was most similar when the variable percentage threshold was applied to standardize the data. The relative dispersion index, a measure of the closeness of grouping, was smallest for the variable percentage threshold method (Table 1). At the other extreme, the constant percentage (different) threshold method (where each data set had its own unique threshold value determined) gave rise to the least consistent grouping of the samples (Fig. 4A) and had the largest relative dispersion index (Table 1). These results are generally consistent with the recovery of groupings of pairs of samples in consensus dendrograms (Table 1). The constant percentage (different) method resulted in a large loss of T-RFs (Table 1) (see the supplemental material), and this loss of information, and consequent simplification of the profiles, seems to have resulted in less-accurate detection of similarities between communities because the analyses are made on a small number of characters (T-RFs). Since this method was nearly as poor as the use of the raw data with no threshold, this suggests that overmanipulation of the data must be avoided. The three methods that resulted in the smallest dispersion all resulted in intermediate numbers of peaks (mean of 54 or 55 peaks) remaining after application of thresholds.

To test the effect of suboptimal threshold values on the data, and so the effects of under- and overmanipulating the data sets, we applied different thresholds to the data sets and compared the effects on the relative dispersion indices in 2sMDS. We applied threshold values at 0.5, 1.0, 1.5, and 2.0 times the optimal value that had been calculated using the variable percentage threshold method. As expected, the number of peaks decreased more rapidly than remaining total area as the threshold value increased (Fig. 5A and B), since application of the threshold eliminated the smallest peaks. The dispersion index decreased as the threshold value increased (Fig. 5C), but the consensus of agreement between the six different REs was greatest at the optimal threshold. In this instance, the dispersion indices are showing one outcome of steadily removing peaks from a given data set. As rare or small peaks are steadily removed from the data sets, the resultant data sets are becom-

TABLE 2. Effect of RE choice on consistency of groupings in 2sMDS compared with information content measured as mean number of T-RFs per profile, Shannon-Wiener diversity indices, and mean difference between profiles

Restriction endonuclease	Relative dispersion index in 2sMDS[a]	Mean no. of T-RFs/profile (SD)[b]	Shannon-Wiener diversity index[b]	Mean difference between profiles (SD)[c]
HaeIII	0.43	64 (10)	3.05	0.501 (0.105)
MspI	0.75	66 (12)	3.45	0.512 (0.122)
HinfI	0.79	60 (11)	3.26	0.582 (0.104)
Sau96I	1.31	68 (16)	3.50	0.545 (0.105)
HhaI	1.35	61 (9)	3.12	0.529 (0.090)
BstUI	1.39	61 (17)	3.12	0.535 (0.092)

[a] From Fig. 4B; lower dispersion indices indicate a higher degree of similarity in the data sets after the application of different threshold methods.
[b] On unmanipulated data sets, $n = 20$ for each RE.
[c] Calculated from Sorensen-Dice similarity coefficients on the unmanipulated data sets; $n = 190$ comparisons for each RE.

ing more similar; hence, there is less scatter with the ordinations, seen here as smaller dispersion indices. The optimum threshold identified by the variable percentage threshold method therefore results in the highest level of agreement between the different REs. Since we can expect that each RE will tend to produce similar patterns when the starting community is similar but also that experimentally induced noise will tend to obscure the true pattern of relationships between samples, we can assume that the highest consensus values occur when the optimum threshold is applied and a balance between noise elimination and information retention is achieved.

These outcomes suggest that the use of any method to calculate a threshold area allows the subsequent analysis to detect real groupings by eliminating noise. Overall, the variable percentage method appeared to be the most useful of the methods tested, probably because of its ability to set an optimal threshold area value.

Choice of restriction endonuclease. The profiles generated by the different REs contained similar numbers of T-RFs, and measures of diversity also suggested that the amounts of information in the profiles were not greatly different when they were generated using different REs (Table 2) (see the supplemental material). However, the 2sMDS ordinations plotted by RE showed that individual enzymes can generate different community patterns (Fig. 4B), as evident from the scatter of the enzyme clustering throughout the 2sMDS (Table 2). It is clear that data sets generated using some enzymes, for example BstUI, are more strongly affected by the choice of threshold method than others, such as those generated using HaeIII (Fig. 4B). It is important to note that REs that resulted in more consistent groupings, like HaeIII and HinfI, did not generate profiles that looked more similar to each other (Table 2); instead, the relationships between samples were more consistent when these REs were used, regardless of which standardization method was used (Fig. 4B). This suggests that data sets generated using some enzymes are less susceptible to noise. In contrast, the choice of standardization method has a large impact on the outcomes of analyses of data generated using other REs. REs that generated profiles in which many small peaks were common to most profiles and large peaks were mainly unique should result in more differentiation of profiles when increasing thresholds were applied, while REs that generated many large peaks that were common to most profiles should be less sensitive to increasing thresholds. However, there were no obvious differences in the distribution of peak sizes in profiles generated using different REs (see the supplemental material). Instead, the distribution of the number of larger peaks relative to smaller peaks was almost identical for profiles generated with each of the six REs. When the six REs are ranked by relative dispersion index across all threshold methods (Table 2), there is no correlation with the characteristics listed for the same REs by Engebretson and Moyer (11). Other studies report sometimes-contradictory conclusions regarding the best REs in experimental comparisons (3, 6, 10, 20, 22, 34). We suggest that the analysis of data generated using multiple REs reduces the likelihood of a lack of resolving power due to RE selection.

Generation of confidence values using multiple REs. Computer simulations of T-RFLP analyses have shown that the choice and number of REs used affects the profile's representation of the community, and the use of multiple REs was suggested to allow a more accurate assignment of peaks to 16S rRNA gene sequences (11). We have extended this by generating multiple data sets and constructing a consensus tree depicting relationships from the results of the individual data sets for six REs (Fig. 3). This allows an estimate of the confidence of different groupings to be made. Many studies employing T-RFLP use only one RE. Conclusions about the similarity of different samples could be attempted based on data sets generated from just one RE. For example, profiles of samples C and HL generated using BstUI appeared similar (Fig. 2C), allowing a conclusion to be made that their bacterial communities were similar. However, the consensus of the six separate data sets indicates that this is likely to be an incorrect conclusion (Fig. 3B), since these samples were found to be similar in only two of the six data sets (generated with BstUI and HinfI). This agrees with results found in a previous study (9), where the combined data from REs were not consistent, and indicates that conclusions about complex communities based on T-RFLPs generated from just one RE may be erroneous.

Each peak in a T-RFLP profile generated from a complex bacterial community such as those found in soils (11, 33) is likely to represent more than one species. The use of different REs will group these species together in different combinations (into different peaks), so that each RE produces a different simplified representation of the community. DNA from different soils may generate the same-sized T-RF that originates from different sets of 16S rRNA genes. These sets can be expected to be distributed in different peaks if a different RE is used. In our analyses, samples that were grouped together by all six REs are therefore more likely to be truly similar, while the confidence in groupings found in data sets generated by fewer REs is lower. The use of multiple REs generates a consensus value, which estimates a confidence level of the grouping. This is different from using a bootstrap analysis to generate confidence values for groups. During a bootstrap analysis, a new data set is generated by sampling characters randomly, with replacement, so that the resulting bootstrapped data set is of the same size as the original, but with some

characters removed and others duplicated (12). The random variation of the results from analyzing these bootstrapped data sets is considered to be typical of the variation that might arise when collecting new data sets. As such, it tests the size and robustness of the data set. Our confidence values express the actual consensus of multiple real data sets. This represents a useful means of assessing the significance of any groupings observed.

Although some studies applying T-RFLP to complex microbial communities have used a number of different REs to generate more information for sample comparisons (2–4, 6, 9, 10, 26, 34), this is the first study to generate consensus values from the data generated using multiple REs. Ayala-del-Río et al. (1) initially tested a subset of samples with six different REs, but carried out community analysis on profiles from two REs, which were selected as the best. Within this study we have been unable to determine criteria for selecting the best REs for comparing complex microbial communities, and we believe the information from all six REs is valuable in our community analyses. Using six REs will also allow the detection of some problems with T-RFLP, such as the generation of artifacts that may occur during PCR (31).

Conclusions. The factors governing the effect of RE choice are poorly understood, and the outcomes of computer-generated simulations (11) did not correlate well with the outcomes of our studies on real samples. We conclude that the use of multiple REs, employed individually, overcomes possible effects of RE choice on generating useful T-RFLP information for complex bacterial communities. Our results suggest that our variable percentage threshold method is a useful addition to the limited range of standardization methods available, because it allows determination of an optimal threshold for each profile, and so minimizes information loss. We also suggest that generating multiple profiles using different REs for each sample, and so giving confidence values for sample groupings, may be a means of determining the significance of relationships detected. We used six REs, but this could be varied. In conjunction with pooling PCRs (6) and generating consensus profiles from multiple separations (10), more confident interpretation of T-RFLP analyses can be made. The consensus approach could also be used with other gene profiling techniques (21), where multiple profiles could be generated using different primer sets, and consensus values generated as indicators of confidence.

ACKNOWLEDGMENTS

This work was supported by grants from the Australian Research Council and from the Grains Development Research Corporation.

We thank Kelly Ewen-White, Karl Eisler, and Melinda Ziino at the Melbourne Division of the Australian Genome Research Facility for carrying out electrophoretic separations and Maja Galic, Shayne Joseph, Leanne Sait, and Michelle Sait for valuable advice.

REFERENCES

1. **Ayala-del-Río, H., S. J. Callister, C. S. Criddle, and J. M. Tiedje.** 2004. Correspondence between community structure and function during succession in phenol- and phenol-plus-trichloroethene-fed sequencing batch reactors. Appl. Environ. Microbiol. **70:**4950–4960.
2. **Blackwood, C. B., T. Marsh, S.-H. Kim, and E. A. Paul.** 2003. Terminal restriction fragment length polymorphism data analysis for quantitative comparison of microbial communities. Appl. Environ. Microbiol. **69:**926–932.
3. **Braker, G., H. L. Ayala-del-Río, A. H. Devol, A. Fesefeldt, and J. M. Tiedje.** 2001. Community structure of denitrifiers, bacteria, and *Archaea* along redox gradients in Pacific Northwest marine sediments by terminal restriction fragment length polymorphism analysis of amplified nitrite reductase (*nirS*) and 16S rRNA genes. Appl. Environ. Microbiol. **67:**1893–1901.
4. **Buckley, D. H., and T. M. Schmidt.** 2001. The structure of microbial communities in soil and the lasting impact of cultivation. Microb. Ecol. **42:**11–21.
5. **Clarke, K. R., and R. M. Warwick.** 2001. Change in marine communities: an approach to statistical analysis and interpretation, 2nd ed. Primer-E Ltd., Plymouth, United Kingdom.
6. **Clement, B. G., L. E. Kehl, K. L. DeBord, and C. L. Kitts.** 1998. Terminal restriction fragment patterns (TRFPs), a rapid, PCR-based method for the comparison of complex bacterial communities. J. Microbiol. Methods **31:** 135–1142.
7. **Curtis, T. P., and W. T. Sloan.** 2004. Prokaryotic diversity and its limits: microbial community structure in nature and implications for microbial ecology. Curr. Opin. Microbiol. **7:**221–226.
8. **Dice, L. R.** 1945. Measures of the amount of ecologic association between species. Ecology **26:**297–302.
9. **Dunbar, J., L. O. Ticknor, and C. R. Kuske.** 2000. Assessment of microbial diversity in four southwestern United States soils by 16S rRNA gene terminal restriction fragment analysis. Appl. Environ. Microbiol. **66:**2943–2950.
10. **Dunbar, J., L. O. Ticknor, and C. R. Kuske.** 2001. Phylogenetic specificity and reproducibility and new method for analysis of terminal restriction fragment profiles of 16S rRNA genes from bacterial communities. Appl. Environ. Microbiol. **67:**190–197.
11. **Engebretson, J. J., and C. L. Moyer.** 2003. Fidelity of select restriction endonucleases in determining microbial diversity by terminal-restriction fragment length polymorphism. Appl. Environ. Microbiol. **69:**4823–4829.
12. **Felsenstein, J.** 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution **39:**783–791.
13. **Felsenstein, J.** 1989. PHYLIP—Phylogeny Inference Package (version 3.2). Cladistics **5:**164–166.
14. **Fitch, W. M., and E. Margoliash.** 1967. Construction of phylogenetic trees. Science **155:**279–284.
15. **Grant, A., L. A. Ogilvie, C. B. Blackwood, T. Marsh, S.-H. Kim, and E. A. Paul.** 2003. Terminal restriction length polymorphism data analysis. Appl. Environ. Microbiol. **69:**6342–6343.
16. **Hackl, E., S. Zechmeister-Boltenstern, L. Bodrossy, and A. Sessitsch.** 2004. Comparison of diversities and compositions of bacterial populations inhabiting natural forest soils. Appl. Environ. Microbiol. **70:**5057–5065.
17. **Hugenholtz, P., B. M. Goebel, and N. R. Pace.** 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J. Bacteriol. **180:**4765–4774.
18. **Kennedy, N., E. Brodie, J. Connolly, and N. Clipson.** 2004. Impact of lime, nitrogen and plant species on bacterial community structure in grassland microcosms. Environ. Microbiol. **6:**1070–1080.
19. **Kitts, C. L.** 2001. Terminal restriction fragment patterns: a tool for comparing microbial communities and assessing community dynamics. Curr. Issues Intest. Microbiol. **2:**17–25.
20. **LaMontagne, M. G., J. P. Schimel, and P. A. Holden.** 2003. Comparison of subsurface and surface soil bacterial communities in California grassland as assessed by terminal restriction fragment length polymorphisms of PCR-amplified 16S rRNA genes. Microb. Ecol. **26:**216–227.
21. **Liesack, W., and P. F. Dunfield.** 2002. Biodiversity in soils: use of molecular methods for its characterization, p. 528–544. *In* G. Bitton (ed.), Encyclopedia of environmental microbiology. John Wiley & Sons, Inc., New York, N.Y.
22. **Liu, W.-T., T. L. Marsh, H. Cheng, and L. J. Forney.** 1998. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. Appl. Environ. Microbiol. **63:**4516–4522.
23. **Lueders, T., and M. W. Friedrich.** 2003. Evaluation of PCR amplification bias by terminal restriction fragment length polymorphism analysis of small-subunit rRNA and *mcrA* genes by using defined template mixtures of methanogenic pure cultures and soil DNA extracts. Appl. Environ. Microbiol. **69:** 320–326.
24. **Lukow, T., P. F. Dunfield, and W. Liesack.** 2000. Use of T-RFLP technique to assess spatial and temporal changes in the bacterial community structure within an agricultural soil planted with transgenic and non-transgenic potato plants. FEMS Microbiol. Ecol. **32:**241–247.
25. **Marsh, T. L.** 1999. Terminal restriction length polymorphism (T-RFLP): an emerging method for characterizing diversity among homologous populations of amplification products. Curr. Opin. Microbiol. **2:**323–327.
26. **Moeseneder, M. M., J. M. Arrieta, G. Muyzer, C. Winter, and G. J. Herndl.** 1999. Optimization of terminal-restriction fragment length polymorphism analysis for complex marine bacterioplankton communities and comparison with denaturing gradient gel electrophoresis. Appl. Environ. Microbiol. **65:** 3518–3525.
27. **O'Farrell, K. A., and P. H. Janssen.** 1999. Detection of verrucomicrobia in a pasture soil by PCR-mediated amplification of 16S rRNA genes. Appl. Environ. Microbiol. **65:**4280–4284.
28. **Olsgard, F., P. J. Somerfield, and M. R. Carr.** 1997. Relationships between taxonomic resolution and data transformation in analysis of a macrobenthic

community along an established pollution gradient. Mar. Ecol. Prog. Ser. **149:**173–181.

29. **Olsgard, F., P. J. Somerfield, and M. R. Carr.** 1998. Relationships between taxonomic resolution, macrobenthic community patterns and disturbance. Mar. Ecol. Prog. Ser. **172:**25–36.
30. **Osborn, A. M., E. R. B. Moore, and K. N. Timmis.** 2000. An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. Environ. Microbiol. **2:**39–50.
31. **Osborne, C. A., M. Galic, P. Sangwan, and P. H. Janssen.** 2005. PCR-generated artefact from 16S rRNA gene-specific primers. FEMS Microbiol. Lett. **248:**183–187.
32. **Rees, G. N., D. S. Baldwin, G. O. Watson, S. Perryman, and D. L. Nielsen.** 2004. Ordination and significance testing of microbial community composition derived from terminal restriction fragment length polymorphisms: application of multivariate statistics. Antonie Leeuwenhoek **86:**339–347.
33. **Rösch, C., and H. Bothe.** 2005. Improved assessment of denitrifying, N_2-fixing, and total-community bacteria by terminal restriction fragment length polymorphism analysis using multiple restriction enzymes. Appl. Environ. Microbiol. **71:**2026–2035.
34. **Saikaly, P. E., P. G. Stroot, and D. B. Oerther.** 2005. Use of 16S rRNA gene terminal restriction fragment analysis to assess the impact of solids retention time on the bacterial diversity of activated sludge. Appl. Environ. Microbiol. **71:**5814–5822.
35. **Sait, L., M. Galic, R. A. Strugnell, and P. H. Janssen.** 2003. Secretory antibodies do not affect the composition of the bacterial microbiota in the terminal ileum of 10-week-old mice. Appl. Environ. Microbiol. **69:**2100–2109.
36. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
37. **Sessitsch, A., A. Weilharter, M. H. Gerzabek, H. Kirchmann, and E. Kandeler.** 2001. Microbial population structures in soil particle size fractions of a long-term fertilizer field experiment. Appl. Environ. Microbiol. **67:**4215–4224.
38. **Somerfield, P. J., and K. R. Clarke.** 1995. Taxonomic levels, in marine community studies, revisited. Mar. Ecol. Prog. Ser. **127:**113–119.
39. **Somerfield, P. J., and K. R. Clarke.** 1997. A comparison of some methods commonly used for the collection of sublittoral sediments and their associated fauna. Mar. Environ. Res. **43:**145–156.
40. **Sorensen, T.** 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. K. Dan. Vidensk. Selsk. Biol. Skr. **5:**1–34.
41. **Spellerberg, I. F., and P. J. Fedor.** 2003. A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the "Shannon-Wiener" index. Glob. Ecol. Biogeogr. **12:**177–179.
42. **Tom-Petersen, A., T. D. Leser, T. L. Marsh, and O. Nybroe.** 2003. Effects of copper amendment on the bacterial community in agricultural soil analyzed by the T-RFLP technique. FEMS Microbiol. Ecol. **46:**53–62.
43. **Van der Gucht, K., T. Vandekerckhove, N. Vloemans, S. Cousin, K. Muylaert, K. Sabbe, M. Gillis, S. Declerk, L. De Meester, and W. Vyverman.** 2005. Characterization of bacterial communities in four freshwater lakes differing in nutrient load and food web structure. FEMS Microbiol. Ecol. **53:**205–220.
44. **Yeager, C. M., D. E. Northup, C. C. Grow, S. M. Barns, and C. R. Kuske.** 2005. Changes in nitrogen-fixing and ammonia-oxidizing bacterial communities in soil of a mixed conifer forest after wildfire. Appl. Environ. Microbiol. **71:**2713–2722.