

# Introducing TreeClimber, a Test To Compare Microbial Community Structures

Patrick D. Schloss and Jo Handelsman\*

*Department of Plant Pathology, University of Wisconsin—Madison, Madison, Wisconsin 53706*

Received 22 October 2005/Accepted 17 January 2006

**The phylogenetic and ecological complexity of microbial communities necessitates the development of new methods to determine whether two or more communities have the same structure even though it is not possible to sample the communities exhaustively. To address this need, we adapted a method used in population genetics, the parsimony test, to determine the relatedness of communities. Here we describe our implementation of the parsimony test, TreeClimber, in which we reanalyzed six previously published studies and compared the results of the analysis to those obtained using  $\beta$ -LIBSHUFF.**

Microbial communities are diverse, containing as many as  $10^{12}$  individuals per g that represent thousands of species. In light of this enormous biodiversity, comparing microbial communities is challenging. Although many aspects of microbial community analysis have been the foci of recent efforts to bring statistical rigor to analysis of bacterial communities (12, 13, 17, 20, 23, 26–29), we continue to lack methods to compare communities. Comparisons of community structure, which is the distribution of the total number of individuals among the species in the community (1), are essential for correlating the relative abundance of a taxonomic group with a specific activity. For example, a comparison of soil communities that suppress plant pathogens with soil communities that do not suppress plant pathogens might identify the features of a community that make a soil suppressive or suggest biological control agents and focus efforts on their cultivation. In communities that are too complex to sample more than a small proportion of the members, it is not possible to state with confidence using descriptive analyses whether the observed differences in sequence representation are due to random chance or to a biologically significant difference. Therefore, quantitative tools are needed to support statistical inferences based on samples of a population.

We previously modified an analysis based on the Cramér-von Mises statistic as implemented in LIBSHUFF (29) and designated our implementation  $\beta$ -LIBSHUFF (27). This test uses a pairwise distance matrix to determine whether two samples are drawn from the same population or whether one is a subset of the other. One limitation of the LIBSHUFF-type of analysis is that every pairwise comparison actually requires two hypothesis tests. If two samples are to be compared, then two comparisons are made; if there are four samples, then there are 12 comparisons (number of comparisons =  $N_{\text{treat}} [N_{\text{treat}} - 1]$ ). This approach is comparable to performing numerous pairwise  $t$  tests instead of performing a single analysis of variance (ANOVA). Multiple comparisons require adjusting the family type I error rate, which is the probability of incorrectly detect-

ing a significant difference, in order to maintain a desired experiment-wise type I error rate that is typically set at 0.05. Adjusting for the effects of multiple comparisons reduces the statistical power to detect differences that actually exist (33). A second limitation is that the user must use distance-based approaches to perform the analysis instead of potentially more robust Bayesian, parsimony, or maximum-likelihood methods. Since all phylogenetic approaches have limitations (9), it would be advantageous if investigators had the flexibility to compare their results using a variety of methods.

Recently, Martin (20) proposed adapting the parsimony test, which is used in population genetics to describe gene flow (19, 30, 31), to make comparisons between 16S rRNA gene libraries. This method has been adapted to perform a similar analysis called UniFrac (17). Microbial ecologists have used the parsimony test to make comparisons using 16S rRNA (16, 34), 18S rRNA (25), ribulose 1,5-bisphosphate carboxylase/oxygenase (22), aerobic carbon monoxide dehydrogenase (7), and viral (3) sequence collections from a diverse collection of environments. Although the parsimony test has received broad application in microbial ecology, no one has clearly articulated the exact hypothesis tested by the method in statistical or biological terms.

In previous implementations of the parsimony test workers have used the MacClade (<http://www.macclade.org>) software package, which is an expensive platform-specific computer program that is not easily adaptable to this test. We developed a free computer program, TreeClimber, which implements the parsimony test in a rapid, easy-to-use, and streamlined manner. Here we state the hypothesis evaluated in the parsimony test, demonstrate TreeClimber's ability to implement the parsimony test, and discuss several aspects of the test's implementation that have not been addressed previously.

## MATERIALS AND METHODS

**Parsimony test algorithm.** The parsimony test consists of three parts (19, 30, 31). First, a composite phylogeny for a set of communities is reconstructed and the taxa are associated with the treatment from which they originated. Second, the minimum number of steps or transitions necessary to explain the covariation of phylogeny with the treatments is determined. Finally, scores are calculated for random phylogenetic trees generated by randomly linking taxa. This step is repeated many times ( $n \geq 1,000$ ) to obtain a probability distribution describing the probability that each score is observed by a random process of assembling

\* Corresponding author. Mailing address: Department of Plant Pathology, University of Wisconsin—Madison, Madison, WI 53706. Phone: (608) 263-8783. Fax: (608) 265-5289. E-mail: joh@plantpath.wisc.edu.

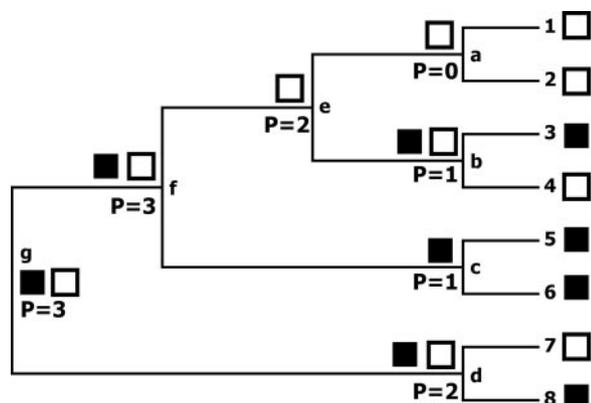


FIG. 1. Examples of using Fitch's algorithm (11) to score a phylogenetic tree when samples are obtained from two communities. The letters represent treatments for the nodes, and the numbers represent the treatment label for each sequence considered. The parsimony scores ( $P$ ) and treatment labels for each node as they aggregate through the nodes are indicated.

sequences into groups. TreeClimber performs the last two parts, producing a probability distribution for observing each parsimony score using the tree(s) submitted by the user and the random trees.

**Phylogenetic reconstruction.** Sequence alignments were constructed in ClustalW (version 1.83) using a gap-opening penalty of 10.0 and a gap extension penalty of 0.10. Trees constructed using the neighbor-joining, bootstrapped neighbor-joining, and parsimony algorithms were calculated using PAUP\* (35). Because an exhaustive search of the most parsimonious tree was not possible for all of the data sets that we analyzed, we retained the 1,000 most parsimonious trees. Trees constructed using MrBayes were calculated by assuming a general time-reversible model and gamma-distributed rates. The MrBayes analysis used four Markov chains, which were run for at most 500,000 generations so that the maximum likelihood of each generation was relatively constant. The final 100,000 trees were sampled every 100 generations to obtain 1,000 trees. The most recent releases of ARB (18), PHYLIP (10), PAUP\* (35), MEGA3 (14), and MrBayes (24) were obtained from the appropriate websites. NEXUS-formatted files used to perform many of the simulations presented here are available on the TreeClimber website (<http://www.plantpath.wisc.edu/fac/joh/treeclimber.html>).

**Scoring a tree by parsimony.** Given a phylogenetic tree in which each taxon is associated with a treatment, TreeClimber determines the most parsimonious number of transitions between the treatments using Fitch's algorithm (11). The parsimony test algorithm uses a phylogenetic tree in which each exterior branch (Fig. 1, leaves 1 to 8) is labeled according to the treatment from which the

sequence originated. The algorithm compares each leaf to determine the appropriate label for each of the interior nodes (Fig. 1, nodes a to g) by comparing the labels of the two branches that share the same node. If there is an intersection between the two sets of labels, then the node is labeled with the result of the intersection and a penalty is not assessed to the parsimony score. If there is not an intersection between the two labels, then the union of the two labels is applied and a penalty of 1 is assessed to the parsimony score. This scoring algorithm can be applied to any bifurcating tree with as many different labels as are desired. Figure 1 shows an example for a phylogenetic tree from an analysis with two treatments. The parsimony score for this tree was three.

**Significance of parsimony score.** TreeClimber constructed 10,000 randomly joining trees for each test to generate each random probability distribution. TreeClimber generates random joining trees by using the same number of individuals for each treatment that was found in the actual data set. The program randomly selects either two taxa, two nodes, or a node and a taxon and joins them at a new node. This process is repeated until a complete tree is constructed. The parsimony score for each random tree is calculated. The fraction of the 10,000 randomly joining trees with that parsimony score is the probability of observing that score by chance, assuming that the hypothesis being evaluated is correct. A comparison of the parsimony score for the user-inputted tree or trees to the cumulative probability distribution indicates the probability of observing the same or a lower parsimony score by chance.

**TreeClimber.** TreeClimber is available free as C++ source code for compilation in Linux or Mac OSX and as a binary executable for the Windows operating system (<http://www.plantpath.wisc.edu/fac/joh/treeclimber.html>). Two input files are required to run the program: a PHYLIP or NEXUS-formatted Newick file containing one or more phylogenetic trees and a file containing the sequence name and its corresponding treatment. The user can select the number of randomly joining trees desired to construct the test distribution. A probability distribution for any number of treatments and replicates can be determined without input of an observed set of phylogenies or set of sequence identifiers. Finally, TreeClimber generates a file containing the probability of observing each parsimony score for the user and randomly joining trees.

## RESULTS AND DISCUSSION

**TreeClimber evaluation.** We developed TreeClimber to provide a streamlined method for performing the parsimony test. To test our implementation of the parsimony test, we evaluated TreeClimber using a variety of data sets obtained from the literature (Table 1). The results of our implementation were consistent with the results observed in the original studies. Phylogenetic trees generated in PHYLIP, ARB, MEGA3, PAUP\*, and MrBayes have been successfully used in TreeClimber (Fig. 2).

TABLE 1. Comparison of  $P$  values obtained using  $f$ -LIBSHUFF and TreeClimber for various methods of phylogenetic tree construction and 16S rRNA sequence collections<sup>a</sup>

Site (reference)	Library name (no. of sequences)	$f$ -LIBSHUFF $P$ value	TreeClimber $P$ value at 95th percentile (proportion of $P$ values that are $\leq 0.05$ )			
			Neighbor joining	Parsimony	Neighbor joining bootstrap	MrBayes
Sequencing batch reactors (2)	SBR1 (97), SBR2 (92)	0.065, 0.747 <sup>b</sup>	0.103 (0.000)	0.159 (0.003)	0.522 (0.032)	0.409 (0.005)
Arid soils (6)	C0 (59), S0 (53)	0.012*, 0.187 <sup>b</sup>	0.147 (0.000)	0.147 (0.000)	0.244 (0.355)	0.147 (0.483)
Scottish soil (21)	SAF (138), SL (137)	0.030, 0.073 <sup>b</sup>	0.004* (1.000)	0.018* (1.000)	0.031* (0.989)	0.047* (0.982)
Wisconsin soil (15)	1997 (139), 2000 (129)	<0.001*, <0.001* <sup>b</sup>	<0.001* (1.000)	<0.001* (1.000)	<0.001* (1.000)	<0.001* (1.000)
Alpine tundra fungi (25)	Spring (56), summer (27), winter (42)	<0.001* <sup>c</sup>	<0.001* (1.000)	<0.001* (1.000)	<0.001* (1.000)	<0.001* (1.000)
Feces from patient receiving antibiotics (36)	Preantibiotic (84), during treatment (74), postantibiotic (84)	<0.001* <sup>c</sup>	<0.001* (1.000)	<0.001* (1.000)	<0.001* (1.000)	<0.001* (1.000)

<sup>a</sup> Asterisks indicate  $P$  values that are less than 0.05 after correction for multiple comparisons. All  $P$  values are based on 10,000 randomizations. The  $P$  value at the 95th percentile is the  $P$  value where 95% of the user-supplied trees had a  $P$  value equal to or less than 0.05.

<sup>b</sup> The first value is the value for the analysis performed using the first library as the homologous library, and the second value is the value obtained when the second library was used as the homologous library.

<sup>c</sup> The  $P$  value for all six comparisons was <0.001.

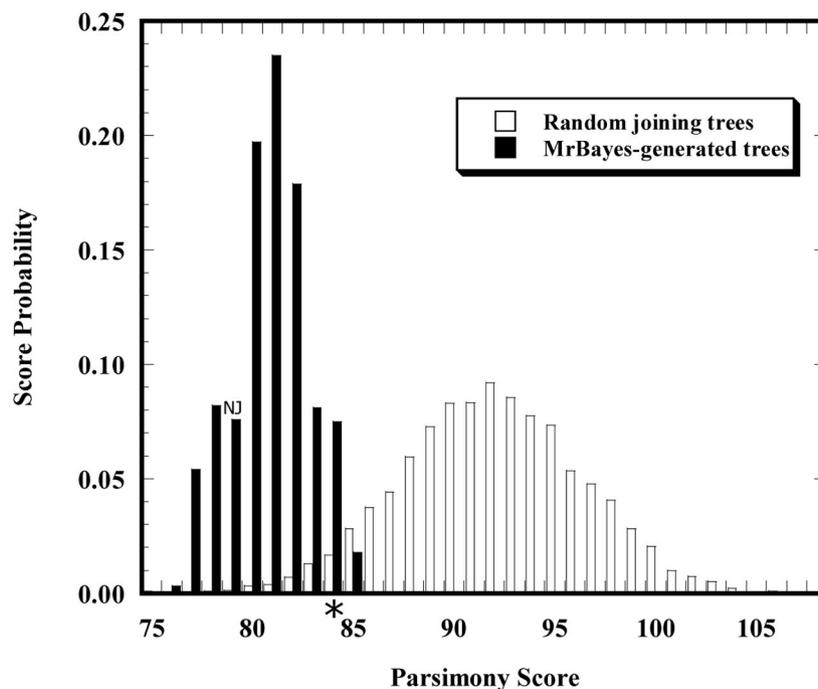


FIG. 2. TreeClimber-generated probability distribution of MrBayes-generated and neighbor-joining (NJ) phylogenetic trees for the data of McCaig et al. (21). A Scottish soil data set was compared to the distribution of scores obtained from 10,000 random joining trees when there were 137 and 138 samples for each treatment. The asterisk at a parsimony score of 84 indicates the score at which the cumulative random probability was equal to 0.047. When MrBayes-generated trees were used, 98.2% of the trees had a  $P$  value less than 0.05 and the  $P$  value for the neighbor-joining tree was 0.004.

A parallel analysis of each study was performed using  $f$ -LIBSHUFF to determine whether each community investigated was a subsample of the community represented in the other library. In four of the six studies that we analyzed, we obtained low  $P$  values by both techniques (Table 1). For one of the six studies we calculated a low  $P$  value using  $f$ -LIBSHUFF and a high  $P$  value using TreeClimber (6), and for another we observed the opposite result (21). Differences in significance between the two methods are likely due to differences in statistical power and the hypotheses tested in the analyses.

Since the parsimony test algorithm was introduced and implemented in other studies (3, 7, 16, 22, 25, 34), we identified five potentially problematic issues. The most significant issue that deserves attention is stating the hypothesis that the parsimony test evaluates. Other issues include choosing a phylogeny, performing the analysis with a subset of the data to identify effects of a treatment on a specific group of organisms, designing appropriate experiments that are amenable to the test, and determining whether to use duplicate sequences in the analysis.

**Hypothesis development.** Martin (20) generically described the application of a parsimony test as a test to determine whether two 16S rRNA sequence collections harbor different lineages. As developed by Maddison and Slatkin (19, 30, 31), however, the hypothesis posits that populations in different treatments are either panmictic, epidemic, or clonal. There are a number of problems with applying the hypothesis to microbial communities, including the lack of classical sexual mating among bacteria, the lack of a single population of one species,

and the concepts of panmictic, epidemic, clonal and populations, which have been defined in microbiology to refer to the association of genes at different loci using multilocus enzyme electrophoresis (32).

For application to community phylogenies, it is necessary to adapt the original hypothesis. Our modified hypothesis is that two communities share an ancestral community structure (i.e., a coalescent community structure) and the phylogenetic differences observed between two or more community structures are due to an accumulation of random variation. Randomly joining trees are an example of the possible phylogenies that are observed if this hypothesis is correct because populations are equally likely to be found in each treatment. Departures from this hypothesis occur when populations in the communities experience some effect (i.e., perturbation or selection) that causes populations to be either gained or lost. Alternatively, some effect may impede communities from demonstrating signs of variation. If populations experience selection pressure, then populations of similar members are more likely to affiliate with a single treatment than with multiple treatments. This would result in a lower parsimony score than is expected by random variation. If no random variation occurs, then the same lineages found in one treatment are found in all of the treatments. This would result in a parsimony score that is higher than the score expected by random variation. Based on the available data (Table 1), it seems unlikely that two or more communities are identical to each other. Therefore, we are interested in testing the probability that a treatment exerts selection pressure on the structure of multiple communities.

**Phylogenetic reconstruction.** We found considerable variability in  $P$  values in an analysis using different phylogenetic methods, indicating that the robustness of the underlying phylogeny is critical to the analysis (Table 1). Comparisons of phylogenies constructed using different methods may increase the confidence that small  $P$  values are due to a biological effect and are not a product of the methodology. We considered the MrBayes-generated trees to be the most robust phylogeny available because they represented an equiprobable random sampling of the true phylogeny. When at least 95% of the 1,000 trees generated had a  $P$  value less than 0.05, we considered the differences between treatments to be significant (Fig. 2).

A salient difference between the parsimony test and  $f$ -LIBSHUFF is that  $f$ -LIBSHUFF calculates a continuous statistic and TreeClimber calculates a discrete statistic whose range is dependent on the number of sequences analyzed. In a  $f$ -LIBSHUFF analysis, if 10,000 randomizations are performed, there are generally 10,000 different values for the statistic. For the same number of randomizations, the distribution is confined to a much narrower window of possibilities, which means that a difference in the parsimony score of 1 or 2 can represent a large difference in the  $P$  value. This results in decreased resolution among  $P$  values and may result in reduced statistical power. For example, when analyzing the data set of Dunbar et al. (6), we obtained a parsimony score of 34 for the neighbor-joining and parsimony trees and a median score of 33 using the bootstrapped neighbor-joining and MrBayes trees; these scores had  $P$  values of 0.147 and 0.081, respectively. Had any of the scores been 32, the  $P$  value would have been 0.040 and considered marginally significant. This indicates that the analysis is sensitive to the quality of the phylogeny.

Ideally, experiments using TreeClimber would be designed to maximize the resolution among  $P$  values, resulting in improved statistical power. The spread in parsimony scores between the 5th and 95th percentiles increased an average of 0.03 for every taxon added to two treatments when there were at least 100 taxa for each treatment (data not shown). When there were at least 100 sequences for each treatment, the 5th percentile parsimony score increased 0.65 for every taxon added to each library (data not shown). Considering that only 53 and 59 sequences were obtained from both communities in the data set of Dunbar et al. (6), it is likely that additional sequencing of community members would result in a high probability of finding evidence of differential selection pressure.

**Experimental design.** The 16S rRNA gene libraries of McCaig et al. (21) from an unimproved pasture and an improved pasture that received fertilizers have been extensively studied by microbiologists developing tools to study microbial ecology (5, 12, 20, 26, 27, 29). Although the complete collection represents three replicate 16S rRNA gene libraries from each of the pastures, each of the previous analyses of the data set ignored the nested experimental design, pooled the three libraries within each site, and implemented a single-classification design. These accommodations were necessary because, unfortunately, the parsimony test and LIBSHUFF-type tests are amenable only to single-classification analysis.

In a classical nested ANOVA, it is permissible to pool replicates and use a single-classification ANOVA when the variability between replicates within a treatment is not significant

(33). We used an analogous approach by considering each of the three replicates for each pasture a separate treatment. A comparison across the improved pasture libraries resulted in a marginally high  $P$  value (87.6% of trees with a  $P$  value less than 0.05), and a low  $P$  value was observed for the comparison across the unimproved libraries (100% of trees with a  $P$  value less than 0.001) with trees obtained using the MrBayes phylogenies. This indicates that the improved soil replicates could be pooled, while the unimproved soil replicates could not be pooled. Pairwise comparisons of the three unimproved replicates produced low  $P$  values (95% of trees in all three comparisons had a  $P$  value less than 0.015), indicating that the community structures represented in the three libraries were significantly different from one another. In light of this result and the inability to account for intrasoil type variability in the parsimony test analysis, it is unclear whether the low  $P$  values observed when the pooled data sets were compared were due to differences between the unimproved soil replicates or differences between replicates of the unimproved soil. Depending on how samples were selected, the data might suggest that fertilization resulted in a more homogeneous distribution of bacteria in the improved soils.

Schadt et al. (25) collected 18S rRNA gene sequences from alpine soils in the spring ( $n = 56$ ), summer ( $n = 27$ ), and winter ( $n = 42$ ) to investigate seasonal variability in fungal communities. They did not report results of a comparison across all three treatments or a pairwise comparison between the summer and winter treatments, but they did describe the results of pairwise comparisons between the summer and spring sequence collections (100% of trees with a  $P$  value of  $<0.05$ ) and between the spring and winter sequence collections (60% of trees with a  $P$  value of  $<0.05$ ). The pairwise comparison approach is problematic because it increases the probability of identifying significant differences that are due to chance (33). However, when we replicated the analysis of Schadt et al. and made one comparison across all three treatments, 100% of 1,000 trees had a  $P$  value less than 0.001, indicating that the original investigators' pairwise comparisons were appropriate. As with a classical ANOVA, it is appropriate to make pairwise comparisons only after showing that there is some experiment-wide difference to mitigate the problems associated with multiple comparisons (33).

**Analysis of individual lineages.** Identification of the lineages responsible for differences in community structure has been accomplished by performing the parsimony test on trees in which specific groups are removed and using trees that contain sequences from only one group. The data set of McCaig et al. (21) was analyzed by investigating the effects of the  $\alpha$ -*Proteobacteria* on community structure for improved and unimproved soil libraries (20). In the previous analysis, in which neighbor-joining trees were used, when the  $\alpha$ -*Proteobacteria* sequences were removed, there was no statistically significant difference between the two communities ( $P > 0.05$ ), and when the  $\alpha$ -*Proteobacteria* from each library were compared, there was a significant difference ( $P = 0.004$ ) (20). The workers then showed that a clade within the  $\alpha$ -*Proteobacteria* having sequences similar to sequences of *Rhodobacter*, *Paracoccus*, and *Aquabacter* spp. accounted for much of the observed difference between  $\alpha$ -*Proteobacteria* lineages. However, when we removed the  $\alpha$ -*Proteobacteria* and used a neighbor-joining tree in

TreeClimber, we obtained a  $P$  value of 0.043, which indicated that exclusion of the  $\alpha$ -*Proteobacteria* lineages still showed marginal, but significant, covariation between phylogeny and treatment. As the original  $P$  value was not stated in the previous analysis, we do not know how different the previous  $P$  value was from our own marginally significant result.

To illustrate the problem with pruning groups of sequences from a phylogenetic tree, we used DOTUR (26) to assign sequences to operational taxonomic units (OTUs). Defining an OTU as a collection of sequences that are at most 0.30 Jukes-Cantor distance unit from each other, we identified 23 OTUs that contained between 1 and 105 sequences. Then for the seven OTUs that contained 10 or more sequences, we serially pruned the sequences from each OTU from the tree and recalculated the  $P$  value. As previously applied (20), if removing the OTU resulted in a high  $P$  value, then the analysis indicated that the organisms represented by the OTU were differentially represented in the two communities. When we repeated the analysis with neighbor-joining-generated trees and removed each OTU, none of the  $P$  values was more than 0.05. When we removed each of the seven OTUs from the overall phylogeny, removing six of the OTUs resulted in less than 95% of the MrBayes-generated trees having  $P$  values less than 0.05; the seventh OTU contained  $\alpha$ -*Proteobacteria* sequences. Although it is probable that shifts among the  $\alpha$ -*Proteobacteria* accounted for the overall differences in community structure, the reduced statistical power to detect differences caused by the removal of 38.2% of the sequence collection probably obscured the detection of such an effect using the parsimony test.

We then calculated the  $P$  values for comparisons using only the sequences in each of the pruned OTUs. After each OTU was pruned, analysis of the MrBayes-generated trees revealed that six of the seven OTUs exhibited random variation (95% of all trees for each OTU had a  $P$  value greater than 0.05). Removal of the seventh OTU, which contained 12 sequences ( $N_{\text{improved}} = 5$ ,  $N_{\text{unimproved}} = 7$ ) related to the  $\gamma$ -proteobacterium *Legionella* sp. sequence, resulted in a low  $P$  value (95% of trees with a  $P$  value less than 0.05). Our inability to detect other OTUs whose representation was the result of selection pressure may have been due to a lack of statistical power in a comparison of 10 to 105 sequences. Alternatively, we may have found an OTU whose representation was the result of selection pressure, because we performed a sufficient number of comparisons to find a difference by chance. Ultimately, studies using domain-level PCR primers evaluate domain-level differences, not differences in specific lineages. Testing hypotheses about specific lineages requires designing PCR primers for the particular group and then testing the comparison (34).

**Community membership or structure?** The generic statement that the parsimony test compares the differences in lineages among groups implies that the test measures community membership, not structure (20). In some analyses, sequences that are identical within a treatment are removed from the community phylogeny (25), and thus, abundance of types is not measured. The hypothesis tested in the parsimony test is a measure of differences in community structure, which takes into account the composition and distribution of sequences. If one lineage is the object of positive selection, it becomes disproportionately more abundant than the same lineage in a treatment that applies no positive selection. Measuring the

effect of selection pressures on a community requires knowing the relative abundance of each lineage within each community.

We reevaluated the pairwise comparisons made by Schadt et al. (25), who removed all duplicate sequences from each treatment prior to their analysis. When we made comparisons across all three treatments using TreeClimber (i.e., ANOVA-style design), all of the MrBayes-generated trees had a  $P$  value that was less than 0.001 when we included or excluded identical sequences. Performing the summer-winter and summer-spring pairwise comparisons with TreeClimber produced  $P$  values for all MrBayes trees with and without the redundant sequences that were less than 0.001. When we used TreeClimber to perform winter-spring pairwise comparisons with and without redundant sequences, 95% of all MrBayes-generated trees had  $P$  values that were less than 0.004 and 0.129, respectively. Using nonparametric estimators of the Jaccard and Sorenson similarity indices (4), we found that the similarity values were significantly less than 1.0 for OTUs when the sequences within an OTU were defined by a group of sequences that were at most 4% different from one another. These results demonstrate that the memberships of the communities were not similar, underscoring the point that the parsimony test compares community structure, not membership.

**Conclusions.** Molecular microbial ecology has advanced beyond cataloging sequences to elucidating ecological mechanisms, which requires answering the general question, "are two communities different?" The complexity of microbial communities makes addressing this question impossible without specialized tools. Many of the statistical tools introduced in the last several years have been retrofitted from macroecology population genetics. Although many tools have been developed in macroecology, it is critical to precisely define the hypothesis when a tool is applied to microbial communities. For example, the original description of the parsimony test described it as a test for classical panmixia, a phenomenon that is not found among the bacteria.

The parsimony test is a powerful tool for measuring mechanisms that force a community structure to change. Our analyses assumed a one-tailed hypothesis test that would determine whether the differences between two communities arose due to random variation or whether lineages from one community had become more dominant through negative or positive selection pressures. This assumption was influenced by data sets that showed that microbial communities are generally not identical. A two-tailed test may be appropriate in other environments in which the environment selects for a particular community structure.

The parsimony test is not a panacea for statistical microbial ecology studies. Experimental design and the discrete nature of the probability distribution limit the types of analyses that can be performed and the statistical power of the analyses. Although LIBSHUFF-type analyses are limited by the requirement of performing two tests for every pairwise comparison, they have the advantage that they are able to determine whether one community structure is a subset of another. Other methods, such as analysis of molecular variance (8, 20), can be useful for determining whether the overall genetic variation between two or more communities is greater than the diversity within a community. Further work is necessary to develop additional tests that are useful for time series analyses, multi-

way ANOVA, and replicated samples. These advances may provide other answers to the question, “are two communities different?”

#### ACKNOWLEDGMENTS

This work was supported by a USDA postdoctoral fellowship in soil biology to P.D.S. (2003-35107-13856), by the NSF Microbial Observatories program (MCB-0132085), by the Howard Hughes Medical Institute, and by the University of Wisconsin—Madison College of Agricultural and Life Sciences.

#### REFERENCES

- Begon, M., J. L. Harper, and C. R. Townsend. 1996. Ecology: individuals, populations, and communities, 3rd ed., p. 829–830. Blackwell Science, Malden, MA.
- Bond, P. L., P. Hugenholtz, J. Keller, and L. L. Blackall. 1995. Bacterial community structures of phosphate-removing and non-phosphate-removing activated sludges from sequencing batch reactors. *Appl. Environ. Microbiol.* **61**:1910–1916.
- Breitbart, M., B. Felts, S. Kelley, J. M. Mahaffy, J. Nulton, P. Salamon, and F. Rohwer. 2004. Diversity and population structure of a near-shore marine-sediment viral community. *Proc. R. Soc. Lond. B Biol. Sci.* **271**:565–574.
- Chao, A., R. L. Chazdon, R. K. Colwell, and T. J. Shen. 2005. A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecol. Lett.* **8**:148–159.
- Curtis, T. P., W. T. Sloan, and J. W. Scannell. 2002. Estimating prokaryotic diversity and its limits. *Proc. Natl. Acad. Sci. USA* **99**:10494–10499.
- Dunbar, J., S. Takala, S. M. Barns, J. A. Davis, and C. R. Kuske. 1999. Levels of bacterial community diversity in four arid soils compared by cultivation and 16S rRNA gene cloning. *Appl. Environ. Microbiol.* **65**:1662–1669.
- Dunfield, K. E., and G. M. King. 2004. Molecular analysis of carbon monoxide-oxidizing bacteria associated with recent Hawaiian volcanic deposits. *Appl. Environ. Microbiol.* **70**:4242–4248.
- Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* **131**:479–491.
- Felsenstein, J. 2004. Inferring phylogenies. Sinauer Associates, Sunderland, Mass.
- Felsenstein, J. 1989. PHYLIP—Phylogeny Inference Package. *Cladistics* **5**:164–166.
- Fitch, W. M. 1971. Toward defining the course of evolution: minimum change for a specific tree topology. *Syst. Zool.* **20**:406–416.
- Hughes, J. B., J. J. Hellmann, T. H. Ricketts, and B. J. M. Bohannan. 2001. Counting the uncountable: statistical approaches to estimating microbial diversity. *Appl. Environ. Microbiol.* **67**:4399–4406.
- Hur, I., and J. Chun. 2004. A method for comparing multiple bacterial community structures from 16S rDNA clone library sequences. *J. Microbiol.* **42**:9–13.
- Kumar, S., K. Tamura, and M. Nei. 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform.* **5**:150–163.
- Liles, M. R., B. F. Manske, S. B. Bintrim, J. Handelsman, and R. M. Goodman. 2003. A census of rRNA genes and linked genomic sequences within a soil metagenomic library. *Appl. Environ. Microbiol.* **69**:2684–2691.
- Lipson, D. A., and S. K. Schmidt. 2004. Seasonal changes in an alpine soil bacterial community in the Colorado Rocky Mountains. *Appl. Environ. Microbiol.* **70**:2867–2879.
- Lozupone, C., and R. Knight. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71**:8228–8235.
- Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, et al. 2004. ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**:1363–1371.
- Maddison, W. P., and M. Slatkin. 1991. Null models for the number of evolutionary steps in a character on a phylogenetic tree. *Evolution* **45**:1184–1197.
- Martin, A. P. 2002. Phylogenetic approaches for describing and comparing the diversity of microbial communities. *Appl. Environ. Microbiol.* **68**:3673–3682.
- McCaig, A. E., L. A. Glover, and J. I. Prosser. 1999. Molecular analysis of bacterial community structure and diversity in unimproved and improved upland grass pastures. *Appl. Environ. Microbiol.* **65**:1721–1730.
- Namba, K., G. M. King, and K. Dunfield. 2004. Analysis of facultative lithotroph distribution and diversity on volcanic deposits by use of the large subunit of ribulose 1,5-bisphosphate carboxylase/oxygenase. *Appl. Environ. Microbiol.* **70**:2245–2253.
- 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.
- Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**:1572–1574.
- Schadt, C. W., A. P. Martin, D. A. Lipson, and S. K. Schmidt. 2003. Seasonal dynamics of previously unknown fungal lineages in tundra soils. *Science* **301**:1359–1361.
- Schloss, P. D., and J. Handelsman. 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl. Environ. Microbiol.* **71**:1501–1506.
- Schloss, P. D., B. R. Larget, and J. Handelsman. 2004. Integration of microbial ecology and statistics: a test to compare gene libraries. *Appl. Environ. Microbiol.* **70**:5485–5492.
- Seguritan, V., and F. Rohwer. 2001. FastGroup: a program to dereplicate libraries of 16S rDNA sequences. *BMC Bioinformatics* **2**:9.
- Singleton, D. R., M. A. Furlong, S. L. Rathbun, and W. B. Whitman. 2001. Quantitative comparisons of 16S rRNA gene sequence libraries from environmental samples. *Appl. Environ. Microbiol.* **67**:4374–4376.
- Slatkin, M., and W. P. Maddison. 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* **123**:603–613.
- Slatkin, M., and W. P. Maddison. 1990. Detecting isolation by distance using phylogenies of genes. *Genetics* **126**:249–260.
- Smith, J. M., N. H. Smith, M. O'Rourke, and B. G. Spratt. 1993. How clonal are bacteria? *Proc. Natl. Acad. Sci. USA* **90**:4384–4388.
- Sokal, R. R., and F. J. Rohlf. 1995. Biometry: the principles and practice of statistics in biological research, 3rd ed. Freeman, New York, NY.
- Stach, J. E. M., L. A. Maldonado, D. G. Masson, A. C. Ward, M. Goodfellow, and A. T. Bull. 2003. Statistical approaches for estimating actinobacterial diversity in marine sediments. *Appl. Environ. Microbiol.* **69**:6189–6200.
- Swofford, D. L. 2003. PAUP\*: phylogenetic analysis using parsimony (\* and other methods), version 4. Sinauer Associates, Sunderland, MA.
- Young, V. B., and T. M. Schmidt. 2004. Antibiotic-associated diarrhea accompanied by large-scale alterations in the composition of the fecal microbiota. *J. Clin. Microbiol.* **42**:1203–1206.