Tsushima, S., Hasebe, A., Komoto, Y., Carter, J. P., Miyashita, K., Yokoyama, K., and Pickup, R. W. (1995). Detection of genetically-engineered microorganisms in paddy soil using a simple and rapid nested polymerase chain-reaction method. *Soil Biol. Biochem.* **27**, 219–227.

Watson, R. J., and Blackwell, B. (2000). Purification and characterization of a common soil component which inhibits the polymerase chain reaction. *Can. J. Microbiol.* **46**, 633–642.

Young, C. C., Burghoff, R. L., Keim, L. G., Minak-Bernero, V., Lute, J. R., and Hintom, S. M. (1993). Polyvinylpyrrolidone-agarose gel electrophoresis purification of polymerase chain-amplifiable DNA from soils. *Appl. Environ. Microbiol.* **59**, 1972–1974.

Yu, Z. T., and Mohn, W. W. (1999). Killing two birds with one stone: Simultaneous extraction of DNA and RNA from activated sludge biomass. *Can. J. Microbiol.* **45**, 269–272.

[17] The Application of Rarefaction Techniques to Molecular Inventories of Microbial Diversity

By JENNIFER B. HUGHES *and* JESSICA J. HELLMANN

Abstract

With the growing capacity to inventory microbial community diversity, the need for statistical methods to compare community inventories is also growing. Several approaches have been proposed for comparing the diversity of microbial communities: some adapted from traditional ecology and others designed specifically for molecular inventories of microbes. Rarefaction is one statistical method that is commonly applied in microbial studies, and this chapter discusses the procedure and its advantages and disadvantages. Rarefaction compares observed taxon richness at a standardized sampling effort using confidence intervals. Special emphasis is placed here on the need for precise, rather than unbiased, estimation methods in microbial ecology, but precision can be judged only with a very large sample or with multiple samples drawn from a single community. With low sample sizes, rarefaction curves also have the potential to lead to incorrect rankings of relative species richness, but this chapter discusses a new method with the potential to address this problem. Finally, this chapter shows how rarefaction can be applied to the comparison of the taxonomic similarity of microbial communities.

Introduction

The increasing ease of inventorying microbial diversity bestows exciting opportunities for microbial ecologists, yet the growing size of molecular inventories challenges researchers to interpret very large datasets in

biologically informative ways. Microbial ecologists, like other ecologists, seek to understand the distribution of biodiversity. To identify these patterns and the biotic and abiotic factors that drive them, methods are needed to compare microbial communities across time, space, and experimental treatments. As a result, a number of papers address the topic of statistical approaches for microbial community comparisons (Curtis *et al.*, 2002; Dunbar *et al.*, 2001; Hughes *et al.*, 2001; Martin, 2002).

Currently, most molecular inventories use polymerase chain reaction (PCR) amplification of a gene, such as the 16S ribosomal gene, to assess the diversity of a microbial community from a sample of environmental DNA. The molecular methodologies have numerous pitfalls, among them gene duplications, PCR biases, and primer biases. Many authors have noted these biases and have discussed how to minimize these problems (Thompson *et al.*, 2002; von Wintzingerode *et al.*, 1997). Still others have suggested correction factors (Acinas *et al.*, 2004) or new molecular sampling approaches to skip PCR methods all together (Tyson *et al.*, 2004; Venter *et al.*, 2004). These advances are already yielding invaluable information about the extent and consequences of sampling biases for diversity comparisons.

This chapter concentrates on the problem of undersampling of microbial communities, a problem that seems less likely to be alleviated in the near future than the problem of PCR-related biases. For instance, Sargasso Sea data collected by Venter and colleagues (2004) used shotgun sequencing to assess the molecular diversity of seawater microbes. This technique removes PCR and primer biases; however, even with sequencing 1 billion bp and 1164 16S genes, the study still undersampled the microbial community. More than 70% of the ''species'' of six protein-coding phylogenetic markers in the database were singletons, i.e., they were seen only once. Thus, for microbial ecologists who cannot generate nearly such large datasets, undersampling will certainly present a problem. In contrast, the statistical approaches discussed in this chapter can be applied to samples that knowingly contain methodological biases; as long as these biases are similar (or random) across samples within a study, one can statistically compare community diversity and composition.

This article focuses on one approach, rarefaction analyses, for comparing diversity among communities. Rarefaction is by no means the single best diversity measurement; however, it is probably the most commonly used statistical method in recent microbial diversity studies. This use is for good reason, as it is usually a very good place to begin analysis of a new dataset. We review other diversity statistics used commonly in microbial ecology elsewhere (Bohannan and Hughes, 2003; Hughes and Bohannan, 2004; Hughes *et al.*, 2001). Furthermore, statistics targeted specifically for

molecular inventories of microbes are quickly being proposed (e.g., Curtis *et al.*, 2002; Dunbar *et al.*, 2001; Martin, 2002; Singleton *et al.*, 2001).

What Is Rarefaction?

Background

Rarefaction accounts for the fact that large samples have more species (or any taxonomic unit) than small samples even if they are drawn from the same community. Hurlbert (1971) and Sanders (1968) first introduced the idea of scaling down samples of community diversity to the same number of individuals so that richness could be compared across samples. These authors proposed using $E(S_n)$ as a measure of community diversity, i.e., the expected number of species in a sample of n individuals, from a larger collection of N individuals containing S species.

Since then, community ecologists have broadened the idea of rarefaction as a statistical procedure to standardize for sampling effort (Gotelli and Colwell, 2001). Sampling effort can be represented by individuals sampled, as first suggested, or other units such as number of samples or sampling time. Because of sampling constraints, analyses of microbial diversity so far use individual-based rarefaction, thus we concentrate on this approach here. However, it is important to note that the unit of sampling effort used has large consequences for the interpretation of rarefaction analyses (see Gotelli and Colwell, 2001); as microbial diversity inventories begin to include many samples within one study, this issue will become more relevant.

Procedure

An accumulation curve is a plot of the cumulative number of species observed as each individual is sampled and recorded. The curve could be drawn from the notes of a birder walking through a forest and writing down in order the identity of every bird she detects (Fig. 1). Molecular inventories of microbial diversity from one sample usually give a mass capture of individuals, and thus an accumulation curve means very little. For instance, a researcher makes a clone library of PCR products and then randomly picks colonies, or "individuals," to sequence. In contrast to data a birder collects along a transect, the sampling order of the clones from within a clone library does not relate any information about the natural community.

There is, however, useful information for a microbial ecologist in a rarefaction curve: a smoothed, or randomized, accumulation curve.
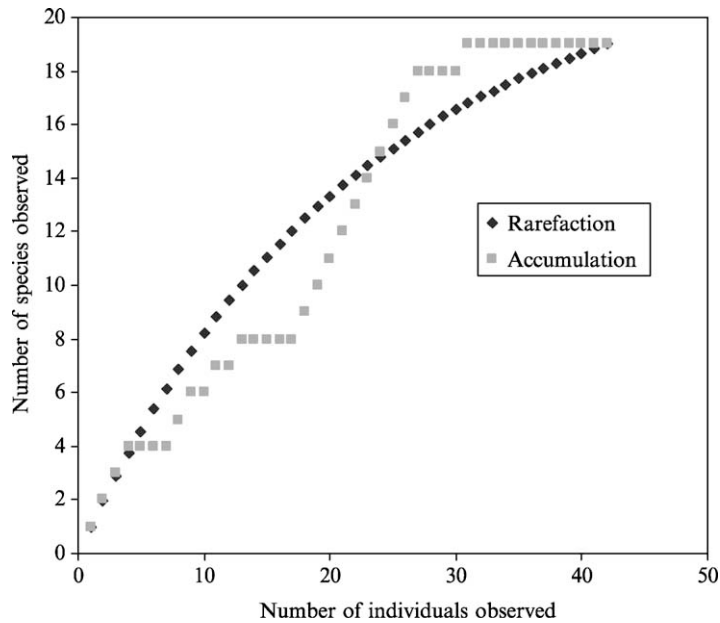
FIG. 1. An example of hypothetical individual-based accumulation and rarefaction curves. The accumulation curve is one possible order of observing the 42 clones. The rarefaction curve was created with the EstimateS program (Colwell, 2004). (See color insert.)

The curve represents the average number of species observed when n individuals are drawn with replacement from the same sample over and over (Fig. 1). In other words, it is the average of all possible accumulation curves. In the case of a clone library where an accumulation curve is an arbitrary ordering of clones, a rarefaction curve is the best way to represent data rather than a random choice of the possible accumulation curves. A rarefaction curve can be estimated by a randomization method or, in the case of an individuals-based sample, by analytic means (Coleman, 1981; Heck *et al.*, 1975).

A key feature of the rarefaction curve is the error bars around the curve. Error bars are so crucial to rarefaction analyses and, at the same time, so often misunderstood. Error bars on a rarefaction curve give a measure of variance around the average accumulation curve; specifically, they represent the variability of the number of species observed [i.e., the variability of $E(S_n)$] when n individuals are drawn from the entire sample. When one individual is drawn, the variance is zero because one species is always observed. Similarly, when all N individuals are drawn, the variance

is zero because all S species in the sample are observed. Error bars can be given as variance or standard deviation, but are most useful when reported as 95% confidence intervals (CIs). The 95% confidence limits for a given sample size are $S_{obs} \pm (1.96)^*$(the standard deviation of S_{obs}), where S_{obs} is the average number of species observed at that sample size. The 95% CIs represent the range in which 95% of all possible accumulation curves of this particular sample fall. (See later for further information about interpretation of rarefaction error bars.)

A number of software programs will perform rarefaction randomizations and/or calculate analytical formulas for rarefaction curves. In particular, EstimateS randomizes and calculates Coleman formulas and is freely available on the Internet (Colwell, 2004). Although EstimateS can be used for a number of different dataset types, the most common for microbial studies are those for which each clone represents the sampling of one individual. In this case, we find it easiest to load input data as a ''Format 3'' input file (Fig. 2).

Interpretation of Rarefaction Curves

General Considerations about Diversity Comparisons

Rarefaction analysis of species richness is just one way among many others to compare community diversity between samples. Given the variety available, one must evaluate the utility of different diversity statistics in light of the question of interest and data at hand. As mentioned earlier, even for large-scale molecular inventories of microbial diversity, data at hand are always a minute fraction of the entire community. This fact limits our ability to estimate the true richness of microbial communities with any large degree of confidence. Moreover, without knowing the correct answer, it is impossible to evaluate thoroughly the success of diversity statistics for any microbial community.

Rarefaction differs from other approaches to diversity measurement. Diversity estimators attempt to extrapolate from a sample to the true diversity of a community. Examples of estimators include Chao1 (Chao, 1984) and the Curtis *et al.* (2002) statistic based on a log-normal assumption. Estimators provide two functions: (1) to estimate true richness and (2) to compare these estimates of true diversity among samples. Rarefaction, in contrast, is performed solely for this second purpose, to compare diversity among samples. To discuss the relative benefits of rarefaction compared to the use of diversity estimators, we first discuss some general ideas behind diversity estimation.

	A	B	C	D
1	figure 1 data			
2	19	42		
3	1	1	1	
4	2	2	1	
5	3	3	1	
6	4	4	1	
7	1	5	1	
8	3	6	1	
9	1	7	1	
10	6	8	1	
11	7	9	1	
12	3	10	1	
13	5	11	1	
14	4	12	1	
15	8	13	1	
16	1	14	1	
17	4	15	1	
18	3	16	1	
19	6	17	1	
20	12	18	1	
21	10	19	1	
22	11	20	1	
23	19	21	1	
24	18	22	1	

	A	B	C	D
22	11	20	1	
23	19	21	1	
24	18	22	1	
25	17	23	1	
26	16	24	1	
27	15	25	1	
28	14	26	1	
29	13	27	1	
30	12	28	1	
31	11	29	1	
32	10	30	1	
33	9	31	1	
34	8	32	1	
35	2	33	1	
36	2	34	1	
37	2	35	1	
38	12	36	1	
39	12	37	1	
40	15	38	1	
41	16	39	1	
42	17	40	1	
43	15	41	1	
44	16	42	1	
45	-1	-1	-1	

FIG. 2. The EstimateS "Format 3" data file for example data in Fig. 1 Data are entered into three columns in a spreadsheet: an index of the operational taxonomic units (OTU), a nonrepeating index of the number of individuals, and a "1" to indicate a sample size of 1 for each colony. The title of the file must be in the first line, the number of OTU types and the number of clones in the second line, and the final line after data must be "-1,-1,-1." The file is then saved as a tab-delimited text file and is imported into EstimateS. The right-hand screen shows continuation of the columns on the left-hand side.

Any sample comparisons, whether of biomass or species richness of microbial or macrobial communities, must come to terms with three statistical parameters: bias, precision, and accuracy. Ideally, one would like an accurate estimate of species richness, a measurement that yields a very small difference between estimated richness and the true, unknown richness and a consistent estimate of that truth with every sample taken of the community (Hellmann and Fowler, 1999). Often we can only achieve a component of accuracy, either bias or precision. Bias is the difference between the expected value of the estimator (the mean of the estimates from all possible samples of the community) and the true

unknown richness of the community being sampled. This difference reveals whether the estimator consistently under- or overestimates the true richness. Precision is the variation of the estimates from all possible samples of the community. This variation represents the repeatability of the richness estimate; i.e., how similar a richness estimate is from one sample to another of the same community. Figure 3 illustrates bias and precision in terms of a dart game. A good dart thrower is accurate because she is both unbiased and precise and therefore always hits the bulls-eye. Less-talented dart throwers are biased, imprecise, or some combination thereof.

Unlike scoring a dart game between two players, evaluating diversity statistics for microbial communities is difficult because we do not know where the bulls-eye is. To test for bias, one needs to know the true richness to compare against the sample estimates. In contrast, precision is relatively easy to assess. With multiple samples, the variance of richness estimates can be calculated and compared. In some cases, an
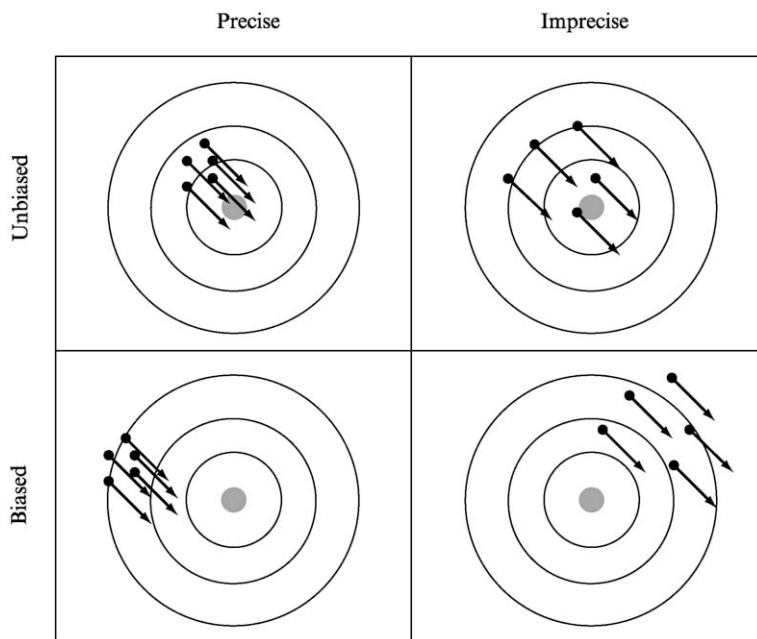
Fig. 3. An illustration of precision and bias. The bulls-eye represents a true value that is trying to be determined. Arrows are individual estimates of the true value. An ideal estimator statistic is precise and unbiased (i.e., accurate), as in the upper left-hand corner. (See color insert.)

estimator may have a closed-form variance that evaluates the precision of the estimate.

What Rarefaction Curves Do and Do Not Tell You

By comparing observed species richness, rarefaction is designed expressly to ignore the issue of bias with respect to true community diversity. Observed richness is always negatively biased, thus rarefaction curves do not say much about the true richness of a community. However, most ecological questions require comparisons of relative diversity (e.g., whether richness is higher or lower in one community or another) rather than an exact number of true richness. For these questions, a diversity statistic that is consistent with repeated sampling (is precise) but biased can be more useful than one that, on average, correctly predicts true richness but is very imprecise (Fig. 4). For relative comparisons, bias is not necessarily a problem as long as the measure is consistently biased. Thus, in theory, rarefaction might be a good approach for relative richness questions if the precision issue is addressed.
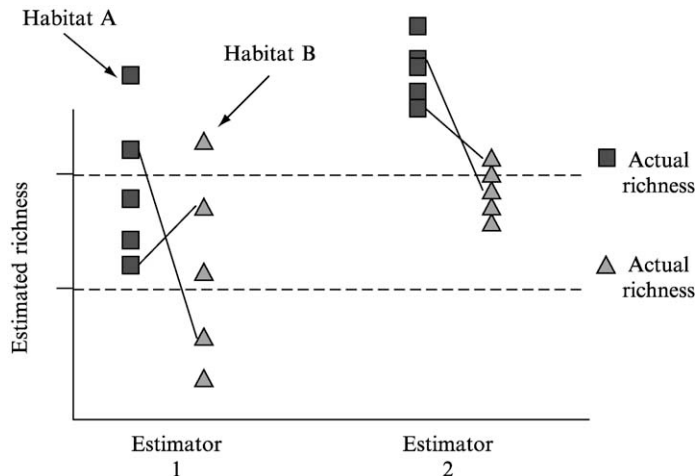

FIG. 4. An illustration of the importance of the precision of richness estimators. Estimator 1 is unbiased, but imprecise. On average, estimator 1 correctly estimates the actual richness of habitat A and B, but the variance of the different estimates is large. Thus, if one compares one estimate of richness from one sample of each habitat, it is easy to incorrectly rank the relative richness of the two habitats. Two example comparisons are shown with solid lines; one pair of estimates correctly orders the relative richness of the habitats, the other pair orders it incorrectly. In contrast, estimator 2 is positively biased, but precise. On average, the estimator overestimates the actual richness; however, because the variance in the estimates is small, any pairwise comparison of estimates correctly ranks the richness of the habitats. (See color insert.)

The most common misconception about rarefaction curves is that the confidence intervals around the curves are a measurement of the precision of $E(S_n)$ (the observed species richness for a given effort). In fact, the CIs do not say anything about the value of $E(S_n)$ if one resamples the community. Error bars only describe the variation of the accumulation curves as one reorders subsamples within the original sample. Specifically, the 95% confidence intervals represent the range in which 95% of all reordered accumulation curves will fall.

This detail, that rarefaction compares *samples* instead of communities, is crucial. A comparison of a rarefaction curve from a grassland and a rarefaction curve from a forest tells you whether the richness of the two samples (standardized for sampling effort) is significantly different (Fig. 5), not whether the richness of grassland and forest communities is significantly different. This difference is particularly important when samples represent a very small fraction of total diversity (as is the case for microbial inventories) so that different samples from the same community can have very different taxonomic representation.

Yet all is not lost for rarefaction. The precision of rarefaction curves can be addressed by sampling from multiple sites or treatment replicates (or see Colwell *et al.*, 2004). The key is to sample multiple times so that many different rarefaction curves are produced. Variation in the curves from multiple samples of a community then yield an estimate of the
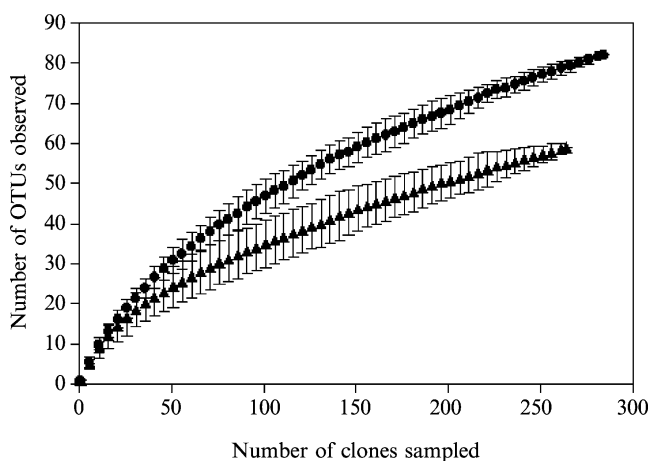

FIG. 5. A comparison of two hypothetical rarefaction curves from a sample of forest soil (●) and a sample of grassland soil (▲). Error bars are 95% confidence intervals. Curves reveal that standardizing for a common sampling effort (at 260 clones), the grassland sample has fewer operational taxonomic units (OTUs) than the forest sample.

precision of $E(S_n)$ for n individuals sampled. For instance, one could sample many grassland and forest sites and draw rarefaction curves for every sample. If the rarefaction curves of the two habitats differ consistently in observed richness for a common sampling effort n, then one could say with some statistical level of confidence whether the richness of grasslands and forest sites differs in richness at that sampling effort. (A technical note: instead of taking multiple samples from the same community, a researcher may prefer to take one very large sample and divide it up. When a large sample is divided into subsamples, the estimated precision of observed richness applies to the number of individuals of the subsamples, not the number of individuals of the total sample. In general, however, it is usually preferable to take independently replicated samples; for instance, independent samples can capture spatial heterogeneity and moderate PCR biases and errors.)

The final nuance of rarefaction is that even with good estimates of precision for rarefaction curves, the precision measure still applies only to observed richness at a particular level of sampling effort (of n individuals). This omission is not a concern if most of the richness of a community has been observed and if the rarefaction curves are asymptotically approaching the true diversity. For microbial inventories, however, sampling effort is low relative to true richness, and the curves are often still steep at the level of sampling effort. As a result, it is possible that the communities differ in their species-abundance patterns so that the rarefaction curves may cross if the sampling effort was increased (Fig. 6). Thus, even if the rarefaction curves are representative of the communities from which they are sampled, at low sample sizes they may suggest an incorrect ordering of relative diversity among communities.

The curve-crossing problem is the primary reason that diversity estimators remain useful even for relative diversity comparisons. Diversity estimators incorporate information about species' abundances in the sample in order to extrapolate true richness. In theory, they predict what happens to the rarefaction curves as one approaches sampling all individuals in the community.

The next section discusses a new technique that might help alleviate the problem of crossing rarefaction curves. However, very little is known overall about the variation in species- or taxa-abundance curves of microbial communities. Therefore, the sampling effort needed to ensure that rarefaction curves do not cross with further sampling is also unknown. Even in the absence of exhaustive surveys of real microbial communities, a few thorough simulation studies could contribute a great deal to this question.
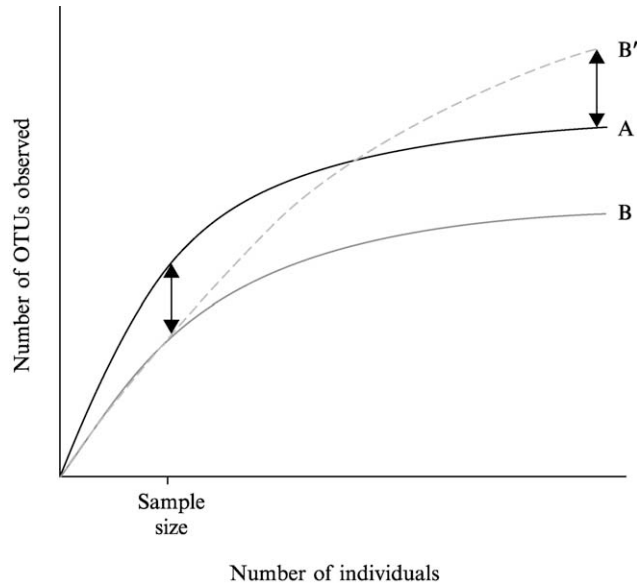
FIG. 6. Hypothetical accumulation curves for two communities: A and B. Rarified curves of the sample size indicated would suggest that OTU richness in A is greater than B. Rarefaction at this sample size cannot distinguish between curves B′ or B″, however. If B′ is the true curve, then richness in B is greater than A.

Future Directions

Rarefaction by Coverage

Cao *et al.* (2002) proposed a modification of rarefaction analysis that attempts to alleviate the problem of crossing rarefaction curves. They suggested that one should standardize by an estimate of the coverage of the sample (the proportion of true richness observed in the sample) rather than sample size.

In any comparison of two samples, samples will vary in their coverage of the communities from which they were drawn. This variation may be due in part to underlying differences between the communities' species-abundance distributions (Brose *et al.*, 2003). Take an example of two jars of colored marbles. The jars contain the same marble richness (i.e., the number of different colors of marbles), but one jar has an even distribution of colors and the other contains 90% blue marbles. If 10 marbles are drawn from each jar, one will almost certainly observe a greater marble richness from the evenly distributed jar than the blue-dominated jar. In this case, standardizing observed richness at 10 marbles falsely suggests that one jar

is richer than the other. This is because the samples vary in their coverage. A greater fraction of total marble richness is observed in a 10-marble sample from the evenly distributed jar than the blue-dominated jar.

To account for differences in sample coverage, one can estimate the coverage of the sample and compare observed richness at similar estimated coverage levels (but different sample efforts). Cao and colleagues (2002) estimated coverage by calculating the ''autosimilarity'' of a sample. [Another commonly used coverage estimator is Good's measure (Good, 1953).] Specifically, they randomly divide the sample in half and estimate the Jaccard coefficient (a similarity index) between the two sample halves. This procedure is repeated at different sampling sizes (for our purposes, the number of individuals) so that one produces a plot of autosimilarity versus individuals sampled (Fig. 7A). In other words, the procedure esti-mates coverage by asking how well one-half of the data reflects the other half of the data at a variety of sample sizes. High coverage will lead to high similarity values. Poor coverage will lead to low similarity values.

The autosimilarity curve can then be used to standardize for coverage on a traditional rarefaction curve. Figure 7B illustrates a case where two rarefaction curves are likely to cross with further sampling. The bottom dashed curve is almost linear, and the top solid curve is leveling off. Under standard rarefaction assumptions, these curves would lead to the conclu-sion that the solid-line community is more diverse than the dashed-line community. Using the Cao method to produce autosimilarity curves (Fig. 7A), one estimates that the highest common coverage of the samples is 30% (a similarity value of 0.3). The autosimilarity curves reveal the sampling effort (in terms of number of individuals) needed to standardize for this coverage value; lines dropping down the x axis in Fig. 7A estimate the number of individuals sampled so that the observed richness represents 30% of the true richness. Reading the observed richness at that sample size on the traditional rarefaction curve in Fig. 7B yields two observed richness values, standardized by coverage. By the coverage-rarefaction method, the dashed curve is now estimated to have a higher observed richness than the solid curve.

How well this method works in general for correctly predicting relative richness remains to be seen. Combined with strategies to sample multiple sites or treatment replicates, it has the potential to improve predictions of relative richness using rarefaction curves.

Rarefaction for Other Community Analyses

The problem of standardizing for sample size rears its head in diversity comparisons other than richness, particularly for comparisons of commu-nity similarity. Community similarity can be calculated by a variety of
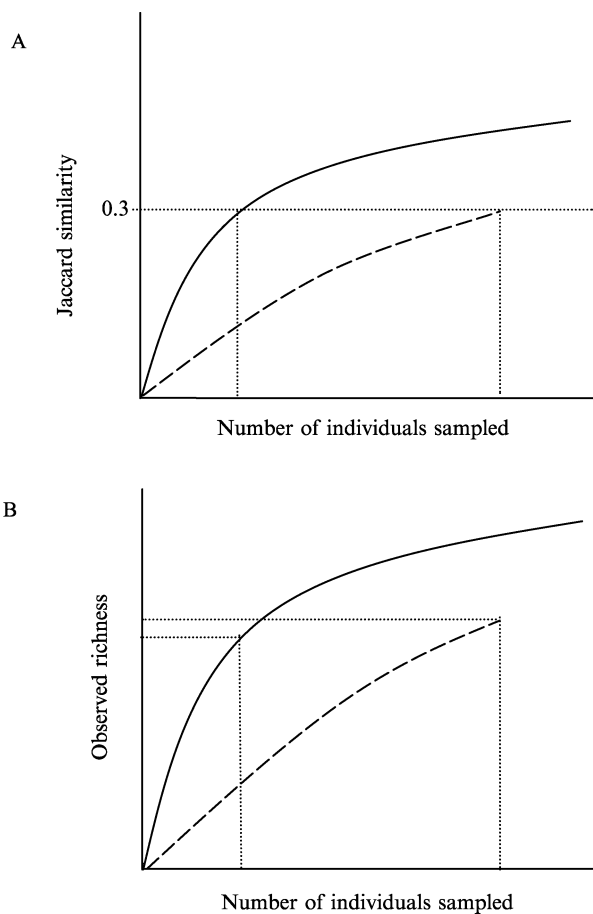
FIG. 7. (A) An autosimilarity versus sample size curve. (B) A traditional rarefaction curve. See text for explanation.

indices, including the Jaccard index, which considers only the presence or absence of taxa, and the Bray–Curtis index, which also considers the abundance of each taxon (Magurran, 1988). This section discusses how similarity indices are biased by sample size and presents a type of rarefaction analysis to account for this bias.

While most ecologists studying large organisms usually ignore the problem of effort standardization for composition comparisons among samples, the problem is especially critical in microbial studies for two reasons. First,

estimates of community similarity are highly negatively biased because microbial inventories represent such small fractions of the total community. Second, molecular inventory techniques often produce uneven number of individuals, usually sequences, between samples. Differences in sampling effort among samples will bias the estimates of true community similarity.

This issue is well demonstrated by considering two samples drawn from the same microbial community. For example, imagine two clone libraries made from the same PCR products from a community that has over 500 bacterial ''species.'' Because the samples are taken from the same community, the true similarity is 1. However, if only 10 clones are analyzed from each library, the similarity value is likely to be low. The low similarity value is not because the communities are actually different, but simply because the overall diversity of the microbial community is so high that the chance of drawing the same composition of 10 clones is very small. If sampling effort is increased and 200 clones are drawn, then the estimate of community similarity will increase toward 1.

As with richness, microbial ecologists are usually more interested in measures of relative community similarity than true community similarity. For instance, one may want to know whether the microbial community of a forest gap is more or less similar to the community of the surrounding forest or a nearby grassland. As long as the samples are standardized for sampling effort, then with a rigorous sampling design, biologically relevant comparisons such as these can be made even without knowing the true community similarities.

As an alternative to randomly throwing out data from samples that are overrepresented, one can perform a randomization procedure like that for typical rarefaction curves. This randomization procedure would estimate $E(C_{ij,n})$, i.e., an expected value of similarity (C) between samples i and j given a sampling effort of n individuals. In brief, at the highest common number of individuals sampled, one draws n individuals at random without replacement from each sample. Then, one calculates all the C_{ij} values (the similarity values between all sample pairs). These randomization and calculation steps are repeated over and over to calculate the $E(C_{ij,n})$ values.

As of yet, we know of no software that offers this type of analysis, but this is a relatively simple programming problem. Horner-Devine *et al.* (2004) used this procedure to compare 26 salt marsh samples of sediment bacteria. A completely unaddressed problem is how likely further sampling might alter relative similarity values among a group of samples; i.e., whether there is an analogous curve-crossing problem as in species richness measurement with rarefaction.

Conclusions

Rarefaction has promise as a reliable method for comparing molecular inventories of microbial communities. The method is easy to perform with freely available software and it is a quick, first-cut approach to surmise potential differences between microbial communities. As with all diversity statistics, however, users must consider carefully the limitations of rarefaction when offering interpretation from its results. First and foremost, rarefaction, like other diversity statistics, performs better as one samples a larger and larger fraction of the diversity of a community. When rarefaction curves are steep and linear, then any diversity statistic calculated from a sample is unlikely to reflect the true community. All diversity analyses are highly suspect when the community is so undersampled that the rarefaction curves are linear and steep (close to a slope of 1).

Specifically with regards to rarefaction, we offer five summary guidelines for its use and interpretation.

- Rarefaction compares observed richness among samples for a given level of sampling effort. It does not attempt to estimate true richness of a community.
- A rarefaction curve must be drawn with confidence intervals to make comparisons against other rarefaction curves. These confidence limits are essential to assess whether variation in the random order of the sampling of individuals may account for apparent differences among the curves.
- Comparisons between two rarefaction curves address whether the observed richness of the *samples* differs, not whether the richness of the *communities* from which the samples were drawn differs. This limitation holds true because the confidence intervals around a rarefaction curve do not give a measure of precision of the observed richness.
- Repeated samples from the same community can be used to estimate the precision of rarefaction and thus compare *communities* from which the samples are taken.
- Finally, rarefaction analyses on small samples do not necessarily yield the correct order of the true richness of the sample, as rarefaction curves may cross with further sampling.

Because of the last point, we recommend that rarefaction be used in concert with other diversity estimators. Diversity measures vary in their performance with respect to bias, precision, and accuracy, and each captures different qualities of a community and has unique benefits and failings (Hellmann and Fowler, 1999; Palmer, 1990; Walther and Morand, 1998). The most robust assessment of a microbial community is one that decides what it aims to accomplish (i.e., comparison of relative diversity or prediction of true diversity) and uses several indices to accomplish that goal. For the

most part, the goal of microbial ecologists is to distinguish between relative diversity measures, such as the differentiation and ordering of richness among communities or treatments. If the chosen indices give different assessments, then more sampling is necessary before strong conclusions can be made. If all measures point to the same ordering of relative diversity among samples, then the statistical interpretation is at least robust under a variety of assumptions.

References

Acinas, S. G., Klepac-Ceraj, V., Hunt, D. E., Pharino, C., Ceraj, I., Distel, D. L., and Polz, M. F. (2004). Fine-scale phylogenetic architecture of a complex bacterial community. *Nature* **430,** 551–554.

Bohannan, B. J. M., and Hughes, J. (2003). New approaches to analyzing microbial biodiversity data. *Curr. Opin. Microbiol.* **6,** 282–287.

Brose, U., Martinez, N. D., and Williams, R. J. (2003). Estimating species richness: Sensitivity to sample coverage and insensitivity to spatial patterns. *Ecology* **84,** 2364–2377.

Cao, Y., Williams, D. D., and Larsen, D. P. (2002). Comparison of ecological communities: The problem of sample representativeness. *Ecol. Monogr.* **72,** 41–56.

Chao, A. (1984). Nonparametric estimation of the number of classes in a population. *Scand. J. Stat.* **11,** 265–270.

Coleman, B. D. (1981). On random replacement and species-area relations. *Math. Biosci.* **54,** 191–215.

Colwell, R. K. (2004). EstimateS: Statistical estimation of species richness and shared species from samples. Version 7. User's Guide and application published at http://purl.oclc.org/estimates.

Colwell, R. K., Mao, C. X., and Chang, J. (2004). Interpolating, extrapolating, and comparing incidenced-based species accumulation curves. *Ecology* **85,** 2717–2727.

Curtis, T. P., Sloan, W. T., and Scannell, J. W. (2002). Estimating prokaryotic diversity and its limits. *Proc. Natl. Acad. Sci. USA* **99,** 10494–10499.

Dunbar, J., Ticknor, L. O., and Kuske, C. R. (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.

Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika* **40,** 337–364.

Gotelli, N. J., and Colwell, R. K. (2001). Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecol. Lett.* **4,** 379–391.

Heck, K. L., Belle, G. v., and Simberloff, D. (1975). Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size. *Ecology* **56,** 1459–1461.

Hellmann, J. J., and Fowler, G. W. (1999). Bias, precision, and accuracy of four measures of species richness. *Ecol. Appl.* **9,** 824–834.

Horner-Devine, M. C., Lage, M., Hughes, J. B., and Bohannan, B. J. M. (2004). A taxa-area relationship for bacteria. *Nature* **432,** 750–753.

Hughes, J. B., and Bohannan, B. J. M. (2004). Application of ecological diversity statistics in microbial ecology. *In* ''Molecular Microbial Ecology Manual'' (G. A. Kowalchuk, F. J. de Bruijn, I. M. Head, A. D. Akkermans, and J. D. van Elsas, eds.), pp. 1321–1344. Springer, Berlin.

Hughes, J. B., Hellmann, J. J., Ricketts, T. H., and Bohannan, B. J. M. (2001). Counting the uncountable: Statistical approaches to estimating microbial diversity. *Appl. Environ. Microbiol.* **67,** 4399–4406.

Hurlbert, S. H. (1971). The nonconcept of species diversity: A critique and alternative parameters. *Ecology* **52,** 577–585.

Magurran, A. E. (1988). ''Ecological Diversity and Its Measurement.'' Princeton University, Princeton.

Martin, A. P. (2002). Phylogenetic approaches for describing and comparing the diversity of microbial communities. *Appl. Environ. Microbiol.* **68,** 3673–3682.

Palmer, M. W. (1990). The estimation of species richness by extrapolation. *Ecology* **71,** 1195–1198.

Sanders, H. (1968). Marine benthic diversity: A comparative study. *Am. Nat.* **102,** 243–282.

Singleton, D., Furlong, M., Rathbun, S., and Whitman, W. (2001). Quantitative comparisons of 16S rRNA gene sequence libraries from environmental samples. *Appl. Environ. Microbiol.* **67,** 4374–4376.

Thompson, J. R., Marcelino, L. A., and Polz, M. F. (2002). Heteroduplexes in mixed-template amplifications: Formation, consequence and elimination by 'reconditioning PCR.' *Nucleic Acids Res.* **30,** 2083–2088.

Tyson, G. W., Chapman, J., Hugenholtz, P., Allen, E. E., Ram, R. J., Richardson, P. M., Solovyev, V. V., Rubin, E. M., Rokhsar, D. S., and Banfield, J. F. (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428,** 37–43.

Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., Wu, D. Y., Paulsen, I., Nelson, K. E., Nelson, W., Fouts, D. E., Levy, S., Knap, A. H., Lomas, M. W., Nealson, K., White, O., Peterson, J., Hoffman, J., Parsons, R., Baden-Tillson, H., Pfannkoch, C., Rogers, Y. H., and Smith, H. O. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304,** 66–74.

von Wintzingerode, F., Gobel, U. B., and Stackebrandt, E. (1997). Determination of microbial diversity in environmental samples: Pitfalls of PCR-based rRNA analysis. *FEMS Microbiol. Rev.* **21,** 213–229.

Walther, B. A., and Morand, S. (1998). Comparative performance of species richness estimation methods. *Parasitology* **116,** 395–405.

[18]   Culture-Independent Microbial Community Analysis with Terminal Restriction Fragment Length Polymorphism

By Terence L. Marsh

Abstract

Terminal restriction fragment length polymorphism is a polymerase chain reaction (PCR)-based technique that has been used to effectively interrogate microbial communities to determine the diversity of both phylogenetic and functional markers. It requires the isolation of community DNA and knowledge of the target sequence. PCR amplification, performed with fluorescently labeled primers, is followed with restriction digestion and