5

FUNGAL BIODIVERSITY PATTERNS

JOHN C. ZAK AND MICHAEL R. WILLIG

No individual, population, or species assemblage exists in ecological isolation, free from interactions with other organisms or populations. Rather, all organisms are integrated within a complex ecological whole, whose higher-order dynamics (i.e., biodiversity) arise from interactions among its abiotic and biotic components (Putman 1994). Consequently, our ability to detect patterns of fungal biodiversity and the mechanisms that shape them is influenced by the approaches we use to evaluate the ecological structure of fungal communities. The parameters that define a community and that are important for assessing aspects of its biodiversity include species composition, along with its functional and genetic correlates; the types and intensities of interspecific interactions and how they regulate species densities and species occurrences; and the dynamics of those attributes over time and space or as they change as a consequence of past or present human activities (e.g., agriculture, atmospheric nitrogen deposition). To truly understand patterns of biodiversity and mechanisms contributing to those patterns, one must address all aspects of community organization and dynamics. Most mycologists, however, focus on only one aspect (or at best two) of communities, thereby providing an incomplete assessment of fungal biodiversity.

Our goal in this chapter is to present analytical approaches, ecological guidelines, and appropriate references that will provide mycologists with the tools required to examine many aspects of ecological communities and, thus, permit a more comprehensive treatment of fungal biodiversity. We begin with a discussion of the types of biological information that mycologists usually collect to assess fungal biodiversity, considering as well the potential problems encountered in such work. In the next section we examine traditional approaches to quantifying biodiversity from a taxonomic perspective and provide guidelines for using the various diversity indices

and analytical procedures. We conclude our chapter with a discussion of the spatial scale of biodiversity and assess the various methods that can be used to quantify that aspect of fungal biodiversity.

THE FUNGAL UNIT

Most fungi, excluding yeasts and some zoosporic taxa, consist of filaments (hyphae) that increase in length by the deposition of cell-wall material from a growing tip. As these tips expand and produce new growing points, a network termed the mycelium develops. Once established, fungal mycelia are capable of essentially unlimited growth and persistence. This form of indeterminate body structure differs significantly from the determinant body plan of most animal and many plant species. Indeed, as a mycelium continues to expand, it may come to occupy a heterogeneous suite of microenvironments or macroenvironments, but its distant segments remain interconnected, facilitating intercellular communication. Because these heterotrophic organisms can occupy heterogeneous environments, they convert organic material and nutrients into biomass at spatial scales that range from several millimeters to entire landscapes (Smith et al. 1992; Anderson et al. 1994).

The mycelial nature of most fungi affects the definition and interpretations of fungal biodiversity and makes the protocols and assumptions used for estimating fungal biodiversity inherently different from those used for most animals and nonclonal plants. To understand the limitations associated with estimations of fungal biodiversity, one must consider methods for counting fungal units and understand how those units differ from units used for most macroorganisms.

The term individual can be used in a numerical, genetic, or ecological context, depending on the focal organism and question of interest (Andrews 1991). From a numerical perspective, an individual is a countable unit of a particular species (e.g., a tree, a mushroom, or a colony that develops from substratum placed on an agar plate). According to Cooke and Rayner (1984), individuals are discrete and functionally independent units. The level of cellular aggregation that fulfills that criterion for fungi and other microorganisms, however, is not clear (Andrews 1991). For example, isolates of a single species of fungus obtained from decomposing leaves at spatially disjunctive locations on the forest floor may represent the mycelium of a single individual or the mycelia of several individuals. As a result, counting individual fungal colonies on an isolation plate, whether the isolate was obtained from a 1-mm^2 piece of substratum or from a dilution series, does not represent the same information obtained from counting the insects that occur within a defined area or in a given amount of leaf litter. Usually, one cannot be sure if fungal isolates of a species from a single habitat have been obtained from one individual or from several individuals. Although the latter situation is common, the former circumstance also occurs.

A genetic perspective does not provide an operational resolution to this problem. Genetically, an "individual" can be defined as a single cell or a collection of cells that exhibit the same genotype or multiple genotypes (i.e., a genet *sensu* Kays and Harper 1974). In addition to being an individual from a quantitative perspective, a squirrel is an individual from a genetic perspective. In contrast, the number of discrete, countable units in fungi is not the same as the number of genets (Andrews 1991). Fot clonal organisms that are capable of asexual growth, the countable units are the ramet (Harper 1977). A ramet, although a member of a specific genet, is capable of essentially independent growth. Thus, sporocarps of one species of basidiomycete on a forest floor can represent multiple ramets of a single genet or the ramets from multiple genets. Only molecular or isozymic analyses of sporocarps can untangle the genetic structure of such a species within a habitat and, thereby, facilitate the quantification of fungal biodiversity in the same manner as for animals and most plants. Unfortunately, that approach is not generally practical. Thus, a clearly stated, operational definition of "individual," relevant to the taxon of interest, must be provided to facilitate unambiguous comparisons among ecosystems.

The presence or absence of species within an area can be used to assess some aspects of biodiversity; other comparisons, however, require the quantification of operational individuals (or units) for each taxon. The operational definition for an individual will differ for different taxonomic or ecological groups of fungi. Those definitions are provided for particular groups in their respective chapters.

CHARACTERIZATION OF COMMUNITIES

Collection Effort Curves

Many, if not most, investigations of populations and communities describe patterns rather than test *a priori* hypotheses. They rely on the collection of presence-absence, biomass, density, or frequency of occurrence data for a suite of species. Although the individual is usually the smallest biological unit for quantitative population and community indices, because of the

fungal growth form, the concept of individual as used for animals and plants does not apply (see "The Fungal Unit," earlier). Population and community metrics are designed, however, to accommodate data that are not based solely on the delineation of individuals. For analyses of fungal biodiversity, each investigator must establish a unit by which presence-absence, biomass, density, or frequency of component species can be estimated over space and time. It is critical that biomass, density, or frequency of occurrence be based on the same unit of collection and obtained in the same manner throughout the spatial and temporal domain of the investigation.

Frequencies of occurrence of fungi can be calculated in two ways:

$$\text{Frequency (\%)} = \frac{\begin{array}{c}\text{sample units in (or on)}\\ \text{which fungal}\\ \text{species occurred}\end{array}}{\begin{array}{c}\text{total number}\\ \text{of sample units}\\ \text{examined}\end{array}} \times 100 \qquad (1)$$

For example, if 50 1-mm × 1-mm particles of organic soil matter are plated onto an agar medium, the total number of samples examined is 50. The maximum number of sample units from which any one species can be isolated is also 50, which would yield a frequency of occurrence of 100% for that species. Note that the sum of the frequencies of all species can exceed 100% when calculating frequency of occurrence in this manner. Standardized frequencies, which sum to 100%, may be desirable. Dividing the frequency of each species by the sum of the frequency of all species provides relative frequency, a straightforward standardization.

$$\text{Relative Frequency (\%)} = \frac{\begin{array}{c}\text{number of isolates}\\ \text{for each species}\end{array}}{\begin{array}{c}\text{total number}\\ \text{of isolates}\end{array}} \times 100 \quad (2)$$

Investigators should present average frequency of occurrence and standard error of the mean for each species when sampling has been replicated. Standard error provides an estimate of the amount of variation that is associated with the mean frequency of each species. Often, the variability among species in frequency conveys more insight into mechanisms that are responsible for occurrence patterns in fungal ecological units than do the means alone.

The choice of sampling design for obtaining information on presence-absence, biomass, density, or frequency of occurrence is determined by the type(s) of fungi under investigation. The chapters in Section II of this volume provide recommended operational definitions of an individual for the different ecological groups treated.

Determination of the most effective size and shape for a sampling plot is based on the total time required to erect and characterize individual plots as well as the variability of the resultant data. The investigator should have some *a priori* knowledge of appropriate plot size and shape so that the range of sizes to be evaluated is not excessive. Practicality is also an important concern when determining plot size. A growing body of evidence (Kolasa and Pickett 1991; Waide et al. 1999; Gross et al. 2000) suggests, however, that the patterns and processes evaluated by a study are scale-dependent, so considerable caution must be exercised when comparing the results of studies based on plots of different size. Likewise, the spatial scale at which data are collected should be related to the spatial scales at which causative mechanisms are thought to operate.

The shape of a plot can influence the ease, accuracy, or efficiency of sampling. In grasslands, shrublands, or even forests, circular plots are delineated easily and quickly with a center pole and a freely swinging radius line. Circular plots have the advantage of minimizing edge effects that can strongly bias estimates of density. Krebs (1989), however, suggested that elongated or rectangular quadrats provide more accurate estimates of the species composition of an area than do circular or square plots of the same area. The rectangular plot is especially suited for assessing environmental gradients within a habitat when the long axis is oriented parallel to the underlying gradient of interest (Cox 1996). The reason is that rectangular quadrats are more effective in detecting habitat heterogeneity and accurately estimating the patchy distribution of organisms within a sampling scheme than are square or circular quadrats. A long quadrat potentially covers more patches than does a circular one of the same area.

Determining the number of sampling units required to provide an accurate estimate of species richness or density is always challenging, whether one is sampling macrofungi fruiting on a forest floor or isolating microfungi from decomposing root pieces. The number of plots assessed or particles of material plated onto an agar medium will depend on the characteristics of the fungal assemblage, the objectives of the investigation, and the magnitude of differences considered to be biologically relevant. In reality, sample size often represents the upper limit of effort that can be expended to categorize an ecosystem or the amount of time necessary to culture and identify isolates on an agar medium. Regardless, the investigator must avoid making decisions concerning sample size in a capricious manner and always should expose the criteria on which the decision was based.

If the biological question of interest concerns the total species richness of an area or domain (in practice this could be a leaf, a patch of soil, or a forest), then a random (or stratified random) series of plots should be placed within the inference space (spatial and temporal limits of the domain) to determine the number of samples required for an accurate estimate of richness. In general, the total number of species identified from the area of interest increases as the number of samples increases but eventually attains a plateau (i.e., asymptote). The number of samples required to reach the plateau will depend on the environmental heterogeneity of the area of interest as well as on the dispersion patterns and fruiting phenology of the focal species. Hence, if one is interested in assessing if mean differences in species richness of fungi on leaves (i.e., the richness of a leaf) occur between shade and sun leaves of oak trees, for example, then collector's curves should be generated for each type of leaf. Because even different leaves of the same type can differ in heterogeneity and species composition, collector's curves should be generated for a number of leaves of each type, and, to be conservative, sample sizes should be chosen beyond the minimal value of effort that corresponds to the plateau. Alternatively, one can assume that the relation between richness and effort is logistic and, via statistical techniques, can associate parameters with variables in the logistic equation and graphically determine a conservative value for sample size associated with the plateau (Fig. 5.1 upper graph). A number of alternative methods (e.g., rarefaction analysis; Magurran 1988; Colwell and Coddington 1994) exist for estimation of species richness and should be considered prior to choosing a particular experimental design. Because human and financial resources are often too limited to enable discovery of species richness at a site, quantitative ecologists have developed methods to estimate total richness from a modest amount of sampling. Colwell and Coddington (1994) discussed some of these methods, and a subset of them is treated as part of "Approaches to Sampling Macrofungi" in Chapter 8 of this volume.

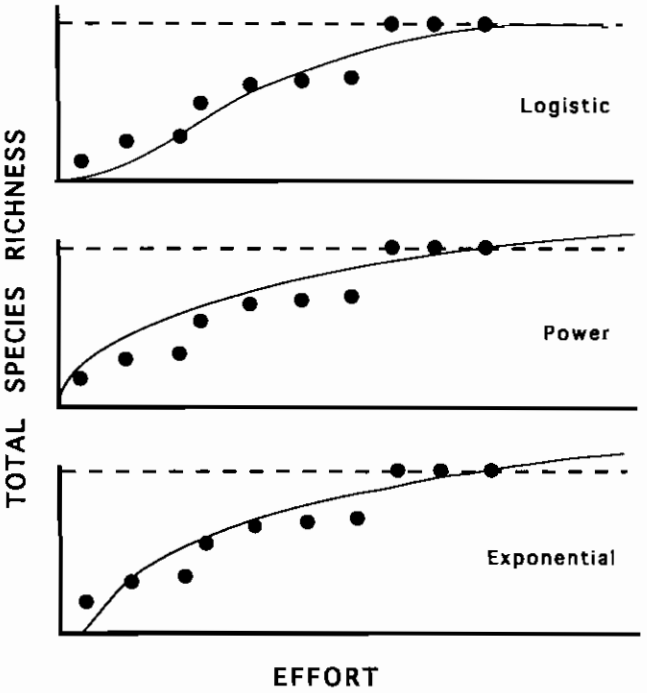

FIGURE 5.1 Graphic representations of the logistic, power, and exponential functions (*solid lines*). Dots represent the way in which total or cumulative species richness increases with effort (cumulative number, area, or volume of samples). The logistic relation approaches an asymptotic value, which is represented by the dashed line, that is equal in value to those of the power and exponential relations. With both the power and exponential curves, richness increases continuously without attaining a plateau.

Quantitative Indices

Richness

Species richness is the most widely used parameter for evaluating aspects of fungal biodiversity. It is deceptively simple to define species richness as an enumeration of the species that are associated with a particular sample, area, habitat, or substratum. In fact, three kinds of species richness can be distinguished: (1) numerical species richness, (2) species density, and (3) total species richness (Hurlbert 1971; Kempton 1979; Brown 1995; Rosenzweig 1995). The number of species in a sample in which the biomass or number of individuals has been standardized is numerical species richness. The number of species in a sample in which the area, volume, or weight of the sampling unit has been standardized is species density. Finally, the cumulative number of species based on a series of samples from a habitat or substratum is the total species richness.

Numerical species richness and species density are defined after complete enumeration of the taxa in a sample, whereas total species richness is estimated from a series of samples. The first two are measured without error, assuming that the size of the sample is sufficiently small and that the techniques of isolation and identification are developed sufficiently to allow an investigator to enumerate all taxa. It is important to note that the effects of natural or experimentally induced variation in environmental characteristics on species density and numerical species richness are scale-dependent—that is, the effect of environmental variation may not be the same for samples that differ in area, volume, weight, biomass, or number of individuals (Waide et al. 1999; Gross et al. 2000). In part, the scale-dependence of patterns occurs because the importance of different causal mechanisms

depends on the scale at which data are collected. Most ecological studies of fungi cover species densities as a consequence of sampling design, although in most cases that aspect of diversity is not stated explicitly. We strongly recommend that future research always include an explicit definition of scale as well as of the attribute of richness that is being evaluated. Once a particular spatial scale has been selected for an investigation in which species density or numerical species richness is the characteristic of diversity under examination, classical statistical considerations of random sampling and power determine the efficacy of a research design.

It would appear that an unambiguous and straightforward index of total species richness, S, would be the cumulative number of unique species present in a series of samples. The magnitude of S, however, depends on the size, number, and dispersion of samples in a particular habitat, biome, or area. Indeed, three mathematical relations (Power, Exponential, and Logistic) have been championed in the literature to predict the way in which S increases with effort, A (nnmber, area, or volume of samples). In both Power ($S = CA^z$; Arrhenius 1921) and Exponential ($S = C + z\ln A$; Gleason 1922, 1925) models, S increases monotonically as effort increases (Fig. 5.1 middle and lower graphs). Those functions are most appropriate when heterogeneous landscapes are sampled and increases in effort result in increases in the heterogeneity of habitats that are inclnded in samples. All three relations are members of the same family of curves (He and Legendre 1996). Consequently, both Power and Exponential models figure prominently in the theory and practice of island biogeography and conservation biology. In contrast, when the domain of interest is circumscribed geographically and random sampling occurs within the borders of that domain, the logistic relation ($S = B/(C + A)^{-z}$; Archibold 1949) more likely characterizes the manner in which S increases with effort. Unlike the other two species-effort curves, the logistic relation predicts that S eventually will reach a plateau or asymptote (Fig. 5.1 upper graph). The value of S at this asymptote is an accurate estimate of the true species richness of the domain (inference space) of interest. The effort required to attain asymptotic values for particular taxa, however, is likely specific to particular substrata, habitats, or biomes. Consequently, comparisons of S at levels of effort not associated with the asymptote can lead to spurious conclusions.

Given the dependence of S on collection effort and the fact that limited financial resources, personnel, or logistic support often prevent the collection of samples sufficient to attain asymptotic values, S is of limited value as a comparative index (Ludwig and Reynolds 1988). Consequently, a number of indices that are independent of the number of samples taken have been developed to estimate species richness. Those indices are based on the relationship between S and n, the total number of individuals in the collection of samples. Ludwig and Reynolds (1988) also cautioned that two well-known species-richness indices, the Margalef index (1958) and the Menhinick index (1964), make specific assumptions concerning the relationship between S and n ($S = kn^{0.5}$, where k is a constant). In many cases, those assumptions may not hold, and as a consequence, the utility of the indices is limited.

Direct counts of species numbers in samples of equal size (i.e., equal number of individuals) may provide an informative alternative to indices of species richness. Implicit in this approach, however, is the assumption that the collector's curves for the treatments of concern are coincident (or at least nonintersecting). For enumeration of fungi from leaf disks or root surfaces, the approach has merit. However, it is not always possible to ensure that all sample sizes are equal. In situations in which the sample size is not constant, rarefaction, a quantitative method (Magurran 1988), facilitates comparison of species richness among areas or habitats as if they were based on a standardized sample size. The number of species that can be expected in a sample of n individuals is:

$$E(S) = \sum \left\{ 1 - \left[\left(\frac{N - N_i}{n} \right) \Big/ \left(\frac{N}{n} \right) \right] \right\} \qquad (3)$$

where $E(S)$ is the expected number of species in a rarified sample, n is the standardized (rarified) sample size (usually chosen to be equal to the smallest sample available for an area or habitat), N is the total number of isolates (individuals) recorded in the set of samples, and N_i is the number of isolates (individuals) in the species. To calculate the expected nnmber of species in a rarified sample, the abundance of each species is inserted into the quantity defined by braces ({}) in formula (3) and summed to provide the expected number of species. To assist in the computation recall that:

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} \qquad (4)$$

where the exclamation point (!) indicates a mathematical operation termed a factorial. Worked examples can be found in Magurran (1988) and Krebs (1989). Polishook and colleagues (1996) used rarefaction to determine the expected number of species of fungi from decaying leaves from a Puerto Rican rain forest.

Diversity Indices

Diversity is a measure of the complexity of structure in an ecological community. It comprises two distinct

TABLE 5.1
Common Indices of Diversity and Evenness

Index	Equation	Terms and comments
Diversity		p_i is the proportion of individuals (isolates) in the ith species. Proportional abundance is
Simpson	$\lambda = \Sigma p_i^2$ $D = \sum \frac{n_i(n_i-1)}{N(N-1)}$	obtained by dividing the density, biomass, or number of plots in which the organism was observed by the toral density, biomass, or observations. Used for infinitely large ecological units. n_i is the number of individuals (isolates) in the ith species, and N is the total number of isolates. The reciprocal form of Simpson's index ($1/D$) usually is presented, ensuring that the index increases with increasing diversity. This index is appropriare for sampling from finite ecologic units.
Shannon	$H' = -\Sigma\, p_i \ln p_i$	p_i (n_i/N) is the proportional abundance of the ith (each) species. ln is the natural log (base e).
McIntosh	$U = \sqrt{(\sum p_i^2)}$	U is the general form of the index; p_i is the proportional abundance of each species.
McIntosh	$D = (N - U)/(N - \sqrt{N})$	D is a dominance measure and is independent of N (total number of isolates).
Berger-Parker	$d = N_{max}/N$	N_{max} is the number of isolares iu the most abundant species. The reciprocal of d is used most commonly.
Hill	$N_1 = e^{H'}$	H' is Shannon diversity based on the natural log (base e).
	$N_2 = 1/D$	D is Simpson diversity.
Log-seriesd α	$\alpha = \frac{N(1-x)}{x}$	N is the total number of isolates. S is the total number of species. x is estimated from the iterative solution of: $S/N = [(1 - x)/x][-\ln(1 - x)]$
Evenness		
Shannon	$E = H'/\ln S$	S is species richness.
Hill	$E = N_2/N_1$	N_1 is Shannon Diversity Index; N_2 is Simpson Diversity Index.
McIntosh's	$E = \frac{N-U}{N-(N/\sqrt{S})}$	N is the total number of isolates. U is McIntosh's diversity index. S is number of species.

attributes: species richness and species evenness. Simpson (1949) proposed the first index of diversity used in ecology (Table 5.1). The index varies from 0 to 1 and is referred to as a dominance measure because it is influenced strongly by the abundance of the most common species. Originally, Simpson's index (λ) was restricted to ecological units in which all members of the unit (in this case community) could be enumerated (Table 5.1). Because we usually sample infinite (in the statistical sense) ecological units for which it is impossible to count all members, Simpson developed an unbiased estimator (D) of diversity based on a sample of N individuals (Table 5.1). Significant differences in Simpson's diversity indices can be tested using parametric analyses (e.g., analysis of variance, regression) or their nonparametric counterparts.

The most widely used measures of diversity are the indices derived from information theory. Those indices are based on the rationale that the diversity of a biological system can be measured in a way similar to that used to estimate the information content of a message. The Shannon Index of Diversity (H') (Shannon and Weaver 1949) is currently the most popular index in community ecology (Table 5.1). The index has sometimes incorrectly been referred to as the Shannon-Weaver index (Krebs 1989). H' is a measure of the average degree of uncertainty in predicting the specific identity of an individual chosen at random from a collection of S species and N individnals. Average uncertainty will increase as the numbers of species increases and as the distribution of individuals among species becomes more even. Ludwig and Reynolds (1988) emphasized two properties of the index that make it popular. First, $H' = 0$ if (and only if) the sample includes only a single species. Second, H' reaches its maximum only when all species are equally abundant. The magnitude of H' is usually between 1.5 and 3.5 and is rarely greater than 4.5 (Margalef 1972). May (1975) calculated that to obtain a value of H' of more than 5.0, the ecological unit would need to include 10^5 species. Fortunately, when the Shannon index is determined for replicate samples of the same ecological unit, it exhibits a normal distribution, facilitating the use of parametric statistics to evaluate differences in central tendency (mean) or dispersion (variance) in different habitats (Taylor 1978).

Other useful diversity indices include McIntosh's measure of diversity (U), the Berger-Parker index (d),

TABLE 5.2
Characteristics of Selected Diversity Indices*

Index	Ability to discriminate among sites	Sensitivity to sample size
Shannon	Moderate	Moderate
Simpson	Moderate	Low
Mcintosh (U)	Good	Moderate
Mcintosh (D)	Poor	Moderate
Berger-Parker	Poor	Low
Fisher	Good	Low

* From Magurran (1988).

and Fisher's log-series alpha (α) (Table 5.1). The McIntosh diversity index is easy to calculate and reflects the Euclidean distance of the sample point (in terms of the abundances of all species) from the origin of an S-dimensional hypervolume. The index, however, is influenced strongly by sample size, limiting its usefulness. Alternatively, the Berger-Parker index reflects the proportional importance of the most abundant species (Table 5.1). This index is independent of S but also is influenced by sample size. Fisher's logarithmic series was the first model to describe mathematically the relationships between the number of species and the number of individuals that compose those species (Fisher et al. 1943). Although the model has been used extensively to examine the distribution patterns of individuals among species within an ecological unit, one parameter of the distribution, alpha (Table 5.1), can be used as an index of diversity (Taylor 1978; Magurran 1988). Taylor (1978) strongly supported the use of the log-series alpha as an index of diversity because of its good discriminant ability. It is not overly sensitive to sample size (Table 5.2), and it is less affected by the abundances of the common species than either the Shannon ot Simpson Index. A disadvantage of log-series alpha is that it is unaffected by the actual distribution of individuals (evenness) among the component species in the community or assemblage (it assumes they adhere to a log-series distribution). Thus, two communities that have the same N and S, but different distributions of individuals among the species in a sample, will have the same alpha value.

Lastly, Ludwig and Reynolds (1988) proposed that the series of diversity indices (N_0, N_1, N_2) presented by Hill (1973) are probably the easiest to interpret ecologically (Table 5.1). The units for these indices are numbers of species, and the indices measure the effective number of species present in a sample. N_0 is the number of all species in the sample, regardless of abundance, whereas N_1 and N_2 measure the numbers of abundant and very abundant species, respectively.

The Brillounin Diversity index (HB) was developed to calculate the diversity of an ecological unit when sampling is not random and when the enumeration of the ecological unit is complete (i.e., all discrete individuals are counted) (Magurran 1988). Giveu the restricted conditions under which the index cau be used (cannot be used with biomass or cover data), the complexity of its calculations, and the number of acceptable alternative indices, we do not recommend that it be used to estimate fungal diversity.

A technique termed jackknifing can be used to improve the accuracy of any diversity index. Zahl (1977) first used this technique to obtain confidence limits for the Simpson and Shannon diversity indices. The method makes no assumptions about the underlying distribution but allows one to test hypotheses statistically. Moreover, Zahl (1977) showed that calculated pseudovalues (see later in this chapter) are normally distributed and that random sampling, which might be difficult to achieve, is not required. The technique can be applied only in studies in which a number of samples have been taken to estimate a diversity index (Magurran 1988). The method involves recalculating overall diversity while disregarding data from each of n constituent samples. This creates a number of jackknifed values (VJ_i), equal to the total number of samples. Each of the jackknifed estimates then is converted to a pseudovalue (VP_i):

$$VP_i = (nV) - [(n-1)(VJ_i)] \tag{5}$$

where n is the number of samples and V is the diversity index based on all samples. The mean of the pseudovalues represents the best estimate of diversity (VP), and the sample influence function is the difference between the jackknifed estimate and the original estimate of diversity. Confidence limits can be calculated as follows:

$$(1-\alpha)\,\mathrm{CI} = \text{mean} \pm t_\alpha\,(\text{standard deviation of } VP_i/\sqrt{n}) \tag{6}$$

where t_α is a value obtained from the critical values of the t distribution for the desired confidence level ($1 - \alpha$) with $n - 1$ degrees of freedom (for t-table values, see Rohlf and Sokal 1995). No confidence intervals should be calculated for datasets in which n is less than 15 (Adams and McCune 1979). Small datasets also may result in overestimation of diversity, and it may be desirable to set degrees of freedom equal to $S - 1$ to provide a more accurate estimate. For large datasets ($n > 100$) and species-rich ecological units (S more than 100), t_{S-1} and t_{n-1} are essentially equal, and use of either $n - 1$ or $S - 1$ degrees of freedom will provide similar results. Magurran (1988) urged caution in using jackknifed

estimates for Shannon and Simpson indices because the technique can produce results that are ecologically absurd.

Little consensus exists on which diversity index gives the best estimate in all situations. Ludwig and Reynolds (1988) favored the use of Hill's (1973) indices, whereas Rosenzweig (1995) recommended the use of Fisher's log-series alpha. Magurran (1988) discussed the sensitivity of common indices to sample size and their ability to discriminate subtle differences between sites; that information is summarized in Table 5.2. Investigators should be aware of the limitations of the diversity measure chosen and of its biological relevance.

Evenness

Diversity indices combine two distinct aspects of a community: the number of species and the degree to which individuals are distributed in an equitable fashion (evenness) among species. Just as species richness can be calculated for any ecological unit, each diversity index also has a corresponding evenness index (Table 5.1). More specifically, indices of evenness measure the degree to which a particular community evinces the maximal diversity possible, given the observed richness—that is,

$$e = \text{Diversity}_{obs} / \text{Diversity}_{max} \qquad (7)$$

where D_{max} is calculated with each taxon assuming $1/S$ of the total abundance.

Species-Abundance Distributions

A limitation of species diversity indices is that they compress the data to a single value that conveys little information about the abundances of the species in the ecological unit. One can examine aspects of the diversity of fungal assemblages, based not only on species composition and richness but also by evaluation of how abundances are partitioned among the component species. Using species-abundance data, one can begin to examine how diversity and the structure and organization of the community are related.

Fisher and colleagues (1943) were the first to recognize that when the number of species and the relative abundances of each species within a community are plotted, a characteristic pattern of species abundances is obtained. These observations led to the development of species-abundance models. Preston (1948) showed that for assemblages of plants or animals, abundance distributions often were described by lognormal functions. Although species-abundance data may adhere to one or more distributions (Pielou 1975), diversity usually is examined in relation to four main species-abundance models: the geometric series, the logarithmic series, the broken-stick model, and the lognormal distribution (Table 5.3). May (1975, 1981) and Southwood (1978) strongly advocated the use of such models as the only sound basis for examining species diversity. Magurran (1988) concurred and emphasized that species-abundance distributions use all available data in quantifying the structure and organization of the ecological unit. Consequently, they provide the most complete assessment of diversity. These distributions can be described mathematically, and perhaps more importantly, each corresponds to specific biological scenarios that account for the form of the distribution (Table 5.3).

Rank-Abundance Plots. Rank-abundance plots are the usual graphic method for presenting species-abundance

TABLE 5.3
Ecological Characteristics of Four Commonly Used Species-Abundance Distributions*

Geometric	Log series	Lognormal	Broken-stick
Abundances proportional to amount of resources utilized	Abundances governed by one or a few factors	Abundances governed by many independent factors	Abundances governed by random division of resources along a continuum
Dominant species preempts largest portion of limiting resource	Dominant species preempts largest portion of limiting resource	Resource utilization characterized as multidimensional	Subdivision of niche space is random and simultaneous
Associated with species-poor habitats	Describes a small assemblage of species	Large and varied communities	Small samples of taxonomically related organisms, with stable populations and long life cycles
Propagules arrive at regular intervals	Propagules arrive at random intervals	—	—
Nonequilibrial assemblages	Nonequilibrial assemblages	Equilibrial assemblages	—

* From May (1975) and Magurran (1988).

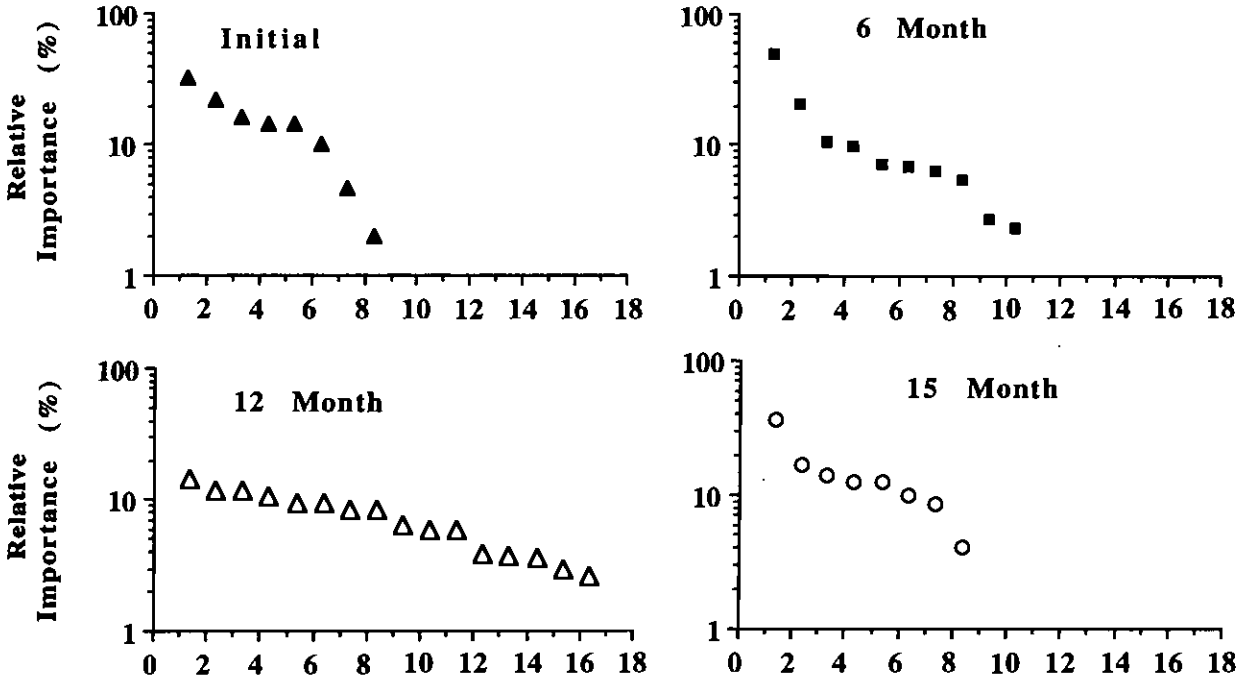

FIGURE 5.2 Abundance distributions for fungal species decomposing oak roots in West Texas at four different times (Zhang 1996). The y-axis (log scale) represents the relative importance of each species in the assemblage as a percent, which was obtained by dividing the number of isolates of each species by the total number of isolates for that particular sample period and multiplying by 100 (see equation 2). The sum of all relative-importance values for each distribution should equal 100. Each species subsequently is ranked from most important to least important along the x-axis, with rank 1 being the most important. The log-series best describes the abundance distributions for all time periods except the 12-month sample, which is best described by the geometric model.

distributions. In these graphs, the abundance or relative importance (expressed as percentages: $P_i = ([n_i/N] \times 100)$ of each species in a collection is plotted on a logarithmic scale against species rank from most abundant to least abundant (Fig. 5.2). Relative importance values are preferred because they allow direct comparison of ecological units with different total numbers of individuals. The four abundance models represent a progression from the geometric series, in which a few species are dominant with the remainder rare, through the lognormal distribution in which most taxa are of intermediate abundance, to the broken-stick model in which species abundances are more equitable.

Geometric Series. The geometric distribution model, sometimes called the niche preemption hypothesis, represents a situation in which the most successful or dominant species preempts some proportion of the total limiting resource. The next most successful species claims the same proportion of the remaining resource, and so on. In general, this model produces the highest dominance and lowest evenness indices compared to other models. The geometric distribution model is typical of species-poor habitats (Zak 1988, 1992) or those that contain stress-tolerant species. In addition, the geometric series can arise when the arrival of species into an unsaturated habitat is regular and subsequent rates of increase are homogeneous (Magurran 1988).

Log-Series. The log-series model is related closely to the geometric model (May 1975). Like the latter, the log-series occurs when one or a few factors control species dominances in an ecological unit, resulting in only a few taxa becoming dominant. The log-series also can arise if the intervals between arrivals of species into a habitat are random rather than regular and when growth rates are homogeneous (May 1975). Thus, the species-abundance distribution of phylloplane fungi growing on *Lolium perenne* can be described by both the geometric and log-series distributions (Thomas and Shattock 1986). Zak (1988) reported that the geometric and log-series models best described the structure of rhizoplane fungal assemblages on a grass growing on reclaimed mine tailings and that the logarithmic species-abundance distribution may be typical of root-surface fungal assemblages and indicative of a nonequilibrial ecological unit. More recently, Polishook and colleagues (1996) examined the species-abundance distributions of

microfungi isolated from decaying leaf litter in a Puerto Rican rain forest. They found that the fungal assemblages were primarily composed of rare species, with only a few moderately abundant and a few highly abundant taxa.

Lognormal. In the lognormal model, occupation of niche space is determined by a number of interacting factors that affect the outcome of interspecific competition. The model is characteristic of ecological units that contain many species (Table 5.3). Lussenhop (1981) reexamined dilution plate data from Wisconsin forest soils as well as from sixteen studies of rhizosphere fungal assemblages. He found that in all cases, the lognormal model best described the fungal species-abundance distributions. Lussenhop (1981) emphasized that if fungi are collected from many microhabitats and mixed during sampling, the resultant distribution is an artifact and will be biased, resembling the lognormal. Nonetheless, he proposed that of the 31 fungal datasets examined, adherence to the lognormal distribution was not an artifact of sampling but the result of a number of independent environmental factors controlling species abundances. Zak (1992) cautioned, however, that in all studies examined by Lussenhop (1981), incomplete fungal abundance data were analyzed. Some rare species were not included in the lists, some taxa were grouped into genera with others identified to the specific level, or all rare taxa were pooled into a single group. When plotting species-abundance distributions, taxonomic units must be applied in the same manner for all ecological units. Therefore, sterile isolates of fungi obtained from litter or soil should be sorted into morphological species groups based on macromorphological and micromorphological characteristics expressed when grown on several agar media. Approaches for fitting each of the four species-abundance distributions are presented in detail along with examples by Magurran (1988). Observed and expected values for each model can be compared using a Chi-square goodness-of-fit test or a G-test (Sokal and Rohlf 1995), thereby determining if a given dataset conforms to model predictions. In addition, these two goodness-of-fit tests can be used to determine whether differences in the proportional abundances of species at different sites are statistically significant.

Spatial Scale of Biodiversity

Whittaker (1977) was the first to realize that ecological diversity was scale dependent or hierarchical in nature. The primary level of diversity, point diversity, reflects the diversity at a particular location. Alpha diversity, sometimes called within-habitat diversity, is the diversity in a patch (MacArthur 1965). Most studies calculate alpha diversity using the indices described previously. Beta diversity describes the contribution of multiple habitats to overall diversity of a site. Gamma diversity represents the number of species within a region or landscape. The final level envisioned by Whittaker, epsilon diversity, is the diversity of a large biogeographic region, such as a biome.

Such a hierarchical arrangement of diversity could be used to describe the fungi that occur on decaying leaves on the forest floor. For example, point diversity would represent the number of fungal species associated with a leaf of a particular species of tree. Alpha diversity would represent the fungi associated with multiple samples of the same leaf type, whereas beta diversity would be obtained by comparing fungal species richness and composition from different leaf types within the same forest. Gamma diversity would be estimated by comparing fungal richness and composition of several forested locations, and epsilon diversity would be obtained by comparing fungal species composition in the deciduous forests across a large geographic area, such as eastern North America.

Differentiation Diversity

Another way to think about beta diversity is to view it as a measure of the degree of similarity or difference in species composition between sites. In other words, beta diversity examines the degree of species turnover as one moves from habitat to habitat, from community to community, or along any ecological gradient (Southwood 1978). The fewer species the various sites or positions along a gradient share, the higher the beta diversity. The term was coined by Whittaker (1977), who devised three levels of differentiation diversity (pattern diversity, beta diversity, and delta diversity). Pattern diversity is the differentiation diversity between samples taken within a homogeneous habitat, and beta diversity is the between-habitat component of diversity. Changes in species composition and abundance between landscape areas are considered delta diversity. Of the three types of differentiation diversity, beta diversity is by far the metric most commonly used to examine the degree to which turnover in species composition characterizes positions along gradients.

The six commonly used metrics of beta diversity (Table 5.4) use presence-absence (binary) data. Each measure has been evaluated regarding sensitivity to community changes, additivity, independence from alpha diversity, independence from excessive sampling, and so forth (Wilson and Shmida 1984). Of the six indices,

TABLE 5.4
Measures of Beta Diversity Using Binary (Presence-Absence) Data

Index	Equation	Terms
Whittaker	$\beta_w = (S/\alpha) - 1$	S is the total number of species (gamma diversity). α is the average sample diversity when each sample is a standardized size, and β diversity is measured as species richness.
Cody	$\beta_c = [g(H) + l(H)]/2$	$g(H)$ is the number of species gained along an entire transect or as one moves from site to site. $l(H)$ is the number of species lost across the same transect.
Routledge	$\beta_r = [S^2/(2r + S)] - 1$	S is the total number of species. r is the number of species pairs with overlapping occurrences. This measure emphasizes species richness and the degree of species overlap.
Routledge	$\beta_l = \ln(T) - [(1/T)\Sigma e_i \ln(e_i)] - [(1/T)\Sigma \alpha_j \ln(\alpha_j)]$	e_i is the number of samples along a transect in which species i is present. α_j is the species richness of sample j. $T = \Sigma e_i = \Sigma \alpha_j$. This measure is based on information theory.
Routledge	$\beta_E = \exp^{\beta_l}$	Exponential form of β_l.
Wilson and Shmida	$\beta_T = [g(H) + l(H)]/2\alpha$	$g(H)$ is the number of species gained along an entire transect or as one moves from site to site. $l(H)$ is the number of species lost across the same transect. α is species richness.

β_w (Whittaker's measure) fulfills most of the criteria for an effective index and has fewest restrictions. The index developed by Wilson and Shmida, β_t is an acceptable alternative (Magurran 1988). In addition to binary data, abundances can be used in some equations of turnover; Wilson and Mohler (1983) have provided guidelines for their use.

Resemblance Functions

Similarity, distance, and dissimilarity coefficients represent an alternative analytical approach that can be used to quantify differences in species composition (i.e., beta diversity) among sites. Approximately 20 similarity indices appear in the literature. Those indices differ in the degree to which shared occurrences or shared absences are weighted. We will discuss only the four most widely used indices (Dice, Ochiai, Jaccard, and Sørenson; Table 5.5). Additional information on a variety of indices, including a discussion of their limitations, can be found in Clifford and Stephenson (1975).

Similarity indices are never metric. Hence, they cannot be used to position objects in relation to each other in geometric space as can principal components analysis (Legendre and Legendre 1983). Information obtained from similarity measures is displayed with a similarity matrix (Table 5.6) or after conversion to measures of dissimilarity (Fig. 5.3). Calculation of the four similarity indices requires presence-absence data from a series of sites. A disadvantage of such indices is that all species contribute equally to the index, regardless of abundance. Citing results from an extensive evaluation of similarity measures by Smith (1986), Magurran (1988) reported that indices that use binary data generally give misleading results that indicate a higher degree of similarity than actually exists. Nonetheless, Magurran recommended that Sørenson's Index (Table 5.5) be used for making comparisons if only binary data can be obtained.

Tulloss (1997) reviewed the properties of 15 similarity indices and concluded that all were unsatisfactory. He found that most indices were insensitive to the relative sizes of the two species lists being compared, to the percentage of entries in the longer species list that were common to the two lists, or to the percentage of entries in the shorter list that were common to both. To deal with these weaknesses, Tulloss (1997) proposed a Tripartite Similarity Index (T) that satisfied six requirements for suitability of a resemblance function. The index is tripartite because it includes a cost function for each of the three main, conflicting, requirements for similarity indices. Those requirements are concerned primarily with sensitivity to the number of species in the ecological unit. When the three cost factors are multiplied together, the resultant similarity index is sensitive to the size of each species list that is being compared (Tulloss 1997). The Tripartite Similarity Index is obtained from the following:

$$T = U \times S \times R \qquad (8)$$

TABLE 5.5
Selected Similarity Coefficients

Index	Equation	Terms
Binary		
Dice	$DI = 2j/(2j + a + b)$	j is the number of species in common between two sites; a is the number of species in site A; b is the number of species in site B.
Jaccard	$JI = j/(a + b - j)$	Same as for Dice.
Ochiai	$OI = j/[(\sqrt{j+a})(\sqrt{j+b})]$	Same as for Dice.
Sorenson	$SI = 2j/(a + b)$	Same as for Dice.
Metric		
Bray–Curtis (modification of the Sørensen Index)	$BCI = 2W/(a + b)$	W is the sum of the lower abundances (number of individuals or other measure of density) of species that occur in each site; a is sum of the abundances in site A; b is sum of the abundances in site B. Frequencies of occurrence also can be substituted for abundances, where W is the sum of the lower frequencies of those species in common between the two assemblages; a is the sum of the frequencies of all species in assemblage A; and b is the sum of the frequencies of all species in assemblage B.
Morisita–Horn	$C_{MH} = \frac{2\Sigma(an_i - bn_i)}{(da + db)(aN\ bN)}$	aN is the number of individuals in site A. bN is the number of individuals in site B. an_i is the number of individuals in the ith species in site A. bn_i is the number of individuals in the ith species in site B. $da = \Sigma an_i^2/aN^2$ $db = \Sigma bn_i^2/bN^2$
Renkonen	$P_{12} = S_{min}\ (p_{1i}, p_{2i})$	p_{1i} is the proportional representation of species i in site 1. p_{2i} is the proportional representation of species i in site 2. S_{min} is the minimum number of species detected from either site.

TABLE 5.6
Hypothetic Matrix of Renkonen Similarity Coefficients Comparing the Compositions and Species Abundances of Fungal Assemblages from Five Sites along an Elevational Gradient

Site	B	C	D	E
A	0.729	0.748	0.748	0.699
B	—	0.733	0.733	0.567
C	—	—	0.965	0.543
D	—	—	—	0.544

where U, S, and R represent cost functions, as follows:

$$\text{Cost Function 1: } U = \frac{\log\left(1 + \frac{\min(b,c) + a}{\max(b,c) + a}\right)}{\log 2} \qquad (9)$$

where a is the number of taxa common to both areas, b is the number of species in the first area that are not in the second, c is the number of species in the second area that are not in the first, $\min(b, c)$ is the smaller of the two values in parentheses, and $\max(b, c)$ is the larger of the two values in parentheses. The value U is designed to reduce the value of the similarity index if the sizes of the species lists of the two areas being compared differ greatly. If two areas have the same number of species, the function reduces to 1. Note that Tulloss (1997) defined a, b, and c differently from the indices presented in Table 5.5.

$$\text{Cost Function 2: } S = \frac{1}{\sqrt{\frac{\log\left(2 + \frac{\min(b,c)}{a+1}\right)}{\log 2}}} \qquad (10)$$

The value S is designed to reduce the value of the index if the size of the smaller assemblage differs substantially from the size of the list of shared species.

$$\text{Cost Function 3: } R = \frac{\log\left(1 + \frac{a}{a+b}\right)\log\left(1 + \frac{a}{a+c}\right)}{(\log 2)^2} \qquad (11)$$

The value R is designed to increase the similarity between two areas as the percentage of species in the two

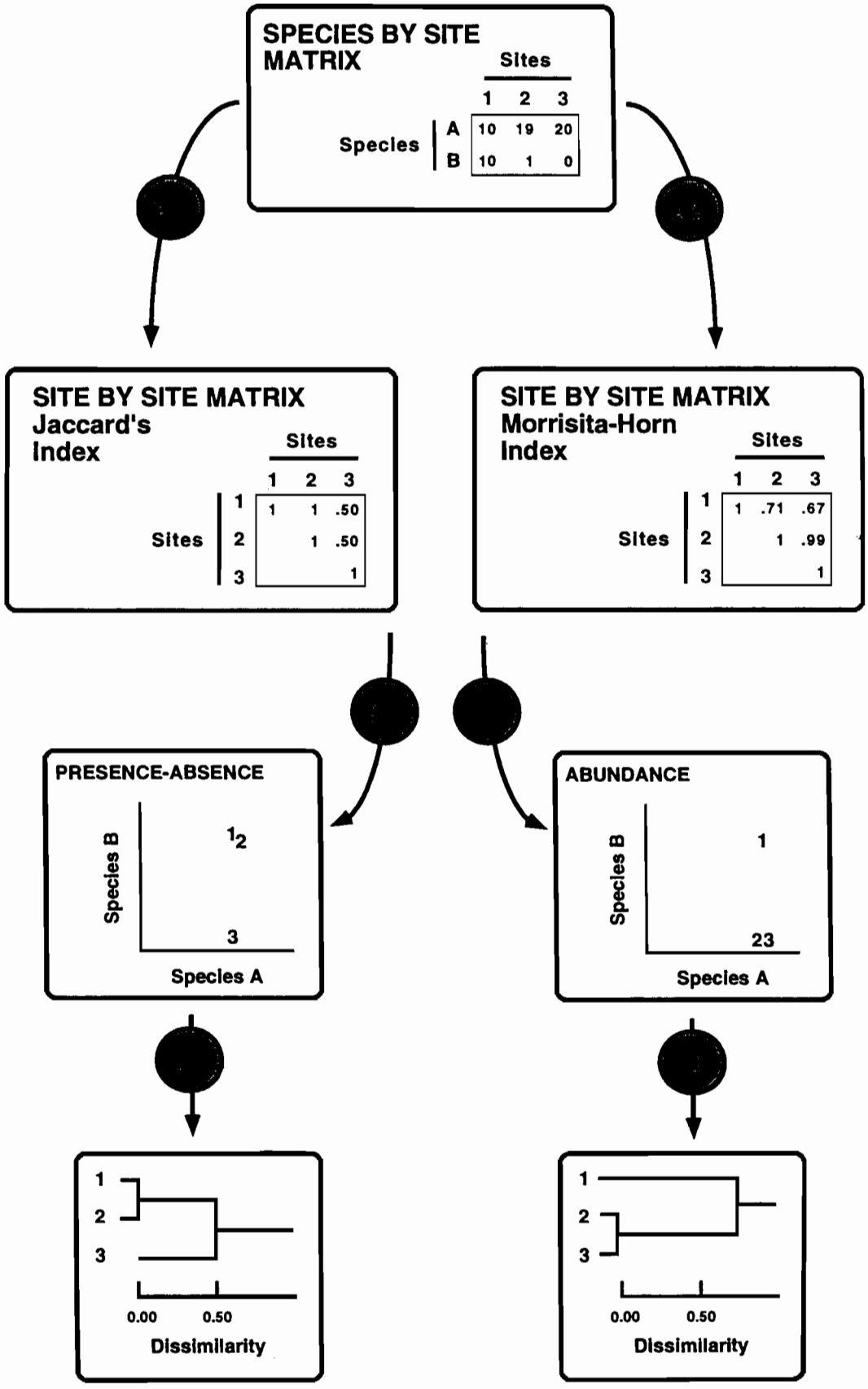

FIGURE 5.3 Schematic representation of the analysis of data on species occurrences (i.e., the species by site matrix) using different resemblance functions. Jaccard's index (A1) based on presence-absence data, and the Morisita-Horn index (A2) based on species' abundances, to give rise to ordinations in species space (B1, B2) and cluster analyses (C1, C2). The species by site matrix was constructed to emphasize the importance of the selected resemblance function to the perceived pattern among sites; sites 1 and 2 are most similar based on presence-absence data, whereas sites 2 and 3 are most similar based on species abundances.

lists that are shared increases. The tripartite index varies roughly as the inverse of the size of the larger of two lists compared (Tulloss 1997). The index should be rounded to two decimal places, except when comparing two very dissimilar lists. For species lists with low similarity, resulting in values with two or more zeros after the decimal point, Tulloss (1997) recommended rounding to the first nonzero digit in the estimate.

If quantitative data can be obtained (i.e., number of isolates, frequencies of occurrence, or other ecologically relevant measure of fungal abundance), they should be used. Similarity measures based on quantitative information are preferable to those based only on binary data. Although frequently used in the literature, the modified Sørensen Index (Table 5.5) of Bray and Curtis (1957) suffers from a number of shortcomings (Wolda 1981), and we do not recommend it for general use.

The Renkonen (1938) index of similarity (P) is based on the proportional abundances of species ($p_i = n_i/\Sigma n_i$) rather than their absolute abundances. The index ranges from 1 (unity; identical proportional abundances of species in both areas) to 0 (no taxa shared in the two areas). Although quite simple, the Renkonen index is affected very little by differences in sample size or diversity between areas (Wolda 1981) and thus is one of the best indices of similarity (Krebs 1989).

Morisita's index, also based on the proportional abundance of species, is best understood in terms of probability. The index is the ratio of two probabilities: the likelihood that an individual randomly obtained from sample 1 will be the same species as an individual obtained at random from sample 2, and the likelihood that two individuals obtained from either area 1 or area 2 will be the same species. Horn (1966) modified the index to allow its use with biomass or cover data as well as numbers of individuals. Like the Renkonen index, Morisita's index ranges from 0 (no similarity) to 1 (total similarity). Except for small samples, the index is essentially independent of sample size. Wolda (1981) recommended it as the best overall measure of similarity.

Distance Coefficients

Distance or dissimilarity coefficients (Ludwig and Reynolds 1988) assume a value of zero when two areas (sites, habitats) are identical in species composition and a maximum value when no species are shared. Both metric and nonmetric coefficients of dissimilarity are available (Pielou 1984). This is important because metric coefficients may be visualized in multidimensional space (ordination procedures) where the distance between each pair of points (sites) is equal to their compositional dissimilarity. Ordination and cluster analysis should be carried out on identical data to facilitate the understanding of diversity patterns (Fig. 5.3). Nonmetric dissimilarity measures cannot be used in this manner (Pielou 1984). We will discuss only metric indices.

Three groups of distance measures (Table 5.7) commonly are used in the ecological literature: (1) Euclidian distance coefficients, (2) the Bray-Curtis (BC) dissimilarity index (see "Resemblance Functions," earlier), and (3) relative Euclidean (RE) distance measures. Kenkel and Booth (1992) provided synopses of some of these measures as they pertain to fungal ecological units. For calculation of distance coefficients, abundance data are arranged in a matrix. Column headings are areas (e.g., leaves, tree root systems, quadrats), and row names are species.

The Euclidean distance (Table 5.7) is the metric with which we are most familiar; it is simply the geometric distance between two points in space. The lower bound is zero; the upper bound is limitless. The absolute distance, also known as the city block metric, expresses distance in terms of absolute differences. For the BC measure, a similarity index is calculated and subtracted from 100 to provide the distance component (PD). The dissimilarity index may be rescaled by subtracting the similarity index from 1 so that the distances obtained are more in line with other distance indices. The RE group (Table 5.7) contains metrics that are expressed on standardized or relative scales. Indices in the RE group compare proportional abundances. Consequently, sample units (e.g., root segments) with similar proportional abundances of species will be in close proximity.

Not all distance metrics are equally suitable for assessing the structure of biodiversity in space or time. Ludwig and Reynolds (1988) recommended against using the Euclidean group distance measures (despite their popularity) because they can produce spurious results. They also noted that the RE group indices perform reasonably well, with one metric having little advantage over another. Nonetheless, they concluded that cord distance performs best over a diverse range of ecological circumstances. Finally, Ludwig and Reynolds (1988) suggested that the Bray-Curtis (PD) index can be used as an alternative to indices in the RE group (Table 5.7).

Classification and Cluster Analysis

Cluster analysis is a classification procedure that groups objects into subgroups that are more similar to each other than to objects in other subgroups. The resulting branching diagram is a classification that provides the

TABLE 5.7
Distance Coefficients for Comparing the Structure of Ecologic Units among Locations or Sampling Times*

Index	Equation	Terms
Euclidean distance coefficients		
Euclidean Distance (ED)	$ED = \sqrt{\Sigma(X_{ij} - X_{ik})^2}$	X_{ij} is the abundance of species i at site or time j. X_{ik} is the abundance of species i at site or time k.
Squared Euclidean Distance (SED)	$SED = \Sigma(X_{ij} - X_{ik})^2$	Same as for ED.
Mean Euclidean Distance (MED)	$MED = \sqrt{\frac{\Sigma(X_{ij} - X_{ik})^2}{S}}$	Same as for ED. S is number of species.
Mean Absolute Distance (MAD)	$MAD = \Sigma \lvert X_{ij} - X_{ik} \rvert / S$	Same as for MED.
Bray-Curtis Dissimilarity Index	1. Calculate $SI = (2W/a + b)(100)$ 2. Calculate $PD = 100 - SI$ PD may be scaled between 0 and 1 by $PD = 1 - SI$	W is the sum of the lower abundances (number of individuals or other measure of density) of species that occur in each site. a is sum of the species abundances in site A, and b is the sum of the species abundances that occur in site B. *SI*, similarity index; *PD*, distance component.
Relative Euclidean Distance coefficients		
Relative Euclidean Distance (RED)	$RED = \sqrt{\Sigma[(X_{ij}/\Sigma X_{ij}) - (X_{ik}/\Sigma X_{ik})]^2}$	Same as for ED.
Relative Absolute Distance (RAD)	$RAD = \Sigma \lvert (X_{ij}/\Sigma X_{ij}) - (X_{ik}/\Sigma X_{ik}) \rvert$	Same as for ED. Values range from 0 to 2.
Cord Distance (CRD)	$CRD = \sqrt{2(1 - ccos_{jk})}$	$ccos_{jk}$ = Chord cosine $ccos_{jk} = \Sigma(X_{ij}\ X_{ik}) - \sqrt{\Sigma X_{ij}\ \Sigma X}$
Geodesic Distance (GDD)	$GDD = \arccos(ccos_{jk})$	This measure is the distance along the arc of a unit circle after projection of the sample units onto a circle of unit radius. Values range from 0 to 1.57.

*Notation follows Ludwig and Reynolds (1988).

sequence of clusters (subgroups) by which a set of objects is subdivided. When a number of ecological units is investigated, cluster analysis is appropriate for representing patterns of species composition among those units (Fig. 5.3). A dissimilarity matrix usually forms the basis of clustering models, although many software packages (e.g., SPSS 1999) allow the input of either similarity or dissimilarity (distance) measures. Essentially, cluster analysis produces a dendrogram or tree whose branches represent each of the ecological units (Fig. 5.3). The data on species composition for these sites determine the branching pattern. The joined branches represent groups or clusters of sites with similar species composition. The length of the branch prior to joining is inversely proportional to the degree of similarity in species composition.

A wide variety of cluster analyses (Sneath and Sokal 1973) are available (e.g., agglomerative and divisive, hierarchic and nonhierarchic, overlapping and nonoverlapping, sequential and simultaneous, local and global, direct and iterative, weighted and unweighted, adaptive and nonadaptive). The choice of a particular method is critical because it determines (in part) the pattern derived from the data on species composition. Nonetheless, most research ecologists and evolutionary biologists follow the unweighted pair group method using arithmetic averages (UPGMA), which provides a good starting point for visualizing the degrees of similarity among sites based on species composition. Like many multivariate statistical analyses, cluster analysis attempts to represent the complex relationships among sites in a simple one-dimensional way. As a result, various amounts of information are lost, and the pattern represented by the dendrogram may not be accurate. The degree to which the observed dendrogram represents the multidimensional relationship among sites based on the similarity or dissimilarity matrix is quantified by the cophenetic correlation coefficient (Sokal and Rohlf 1962). Clustering methods, including considerations of distortion and optimality, are discussed by Sneath and Sokal (1973), Gaugh (1982), and van Tongeren (1995). It is important to remember that the veracity of the pattern detected by cluster analysis is only as reliable as the input data (species presence or absence) and that the pattern is affected strongly by the resemblance function used (Fig. 5.3). The sites to be included in analyses, the suite

of species on which to base resemblance functions, and the resemblance function chosen should be considered with care.

Power Analysis

To understand the effects of abiotic or biotic factors on patterns of biological diversity, differences in species richness among areas must be compared statistically. When the null hypothesis (H_o) is rejected, the likelihood of making an error is the significance level of the test (i.e., P value). Often, researchers believe that "real" biological differences exist between ecological units, although H_o is not rejected. In fact, differences may not be statistically significant for two reasons: (1) no differences exist among the ecological units, or (2) sample sizes are too small to reveal biological differences of a magnitude considered to be important. These problems plague investigators and pose important challenges for the interpretations of patterns in fungal biodiversity. Power analysis can help to distinguish between the two alternatives and should be considered a crucial component of any experimental design and analysis (Taylor and Gerrodette 1993; Thomas and Juanes 1996).

A good discussion of power analysis can be found in Muller and Benignus (1992). The power of a statistical test, such as analysis of variance (ANOVA), is the ability to reject the null hypothesis when an alternative hypothesis is true. Power differs depending on the type of statistical test used to assess differences. For example, parametric tests have more power than nonparametric tests when the assumptions of the former are met. Similarly, a priori tests have more power than a posteriori for detecting differences. Critically, power increases as the number of samples increases (decreasing standard error and increasing the degrees of freedom), as the size of samples increases, if a one-tailed rather than a two-tailed t test is used (if the alternate hypothesis is in the direction of the true difference), or if a higher alpha level (probability of rejecting the null hypothesis) is used. It is important to note that when the sample variance is large, power is low and ability to detect differences among ecological units is small.

Power analysis is most useful when designing and planning a study because it allows the investigator to evaluate the relationships among all five factors, including (1) the range of sample sizes, quadrats, or plots that feasibly can be examined, (2) the magnitude of differences among ecological units considered to be biologically important, (3) the magnitude of variation, (4) desired levels of rejecting the null hypothesis (alpha), and (5) statistical power (Thomas and Krebs 1997). Such preliminary investigations lead to optimal use of financial resources and personnel. Even after a study is completed, power analysis can be useful for interpreting the biological meaning of nonsignificant results.

TABLE 5.8
Selected, Stand-alone Statistical Software for Performing Power Analysis*

Software	Version	Vendor†	Operating System
Nquery Advisor	1.0	Statistical Solutions	Windows
PASS	6.0	NCSS Statistical Software	Windows
SPSS	10	SPSS, Inc.	Windows or Mac

* Information modified from Thomas and Krebs (1997).
† See Appendix IV for contact information.

Although power analysis can be performed using charts and tables provided in several texts (e.g., Cohen 1988; Lipsey 1990; Zar 1996), interpolation between tabled values can lead to errors. Thomas and Krebs (1997) reviewed 29 programs and five statistical packages that perform power analyses (Table 5.8) and are easy to use. In addition, all of the programs cover a wide range of tests and have a number of useful utilities.

EPILOGUE

Assessments of fungal biodiversity—whether in tussock grasslands of arctic tundra or in steamy, lowland rain forests of equatorial regions—must be grounded firmly in current methodologies of sampling design that account for effects of scale on patterns of biodiversity. Moreover, research programs should account for both temporal variation and spatial heterogeneity; this is particularly important because of the growing recognition that previous land use and disturbance can have pervasive effects on both. The ability to understand patterns of diversity across landscapes is predicated on a research design that effectively matches the temporal and spatial scales at which processes operate, with a corresponding scale of pattern detection.

Our goal has been to expose mycologists and others examining the diversity of fungi to analytical tools and approaches for understanding how biodiversity is partitioned across space and time. Because diversity is hierarchical, attention to the level of assessment is critical for teasing apart mechanisms that contribute to patterns. Ultimately, we must accept that diversity is a dynamic aspect of ecosystems and landscapes whose importance and contributions to system function are still not well

understood. If we simply enumerate long lists of fungal species from various regions without attempting to evaluate and understand the mechanisms that contribute to observed patterns of diversity, we will only have described the shadows of the ecological actors that are participating in a complex and interesting evolutionary play. The challenge is to develop and maintain the taxonomic expertise necessary to identify the fungal taxa that are present in an ecosystem and at the same time to ensure that students and researchers have the strong ecological background necessary to evaluate fungal biodiversity patterns.