

REVIEW

Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness

Nicholas J. Gotelli¹

and Robert K. Colwell²

¹*Department of Biology,*

University of Vermont,

Burlington,

Vermont 05405, U.S.A.

E-mail: ngotelli@zoo.uvm.edu

²

In principle, for a survey of some well-defined spatial scope, an asymptote will eventually be reached and no further taxa will be added.

We distinguish four kinds of *assessment*, based on two dichotomies (Fig. 1). Although we will present these curves in terms of species richness, they apply just as well to richness of higher taxa.

The first dichotomy concerns the sampling protocol used to assess species richness. Suppose one wishes to compare the number of tree species in two contrasting 10-ha forest plots. One approach is to examine some number of individual trees at random within each plot, recording sequentially the species identity of one tree after another. We refer to such an assessment protocol as *individual-based* (Fig. 1). Alternatively, one could establish a series of quadrats in each plot, record the number and identity of all the trees within each, and accumulate the total number of species as additional quadrats are censused (e.g. Cannon *et al.* 1998; Chazdon *et al.* 1998; Hubbell *et al.* 1999; Vandermeer *et al.* 2000). This is an example of a *quadrat-based* assessment

protocol (Fig. 1). The relative merit of these approaches for estimating species richness of trees is not the point here. Rather, we emphasize that species richness censuses can be validly based on datasets consisting either of individuals or of replicated, multi-individual samples. The key distinction is

In contrast, a *rarefaction curve* is produced by repeatedly re-sampling the pool of N individuals or N samples, at random, plotting the average number of species represented by $1, 2, \dots, N$ individuals or samples (Fig. 1). Sampling is generally done without replacement, within each re-sampling. Thus, rarefaction generates the expected number of species in a small collection of n individuals (or n samples) drawn at random from the large pool of N individuals (or N samples; Simberloff 1978).

These two dichotomies jointly define four kinds of taxon sampling curves, as shown in Fig. 1. Accumulation curves, in effect, move from left to right, as they are further extended by additional sampling. In contrast, rarefaction curves move from right to left, as the full dataset is increasingly "rarefied". Because the entire rarefaction curve

distributions, individual-based rarefaction inevitably overestimates the number of species (or higher taxa) that would have been found with less effort. In fact, the difference between the sample-based and individual-based rarefaction curves can be used as a measure of patchiness (Colwell & Coddington 1994).

Regardless of which approach is used, it is the individual that carries taxonomic information. When sample-based rarefaction curves are used to compare taxon richness at comparable levels of sampling effort, the number of taxa should be plotted as a function of the accumulated number of individuals, not accumulated number of samples, because datasets may differ systematically in the mean number of individuals per sample. (Here, we are assuming that taxon richness is the question, not taxon density; see below.)

An example makes this pitfall clear. Suppose you wish to know whether tropical old-growth forest or nearby tropical second-growth forest is richer in tree species. You identify all individual stems in 10 × 10 m randomly placed quadrats in each forest type. The sample rarefaction curve for second-growth forest, plotted as a function of samples, lies above the corresponding curve for old-growth forest, but neither has reached an asymptote (Fig. 2a). The mean number of stems per quadrat is considerably greater in the second-growth forest, as would be expected. Are there really more species in the second-growth forest? Not even an approximate answer can be given to this question without re-scaling the x-axis to number of individuals (based on the average number of individuals per sample). Once re-scaled, the second-growth forest curve will drop relative to the old-growth forest curve; it may (still) lie above it, coincide, or fall below it (Fig. 2b). (Cannon *et al.* 1998 demonstrated this pitfall for logged vs. unlogged forests, which differ in stem density and in quadrat-based richness, but have similar species richness when re-scaled to individuals.) This example illustrates the importance of using taxon sampling

for two communities with different patterns of relative abundance may cross once or even twice. Likewise, sample-based rarefaction can cross, if based on communities that differ sufficiently in patchiness. Thus, the sample size to which one rarefies can potentially change the rank order of

taxonomically very different, the sampling may not adequately characterize each taxon (Simberloff 1978). If the sampling methods are not identical, different kinds of species may be over- or under-represented in different samples, because no sampling method is completely random and unbiased (Boulinier *et al.* 1998). In addition, the shape of individual-based rarefaction curves depends upon relative abundance – the greater the evenness of the relative abundance distribution, the steeper the rarefaction curve (Gotelli & Graves 1996). For this reason, rarefaction curves

number of accumulated individuals is also computed, to allow re-scaling of sample-based rarefaction curves. Free software is also available for the construction of individual-based rarefaction curves and confidence intervals for species richness and other diversity indices (Gotelli & Entsminger 2001).

CATEGORY-SUBCATEGORY RATIOS AND THEIR PITFALLS

Individuals and species

To introduce the concept, and the perils, of what we call *ca eg, , , bca eg, , , ai, ,*, let us return to the example (above) of assessing tree species richness in old-growth vs. second-growth forest. Recall that the problem with comparing sample-based rarefaction curves scaled by number of samples was that second-growth quadrats each had more stems than equal-sized old-growth quadrats, on average. Why not simply compare average species per stem, among quadrats, for each forest type, to remove the effect of stem density? This index is the species-per-individual ratio, a particular class of category-subcategory ratios.

Figure 4 illustrates the hazards of using the species-per-individual ratio to compare samples. Each panel in Fig. 4 shows hypothetical, sample-based rarefaction curves for contrasting forest habitats. Each curve is based on the same number of quadrats, but each is re-scaled to the number of individuals on the x -axis. The solid dots indicate total richness for the pooled quadrats in each forest habitat. The slopes of the lines connecting these points to the origin equal the ratio of species to individuals for the dots. In Fig. 4(a), old-growth and second-growth forest have identical species richness (at least as far as the curves extend), yet the number of species per individual is much lower for the second-growth forest. In Fig. 4(b), species richness is higher in forest gaps than in non-gaps (forest matrix), yet the number of species per individual is identical for total richness in gaps and non-gaps.

An example from the recent literature illustrates the perils of "normalizing" richness by dividing the number of species by the number of individuals. In support of their inference that tree species richness does not differ between gaps and non-gaps, Hubbell *et al.* (1999) showed that number of species divided by number of stems did not differ for saplings in gaps vs. non-gaps in a Panamanian forest. Using Hubbell's reported stem densities and richness values for saplings in 20×20 -m quadrats, Chazdon *et al.* (1999) showed that true sapling species richness might in fact fit curves such as those in Fig. 4(b) (see also Kobe 1999; Vandermeer *et al.* 2000), with greater total richness in gaps. In his reply, Hubbell (1999) failed to provide the individual-based species accumulation curves to disprove Chazdon's

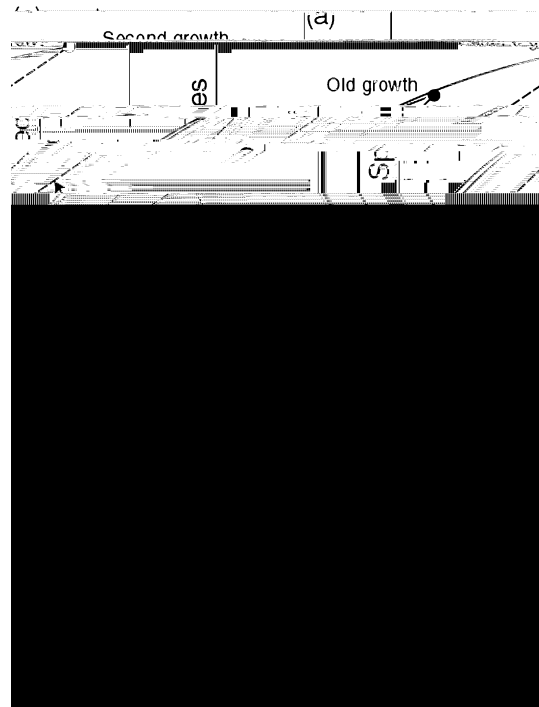


Figure 4 Pitfalls of using species/individual ratios to compare datasets. In (a), an old-growth and a second-growth forest stand are compared. The 2 stands have identical individual-scaled rarefaction curves, and thus do not differ in species richness. The second growth curve extends farther simply because stem density is greater, so that more individuals have been examined for the same number of samples. However, when the ratio of species/individual is computed for each, the ratio is much higher for the old-growth stand. In (b), species richness in treefall gap quadrats is compared with richness in non-gap (forest matrix) quadrats. In this case, species/individual ratios are identical, yet the true species richness is higher in gaps.

conjecture for the sapling dataset at issue. Instead, Hubbell *et al.* (1999) provided individual-based accumulation curves for a quite different dataset (no size class specified) and cited the fact that *q*-based accumulation curves do not differ for gaps and non-gaps, leaving the debate unresolved. Our point here is simply that, had individual-based accumulation curves been published for the sapling dataset at issue in the first place, the ambiguity that instigated the debate would never have arisen.

Using the species-per-individual ratio to correct for unequal numbers of individuals is invalid because it assumes that richness increases linearly with abundance – true only for the idealized case of extreme unevenness, in which one species is maximally dominant (Gotelli & Graves 1996). Because abundances are rarely this extreme, the species-per-abundance ratio will distort patterns of species richness.

communities. For animal communities, Williams (1947, 1964) elucidated these same patterns using species-abundance models and computer simulations. Although their work was ignored by ecologists for several decades (Järvinen 1982), re-analyses of species-to-genus ratios now suggest that island communities harbour slightly *more* species per genus than expected by chance, in spite of the lower absolute number of species per genus expected in smaller samples (Simberloff 1970). This finding is the opposite of what competition theory predicts, perhaps reflecting instead the similar dispersal potential and ecological requirements of congeneric species (the Icarus Effect of Colwell & Winkler 1984). Despite the periodic rediscovery of this classic pitfall, sample-size dependence of taxonomic ratios continues to trap the unwary (e.g. Ashton 1998).

SPECIES RICHNESS VS. SPECIES DENSITY

We have emphasized the importance of using taxon sampling curves (both individual- and sample-based) to standardize datasets to a common number of individuals for the purposes of comparing species richness. In contrast, most community ecology studies standardize on the basis of area or sampling effort. Thus, most ecological comparisons of biodiversity are actually comparisons of *species density*: the number of species per unit area (Simpson 1964). Such studies hinge on the assumption that samples are drawn from populations of individuals that are at comparable densities. However, species density depends on both species richness and on the mean density of individuals (disregarding species), as discussed in relation to the example of old-growth vs. second-growth forest above (Fig. 2). Consequently, the ordering of communities may differ when ranked by species richness vs. species density (James & Wamer 1982; McCabe & Gotelli 2000).

Both species richness and species density can be compared using sample- and individual-based rarefaction curves (Fig. 7). Individual-based rarefaction curves standardize each of two or more samples on the basis of the number of individuals, for the purpose of comparing species richness. Sample-based rarefaction curves can be used to compare richness in the same way, as long as the *y*-axis is re-scaled in units of individuals. In contrast, to compare

to area is not a valid measure of species density, because the number of species increases nonlinearly with area. Instead, species density is validly compared only with the appropriate taxon sampling curves (e.g. James & Wamer 1982).

Which measure is more appropriate, species richness or species density? In other words, should communities be compared on the basis of a standardized number of individuals (species richness) or a standardized area or

most tropical arthropod datasets (e.g. Stork 1991; Wolda *et al.* 1998; Fisher 1999; Novotny & Basset 2000). The tricky issue is whether the performance of the estimators on benchmark datasets – which usually consist of relatively small numbers of species – accurately predicts the performance of the same estimators on not-yet-asymptotic datasets, which usually consist of very large numbers of species. One indication of the failure of the existing catalogue of estimators for hyperdiverse taxa is that they often fail to reach any asymptote at all, rising more or less in parallel with the still-steep sample-based rarefaction curve (e.g. Fisher 1999). In these cases, the estimators must be viewed as providing only lower-bound estimates of species richness (Anne Chao, personal communication). On the other hand, restricting datasets to ecologically more homogenous subsets of samples sometimes does produce well-behaved, asymptotic richness estimates (J. Longino *et al.*, in press). This is still an ongoing area of research, and there is much need for comparative studies of the performance of asymptotic species estimators on different empirical and theoretically derived data sets.

CONCLUSIONS

The principles of species accumulation, rarefaction, species richness, and species density have been established for many decades. However, ecologists have only recently begun in earnest to incorporate these concepts into their measurements of species diversity patterns and evaluation of theory in community ecology and biogeography. These tasks are especially important as ecologists attempt to inventory species-rich communities and document the loss of species diversity from habitat destruction and global climate change. Ecologists may have avoided individual-based and sample-based rarefaction curves because they are computationally intensive, but public-domain software is now available for these calculations (Colwell 2000a; Gotelli & Entsminger 2001).

ACKNOWLEDGEMENTS

We thank J. Grover for inviting us to write this review. EcoSim software development supported by NSF grants BIR-9612109 and DBI 9725930 to NJG. EstimateS software development supported by NSF grants BSR-9025024, DEB-9401069 and DEB-9706976 to RKC. Preparation of this paper was supported by NSF grant DEB-0072702 to RKC.

REFERENCES

Abele, L.G. & Walters, K. (1979). The stability-time hypothesis: Reevaluation of the data. *Acta Numerica*, 114, 559–568.

Abrams, P. (1995). Monotonic or unimodal diversity-productivity gradients: what does competition theory predict? *Ecology*, 76, 2019–2027.

Anderson, R.S. & Ashe, J.S. (2000). Leaf litter inhabiting beetles as surrogates for establishing priorities for conservation of selected tropical montane cloud forests in Honduras, Central America (Coleoptera; Staphilidae, Curculionidae). *Biodiversity Conservation*, 9, 617–653.

- Colwell, R.K. & Winkler, D.W. (1984). A null model for null models in biogeography. In: *Ecological Communities: Concepts and Methods*, eds Strong D.R. Jr, Simberloff D., Abele L.G. & Thistle A.B.), pp. 344–359. Princeton University Press, Princeton, U.S.A.
- Connell, J.H. (1978). Diversity in tropical rain forests and coral reefs. *Science*, 199, 1302–1303.
- Cook, R.E. (1969). Variation in species density of North American birds. *Systematic Zoology*, 18, 63–84.
- Cornell, H.V. (1999). Saturation and regional influences on species richness in ecological communities: a review of the evidence. *Ecology*, 6, 303–315.
- Darwin, C. (1859). *The Origin of Species by Means of Natural Selection*. Murray, London, U.K.
- Denslow, J. (1995). Disturbance and diversity in tropical rain forests: the density effect. *Ecological Applications*, 5, 962–968.
- DiTomasso, A. & Aarssen, L.W. (1989). Resource manipulations in natural vegetation: a review. *Vegetatio*, 84, 9–29.
- Elton, C. (1946). Competition and the structure of ecological communities. *Journal of Ecology*, 15, 54–68.
- Engstrom, R.T. & James, F.C. (1981). Plot size as a factor in winter bird-population studies. *Condor*, 83, 34–41.
- Fisher, B.L. (1999). Improving inventory efficiency: a case study of leaf-litter ant diversity in Madagascar. *Ecological Applications*, 9, 714–731.
- Gaston, K.J. (2000). Global patterns in biodiversity. *Nature*, 405, 220–227.
- Gotelli, N.J. (2001). *Applied Ecology*, 3rd edn. Sinauer Associates, Inc, Sunderland, MA, U.S.A.
- Gotelli, N.J. & Arnett, A.E. (2000). Biogeographic effects of red fire ant invasion. *Ecological Letters*, 3, 257–261.
- Gotelli, N.J. & Entsminger, G.L. (2001). *Ecological Niche Modeling Software*, Version 6.0. Acquired Intelligence Inc, & Kesey-Bear <http://homepages.together.net/gentsmin/ecosim.htm>
- Gotelli, N.J. & Graves, G.R. (1996). *Niche Modeling in Ecology*. Smithsonian Institution Press, Washington, DC, U.S.A.
- Grassle, J.F. & Maciolek, N.J. (1992). Deep-sea species richness: regional and local diversity estimates from quantitative bottom samples. *Deep-Sea Research*, 139, 313–341.

- Rudran R. & Foster, M., pp. 226–234. Smithsonian Institution Press, Washington, DC, U.S.A.
- Novotny, V. & Basset, Y. (2000). Rare species in communities of tropical insect herbivores: pondering the mystery of singletons. *Oikos*, 89, 564–572.
- Oksanen, J. (1996). Is the humped relationship between species richness and biomass an artifact due to plot size? *J. Ecology*, 84, 293–295.
- Palmer, M.W. (1990). The estimation of species richness by extrapolation. *Ecology*, 71, 1195–1198.
- Pielou, E.C. (1966). The measurement of diversity in different types of biological collection. *J. Theoretical Biology*, 13, 131–144.
- Pielou, E.C. (1975). *Ecological Diversity*. Wiley Interscience, New York, NY, U.S.A.
- Poulsen, B.O., Krabbe, N., Frølander, A., Hinojosa, M.B. & Quiroga, C.O. (1997). A rapid assessment of Bolivian and Ecuadorian montane avifaunas using 20-species lists: efficiency, biases and data gathered. *Biodiversity Letters*, 7, 53–67.
- Purvis, A. & Hector, A. (2000). Getting the measure of biodiversity. *Nature*

